



THE UNIVERSITY *of* EDINBURGH

Edinburgh Research Explorer

A novel framework to build saliva-based DNA methylation biomarkers: quantifying systemic chronic inflammation as a case study

Citation for published version:

Schmunk, LJ, Call, TP, McCartney, DL, Javaid, H, Hastings, WJ, Jovicevic, V, Kojadinovic, D, Tompkinson, N, Zlamalova, E, McGee, KC, Sullivan, J, Campbell, A, McIntosh, AM, Ovari, V, Wishart, K, Behrens, CE, Stone, E, Gavrilov, M, Thompson, R, Jackson, T, Lord, JM, Stubbs, TM, Marioni, RE & Martin-Herranz, DE 2025, 'A novel framework to build saliva-based DNA methylation biomarkers: quantifying systemic chronic inflammation as a case study', *Aging Cell*, vol. 24, no. 4, e14444, pp. e14444.
<https://doi.org/10.1111/accel.14444>

Digital Object Identifier (DOI):

[10.1111/accel.14444](https://doi.org/10.1111/accel.14444)

Link:

[Link to publication record in Edinburgh Research Explorer](#)

Document Version:

Publisher's PDF, also known as Version of record

Published In:

Aging Cell

General rights

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.





A novel framework to build saliva-based DNA methylation biomarkers: Quantifying systemic chronic inflammation as a case study

Lisa J. Schmunk¹ | Toby P. Call¹ | Daniel L. McCartney² | Hira Javid¹ |
Waylon J. Hastings³ | Vanja Jovicevic¹ | Dragoljub Kojadinović¹ |
Natacha Tomkinson¹ | Eliska Zlamalova¹ | Kirsty C. McGee⁴ | Jack Sullivan⁴ |
Archie Campbell² | Andrew M. McIntosh^{2,5} | Veronika Óvári⁶ | Karl Wishart⁶ |
Christian E. Behrens⁷ | Emma Stone¹ | Miloš Gavrilov¹ | Rob Thompson¹ |
Hurdle bio-infrastructure team¹ | Thomas Jackson⁴ | Janet M. Lord^{4,8} |
Thomas M. Stubbs¹ | Riccardo E. Marioni² | Daniel E. Martin-Herranz¹

¹Hurdle.Bio/Chronomics Ltd., London, UK

²Centre for Genomic and Experimental Medicine, Institute of Genetics and Cancer, University of Edinburgh, Edinburgh, UK

³Department of Psychiatry and Behavioral Sciences, Tulane University School of Medicine, New Orleans, Louisiana, USA

⁴MRC-Versus Arthritis Centre for Musculoskeletal Ageing Research, Institute of Inflammation and Ageing, University of Birmingham, Birmingham, UK

⁵Division of Psychiatry, Centre for Clinical Brain Sciences, University of Edinburgh, Edinburgh, UK

⁶Bayer Consumer Care AG, Basel, Switzerland

⁷Bayer HealthCare LLC, Whippany, New Jersey, USA

⁸NIHR Birmingham Biomedical Research Centre, University Hospitals Birmingham, Birmingham, UK

Correspondence

Lisa J. Schmunk and Daniel E. Martin-Herranz, Hurdle.bio/Chronomics Ltd., London, UK.

Email: lisa.schmunk@hurdle.bio and daniel@hurdle.bio

Abstract

Accessible and non-invasive biomarkers that measure human ageing processes and the risk of developing age-related disease are paramount in preventative healthcare. Here, we describe a novel framework to train saliva-based DNA methylation (DNAm)

Abbreviations: AgeAccel, epigenetic age acceleration; ANOVA, one-way analysis of variance; BEST, NIH-FDA Biomarkers EndpointS and other Tools classification framework; BMI, body mass index; BMP1, Bone Morphogenetic Protein 1; CCL11, C-C Motif Chemokine Ligand 11; CCL18, C-C Motif Chemokine Ligand 18; CI, confidence interval; COPD, chronic obstructive pulmonary disease; COVID-19, coronavirus disease 2019; CpG, 5'-cytosine-phosphate-guanine-3'; CRP, C-reactive protein; CXCL10, C-X-C Motif Chemokine Ligand 10; CXCL9, C-X-C Motif Chemokine Ligand 9; DNA, Deoxyribonucleic acid; DNAm, DNA methylation; EHR, electronic health record; Eos, eosinophils; EpiScore, Epigenetic proxy score to predict analyte levels from DNAm data; ESR, erythrocyte sedimentation rate; EWAS, epigenome-wide association study; FAP, Fibroblast Activation Protein Alpha; FDR, false discovery rate; GDF-15, Growth Differentiation Factor 15; GEO, Gene Expression Omnibus; GO, Gene ontology; GS, Generation Scotland; hg19, Human genome version 19, equivalent to Homo sapiens (human) genome assembly GRCh37; HR, Hazard Ratio; iAge, InflammationAge (Sayed et al 2021) proxy of inflammageing based on protein and cytokine measurements; ICC, intraclass correlation coefficient; IgG, Immunoglobulin G; IL-18R1, Interleukin 18 receptor 1; IL-6, Interleukin 6; InflammAge, Novel epigenetic biomarker to measure DNAm inflammation age (this work); InflammAgeAccel, InflammAge acceleration; INF- γ , Interferon-gamma; MAE, median absolute error; MET, metabolic equivalents; ML, machine learning; MMP12, Matrix Metalloproteinase 12; Mono, monocytes; N, sample size; Neu, Neutrophils; NHL, Non-Hodgkin's lymphoma; NK cell, Natural killer cell; NOTCH1, Notch Receptor 1; NS, non-significant; NTRK3, Neurotrophic Receptor Tyrosine Kinase 3; PCA, principal component analysis; pQTL, genetic protein quantitative trait loci; Probe, Methylation array probe that hybridises with select genomic regions to determine CpG methylation status; R_p , Pearson correlation coefficient; R_s , Spearman correlation coefficient; SCI, systemic chronic inflammation; TNFRSF1B, TNF receptor superfamily member 1B; TNF- α , tumour necrosis factor-alpha; UTR, untranslated regions.

See [Appendix](#) for the Hurdle bio-infrastructure team.

This is an open access article under the terms of the [Creative Commons Attribution](#) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2025 Chronomics Ltd. Bayer and The Author(s). *Aging Cell* published by Anatomical Society and John Wiley & Sons Ltd.

**Present address**

Eliska Zlamalova, Pale Fire Capital SE,
Prague, Czech Republic

Funding information

National Alliance for Research on
Schizophrenia and Depression, Grant/
Award Number: 27404; Chief Scientist
Office, Scottish Government Health and
Social Care Directorate, Grant/Award
Number: CZD/16/6; Scottish Funding
Council, Grant/Award Number: HR03006;
Royal College of Physicians of Edinburgh;
Bayer Consumer Health; Wellcome Trust,
Grant/Award Number: 104036/Z/14/Z,
216767/Z/19/Z and 220857/Z/20/Z;
Medical Research Council

biomarkers that are reproducible and biologically interpretable. By leveraging a reliability dataset with replicates across tissues, we demonstrate that it is possible to transfer knowledge from blood DNAm to saliva DNAm data using DNAm proxies of blood proteins (EpiScores). We apply these methods to create a new saliva-based epigenetic clock (InflammAge) that quantifies systemic chronic inflammation (SCI) in humans. Using a large blood DNAm human cohort with linked electronic health records and over 18,000 individuals (Generation Scotland), we demonstrate that InflammAge significantly associates with all-cause mortality, disease outcomes, lifestyle factors, and immunosenescence; in many cases outperforming the widely used SCI biomarker C-reactive protein (CRP). We propose that our biomarker discovery framework and InflammAge will be useful to improve understanding of the molecular mechanisms underpinning human ageing and to assess the impact of gero-protective interventions.

KEYWORDS

ageing, biomarker, DNA methylation, epigenetic clock, inflammaging, machine learning, systemic chronic inflammation (SCI)

1 | INTRODUCTION

DNA methylation (DNAm) is a powerful data type to build biomarkers that quantify the risk for developing age-related diseases and the ageing process itself (Horvath & Raj, 2018). This may be due to the ability of DNAm to capture information from both genetics and lifestyle/environmental factors. DNAm patterns associated with ageing and maximum lifespan are evolutionarily conserved (Lu et al., 2023), suggesting a critical role in the biology of ageing.

Epigenetic ageing clocks are becoming more widespread in clinical research, including in the context of quantifying the effect of candidate healthspan extending interventions in human trials (Moqri et al., 2023), and in the consumer market for preventative healthcare applications. This has been driven by significant improvements in the design and performance of these biomarkers, both from a technical and clinical perspective. Early examples include the multi-tissue Horvath clock (Horvath, 2013), the blood-specific Hannum clock (Hannum et al., 2013) or the skin-blood clock (which has good predictive performance in saliva samples and in vitro cell culture) (Horvath et al., 2018). These were trained directly on chronological age and are able to predict it with remarkable accuracy, which has applications in the forensic field (Horvath & Raj, 2018).

Second-generation epigenetic clocks have improved associations with clinical outcomes. Instead of training DNAm biomarkers directly on chronological age, it is possible to train them on intermediate phenotypic proxies that better capture morbidity and all-cause mortality risk. These include phenotypic ageing measurements derived from clinical biomarkers (PhenoAge) (Levine et al., 2018), a proxy that quantifies the pace of ageing across multiple organs/systems (DunedinPACE) (Belsky et al., 2022) or DNAm-based predictors of plasma proteins and smoking exposure that accurately predict all-cause mortality (GrimAge) (Lu et al., 2019). Further methodological innovations, such as leveraging principal component

analysis (PCA) to reduce technical variability of CpG probes, have enhanced the test-retest reliability of DNAm biomarkers, which is critical to capture longitudinal trajectories and intervention effects (Higgins-Chen et al., 2022).

Despite these improvements, two problems remain for existing DNAm-based biomarkers: (Aps et al., 2002) the lack of biological interpretability with current epigenetic clocks behaving like a “black box” (with the exception of perhaps GrimAge, where features can be interpreted), and (Aryee et al., 2014) the paucity of second-generation reproducible epigenetic clocks in saliva (which is a more accessible and scalable sample type for preventative healthcare propositions when compared to blood). Here, we outline a novel framework to build saliva-based DNAm biomarkers that addresses these challenges by leveraging a library of blood-based DNAm predictors of different plasma proteins (protein epigenetic scores or EpiScores) (Aslibekyan et al., 2018; Gadd et al., 2022; Hillary et al., 2020; Ligthart et al., 2016; Stevenson et al., 2020, 2021).

Ageing is driven by many molecular processes known as “Hallmarks of Ageing” (López-Otín et al., 2023), with the geroscience hypothesis postulating that measuring and targeting these processes rather than individual diseases will lead to a more successful impact on human healthspan (Kennedy et al., 2014). Low-grade systemic chronic inflammation (SCI), also known as “inflammaging”, has recently emerged as one of these hallmarks (López-Otín et al., 2023). SCI is characterised by persistent and non-resolving inflammation in the absence of an acute trigger such as infection, leading to collateral tissue damage and an increase in disease risk (Furman et al., 2019).

A subset of blood-based biomarkers traditionally associated with acute inflammatory responses also consistently change their baseline levels during human ageing; thus they have been proposed as SCI biomarkers. This includes biomarkers such as C-reactive protein (CRP) (Tang et al., 2018), erythrocyte sedimentation rate (ESR) (Siemons et al., 2014), IL-6 (Stevenson et al., 2021), TNF- α (Koelman



et al., 2019), interferon-gamma (IFN- γ) (Koelman et al., 2019), CXCL9 (Sayed et al., 2021), and other cyto- and chemokines (Koelman et al., 2019). However, due to the high biological variability of these biomarkers with acute inflammation and infection status, single measurements are normally insufficient to determine the SCI status of an individual.

Recently, omics-based approaches have used machine learning (ML) algorithms to predict SCI status, including the immunoglobulin G (IgG) glycome (Štambuk et al., 2020) or iAge (Sayed et al., 2021). However, these biomarkers still require invasive procedures to access blood, they may not be very stable (e.g., cytokines in iAge) and they require access to expensive and complex molecular assays (which are not globally available across laboratories); all of which increase the barriers for widespread adoption in preventative health-care settings.

In this work, we demonstrate the utility of our framework to train a new saliva-based DNAm biomarker for SCI (hereafter termed InflammAge). InflammAge addresses the hurdles outlined above for SCI quantification while demonstrating high technical reliability, significant associations with clinical outcomes and increasing biological interpretability when compared to other epigenetic clocks.

2 | RESULTS

2.1 | A novel framework to train biologically-interpretable and reproducible DNAm biomarkers in saliva

To harness the power of DNAm as a data type and saliva as an easily accessible human sample, we created a framework for developing novel saliva-based DNAm biomarkers informed by biological interpretability.

Genome-wide DNAm can be quantified using Illumina Infinium methylation microarrays. The most widely used technology is the Illumina EPIC array, which quantifies >850,000 CpG sites and was designed to enhance novel discovery. We aimed to create a more cost-effective approach to DNAm assessment at sites relevant to geroscience. To generate this custom array, termed the Hurdle DNAm platform, we selected CpG sites with strong phenotypic associations established by published epigenome-wide association studies (EWAS) (Battram et al., 2022; Li et al., 2019). We enriched for actionable traits, lifestyle and environmental exposures, epigenetic clocks and complex diseases (Tables S1 and S2); resulting in ~30,000 CpG sites that are good candidate features for biomarker development (see Section 4 and Figure S1). Genomic location analysis (see Section 4) revealed a higher proportion of probes in functional genomic sites (such as promoters, gene bodies, CpG islands and shores, and DNase Hypersensitivity Sites) in the Hurdle DNAm platform compared to all EPIC probes (Figure S2).

EpiScores are DNAm-based predictors of varied traits with direct biological interpretability, such as blood protein levels (Gadd et al., 2022; Stevenson et al., 2020, 2021). As DNAm captures

cellular states upstream of gene regulation, these EpiScores may act as de-noised proxies for blood proteins, better capturing chronic disease risk. This may prevent confounding with short-term fluctuations in blood concentrations, for example, due to circadian rhythms or acute infection. We propose that these EpiScores can be transferred from blood (where they were originally trained) to saliva; and then used as improved biological features to train ML models able to predict complex phenotypes from DNAm data (while still allowing for biological interpretability).

Our DNAm biomarker discovery framework consists of the following steps (Figure 1a):

1. generate genome-wide DNA methylation data in all datasets;
2. calculate a library of published blood EpiScores as features in all datasets;
3. filter the calculated features in the training dataset by reproducibility across tissues using a reliability dataset with matched blood and saliva DNAm data;
4. filter the remaining features in the training dataset by technical reliability of each feature within the tissue (within blood/saliva), using technical replicates from a reliability dataset;
5. select final features in the training dataset based on biological criteria (e.g., association with physiological ageing, association with a disease process, etc);
6. train the final ML model on the desired phenotypic outcome (e.g., chronological age, all-cause mortality, time-to-disease, etc) using the selected features in the training dataset in order to generate the DNAm biomarker; and
7. validate the final DNAm biomarker in independent test datasets.

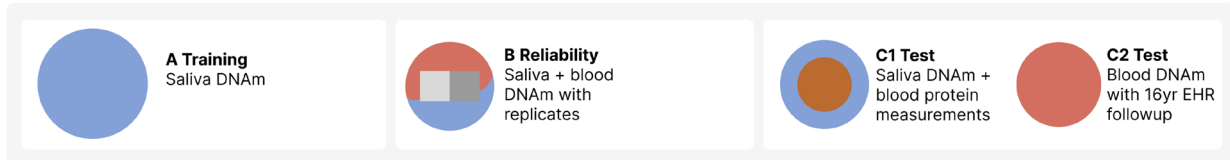
Next, we describe how this hypothesis-driven framework can be applied to build a novel saliva-based DNAm biomarker of systemic chronic inflammation (SCI) in humans (InflammAge).

2.2 | Applying the framework to train a saliva-based DNAm biomarker for systemic chronic inflammation (InflammAge)

Different proteins (including cytokines) have been associated with SCI, providing evidence that this ageing hallmark can be quantified at the molecular level in blood. We observed the age-related association of many of these biomarkers in one of our healthy ageing cohorts (see Alpha test dataset in Section 4), which included an increase with age in baseline levels of GDF-15 (Spearman correlation coefficient $R_s=0.76$, FDR-adj. p -value $p_{FDR}=3.8\times 10^{-12}$), CXCL9 ($R_s=0.67$, $p_{FDR}=6.6\times 10^{-9}$), TNF- α ($R_s=0.40$, $p_{FDR}=4.9\times 10^{-3}$), IL-6 ($R_s=0.36$, $p_{FDR}=0.013$), CRP ($R_s=0.33$, $p_{FDR}=0.027$), or CCL11 ($R_s=0.30$, $p_{FDR}=0.0498$) (Figure S3). Based on this, we collated a library of published blood DNAm CpG sites and EpiScores associated with cytokines and other inflammation-related proteins. This includes CRP (Ligthart et al., 2016; Stevenson et al., 2020), IL-6 (Stevenson et al., 2021), TNF- α (Aslibekyan et al., 2018), CXCL9,



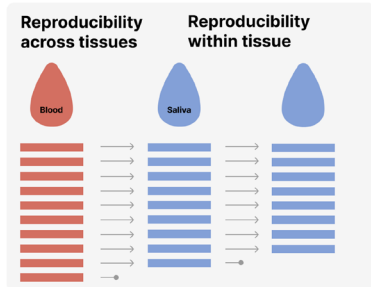
(a) Step 1 - generate or access DNA methylation data in all datasets



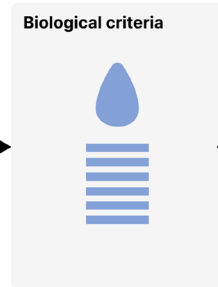
Step 2 - calculate EpiScores in all datasets



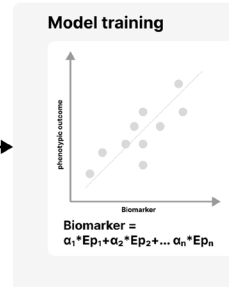
Step 3 & 4 - filter EpiScores for reproducibility across and within tissues



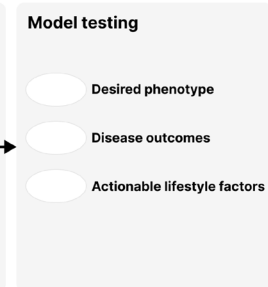
Step 5 - select features for biological criteria



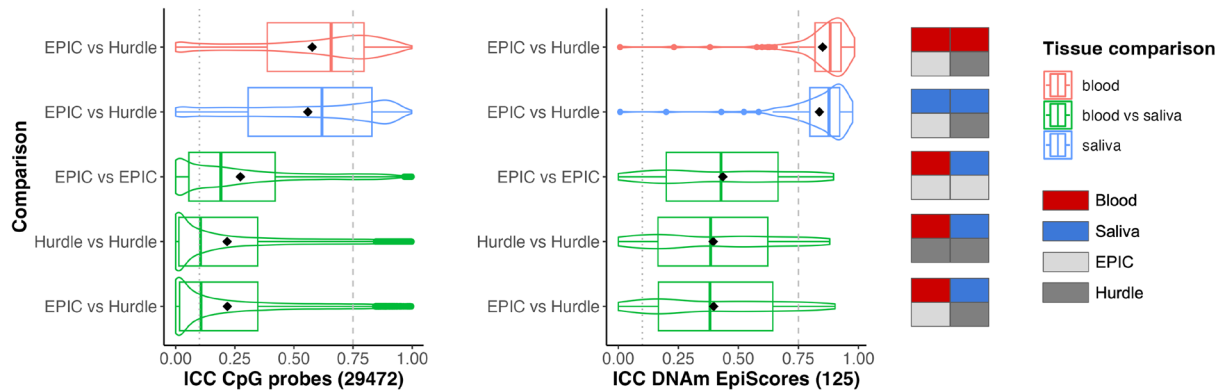
Step 6 - train ML model on desired outcome



Step 7 - test performance in independent datasets



(b)



(c)

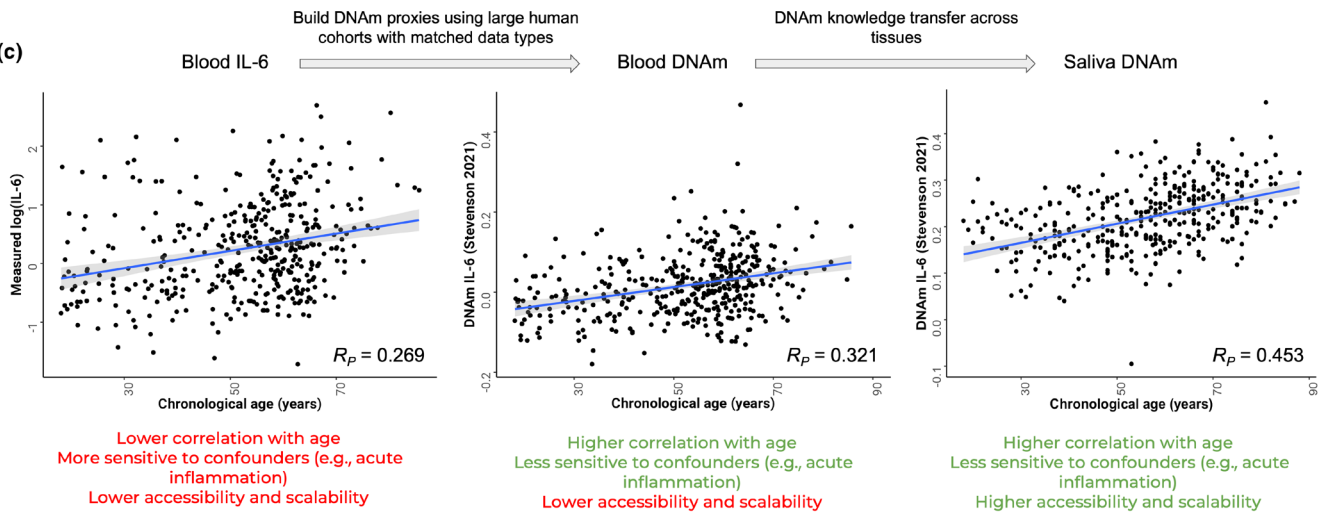




FIGURE 1 A novel framework to train biologically-interpretable and reproducible DNAm biomarkers in saliva. (a) Framework steps (Aps et al., 2002; Aryee et al., 2014; Aslibekyan et al., 2018; Battram et al., 2022; Belsky et al., 2022; Bocklandt et al., 2011; Costantini et al., 2018) and datasets (training, reliability, test) required to train novel DNAm biomarkers in saliva. CRP, C-reactive protein; EHR, electronic health records; IL-6, Interleukin 6. (b) Reproducibility analysis for CpG probes (left) and EpiScore (right) features considered during training of InflammAge. Distributions of intraclass correlation coefficients (ICC) from the reliability dataset (V1 samples) are shown as box and violin plots for five comparisons (within/between blood and saliva); interquartile range = squares, mean = black diamond. Grey vertical lines indicate ICC filtering thresholds (0.75 for within and 0.10 between tissues). (c) Transfer of biological information from blood proteins (left, data: Stevenson et al., 2021), to blood DNA methylation EpiScores (center, data: Stevenson et al., 2021), to saliva DNAm EpiScores (right, data: Saliva training dataset). For InflammAge, saliva DNAm SCI features selected based on the association with ageing (chronological age) included the SCI biomarker IL-6 which increases with age.

CCL11, IL-18R1 (Hillary et al., 2020), and 109 EpiScores containing inflammatory proteins measured in the Olink and SOMAScan platforms (Gadd et al., 2022). These features were calculated across all DNAm samples in all datasets (training, reliability, test; step 2 of the framework).

To ensure the reproducibility of these features, we generated a reliability dataset with DNAm technical replicates for matched saliva and blood samples by measuring methylation profiles for each tissue on both the EPIC array and the Hurdle DNAm platform. We calculated intraclass correlation coefficients (ICCs) between individual CpG probes present on both platforms (29,472) and for each summary EpiScore (125 in total) for the following comparisons: blood EPIC versus blood Hurdle, saliva EPIC versus saliva Hurdle, blood EPIC versus saliva EPIC, blood Hurdle versus saliva Hurdle, and blood EPIC versus saliva Hurdle (Figure 1b). Using this approach, we assessed which CpG probes and EpiScores could be measured reliably across platforms and tissues. We observed that EpiScores showed higher reliability both within tissues (mean values of $ICC_{\text{Blood}} = 0.85$, 95% confidence interval [0.76–0.90]; $ICC_{\text{Saliva}} = 0.84$ [0.74–0.90]) and between tissues measured on the same platform (mean $ICC_{\text{Blood-vs-Saliva}} = 0.42$ [0.13–0.61]) compared to means of individual CpG probes ($ICC_{\text{Blood}} = 0.58$ [0.42–0.69]; $ICC_{\text{Saliva}} = 0.56$ [0.40–0.68]; $ICC_{\text{Blood-vs-Saliva}} = 0.25$ [0.03–0.42]). Furthermore, many EpiScores showed strong positive correlations between blood EPIC and saliva Hurdle replicates collected from the same individual at the same time (Figure S4a), adding robust evidence to the ability to transfer EpiScores information from standard blood DNAm data to saliva DNAm data generated using the novel Hurdle platform. We therefore decided to train InflammAge using the EpiScores instead of the individual CpGs that comprise the EpiScores.

Using the V1 samples from the reliability dataset (see Appendix S1), different ICC thresholds for the EpiScores were benchmarked to optimise the ICC for the final InflammAge biomarker (Figure S5). This resulted in selecting EpiScores with $ICC > 0.10$ between blood and saliva (step 3 of the framework; 101/125 EpiScores remained) and an $ICC > 0.75$ within saliva and within blood replicates (step 4 of the framework; 86/125 EpiScores remained). The 86 EpiScores with high reliability were used downstream for training.

To train our InflammAge biomarker in saliva, we assembled a cohort with human saliva DNAm samples from published and in-house data (training dataset, see Appendix S1). This dataset ($N = 338$,

27% female) covers the average adult human lifespan (age range of 18–88 years, median 58.5 years, mean = 56.8 years), including individuals from diverse genetic backgrounds (Table S3). In this cohort, we calculated the correlation (Pearson correlation coefficient, R_p) of each EpiScore with chronological age in saliva (see Figure 2a for top 15 positively and negatively correlated and Figure S6 for all EpiScores). The biological assumption is that if the EpiScore for an inflammatory protein increases or decreases in the human population (i.e., cross-sectionally) with age, then it is more likely to be a marker for SCI. We kept those features with associations with age in saliva that had $p_{FDR} < 0.05$ (step 5 of the framework, see Figure 1c).

This retained 63/125 EpiScores (dark grey in Figure S6); with 42 positively and 21 negatively correlated with age. EpiScores with positive correlation with age in saliva included well characterised pro-inflammatory markers like TNFRSF1B ($R_p = 0.56$, $p_{FDR} = 4.8 \times 10^{-27}$), CCL18 ($R_p = 0.56$, $p_{FDR} = 4.8 \times 10^{-27}$), MMP12 ($R_p = 0.55$, $p_{FDR} = 6.8 \times 10^{-27}$), CXCL10 ($R_p = 0.55$, $p_{FDR} = 2.8 \times 10^{-26}$), or IL-6 ($R_p = 0.43$, $p_{FDR} = 1.7 \times 10^{-15}$) (see Section 3). EpiScores with negative correlation with age in saliva included NOTCH1 ($R_p = -0.43$, $p_{FDR} = 2.1 \times 10^{-15}$), BMP1 ($R_p = -0.36$, $p_{FDR} = 4.3 \times 10^{-11}$), FAP ($R_p = -0.31$, $p_{FDR} = 1.9 \times 10^{-8}$) and NTRK3 ($R_p = -0.31$, $p_{FDR} = 3.9 \times 10^{-8}$). Positively age-associated EpiScore genes were significantly enriched for inflammatory processes, cytokine activity, and response to stress in a gene ontology analysis (Figure S7a). Negatively age-associated EpiScore genes were significantly associated with serine peptidase activity, humoral immune, and biotic stimulus response (Figure S7b). These results give confidence that the proteins quantified by the selected EpiScores are enriched for inflammatory and immune response pathways.

Next, we trained our DNA methylation biomarker (InflammAge) in our saliva training dataset using an elastic net algorithm (step 6 of the framework). We selected the 63 DNAm features enriched for inflammation-related pathways as the input and chronological age as the outcome. In the training dataset, InflammAge predicted chronological age with a performance in range with gold standard epigenetic clocks ($R_p = 0.93$, median absolute error MAE = 3.83 years, Figure 2b). A total of 42 DNAm features (containing a total of 3788 unique CpG sites in the Hurdle DNAm platform) constitute the final InflammAge model. We calculated age-adjusted InflammAge acceleration (InflammAgeAccel), which represents the difference between measured InflammAge and the average InflammAge expected for someone with the chronological

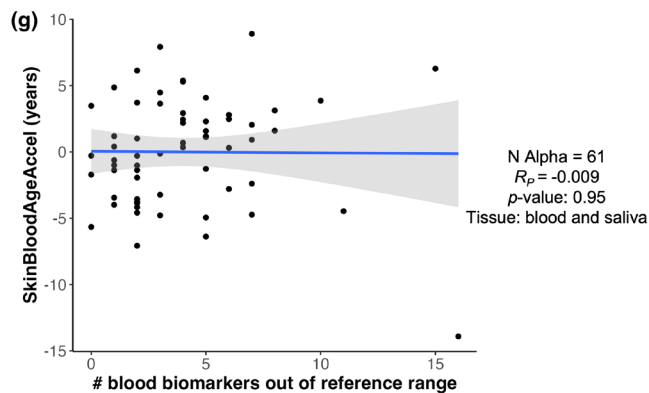
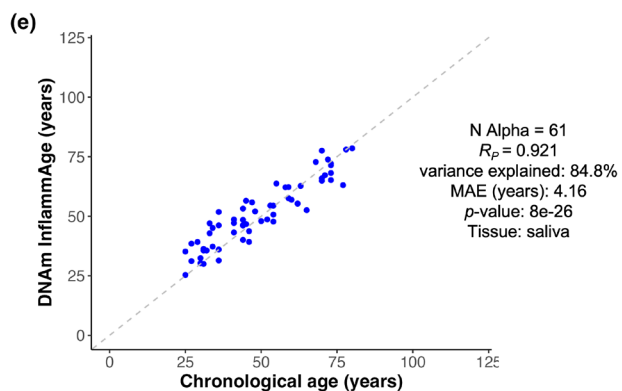
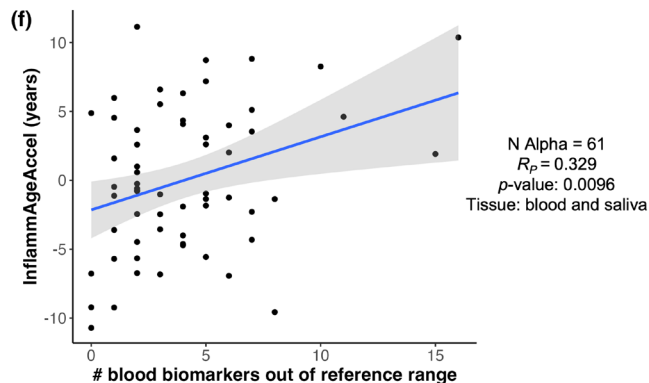
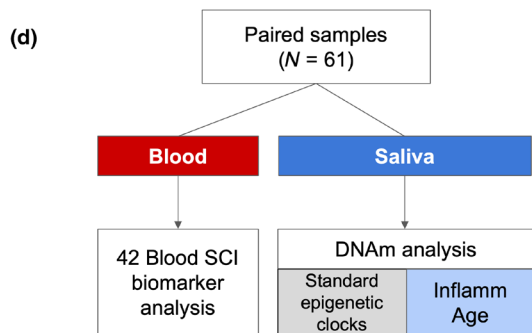
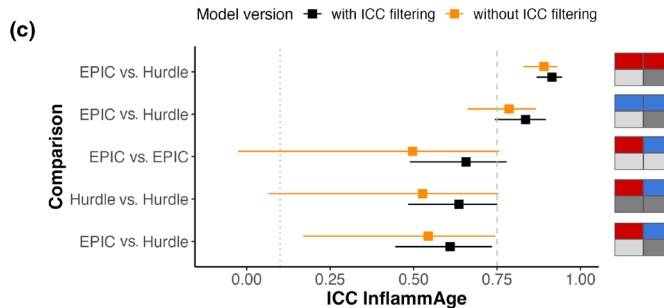
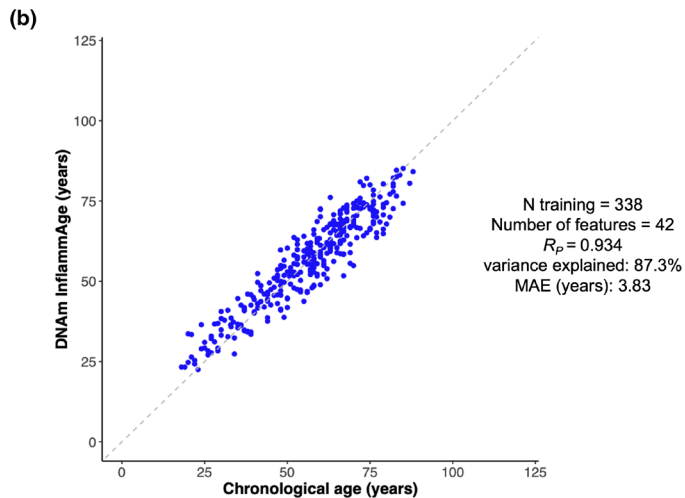
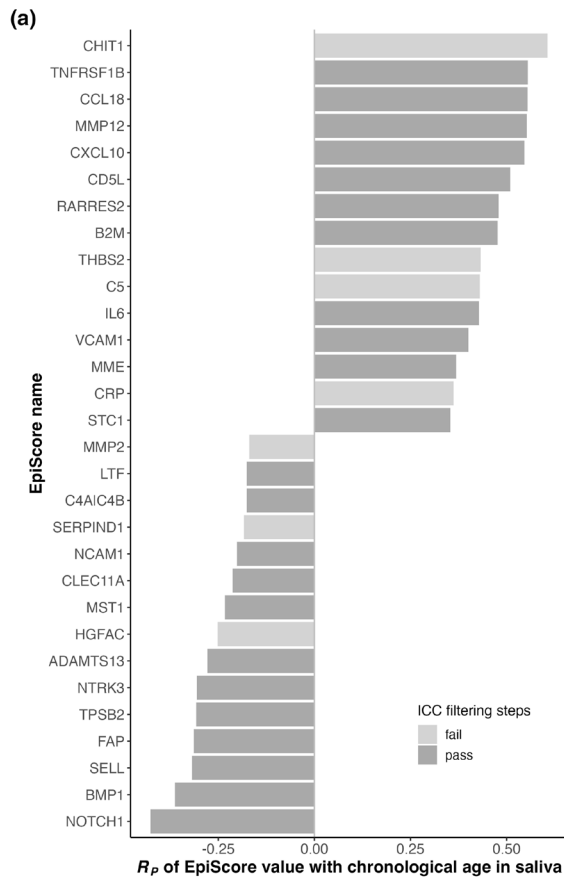




FIGURE 2 Train and test results of a saliva-based DNAm biomarker for systemic chronic inflammation (InflammAge). (a) Barplot of top 15 positive and negative SCI DNAm EpiScore features significantly associated with chronological age (FDR-adj. p -value <0.05) in the saliva training dataset, see [Figure S6](#) for the full list. Features passing all technical filtering steps (dark-grey) were used to train InflammAge. R_p , Pearson correlation coefficient. (b) Scatterplot of InflammAge performance (training dataset). (c) Reproducibility of InflammAge within and between tissues measured by intraclass correlation coefficient (ICC) in the reliability dataset (V2 samples), squares = mean, bars = 95% confidence interval. Values in black show improvement after filtering by ICC before training (framework steps 3 and 4). Orange = Results for a model without ICC-based feature filtering. Red = blood, blue = saliva; light-grey = EPIC, dark-grey = Hurdle DNAm platform). (d–g) Testing InflammAge performance in the Alpha test dataset. (d) Cohort design ($N=61$ individuals) with matched saliva DNAm and gold standard blood SCI biomarkers. (e) Scatterplot of InflammAge performance in the Alpha test dataset. (f, g) Scatterplots of association between SCI status in blood (quantified as number of SCI blood biomarkers outside of normal reference ranges in an individual) with (f) age-adjusted InflammAge acceleration (InflammAgeAccel) and (g) age-adjusted acceleration in the skin-blood clock (SkinBloodAgeAccel) (Horvath et al., 2018). Blue lines = linear model fit, grey shading = confidence interval.

age of the measured individual. A positive (>0) InflammAge acceleration should be associated with higher SCI when compared to other individuals of the population with the same chronological age (i.e., that person should be “inflammageing” faster). Saliva comprises a mix of immune cells, epithelial cells, and fibroblasts, and these proportions may change during ageing or disease (e.g., oral disease) (Aps et al., 2002). To study how the InflammAge biomarker may be affected by changes in saliva cell composition, we assessed the association of InflammAgeAccel with DNAm-predicted cell type proportions (see Section 4). All these associations are non-significant ([Figure S8](#)), which highlights that a higher InflammAgeAccel is not mainly driven by changes in cell composition. Of note, other gold standard blood- and buccal cell-trained epigenetic clocks (skin-blood clock (Horvath, 2013), multi-tissue Horvath, Hannum, paediatric clock (PedBE, McEwen et al., 2019), PhenoAge, PCGrimAge (Higgins-Chen et al., 2022), DunedinPACE) exhibited strong associations between age-adjusted epigenetic age acceleration and the proportions of different cell types in saliva ([Figure S9](#)).

We then tested the performance of InflammAge in two independent datasets (step 7 of the framework). In the reliability dataset (V2 samples), reliability of InflammAge was good in saliva (ICC = 0.84 [0.74–0.90]) and excellent in blood (ICC = 0.91 [0.87–0.95]) ([Figure 2c](#)), showing that saliva DNAm biomarkers are reproducible when trained using our framework. Moderate reliability (ICC = 0.61 [0.44–0.73]; for saliva Hurdle vs. blood EPIC) between tissues was also observed, which highlights that even though InflammAge was trained in saliva DNAm samples, the results can be extrapolated to blood DNAm samples assessed on alternative platforms ([Figures S4](#) and [S5](#)). Importantly, training an alternative InflammAge model using all the features associated with age as input but without filtering for ICC (i.e., skipping steps 3 and 4 in the framework) resulted in a less reliable biomarker with wide confidence intervals across all comparisons ([Figure 2c](#)). This validates the importance of the reliability dataset and steps 3 and 4 in our framework.

Finally, we tested the performance of InflammAge in the Alpha test dataset (see [Appendix S1](#)). This cohort was recruited for this work ($N=61$, 46% female) and also broadly covers the average adult human lifespan (age range of 25–80 years, median = 46 years, mean = 49.6 years), including only self-reported healthy individuals.

Importantly, this dataset included matched saliva DNAm and blood SCI biomarkers ([Figure 2d](#)), which allows for assessing whether InflammAge is capturing changes in the SCI status of the individuals (besides tracking chronological age). InflammAge-predicted chronological age with a correlation (R_p) of 0.94 and MAE of 4.2 years ([Figure 2e](#)). Importantly, InflammAgeAccel was associated with the inflammatory biomarker CRP ($R_p=0.30$, $p=0.02$, [Figure S10](#)) and also associated with the total number of blood SCI biomarkers outside of reference ranges ($R_p=0.33$, $p=0.01$, [Figure 2f](#)). On the contrary, age-adjusted epigenetic age acceleration of another saliva-based epigenetic clock (skin-blood clock, Horvath et al., 2018), was not correlated ($R_p=-0.01$, $p=0.95$, [Figure 2g](#)). We further benchmarked performance against other gold standard epigenetic clocks including multi-tissue Horvath, blood-trained Hannum, PhenoAge, PCGrimAge, DunedinPACE, and the buccal-swab trained PedBE clock ([Figure S11](#) and [Table S4](#)). Only PhenoAgeAccel showed a significant and positive correlation with the number of blood SCI biomarkers ($R_p=0.26$, $p=0.04$) in the same direction as InflammAgeAccel, although with a smaller effect size ([Figure S11f](#)). This demonstrates the unique ability of InflammAge to capture SCI status from a saliva DNAm sample.

2.3 | InflammAge is accelerated in SCI-related clinical outcomes

Since saliva DNAm cohorts with mapped clinical outcomes are rare, we tested InflammAge in Generation Scotland (GS), one of the world's largest biobanks with access to blood DNAm samples linked to electronic health records ([Table S5](#)). GS has blood DNAm samples from 18,865 individuals (58.8% female) covering the adult human lifespan (age range of 17–99 years, median = 49.2 years, mean = 47.8 years).

InflammAge calculated in the GS dataset showed a strong correlation with chronological age, although an intercept shift was observed (i.e., a systematic underestimation of InflammAge in blood when compared to saliva DNAm samples; $R_p=0.904$, MAE = 10.42 years; [Figure S12](#)). We decided to proceed to use GS as a test dataset using InflammAgeAccel, which as mentioned before is age-adjusted and would correct for such a shift in the downstream analyses.

We tested the Spearman correlation coefficients (R_s) between InflammAgeAccel and basic demographic and biochemistry



variables (Table S6). Despite sex chromosome probes having been removed from all datasets in this study, InflammAgeAccel showed a significant difference when stratified by sex, where males had a mean InflammAgeAccel of 0.198 compared to -0.139 years in females (t test, $p = 3.1 \times 10^{-4}$). This is consistent with reports in the literature that highlight higher SCI baseline levels in males (Martínez de Toda et al., 2023). InflammAgeAccel (from samples taken between 2006 and 2011) was also associated with positive COVID-19 disease diagnosis, ascertained from primary and secondary care records (hazard ratio $HR = 1.24$, 95% confidence interval [1.1–1.4], $p = 3.1 \times 10^{-4}$).

Next, we screened for an association between InflammAgeAccel and 308 diseases that were annotated using disease codes from Kuan et al. (2023). This was performed using Cox proportional hazards regression analyses that included demographic risk factors (age, sex, BMI, years of education, socioeconomic status, smoking pack-years, and DNAm processing batch) as covariates. This allows us to assess if InflammAgeAccel adds significant value when modelling the risk of disease when compared with traditional risk factors. Figure 3a summarises results for 36 diseases with >30 cases that showed significant association across three biomarkers (InflammAgeAccel, SkinBloodAgeAccel and blood CRP—a gold standard inflammatory biomarker) with the best performing one highlighted per disease.

InflammAgeAccel was associated with the incidence of 18 diseases after multiple testing correction and without violation of the model assumptions (shown as filled circles in Figure 3a); 33 diseases were significant when using unadjusted p -values (Table S7). The 18 hits included strong associations with the risk of developing diseases where SCI plays an important role in the aetiology, such as cancer (e.g., non-Hodgkin's lymphoma, $HR = 1.47$ [1.20–1.81], $p_{FDR} = 5.78 \times 10^{-4}$), diseases of the circulatory system (e.g., peripheral arterial disease, $HR = 1.24$ [1.09–1.41], $p_{FDR} = 3.37 \times 10^{-3}$), diseases of the digestive system (e.g., cholecystitis, $HR = 1.25$ [1.13–1.40], $p_{FDR} = 1.01 \times 10^{-4}$), and diseases of the respiratory system (e.g., COPD, $HR = 1.23$ [1.13–1.35], $p_{FDR} = 2.06 \times 10^{-5}$). InflammAgeAccel outperformed CRP in 15 out of these 18 diseases (based on a comparison of FDR-adj. p -values). Furthermore, InflammAgeAccel outperformed the only widely available saliva-based ageing DNA methylation biomarker (skin-blood clock, SkinBloodAgeAccel) in all 18 diseases. CRP generally outperformed InflammAgeAccel in the “diseases of the circulatory system” category, and skin-blood was mainly predictive for some cancer types. This confirms that InflammAge captures SCI-related disease outcomes from an accessible tissue.

Covariate-adjusted cox proportional hazards regression models were performed to assess the relationship between InflammAgeAccel and all-cause mortality. This was assessed at intervals of 2 years ($N = 66$ events), 5 years ($N = 241$ events), 10 years ($N = 683$ events), and the entire follow-up period (up to 16 years, $N = 1031$ events). Except for the 2-year interval, there was a significant relationship between InflammAgeAccel and all-cause mortality (Figure 3b). The performance of InflammAgeAccel was comparable to that of blood-based CRP measurements and outperformed the skin-blood clock when predicting mortality

within 5 years ($HR_{CRP} = 1.29$ [1.14–1.45], $HR_{InflammAgeAccel} = 1.23$ [1.11–1.37], $HR_{SkinBloodAgeAccel} = 1.15$ [1.05–1.27]), within 10 years ($HR_{CRP} = 1.22$ [1.14–1.32], $HR_{InflammAgeAccel} = 1.20$ [1.13–1.29], $HR_{SkinBloodAgeAccel} = 1.04$ [0.97–1.12]), and the entire follow-up period (up to 16 years: $HR_{CRP} = 1.18$ [1.11–1.25], $HR_{InflammAgeAccel} = 1.18$ [1.12–1.24], $HR_{SkinBloodAgeAccel} = 1.07$ [1.01–1.13]). When compared to blood-based epigenetic clocks, InflammAgeAccel performed in the range of PhenoAgeAccel, while GrimAgeAccel and DunedinPACE outperformed all other predictors (Figure S13). This is expected since InflammAge was not trained specifically to predict all-cause mortality, but rather to capture the SCI component of human ageing.

2.4 | InflammAge acceleration is associated with lifestyle factors

Lifestyle factors can have a profound influence on human healthspan, lifespan, and SCI status (Furman et al., 2019). We tested the association of InflammAgeAccel with lifestyle factors such as self-reported smoking status, alcohol consumption, physical activity, and dietary variables in the GS cohort.

Smoking status was significantly associated with InflammAgeAccel (ANOVA $p = 1.1 \times 10^{-18}$) in a dose-response manner, with non-smokers showing the lowest acceleration (ex-smoker vs. non-smoker: Diff = -0.70 [-0.95 to 0.44], Tukey adj. p -value $p_T = 3.2 \times 10^{-8}$; smoker vs. non-smoker: Diff = 1.03 [0.73 – 1.33], $p_T = 3.2 \times 10^{-8}$) (Figure 4a). For alcohol consumption, InflammAgeAccel was higher in heavy (≥ 14 units/week) compared to moderate drinkers (< 14 units/week; Diff = 0.62 [0.30 – 0.95], $p_T = 1.8 \times 10^{-5}$), while lower in moderate compared to non-drinkers (Diff = -0.48 [-0.83 to 0.12], $p_T = 4.4 \times 10^{-3}$) (Figure 4b).

Self-reported physical activity was not associated with InflammAge acceleration ($R_5 = -0.01$, $p = 0.16$). This was unexpected, given the links between exercise and reductions in systemic chronic inflammation (Furman et al., 2019). However, this could be due to the nature of self-reported exercise data, which tends to be less accurate than wearable-based data.

When analysing the association of dietary consumption patterns with InflammAgeAccel, we ran linear models adjusted for the same confounding factors used in the disease outcome analysis (Table 1). An intake of oily fish at any frequency was significantly associated with lower InflammAgeAccel compared to never having oily fish (Effect size = -0.31 , standard error = 0.13 , $p = 0.02$). Intake of other dietary constituents at any frequency compared to never showed no significant difference when adjusted for confounding factors. However, individual dose groups showed significance in the unadjusted ANOVA model (see Table S8). This included lower InflammAgeAccel for individuals with a higher intake of brown bread (5–6 days/week vs. never: Diff = -0.94 [-1.66 to 0.23], $p_T = 2 \times 10^{-3}$), higher fruit (daily vs. once per week: Diff = -0.84 [-1.51 to 0.17], $p_T = 4.4 \times 10^{-3}$) and higher green vegetables intake (5–6 days/week vs. never: Diff = -1.08 [-2.00 to 0.17], $p_T = 8.8 \times 10^{-3}$). On the contrary, a higher intake of meat showed a trend towards higher

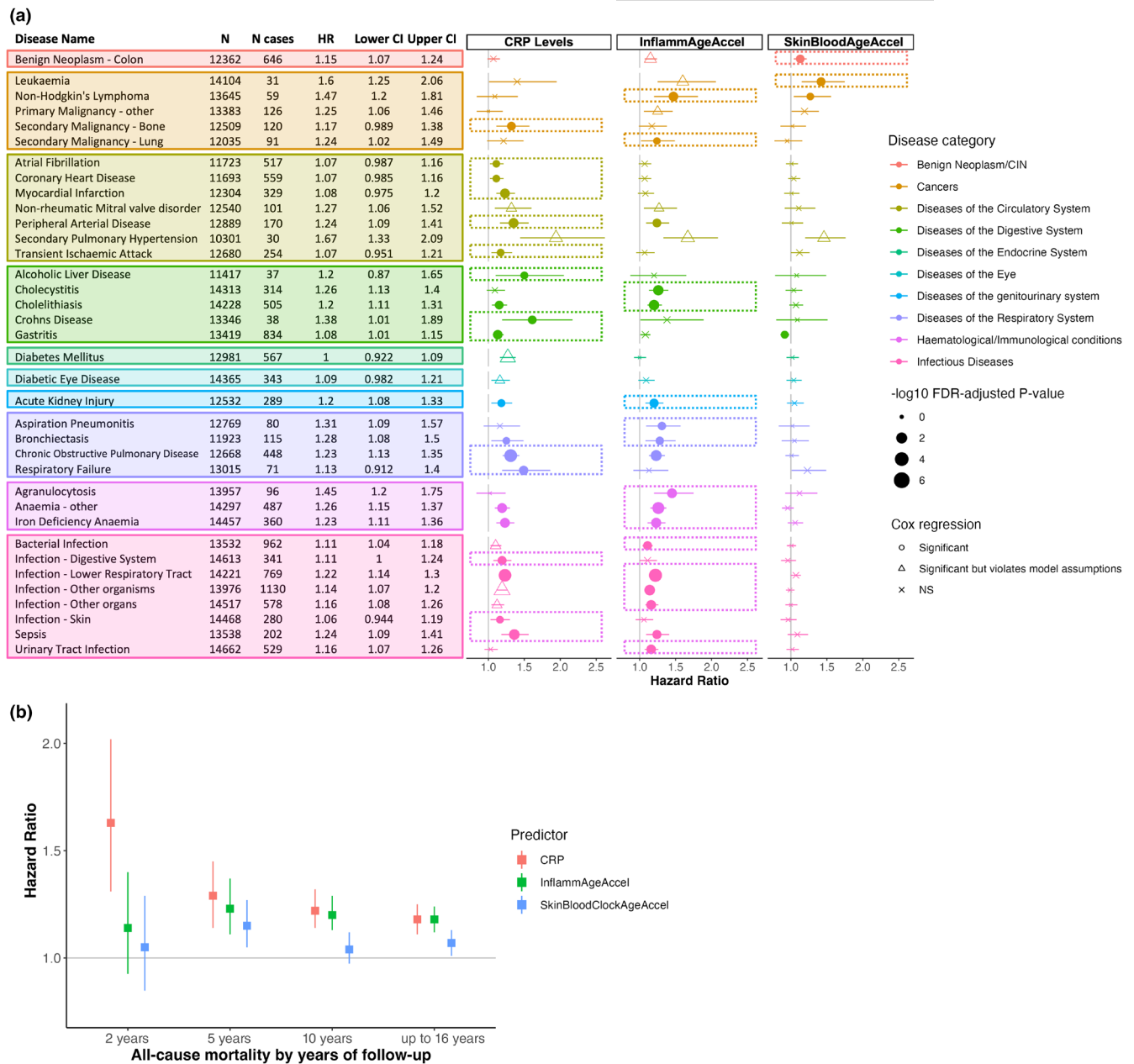


FIGURE 3 InflammAge acceleration performance in clinical endpoints compared to gold standard SCI markers in Generation Scotland. Incidence of clinical endpoint predicted for InflammAge acceleration (InflammAgeAccel), skin-blood clock age acceleration (SkinBloodAgeAccel) and CRP using covariate-adjusted Cox regression models. (a) Disease incidence Cox regression analysis. Displayed are 36 diseases with >30 cases in GS and association FDR-adj. p -value <0.05 for at least one predictor. The table shows hazard ratio (HR) and 95% confidence intervals (CI) per disease for InflammAgeAccel (full summary statistics in Table S6). The plot shows HR and 95% CI across all three predictors (best performed highlighted by dashed box). Colour=disease category (Kuan et al., 2023), shape=significance (NS, nonsignificant), filled=no model assumption violations. Point size is proportional to $-\log_{10}$ of FDR-adj. p -value. (b) Mortality prediction in GS for 2 ($N=66$ events), 5 ($N=241$ events), 10 years ($N=683$ events) and the entire follow-up period ($N=1031$ events). Squares=HR, lines=95% CI.

InflammAgeAccel (intake daily vs. < once per month: Diff=1.15 [0.17–2.12], $p_T=9.7 \times 10^{-3}$) although this was not observed across all dietary consumption amounts. These effects are consistent with reports for other blood-based epigenetic clocks (Levine et al., 2018; Lu et al., 2019; Quach et al., 2017).

Overall, our results suggest that InflammAge has potential to monitor the impact of lifestyle factors on SCI.

2.5 | InflammAge recapitulates immunosenescent cell type composition in blood

We performed cell type deconvolution in blood DNAm data from GS and calculated the correlation of InflammAgeAccel with the estimated proportions of each cell type. InflammAgeAccel was positively correlated with monocytes, neutrophils and

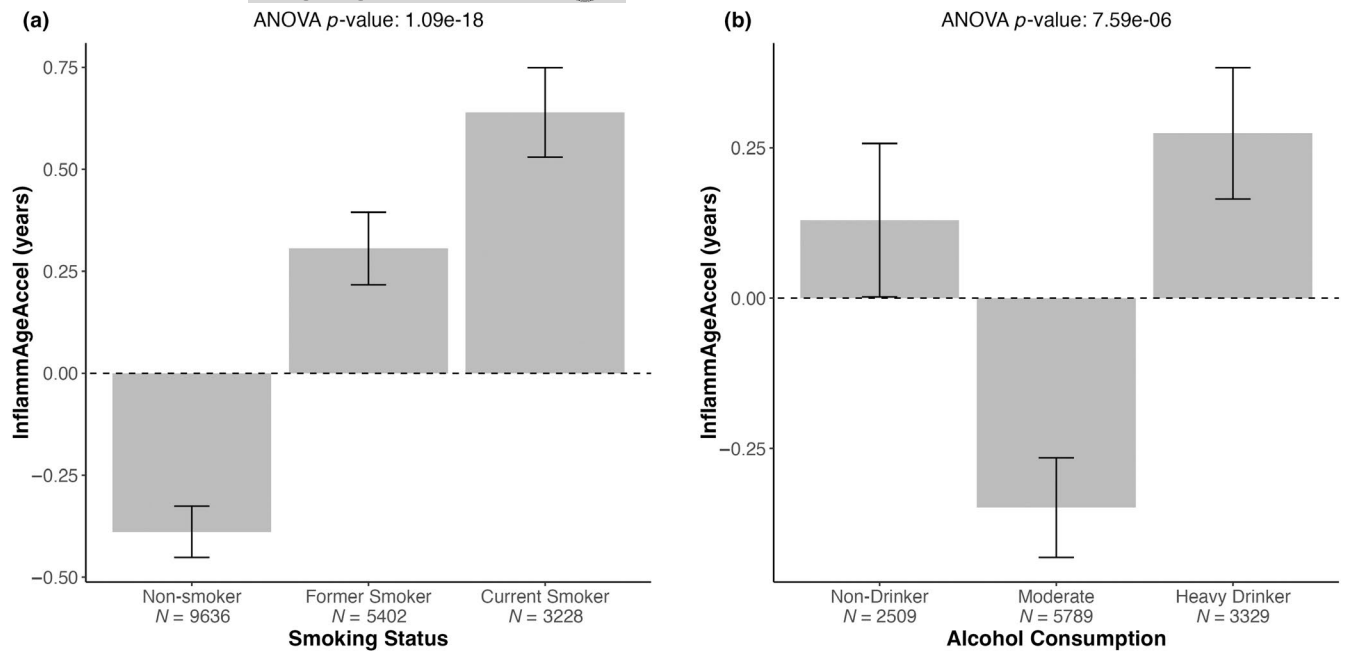


FIGURE 4 InflammAge acceleration (InflammAgeAccel) associations with smoking status and alcohol intake in GS. Barplot with ANOVA p -value for (a) self-reported smoking status, and (b) self-reported alcohol consumption in the week before the blood sample for DNAm measurement was taken. Plots show mean InflammAgeAccel per group, error bars, standard errors of the mean.

TABLE 1 Association of InflammAgeAccel with dietary variables in GS.

Dietary category	Effect size	Standard error	P-value
Oily fish	-0.305	0.134	0.023
Dairy	-1.017	0.664	0.126
Brown bread	-0.315	0.216	0.145
Other fish	0.232	0.213	0.275
Eggs	-0.223	0.358	0.534
Liver	-0.053	0.105	0.613
Green vegetables	-0.147	0.302	0.627
Poultry	0.120	0.269	0.656
Other meat	0.089	0.251	0.724
Fruit	-0.150	0.470	0.749
Other vegetables	-0.128	0.570	0.823

Note: Summary statistics for covariate-adjusted linear models testing association between InflammAgeAccel and dietary variables (ever versus never categories, collected at baseline during GS participant recruitment). Full summary statistics of all ANOVA comparisons in [Table S5](#).

CD8+CD28-CD45RA- T cells (memory/effector CD8+ T cells, Tomiyama et al., 2002). In contrast, InflammAgeAccel was negatively correlated with B cells, CD4+ T cells, total CD8+ T cells, and naïve CD8+ T cells. Eosinophils and NK cells showed almost no change with InflammAgeAccel (Figure 5). This is consistent with changes observed in the immune system during human ageing and immunosenescence (Saule et al., 2006). Thus, individuals that have positive InflammAgeAccel (InflammAge results higher than expected) display a more immunosenescent profile.

3 | DISCUSSION

Preventative healthcare and interventions that target ageing directly can have a profound health and economic impact on our societies. A slowdown in ageing that results in an increase in human life expectancy by 1 year is understood to be worth US\$38 trillion (Scott et al., 2021). This can be transformational for healthcare systems and patients who currently live carrying the burden of multi-morbidity. In order to understand the ageing process and the risk of developing age-related diseases, accessible and actionable biomarkers need to be developed and made widely available.

In this work, we created a framework that facilitates the creation of novel saliva-based DNA methylation biomarkers. Saliva as an accessible and non-invasive sample type comprises ~80% immune cells and ~20% epithelial buccal cells and fibroblasts. This overlap in cell composition substantiates the strong correlations ($r \sim 0.97$) between saliva and blood DNAm patterns (Nishitani et al., 2023), and alongside the results of this study, supports the case for saliva DNAm as a clinically meaningful and reproducible data type to build biomarkers. In fact, the first ever-developed epigenetic clock, which explained 73% variance in chronological age using two CpG sites, was developed using saliva DNA methylation data (Bocklandt et al., 2011). Recent studies have also leveraged the saliva's non-invasiveness to develop the first DNA methylation age predictor specialised for paediatric samples (McEwen et al., 2019), or sought to adapt epigenetic clocks trained in blood for application in saliva by accounting for differences in cell-type composition (Galkin et al., 2021). Given that most biological data in humans, including DNAm data, has been generated in blood samples, the ability to transfer biological information from blood to saliva DNAm opens the door for preventative

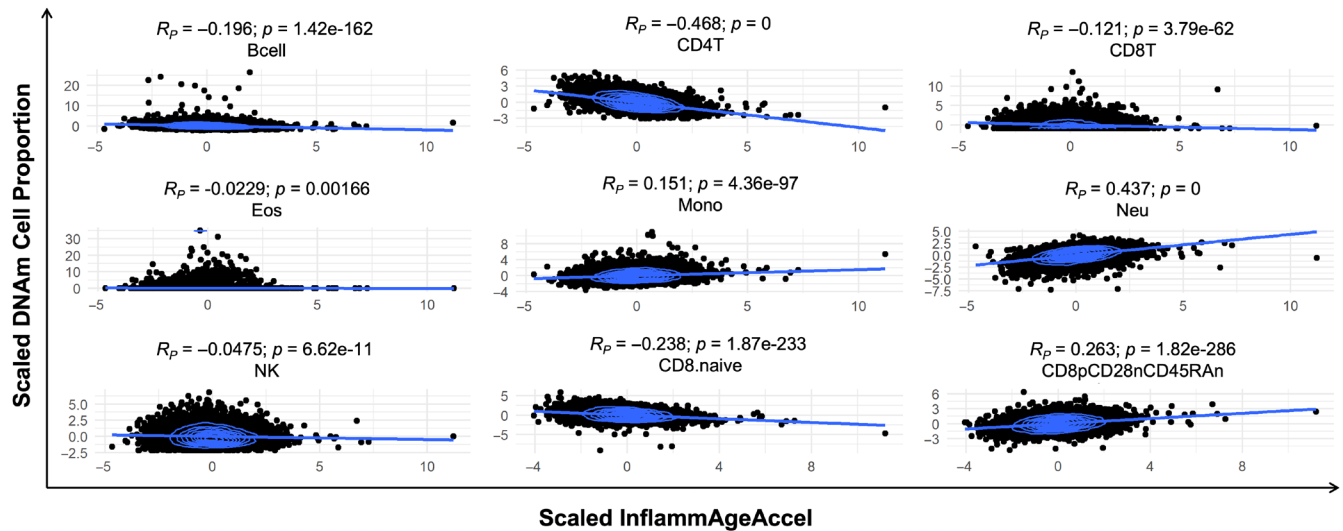


FIGURE 5 InflammAge acceleration (InflammAgeAccel) recapitulates immunosenescent cell type composition. Scatterplot of InflammAgeAccel and blood cell type proportions estimated from GS blood DNAm data (scaled and centred around zero with standard deviation = 1). Bcell, B-cell; CD4T, CD4+ T-cells; CD8.naive, naive CD8+ T-cells; CD8pCD28nCD45RAn, memory/effector CD8+ T-cells; CD8T, CD8+ T-cells; Eos, eosinophils; Mono, monocytes; Neu, neutrophils; NK, natural killer cells; p , p -value; R_p , pearson correlation coefficient.

healthcare and population screening applications at scale, including at-home testing and reaching isolated populations.

We illustrated the value of our biomarker discovery framework by creating InflammAge, a novel saliva-based DNAm biomarker for SCI. InflammAge was built bottom-up by selecting DNAm features that are associated with chronic inflammatory markers, which allows for biological interpretability of the results. This may also benefit clinical trials to improve patient stratification since individuals may age differently, that is, have different “ageotypes” that personalised interventions should recognise. Using DNAm EpiScores that aggregate multiple CpG probes, we also benefit from dimensionality reduction as a de-noising mechanism, which has been reported to significantly improve the reliability of epigenetic clocks (Higgins-Chen et al., 2022).

During feature selection for InflammAge, we uncovered DNAm predictors of blood inflammatory proteins that change consistently during ageing. Importantly, many of these EpiScores were trained on protein levels adjusted for age, sex, cohort-specific variables, and potential genetic protein quantitative trait loci (pQTL) effects (Gadd et al., 2022); which requires careful interpretation when comparing them to direct protein concentration measurements and drawing biological conclusions from these features. Nevertheless, EpiScores that show age-associated increase include several examples of pro-inflammatory proteins, such as IL-6, a classic biomarker for inflammaging, TNFRSF1B, one of the receptors of TNF-alpha, or CXCL10, a cytokine induced by TNF- α and IFN- γ .

Monocytes and macrophages are considered key effector cellular populations for inflammaging. Traditionally, M1 macrophages are thought to be pro-inflammatory while M2 macrophages are anti-inflammatory and involved in tissue repair. Interestingly, M1/M2 macrophage populations change during human ageing (Costantini

et al., 2018) and we observed evidence that InflammAge may be identifying this phenomenon, for example, through an increase in the DNAm proxy for proteins like CCL18, which causes M2 macrophage maturation, or MMP12, a matrix metalloprotease secreted by M2 macrophages. Relatedly, NOTCH1, whose inhibition increases M2 macrophage polarisation (Wang et al., 2010), was the top EpiScore negatively associated with age. Notch signalling has been proposed as a target in chronic inflammatory diseases, and further studies investigating this phenomenon may benefit from the application of InflammAge and/or related EpiScores to enhance translation of findings.

Thanks to the rich electronic health record data available in GS we performed the most comprehensive screening to assess the ability of a new epigenetic clock to predict clinical outcomes. We demonstrated that InflammAgeAccel captures associations with all-cause mortality, disease outcomes, and lifestyle factors; in many cases outperforming CRP, a widely-used gold standard SCI biomarker. For most of the cancer outcomes, InflammAgeAccel showed a higher hazard ratio (e.g., for non-Hodgkin's lymphoma (NHL)) when compared to CRP. A meta-analysis of 17 studies identified that elevated levels of several inflammatory biomarkers (such as TNF- α) are associated with a higher risk of developing NHL, potentially implicating them in the aetiology of the disease (Makgoeng et al., 2018). Albeit violation of some model assumptions, InflammAgeAccel improved results from CRP also for leukaemia. A recent single-cell multi-omics analysis identified a role for chronic inflammation as a driver of leukemic evolution (Rodríguez-Meira et al., 2023). InflammAgeAccel similarly outperformed CRP in the disease categories “haematological/immunological conditions” and “infectious diseases”. The latter could be reflective of the ability of InflammAge to quantify immunosenescence in blood and the link between an increase in SCI and



the decreased ability of the immune system to fight infection with age. While CRP was better at capturing all-cause mortality in the short term (2 years), InflammAgeAccel-matched CRP in the longer term (up to 16 years), perhaps indicating that InflammAge captures the medium to long-term (as opposed to acute) risk of disease.

This study has several limitations. First, the reliability ($N=83$) and saliva DNAm training ($N=338$) datasets were relatively small. As more saliva DNAm data from consenting individuals becomes available, our biomarker discovery framework and performance of new versions of InflammAge are likely to improve. Interestingly, unlike what we observed in blood, InflammAgeAccel was not significantly associated with cell composition in saliva in our analysis; which could be related to the fact that InflammAge was trained using blood DNAm proxies. These conclusions should be further tested in larger and independent datasets. Additionally, it would be important to study whether oral disease can confound the results for InflammAge in future studies. Furthermore, the current saliva DNAm datasets do not have linked mortality or disease outcome data, which precluded us from a saliva-only training and testing set-up, necessitating training on chronological age and testing in a large blood DNAm dataset like GS instead. A direct comparison against blood-trained epigenetic clocks (like PhenoAge, DunedinPACE, or GrimAge) should be performed against a blood-based InflammAge biomarker to avoid bias based on the training tissue type. Additionally, sex differences were observed for our biomarker, with the male population displaying a higher InflammAge acceleration. Given that organs and systems display sex-specific ageing processes (including immune system ageing), investigating this in more detail (or even training sex-specific biomarkers) is necessary. When looking at the proposed extension of the NIH-FDA Biomarkers EndpointS and other Tools (BEST) classification framework recently suggested by the “Biomarkers of Ageing Consortium” (Moqri et al., 2023), InflammAge is best suited to be classified as a predictive or prognostic biomarker. Understanding how InflammAge changes longitudinally and how responsive it is upon intervention is currently underway. For example, InflammAgeAccel showed an improvement for individuals with high SCI at baseline in a 3-month nutraceutical intervention study (McGee et al., 2024).

Future work will continue to explore if our saliva-based DNAm framework can be used to deconvolute the ageing process and quantify other hallmarks of ageing with accessible, biologically interpretable biomarkers. Ultimately, we envision a saliva-based and cost-effective DNAm test that is able to personalise follow-up diagnostics and interventions and that becomes the entry point for a preventative healthcare journey for the global asymptomatic population.

4 | MATERIALS AND METHODS

4.1 | Hurdle DNAm platform

We compiled CpG probes associated with relevant health and disease phenotypes from the EWAS catalogue (Battram et al., 2022)

and EWAS Atlas (Li et al., 2019) (accessed March 2022). After harmonisation and filtering for genome-wide significance ($p < 3.6 \times 10^{-8}$), a core set of high-confidence associations was formed by intersecting studies and including up to 1000 CpG probes per study (Figure S1a). Additionally, relevant literature reporting EWAS results for phenotypes of interest was manually curated, resulting in the inclusion of approximately 30,000 CpG probes from 212 publications (Figure S1b). These probes were deployed in a cost-effective Illumina methylation array termed the Hurdle DNAm platform.

4.2 | Genomic annotation of CpG sites

We annotated the genomic location of the Hurdle CpG sites using the manifest from the Illumina EPIC 850k array (mapped to hg19, Infinium MethylationEPIC v1.0 B5 Manifest file). We obtained gene distribution categories using the UCSC_RefGene_Group annotation column: Promoter <- c(“TSS200”, “TSS1500”), Gene_body <- c(“Body”, “UTR”, “Exon”), Intergenic (no annotation present); where a probe overlapped multiple genes in both the Promoter and Gene_body category, it was annotated as a Promoter in this analysis. We defined CpG Island distribution categories using the Relation_to_UCSC_CpG_Island annotation column: CpG_Islands <- c(“Island”), CpG_Shores <- c(“N_Shore”, “S_Shore”), “>2kb from Islands” (no or “Shelf” annotation). We defined DNase Hypersensitivity categories based on having a named DHS in the DNase_Hypersensitivity_NAME annotation column. We then plotted the percentage of CpG probes with an annotation in each category compared to all Hurdle or all EPIC probes.

4.3 | Dataset description

Construction and processing of the training, reliability, and test datasets is described in Appendix S1.

4.4 | DNA methylation data processing pipeline

This applies to all DNAm data presented in this manuscript except the GS cohort. Data processing was performed using RStudio (Version: 2023.09.1+494). IDAT files were processed using a custom pipeline leveraging the *minfi* (Aryee et al., 2014) and *SeSaMe* (Zhou et al., 2018) packages. The pipeline uses *noob* background correction and BMIQ within-array normalisation as previously described (Martin-Herranz et al., 2019). We removed cross-reactive, SNP- and sex chromosome-overlapping probes. Samples with overall call rates <70% or low mean intensity values were removed from analyses. We subset probes to a set of ~30K included in the Hurdle DNAm platform before or after normalisation as indicated. GS DNAm data was processed as described in Appendix S1.



4.5 | Reliability analyses

ICCs were calculated using a customised version of the *dupicc* function of the *ENmix* R package version 1.36.08 (Xu & Taylor, 2021) for a single-rater, absolute-agreement, two-way random-effects model (type=agreement, model=c(two-way), unit=single). Only probes present in both platforms (EPIC & Hurdle) after quality control were used for probe level ICC (29,472). V1 samples of the reliability dataset were used for ICC calculations at feature selection (steps 3+4 in the framework), and V2 samples for testing of the final InflammAge biomarker (step 7) to avoid data leakage. EpiScores and InflammAge calculated in each dataset were used as input for the *dupicc* function for feature and biomarker level ICC.

4.6 | Calculating EpiScores in saliva DNAm data

EpiScores were selected by literature review and calculated using coefficients and methods originally provided (Aslibekyan et al., 2018; Gadd et al., 2022; Hillary et al., 2020; Lighthart et al., 2016; Stevenson et al., 2020, 2021). For 109 EpiScores containing inflammatory proteins measured in the Olink and SOMAScan platforms (Gadd et al., 2022), the function *scale* (center=TRUE, scale=TRUE) from the *base* R package was used to normalise the beta-values across samples in the dataset before calculating the EpiScores.

4.7 | EpiScore gene enrichment analysis

Gene Ontology (GO)-term enrichment was performed using the *g:Profiler* platform (Kolberg et al., 2023). Gene names of EpiScore proteins selected after ICC filtering were analysed in two sets (positively/negatively associated with chronological age in saliva). All annotated genes were used as background for enrichment. Input gene names were ranked by using the absolute Pearson correlation coefficient with age in saliva.

4.8 | Training the InflammAge biomarker

Elastic net regression was used to train a model that would predict chronological age (*y_fit*, dependent/output variable) using the selected SCI-related DNAm features (*x_fit*, independent variables). We used 10-fold cross-validation and $\alpha=0.5$ as implemented in the *cv.glmnet* function in the *glmnet* R package (Friedman et al., 2010):

```
cvfit_model <- cv.glmnet(x_fit, y_fit, alpha = 0.5).
```

We calculated age-adjusted InflammAge acceleration (InflammAgeAccel) as the residuals from this linear regression model:

$$\text{DNAm_InflammAge} \sim \text{Chronological_age}.$$

4.9 | Saliva cell type composition analysis

Cell type composition of the saliva cohorts was estimated from DNA methylation data with the EpiDISH (Epigenetic Dissection of Intra-Sample Heterogeneity) R package, using the Robust Partial Correlations-RPC method (Teschendorff et al., 2017).

4.10 | Benchmarking against other epigenetic clocks

This applies to all epigenetic clocks calculated in this manuscript except for the GS cohort samples. We ran a suite of gold standard epigenetic clocks in our saliva training and testing cohorts to benchmark the performance of InflammAge. The epigenetic clocks were calculated using the *dnaMethyAge* R package (Wang, 2021; Wang et al., 2024). These included: SkinBlood: skin-blood clock (Horvath et al., 2018); Horvath: multi-tissue Horvath clock (Horvath, 2013); Hannum: Hannum clock (Hannum et al., 2013); PedBE: paediatric clock (McEwen et al., 2019); PhenoAge: PhenoAge clock (Levine et al., 2018); PCGrimAge: principal components GrimAge clock (Higgins-Chen et al., 2022); DunedinPACE: DunedinPACE clock (Belsky et al., 2022).

For analysis in the GS dataset, DNAm data was uploaded to the UCLA clock calculator (<https://dnamage.genetics.ucla.edu/home>), described in Horvath (2013) to calculate GrimAge and PhenoAge acceleration. Dunedin PACE was calculated using the PACEprojector function in the DunedinPACE R package (<https://github.com/danbelsky/DunedinPACE>).

4.11 | Disease association analysis in GS

A series of Cox proportional hazards regression models were run to relate InflammAgeAccel (residual of InflammAge ~ chronological age) to the risk of developing 308 diseases, as coded by Kuan et al. (2023). Additional covariates included age at blood draw, sex, log-transformed BMI, years of education, socioeconomic status (Scottish Index of Multiple Deprivation), smoking pack years, and DNAm processing batch. Where the proportion of females with an incident diagnosis for each disease was >90% (or <10%), we stratified the model to analyse females (males) only. Given some diseases more likely occur at certain periods of life (e.g., dementia after the age of 65 years), we calculated age of diagnosis for each disease, extracted age of cases at the 2.5 and 97.5 percentiles and filtered age at the event or censoring to fall within this window to retain a set of controls with similar profiles to the cases. Censoring was marked by being unaffected by the disease and either surviving to the latest point of linkage (April 2022) or time to death during follow-up. Regression assumptions were examined by tests of the Schoenfeld residuals, with p -value <0.05 indicating a model violation. These analyses were repeated with log-transformed measured CRP and other age-adjusted epigenetic age acceleration (SkinBloodAgeAccel, PhenoAgeAccel,



GrimAgeAccel, DunedinPACE) metrics for diseases with >30 cases (to preserve participant anonymity and reduce the risk of models not converging).

In addition to the diseases listed in Kuan et al. (2023), we examined the relationship between InflammAgeAccel and a COVID-19 disease diagnosis (see Appendix S1).

4.12 | Lifestyle factor association analysis in GS

Alcohol consumption was binned into non-drinker (0 units in the last week), moderate (<14 units) and heavy drinker groups (≥ 14 units) and filtered to individuals reporting intake as their usual weekly intake. ANOVA tests were performed on smoking and alcohol groups, followed by Post-hoc Tukey tests to assess two-group relationships. Physical activity reported as metabolic equivalents (MET) minutes was scaled and Spearman correlation with scaled InflammAge acceleration was calculated.

Dietary intake data included self-reported information on frequency of brown bread, fruit, green vegetables, other vegetables, liver, poultry, other meat (non-poultry), eggs, dairy, oily fish, and other fish (non-oily). Two types of analyses were performed for dietary intake data:

1. An “ever/never” diet variable was derived per multi-level diet variable. Effect sizes and *p*-values were obtained from linear models adjusted with covariates (Table 1):

$$\text{scale(InflammAgeAccel)} \sim \text{age} + \text{sex} + \log(\text{BMI}) + \text{education_years} \\ + \text{scale(deprivation_rank)} + \text{smoking_status} \\ + \text{DNAm_processing_batch} + \text{diet_variable}$$

2. To test associations between diet levels (doses), ANOVA and post-hoc Tukey tests were performed as described (Table S7).

4.13 | Cell type composition analysis in GS

White blood cell proportions were estimated using meffil's implementation of the Houseman algorithm (Houseman et al., 2012) with Reinius et al.'s peripheral blood reference data (Reinius et al., 2012). Additional measures (CD8pCD28nCD45RAn) were obtained from the Horvath clock output (<https://dnamage.genetics.ucla.edu/home>, Horvath, 2013). Blood cell proportions were scaled to mean 0 and standard deviation 1 using the *scale* function. Correlation between scaled InflammAgeAccel and scaled cell proportions was calculated for each cell type using Pearson correlation coefficient. Figures were plotted using *ggplot2* in R.

AUTHOR CONTRIBUTIONS

LJS, TPC, DLM, HJ, EZ, and DEMH carried out analyses. LJS, TPC, TMS, and DEMH devised the study and secured funding. KMG, JS,

AC, AMM, TJ, and JML performed study design and recruitment for the reliability dataset and contributed data. VJ, DK, NT, ES, MG, RT, and the Hurdle bio-infrastructure team performed and supervised tech work and logistics. WJH, VO, KW, CEB, and REM advised on the experimental and InflammAge validation strategy. WJH assisted in the generation of manuscript figures. LJS, TPC, HJ, TMS, and DEMH wrote the first draft and all authors revised and approved the manuscript.

ACKNOWLEDGEMENTS

We thank all study participants for their involvement. We further thank the following individuals for their support with this study: Ann Jordan, Evgenia Belova, Pavitra Chavan, Orla Codyre, Debora Crimella, Domenico Musto, Nabil El Nouby, Joanne Millar, Swetha Suresh, Harvey Watts, Ruth Atherton, Roughan Sheedy, Gordon McCulloch, Charles Ball, Felice Leung, and Richard Rilstone. Generation Scotland received core support from the Chief Scientist Office of the Scottish Government Health Directorates [CZD/16/6] and the Scottish Funding Council [HR03006] and is currently supported by the Wellcome Trust [216767/Z/19/Z]. Genotyping of the GS:SFHS samples was carried out by the Genetics Core Laboratory at the Edinburgh Clinical Research Facility, University of Edinburgh, Scotland and was funded by the Medical Research Council UK and the Wellcome Trust (Wellcome Trust Strategic Award “STratifying Resilience and Depression Longitudinally” (STRADL) Reference 104036/Z/14/Z). The GS DNA methylation profiling and analysis was supported by Wellcome Investigator Award 220857/Z/20/Z and Grant 104036/Z/14/Z (PI: AM McIntosh) and through funding from NARSAD (Ref: 27404; awardee: Dr. DM Howard) and the Royal College of Physicians of Edinburgh (Sim Fellowship; Awardee: Dr. HC Whalley).

CONFLICT OF INTEREST STATEMENT

LJS, HJ, VJ, DK, NT, EZ, ES, MG, and the Hurdle bio-infrastructure team are employees and TPC, RT, TMS, and DEMH are co-founders and shareholders at Chronomics Ltd., a company focused on multi-omics and epigenetic biomarker development and diagnostic infrastructure. VO and KW are employees at Bayer Consumer Care AG. CEB is an employee at Bayer HealthCare LLC. JML and WJH declare their roles as consultants for Bayer Consumer Care AG. Both report receiving consultancy fees from Bayer Consumer Care AG during the preparation of the study. Bayer Consumer Care AG financially supported this work. REM acts as a scientific consultant for Optima Partners and is an advisor to the Epigenetic Clock Development Foundation. All other authors declare no competing interests.

DATA AVAILABILITY STATEMENT

Public training datasets (listed in Table S3) are available in GEO (Edgar et al., 2002) and ArrayExpress (Parkinson et al., 2007). Data from Hurdle's users or study participants cannot be made publicly available due to consent limitations, but requests for research collaboration can be submitted to the corresponding authors. Generation Scotland allows researchers to apply for access to their data through



their collaboration proposal process. Further details can be found at <https://www.ed.ac.uk/generation-scotland/for-researchers/access>. We invite academics to contact us directly about running InflammAge in their cohorts.

ORCID

Lisa J. Schmunk <https://orcid.org/0009-0009-0879-462X>

Waylon J. Hastings <https://orcid.org/0000-0003-0875-2568>

Janet M. Lord <https://orcid.org/0000-0003-1030-6786>

REFERENCES

- Aps, J. K., Van den Maagdenberg, K., Delanghe, J. R., & Martens, L. C. (2002). Flow cytometry as a new method to quantify the cellular content of human saliva and its relation to gingivitis. *Clinica Chimica Acta; International Journal of Clinical Chemistry*, 321(1–2), 35–41. [https://doi.org/10.1016/s0009-8981\(02\)00062-1](https://doi.org/10.1016/s0009-8981(02)00062-1)
- Aryee, M. J., Jaffe, A. E., Corrada-Bravo, H., Ladd-Acosta, C., Feinberg, A. P., Hansen, K. D., & Irizarry, R. A. (2014). Minfi: A flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. *Bioinformatics*, 30(10), 1363–1369. <https://doi.org/10.1093/bioinformatics/btu049>
- Aslibekyan, S., Agha, G., Colicino, E., Do AN, L. J., Ligthart, S., Marioni, R. E., Marzi, C., Mendelson, M. M., Tanaka, T., Wielscher, M., Absher, D. M., Ferrucci, L., Franco, O. H., Gieger, C., Grallert, H., Hernandez, D., Huan, T., Iurato, S., Joehanes, R., ... Arnett, D. K. (2018). Association of methylation signals with incident coronary heart disease in an epigenome-wide assessment of circulating tumor necrosis factor α . *JAMA Cardiology*, 3(6), 463–472. <https://doi.org/10.1001/jamacardio.2018.0510>
- Battram, T., Yousefi, P., Crawford, G., Prince, C., Babaei, M. S., Sharp, G. C., Hatcher, C., Vega-Salas, M. J., Khodabakhsh, S., Whitehurst, O., Langdon, R., Mahoney, L., Elliott, H. R., Mancano, G., Lee, M. A., Watkins, S. H., Lay, A. C., Hemani, G., Gaunt, T. R., ... Suderman, M. (2022). The EWAS catalog: A database of epigenome-wide association studies. *Wellcome Open Research*, 7, 41. <https://doi.org/10.12688/wellcomeopenres.17598.2>
- Belsky, D. W., Caspi, A., Corcoran, D. L., Sugden, K., Poulton, R., Arseneault, L., Baccarelli, A., Chamarti, K., Gao, X., Hannon, E., Harrington, H. L., Houts, R., Kothari, M., Kwon, D., Mill, J., Schwartz, J., Vokonas, P., Wang, C., Williams, B. S., & Moffitt, T. E. (2022). DunedinPACE, a DNA methylation biomarker of the pace of aging. *eLife*, 11, e73420. <https://doi.org/10.7554/elife.73420>
- Bocklandt, S., Lin, W., Sehl, M. E., Sánchez, F. J., Sinsheimer, J. S., Horvath, S., & Vilain, E. (2011). Epigenetic predictor of age. *PLoS One*, 6(6), e14821. <https://doi.org/10.1371/journal.pone.0014821>
- Costantini, A., Viola, N., Berretta, A., Galeazzi, R., Maticchione, G., Sabbatinelli, J., Storci, G., De Matteis, S., Butini, L., Rippon, M. R., Procopio, A. D., Caraceni, D., Antonicelli, R., Olivieri, F., & Bonafè, M. (2018). Age-related M1/M2 phenotype changes in circulating monocytes from healthy/unhealthy individuals. *Aging*, 10(6), 1268–1280. <https://doi.org/10.18632/aging.101465>
- Edgar, R., Domrachev, M., & Lash, A. E. (2002). Gene expression omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Research*, 30(1), 207–210. <https://doi.org/10.1093/nar/30.1.207>
- Friedman, J. H., Hastie, T., & Tibshirani, R. (2010). Regularization paths for generalized linear models via coordinate descent. *Journal of Statistical Software*, 33(1), 1–22. <https://doi.org/10.18637/jss.v033.i01>
- Furman, D., Campisi, J., Verdin, E., Carrera-Bastos, P., Targ, S., Franceschi, C., Ferrucci, L., Gilroy, D. W., Fasano, A., Miller, G. W., Miller, A. H., Mantovani, A., Weyand, C. M., Barzilai, N., Goronzy, J. J., Rando, T. A., Effros, R. B., Lucia, A., Kleinstreuer, N., & Slavich, G. M. (2019). Chronic inflammation in the etiology of disease across the life span. *Nature Medicine*, 25(12), 1822–1832. <https://doi.org/10.1038/s41591-019-0675-0>
- Gadd, D. A., Hillary, R. F., McCartney, D. L., Zaghlool, S. B., Stevenson, A., Cheng, Y., Fawns-Ritchie, C., Nangle, C., Campbell, A., Flaig, R., Harris, S. E., Walker, R. M., Shi, L., Tucker-Drob, E. M., Gieger, C., Peters, A., Waldenberger, M., Graumann, J., McRae, A. F., ... Marioni, R. E. (2022). Epigenetic scores for the circulating proteome as tools for disease prediction. *eLife*, 11, e71802. <https://doi.org/10.7554/elife.71802>
- Galkin, F., Kochetov, K., Mamoshina, P., & Zhavoronkov, A. (2021). Adapting blood DNA methylation aging clocks for use in saliva samples with cell-type deconvolution. *Frontiers in Aging*, 2, 697254. <https://doi.org/10.3389/fragi.2021.697254>
- Hannum, G., Guinney, J., Zhao, L., Zhang, L., Hughes, G., Sadda, S. V. R., Klotzle, B., Bibikova, M., Fan, J. B., Gao, Y., Deconde, R., Chen, M., Rajapakse, I., Friend, S., Ideker, T., & Zhang, K. (2013). Genome-wide methylation profiles reveal quantitative views of human aging rates. *Molecular Cell*, 49(2), 359–367. <https://doi.org/10.1016/j.molcel.2012.10.016>
- Higgins-Chen, A., Thrush, K., Wang, Y., Minter, C. J., Kuo, P., Wang, M., Niimi, P., Sturm, G., Lin, J., Moore, A. Z., Bandinelli, S., Vinkers, C. H., Vermetten, E., Rutten, B. P. F., Geuze, E., Okhuisen-Pfeifer, C., van der Horst, M. Z., Schreiter, S., Gutwinski, S., ... Levine, M. E. (2022). A computational solution for bolstering reliability of epigenetic clocks: Implications for clinical trials and longitudinal tracking. *Nature Aging*, 2(7), 644–661. <https://doi.org/10.1038/s43587-022-00248-2>
- Hillary, R. F., Baños, D. T., Kousathanas, A., McCartney, D. L., Harris, S. E., Stevenson, A., Patxot, M., Ojavee, S. E., Zhang, Q., Liewald, D. C., Ritchie, C. W., Evans, K. L., Tucker-Drob, E. M., Wray, N. R., McRae, A. F., Visscher, P. M., Deary, I. J., Robinson, M. R., & Marioni, R. E. (2020). Multi-method genome- and epigenome-wide studies of inflammatory protein levels in healthy older adults. *Genome Medicine*, 12(1), 60. <https://doi.org/10.1186/s13073-020-00754-1>
- Horvath, S. (2013). DNA methylation age of human tissues and cell types. *Genome Biology*, 14(10), R115. <https://doi.org/10.1186/gb-2013-14-10-r115>
- Horvath, S., Oshima, J., Martin, G. M., Lu, A. T., Quach, A., Cohen, H. A., Felton, S., Matsuyama, M., Lowe, D., Kabacik, S., Wilson, J. G., Reiner, A. P., Maierhofer, A., Flunkert, J., Aviv, A., Hou, L., Baccarelli, A. A., Li, Y., Stewart, J. D., ... Raj, K. (2018). Epigenetic clock for skin and blood cells applied to Hutchinson Gilford progeria syndrome and ex vivo studies. *Aging*, 10(7), 1758–1775. <https://doi.org/10.18632/aging.101508>
- Horvath, S., & Raj, K. (2018). DNA methylation-based biomarkers and the epigenetic clock theory of ageing. *Nature Reviews Genetics*, 19(6), 371–384. <https://doi.org/10.1038/s41576-018-0004-3>
- Houseman, E. A., Accomando, W. P., Koestler, D. C., Christensen, B. C., Marsit, C. J., Nelson, H. H., Wiencke, J. K., & Kelsey, K. T. (2012). DNA methylation arrays as surrogate measures of cell mixture distribution. *BMC Bioinformatics*, 13, 86. <https://doi.org/10.1186/1471-2105-13-86>
- Kennedy, B. K., Berger, S. L., Brunet, A., Campisi, J., Cuervo, A. M., Epel, E. S., Franceschi, C., Lithgow, G. J., Morimoto, R. I., Pessin, J. E., Rando, T. A., Richardson, A., Schadt, E. E., Wyss-Coray, T., & Sierra, F. (2014). Geroscience: Linking aging to chronic disease. *Cell*, 159(4), 709–713. <https://doi.org/10.1016/j.cell.2014.10.039>
- Koelman, L., Pivovarova-Ramich, O., Pfeiffer, A. F. H., Grune, T., & Aleksandrova, K. (2019). Cytokines for evaluation of chronic inflammatory status in ageing research: Reliability and phenotypic characterisation. *Immunity & Ageing*, 16(1), 11. <https://doi.org/10.1186/s12979-019-0151-1>
- Kolberg, L., Raudvere, U., Kuzmin, I., Adler, P., Vilo, J., & Peterson, H. (2023). g:Profiler—Interoperable web service for functional



- enrichment analysis and gene identifier mapping (2023 update). *Nucleic Acids Research*, 51(W1), W207–W212. <https://doi.org/10.1093/nar/gkad347>
- Kuan, V., Denaxas, S., Patalay, P., Nitsch, D., Mathur, R., González-Izquierdo, A., Sofat, R., Partridge, L., Roberts, A., Wong, I. C. K., Hingorani, M., Chaturvedi, N., Hemingway, H., & Zwierzyńska, M. (2023). Identifying and visualising multimorbidity and comorbidity patterns in patients in the English National Health Service: A population-based study. *The Lancet Digital Health*, 5(1), e16–e27. [https://doi.org/10.1016/s2589-7500\(22\)00187-x](https://doi.org/10.1016/s2589-7500(22)00187-x)
- Levine, M. E., Lu, A. T., Quach, A., Chen, B. H., Assimes, T. L., Bandinelli, S., Hou, L., Baccarelli, A. A., Stewart, J. D., Li, Y., Whitsel, E. A., Wilson, J. G., Reiner, A. P., Aviv, A., Lohman, K., Liu, Y., Ferrucci, L., & Horvath, S. (2018). An epigenetic biomarker of aging for lifespan and healthspan. *Aging*, 10(4), 573–591. <https://doi.org/10.18632/aging.101414>
- Li, M., Zou, D., Li, Z., Gao, R., Sang, J., Zhang, Y., Li, R., Xia, L., Zhang, T., Niu, G., Bao, Y., & Zhang, Z. (2019). EWAS atlas: A curated knowledge-base of epigenome-wide association studies. *Nucleic Acids Research*, 47(D1), D983–D988. <https://doi.org/10.1093/nar/gky1027>
- Ligthart, S., Marzi, C., Aslibekyan, S., Mendelson, M., Conneely, K. N., Tanaka, T., Colicino, E., Waite, L. L., Joehanes, R., Guan, W., Brody, J. A., Elks, C., Marioni, R., Jhun, M. A., Agha, G., Bressler, J., Ward-Caviness, C. K., Chen, B. H., Huan, T., ... Dehghan, A. (2016). DNA methylation signatures of chronic low-grade inflammation are associated with complex diseases. *Genome Biology*, 17(1), 255. <https://doi.org/10.1186/s13059-016-1119-5>
- López-Otín, C., Blasco, M. A., Partridge, L., Serrano, M., & Kroemer, G. (2023). Hallmarks of aging: An expanding universe. *Cell*, 186(2), 243–278. <https://doi.org/10.1016/j.cell.2022.11.001>
- Lu, A. T., Fei, Z., Haghani, A., Robeck, T. R., Zoller, J. A., Li, C. Z., Lowe, R., Yan, Q., Zhang, J., Vu, H., Ablaeva, J., Acosta-Rodriguez, V. A., Adams, D. M., Almunia, J., Aloysius, A., Ardehali, R., Arneson, A., Baker, C. S., Banks, G., ... Mergl, J. (2023). Universal DNA methylation age across mammalian tissues. *Nature Aging*, 3(9), 1144–1166. <https://doi.org/10.1038/s43587-023-00462-6>
- Lu, A. T., Quach, A., Wilson, J. G., Reiner, A. P., Aviv, A., Raj, K., Hou, L., Baccarelli, A. A., Li, Y., Stewart, J. D., Whitsel, E. A., Assimes, T. L., Ferrucci, L., & Horvath, S. (2019). DNA methylation GrimAge strongly predicts lifespan and healthspan. *Aging*, 11(2), 303–327. <https://doi.org/10.18632/aging.101684>
- Makgoeng, S. B., Bolanos, R. S., Jeon, C. Y., Weiss, R. E., Arah, O. A., Breen, E. C., Martínez-Maza, O., & Hussain, S. K. (2018). Markers of immune activation and inflammation, and Non-Hodgkin Lymphoma: A meta-analysis of prospective studies. *JNCI Cancer Spectrum*, 2(4), pky082. <https://doi.org/10.1093/jncics/pky082>
- Martínez de Toda, I., González-Sánchez, M., Díaz-Del Cerro, E., Valera, G., Carracedo, J., & Guerra-Pérez, N. (2023). Sex differences in markers of oxidation and inflammation. Implications for ageing. *Mechanisms of Ageing and Development*, 211, 111797. <https://doi.org/10.1016/j.mad.2023.111797>
- Martin-Herranz, D. E., Aref-Eshghi, E., Bonder, M. J., Stubbs, T. M., Choufani, S., Weksberg, R., Stegle, O., Sadikovic, B., Reik, W., & Thornton, J. M. (2019). Screening for genes that accelerate the epigenetic aging clock in humans reveals a role for the H3K36 methyltransferase NSD1. *Genome Biology*, 20(1), 146. <https://doi.org/10.1186/s13059-019-1753-9>
- McEwen, L. M., O'Donnell, K. J., McGill, M. G., Edgar, R. D., Jones, M. J., MacIsaac, J. L., Lin, D. T. S., Ramadori, K., Morin, A., Gladish, N., Garg, E., Unternaehrer, E., Pokhvisneva, I., Karnani, N., Kee, M. Z. L., Klengel, T., Adler, N. E., Barr, R. G., Letourneau, N., ... Kober, M. S. (2019). The PedBE clock accurately estimates DNA methylation age in pediatric buccal cells. *Proceedings of the National Academy of Sciences of the United States of America*, 117(38), 23329–23335. <https://doi.org/10.1073/pnas.1820843116>
- McGee, K. C., Sullivan, J., Hazeldine, J., Schmunk, L. J., Martin-Herranz, D. E., Jackson, T., & Lord, J. M. (2024). A combination nutritional supplement reduces DNA methylation age only in older adults with a raised epigenetic age. *GeroScience*, 46, 4333–4347. <https://doi.org/10.1007/s11357-024-01138-8>
- Moqri, M., Herzog, C., Poganić, J. R., Justice, J. N., Belsky, D. W., Higgins-Chen, A., Moskalev, A., Fuellen, G., Cohen, A. A., Bautmans, I., Widschwendter, M., Ding, J., Fleming, A., Mannick, J., Han, J. J., Zhavoronkov, A., Barzilai, N., Kaerberlein, M., Cummings, S., ... Gladyshev, V. N. (2023). Biomarkers of aging for the identification and evaluation of longevity interventions. *Cell*, 186(18), 3758–3775. <https://doi.org/10.1016/j.cell.2023.08.003>
- Nishitani, S., Isozaki, M., Yao, A., Higashino, Y., Yamauchi, T., Kidoguchi, M., Kawajiri, S., Tsunetoshi, K., Neish, H., Imoto, H., Arishima, H., Kodera, T., Fujisawa, T. X., Nomura, S., Kikuta, K., Shinozaki, G., & Tomoda, A. (2023). Cross-tissue correlations of genome-wide DNA methylation in Japanese live human brain and blood, saliva, and buccal epithelial tissues. *Translational Psychiatry*, 13(1), 72. <https://doi.org/10.1038/s41398-023-02370-0>
- Parkinson, H., Kapushesky, M., Shojatalab, M., Abeygunawardena, N., Coulson, R. M. R., Farne, A., Holloway, E., Kolesnykov, N., Lilja, P., Lukk, M., Mani, R., Rayner, T., Sharma, A., William, E., Sarkans, U., & Bråzma, A. (2007). ArrayExpress—A public database of microarray experiments and gene expression profiles. *Nucleic Acids Research*, 35, D747–D750. <https://doi.org/10.1093/nar/gkl995>
- Quach, A., Levine, M. E., Tanaka, T., Lu, A. T., Chen, B. H., Ferrucci, L., Ritz, B., Bandinelli, S., Neuhauser, M. L., Beasley, J. M., Snetelaar, L., Wallace, R. B., Tsao, P. S., Absher, D., Assimes, T. L., Stewart, J. D., Li, Y., Hou, L., Baccarelli, A. A., ... Horvath, S. (2017). Epigenetic clock analysis of diet, exercise, education, and lifestyle factors. *Aging*, 9(2), 419–446. <https://doi.org/10.18632/aging.101168>
- Reinius, L. E., Acevedo, N., Joerink, M., Pershagen, G., Dahlén, S. E., Greco, D., Söderhäll, C., Scheynius, A., & Kere, J. (2012). Differential DNA methylation in purified human blood cells: Implications for cell lineage and studies on disease susceptibility. *PLoS One*, 7(7), e41361. <https://doi.org/10.1371/journal.pone.0041361>
- Rodríguez-Meira, A., Norfo, R., Wen, S., Chédeville, A. L., Rahman, H., O'Sullivan, J., Wang, G., Louka, E., Kretzschmar, W. W., Paterson, A., Brierley, C., Martin, J. E., Demeule, C., Bashton, M., Sousos, N., Moralli, D., Subha Meem, L., Carrelha, J., Wu, B., ... Mead, A. J. (2023). Single-cell multi-omics identifies chronic inflammation as a driver of TP53-mutant leukemic evolution. *Nature Genetics*, 55(9), 1531–1541. <https://doi.org/10.1038/s41588-023-01480-1>
- Saule, P., Trauet, J., Dutriez, V., Lekeux, V., Dessaint, J.-P., & Labalette, M. (2006). Accumulation of memory T cells from childhood to old age: Central and effector memory cells in CD4+ versus effector memory and terminally differentiated memory cells in CD8+ compartment. *Mechanisms of Ageing and Development*, 127(3), 274–281. <https://doi.org/10.1016/j.mad.2005.11.001>
- Sayed, N., Huang, Y., Van Nguyen, K., Krejciova-Rajaniemi, Z., Grawe, A. P., Gao, T., Tibshirani, R., Hastie, T., Alpert, A., Cui, L., Kuznetsova, T., Rosenberg-Hasson, Y., Ostan, R., Monti, D., Lehallier, B., Shen-Orr, S. S., Maecker, H. T., Dekker, C. L., Wyss-Coray, T., ... Furman, D. (2021). An inflammatory aging clock (iAge) based on deep learning tracks multimorbidity, immunosenescence, frailty and cardiovascular aging. *Nature Aging*, 1(7), 598–615. <https://doi.org/10.1038/s43587-021-00082-y>
- Scott, A. J., Ellison, M., & Sinclair, D. A. (2021). The economic value of targeting aging. *Nature Aging*, 1(7), 616–623. <https://doi.org/10.1038/s43587-021-00080-0>
- Siemons, L., Ten Klooster, P. M., Vonkeman, H. E., van Riel, P. L. C. M., Glas, C. A. W., & van de Laar, M. A. F. J. (2014). How age and sex affect the erythrocyte sedimentation rate and C-reactive protein in early rheumatoid arthritis. *BMC Musculoskeletal Disorders*, 15, 368. <https://doi.org/10.1186/1471-2474-15-368>



- Štambuk, J., Nakić, N., Vučković, F., Pučić-Baković, M., Razdorov, G., Trbojević-Akmačić, I., Novokmet, M., Keser, T., Vilaj, M., Štambuk, T., Gudelj, I., Šimurina, M., Song, M., Wang, H., Salihović, M. P., Campbell, H., Rudan, I., Kolčić, I., Eller, L. A., ... Lauc, G. (2020). Global variability of the human IgG glycome. *Aging*, *12*(15), 15222–15259. <https://doi.org/10.18632/aging.103884>
- Stevenson, A., Gadd, D. A., Hillary, R. F., McCartney, D. L., Campbell, A., Walker, R. M., Evans, K. L., Harris, S. E., Spires-Jones, T. L., McRae, A. F., Visscher, P. M., McIntosh, A. M., Deary, I. J., & Marioni, R. E. (2021). Creating and validating a DNA methylation-based proxy for interleukin-6. *The Journals of Gerontology: Series A*, *76*(12), 2284–2292. <https://doi.org/10.1093/gerona/glab046>
- Stevenson, A., McCartney, D. L., Hillary, R. F., Campbell, A., Morris, S. W., Bermingham, M. L., Walker, R. M., Evans, K. L., Boutin, T. S., Hayward, C., McRae, A. F., McColl, B. W., Spires-Jones, T. L., McIntosh, A. M., Deary, I. J., & Marioni, R. E. (2020). Characterisation of an inflammation-related epigenetic score and its association with cognitive ability. *Clinical Epigenetics*, *12*(1), 113. <https://doi.org/10.1186/s13148-020-00903-8>
- Tang, Y., Liang, P., Chen, J., Fu, S., Liu, B., Feng, M., Lin, B., Lee, B., Xu, A., & Lan, H. (2018). The baseline levels and risk factors for high-sensitive C-reactive protein in Chinese healthy population. *Immunity & Ageing*, *15*(1), 21. <https://doi.org/10.1186/s12979-018-0126-7>
- Teschendorff, A. E., Breeze, C. E., Zheng, S. C., & Beck, S. (2017). A comparison of reference-based algorithms for correcting cell-type heterogeneity in epigenome-wide association studies. *BMC Bioinformatics*, *18*, 105. <https://doi.org/10.1186/s12859-017-1511-5>
- Tomiya, H., Matsuda, T., & Takiguchi, M. (2002). Differentiation of human CD8+ T cells from a memory to memory/effector phenotype. *The Journal of Immunology*, *168*(11), 5538–5550. <https://doi.org/10.4049/jimmunol.168.11.5538>
- Wang, Y. (2021). dnaMethAge: A user friendly R package to predict epigenetic age and calculate age acceleration from DNA methylation data. <https://github.com/yiluyucheng/dnaMethAge>
- Wang, Y., Grant, O. A., Zhai, X., McDonald-Maier, K. D., & Schlkwyk, L. C. (2024). Insights into ageing rates comparison across tissues from recalibrating cerebellum DNA methylation clock. *GeroScience*, *46*, 39–56. <https://doi.org/10.1007/s11357-023-00871-w>
- Wang, Y., He, F., Feng, F., Liu, X., Dong, G., Qin, H., Hu, X. B., Zheng, M. H., Liang, L., Feng, L., Liang, Y. M., & Han, H. (2010). Notch signaling

determines the M1 versus M2 polarization of macrophages in antitumor immune responses. *Cancer Research*, *70*(12), 4840–4849. <https://doi.org/10.1158/0008-5472.can-10-0269>

Xu, Z., & Taylor, J. A. (2021). Reliability of DNA methylation measures using Illumina methylation BeadChip. *Epigenetics*, *16*(5), 495–502. <https://doi.org/10.1080/15592294.2020.1805692>

Zhou, W., Triche, T. J., Jr., Laird, P. W., & Shen, H. (2018). SeSAMe: Reducing artifactual detection of DNA methylation by Infinium BeadChips in genomic deletions. *Nucleic Acids Research*, *46*(20), e123. <https://doi.org/10.1093/nar/gky691>

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Schmunk, L. J., Call, T. P., McCartney, D. L., Javid, H., Hastings, W. J., Jovicevic, V., Kojadinović, D., Tomkinson, N., Zlamalova, E., McGee, K. C., Sullivan, J., Campbell, A., McIntosh, A. M., Óvári, V., Wishart, K., Behrens, C. E., Stone, E., Gavrilov, M., Thompson, R., ... Martin-Herranz, D. E. (2025). A novel framework to build saliva-based DNA methylation biomarkers: Quantifying systemic chronic inflammation as a case study. *Aging Cell*, *00*, e14444. <https://doi.org/10.1111/accel.14444>

APPENDIX

The Hurdle bio-infrastructure team includes: Pierric Descamps, Emily Taylor, Chuck Paiusi, Leona Mc Girr, Ollie Philpott, Ian Robinson, Hector Watson, Ana Magalia, Dave Russell, and David Woods, all employees at Chronomics Ltd., London, UK.