Quantifying the ageing methylome:
a multi-tissue map in humans

Kirsten Blythe Seale

Thesis submitted in fulfilment of the requirements for the degree of
Doctor of Philosophy

Victoria University

Institute for Health and Sport (IHES)

August 2023

Abstract

The epigenome is a dynamic system of chemical modifications that operate to control chromatin structure and regulate gene activity without altering the underlying DNA sequence. One of the most extensively studied epigenetic marks is DNA methylation (DNAm) that occurs at cytosine-guanine ('CpG') dinucleotides in the genome.

The DNAm landscape ('the methylome') is extensively remodelled during ageing. Studies have described various *features* of the ageing methylome, including linear changes at individual CpGs (i.e. differential and variable methylation) and in the distribution of methylation over all CpGs (i.e. entropy). Current research efforts have contributed greatly to our understanding of the ageing methylome; however, we still do not know the full extent to which the methylome is altered with ageing, and whether differences exist between different tissues. As such, teasing apart the global changes in DNAm that accrue over time, quantifying *differentially* and *variably* methylated positions (DMPs and VMPs) and *entropy*, in various tissues, is needed to obtain a better understanding of the ageing methylome in its entirety. The overarching objective for this PhD project was to quantify the age-related changes in DNAm, including DMPs, VMPs and entropy, in six human tissues, building the largest, *multi-tissue map of DNAm ageing in humans*.

First, we assembled an extensive database of 40,830 human methylomes from 113 datasets in six human tissues, including blood, brain, skin, adipose, buccal and muscle tissue. Then, using a sophisticated statistical pipeline, we conducted a large-scale, multi-tissue Epigenome-Wide Association Study (EWAS) meta-analysis of age to quantify the tissue-specific signatures of DMPs, VMPs and entropy. We also performed functional enrichment analyses to provide biological interpretation of our findings.

In blood, we found that 47% of the CpGs we investigated are DMPs with two-thirds of them (66%) hypomethylated with age, and 37% are VMPs. Entropy increases with age, but only DMPs are driving this increase. The other tissues were variable in the identified age-related changes. Despite varying degrees of statistical power, we detected DMPs in all tissues, and there was no consistent pattern of hyper- vs hypomethylation. We only detected VMPs in blood, brain, saliva, and skin, but not in muscle or adipose tissue. We hypothesise that these age-related changes reflect the proliferative capacities of tissues across the lifespan. However,

this comparison between tissues may be biased by differences in the substructures of the tissue cohorts and varying statistical power, which strongly impacted our ability to detect age-related changes.

By breaking down the methylome into individual components, focusing on DMPs, VMPs and entropy, we obtained a deeper understanding of the ageing methylome in its entirety. This research therefore holds the potential to contribute significantly to the current body of knowledge, by painting a global picture of epigenetic ageing in humans.

Student Declaration

"I, Kirsten Seale, declare that the PhD thesis entitled "Quantifying the ageing methylome: a multi-tissue map in humans" is no more than 80,000 words in length including quotes and exclusive of tables, figures, appendices, references and footnotes. This thesis contains no material that has been submitted previously, in whole or in part, for the award of any other academic degree or diploma. Except where otherwise indicated, this thesis is my own work."

"I have conducted my research in alignment with the Australian Code for the Responsible Conduct of Research and Victoria University's Higher Degree by Research Policy and Procedures."
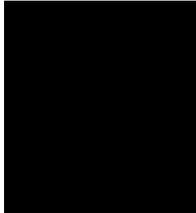
Signature: Date: 25/07/2023

Acknowledgements

Thank you to my primary supervisor, Professor Nir Eynon, for giving me this opportunity to pursue my passion for research in this field. I am deeply grateful for your support, guidance, and experience. The day I found out that I had received a scholarship to pursue a PhD in Australia is one of my fondest memories. Secondly, I would like to thank my co-supervisor Dr Sarah Voisin. Sarah, you have been a mentor, an inspiration, a pioneer in the field and a friend. I am so grateful to have spent 3 and a half years being trained by the best. Thank you for imparting your knowledge and experience, encouraging me to think critically and allowing me to grow into my own researcher. My learning curve over the last few years has been exponential, and I am so thankful that you pushed me beyond what I thought I was capable of.

To the broader Genetics and Epigenetics group at VU, what a weird but equally wonderful time it has been working with you all during the pandemic. Never would I have imagined to be living in a new country locked up in an apartment pursuing a PhD from behind my laptop. It was thanks to your ongoing support that I got through some of the most challenging days. A special mention to my fellow PhD student Andrew Palmer, who started his PhD with me in 2020. The road has been windy, with many unforeseen twists and turns. Thank you for remaining a loyal friend and confidant.

To my mom, there are no words to adequately capture the deepest gratitude that I have for your love and support. I credit every bit of my success to you, and the sacrifices you made for me to be where I am. Thank you from the depths of my heart. Equally, Jen, Luke, and Matt, I couldn't ask for more supportive and caring siblings. I cherish you all deeply and I could not do hard things without you three keeping me grounded. To my dad, I have drawn huge amounts of strength and comfort in knowing that this is a path you would've chosen for me as your daughter, so thank you for guidance from above.

Finally, to my husband Callum, we did this together. If someone had told me ten years ago when we met on the side of the hockey field that we would be adventuring together on the other side of the world I wouldn't have believed them. Thank you for your love, your support, your pep talks, and your friendship. You kept me sane when I was going

insane, always reminding me to never take life so seriously. You stood by side when I was deep in the trenches, and you celebrated my successes. You are the best person I could've chosen to do life with. Thank you.

List of Publications

The following work has been published in peer-reviewed journals in support of this thesis:

Seale KB, Horvath S, Teschendorff A, Eynon N and Voisin S. Making sense of the ageing methylome. *Nat Rev Genet* 2022;**23**(10):585-605. (Chapter 2)

The following work is being prepared for publication in peer-reviewed journals in support of this thesis:

Seale KB, Teschendorff AT, Voisin S, and Eynon N. A comprehensive map of the ageing blood methylome. *(Being prepared for submission)* (Chapter 4)

Voisin S, **Seale KB,** and Eynon N. A multi-tissue atlas of DNA methylation and ageing in humans. *(Being prepared for submission)* (Chapter 5)

The following work has been published or submitted in peer-reviewed journals during my candidature, and is outside the scope of this thesis:

Voisin SV, **Seale KB,** …, Eynon N. Exercise rejuvenates the skeletal muscle methylome and transcriptome in humans. *Aging Cell* 2023;e13859.

Alvarez-Romero J, Laguette MN, **Seale KB**, …, Eynon N. Genetic variants within the COL5A1 gene are associated with ligament injuries in physically active population from Australia, South Africa, and Japan. *Eur J Sport Sci* 2021;1-10.

Jones P, Voisin S, Nolan BJ, …, **Seale KB**, …, Eynon N. Uncovering the effects of gender affirming hormone therapy on skeletal muscle and epigenetics: protocol for a prospective matched cohort study in transgender individuals (The GAME Study). *BMJ Open* 2022;**12**:e060869.

Luo Q, … **Seale KB**, … Teschendorff A. A meta-analysis of immune cell fractions at high resolution reveals novel associations with common phenotypes and health outcomes. *Genome Med* 2023;**15**(1):59.

Details of Included Papers

DETAILS OF INCLUDED PAPERS: THESIS WITH PUBLICATION

Please list details of each scholarly publication and/or manuscript included in the thesis submission. Copies of published scholarly publications and/or manuscripts submitted and/or final draft manuscripts should also be included in the thesis submission.

This table must be incorporated in the thesis before the Table of Contents.

Chapter No.	Publication Title	Publication Status • Published • Accepted for publication • In revised and resubmit stage • Under review • Manuscript ready for submission	Publication Details • Citation, if published • Title, Journal, Date of acceptance letter and Corresponding editor's email address • Title, Journal, Date of submission
2	Making sense of the ageing methylome	Published	Seale, K., Horvath, S., Teschendorff, A. et al. Making sense of the ageing methylome. Nat Rev Genet 23, 585–605 (2022). https://doi.org/10.1038/s41576-022-00477-6

Declaration by [candidate name]:

Kirsten Seale

Signature: Kirsten Seale

Digitally signed by Kirsten Seale
Date: 2023.04.14
08:28:36 +10'00'

Date:

14/04/2023

Updated: September 2020

OFFICE FOR RESEARCH TRAINING, QUALITY AND INTEGRITY

DECLARATION OF CO-AUTHORSHIP AND CO-CONTRIBUTION: PAPERS INCORPORATED IN THESIS

This declaration is to be completed for each conjointly authored publication and placed at the beginning of the thesis chapter in which the publication appears.

1. PUBLICATION DETAILS (to be completed by the candidate)

Title of Paper/Journal/Book: Making sense of the ageing methylome

Surname: Seale   First name: Kirsten

Institute: Institute for Health and Sport   Candidate's Contribution (%): 75

Status:

Accepted and in press: ☐   Date:

Published: ☑   Date: 02 May 2022

2. CANDIDATE DECLARATION

I declare that the publication above meets the requirements to be included in the thesis as outlined in the HDR Policy and related Procedures – policy.vu.edu.au.

Kirsten Seale   Digitally signed by Kirsten Seale
Date: 2023.05.23 11:56:32 +10'00'   23/05/2022

Signature   Date

3. CO-AUTHOR(S) DECLARATION

In the case of the above publication, the following authors contributed to the work as follows:

The undersigned certify that:

1. They meet criteria for authorship in that they have participated in the conception, execution or interpretation of at least that part of the publication in their field of expertise;

2. They take public responsibility for their part of the publication, except for the responsible author who accepts overall responsibility for the publication;

VICTORIA UNIVERSITY

THE NEW WAY TO DO UNI

3. There are no other authors of the publication according to these criteria;

4. Potential conflicts of interest have been disclosed to a) granting bodies, b) the editor or publisher of journals or other publications, and c) the head of the responsible academic unit; and

5. The original data will be held for at least five years from the date indicated below and is stored at the following **location(s)**:

N/A

Name(s) of Co-Author(s)	Contribution (%)	Nature of Contribution	Signature	Date
Sarah Voisin	15	Literature research, ideation, writing and editing		24/05/2023
Nir Eynon	5	Reviewing and editing		24/05/2023
Steve Horvath	2.5	Reviewing and editing		5/062023
Andrew Teschendorff	2.5	Reviewing and editing		24/05/2023

Updated: September 2019

Conferences

Oral Presentations

2020: Australian Physiology Society (AuPS) Student and Early-Career Research Forum. *Uncovering Robust DNAm Markers for Skeletal Muscle Biological Ageing.*

2020: Victoria University Higher Degree Researcher Conference. *Uncovering Robust DNAm Markers for Skeletal Muscle Biological Ageing.*

Poster Presentations

2021: The 8th Australian Epigenetics Conference. *Uncovering Robust DNAm Markers for Skeletal Muscle Biological Ageing.*

2022: The 9th Aging Research and Drug Discovery Meeting (ARDD 2022). *A multi-tissue map of the ageing methylome in humans.*

2022: EuroGeroscience. *A Multi-Tissue Meta-Analysis of Methylation Variability: Uncovering Robust Markers for Biological Ageing.*

Media Appearances

Sheeky Science Show

Victoria University News

Table of Contents

List of Tables

List of Figures

List of Abbreviations

5hmC = 5-Hydroxymethylcytosine

5mC = 5-Methylcytosine

ABOS = Biological Atlas of Severe Obesity

AD = Alzheimer's disease

α-KG = α-ketoglutarate

AMPK = AMP-dependent kinase

BMAL1 = brain and muscle ARNT-Like 1

BMI = body mass index

BP = blood pressure

CALERIE = Comprehensive Assessment of Long-term Effects of Reducing Intake of Energy

CGI = CpG island

CGP = expression signatures of genetic and chemical perturbations

CLOCK = circadian locomotor output cycles kaput

COPD = chronic obstructive pulmonary disease

COVID-19 = coronavirus disease 2019

CpG = cytosine-phosphate-guanine dinucleotide

CR = caloric restriction

CVD = cardiovascular disease

dbGaP = database of Genotypes and Phenotypes

DDR = DNA damage and repair

DHEA = dehydroepiandrosterone

DMG = differentially methylated gene

DMP = differentially methylated position

DMR = differentially methylated region

DNA = deoxyribose nucleic acid

DNAm = DNAm

DNMT = DNA methyltransferase

DR = dietary restriction

DSB = double-stranded break

DunedinPACE = Dunedin Pace of Aging Calculated from the Epigenome

ECM = extracellular matrix

EGA = European Genome-Phenome Archive

ELOVL2 = elongation of very long chain fatty acids protein 2

EPIK = Epigenetica & Kracht

ESC = embryonic stem cells

EWAS = epigenome-wide association study

FDR = false discovery rate

FOXO = forkhead transcription factor class O

FTC = Finnish Twin Cohort

GeneSMART = Gene Skeletal Muscle Adaptive Response to Training

GEO = gene expression omnibus

GO = gene ontology

GPCR = G protein coupled receptors

H3K27me3 = trimethylation of lysine 27 on histone H3 protein subunit

H3K4me3 = trimethylation of lysine 4 on histone H3 protein subunit

HPO = human phenotype ontology

IDAT = intensity data files

IGF-1 = insulin-like growth factor-1

LITER = Limb Immobilisation and Transcriptional/Epigenetic Responses

ME = module eigenene

MF = methylation fraction

MSigDB = molecular signature database

mTORC1 = mammalian target of rapamycin complex 1

NAD+ = nicotinamide adenine dinucleotide

NAFLD = non-alcoholic fatty liver disease

NAMPT = Nicotinamide phosphoribosyltransferase

NASH = non-alcoholic steatohepatitis

NK = natural killer

NMN = nicotinamide mononucleotide

NR = nicotinamide riboside

OA = osteoarthritis

OSKM = Oct4, Sox2, Klf4, and c-Myc transcription factors

PARP = poly(ADP-ribose) polymerases

PBMC = peripheral blood mononuclear cell

PRC = Polycomb repressive complex

RRBS = reduced representation bisulphite sequencing

SAM = S-adenosylmethionine

SIRT1 = sirtuin-1

SNP = single nucleotide polymorphism

SVD = singular value decomposition

T2D = type 2 diabetes

TAME = Targeting Ageing with Metformin

TET = ten-eleven translocation

TF = transcription factor

TOM = topological overlap measure

TTFL = transcription-translation feedback loop

UV = ultraviolet

VMG = variably methylated gene

VMP = variably methylated position

VMR = variably methylated region

WGBS = whole genome bisulphite sequencing

WGCNA = weighted gene correlation network analysis

Chapter 1: Introduction

For centuries, humans have searched for the 'secret' to eternal youth. Some of the earliest literary texts describe tales of the Epic of Gilgamesh, and legends tell of mythical substances such as the Philosopher's stone or the Fountain of Youth. Fiction aside, ageing research has garnered significant interest over the years. This unique biological phenomenon is accompanied by a decline in functional capacity across the lifespan and an age-associated burden of multiple diseases and disability[1]. Thus, mitigating this burden is of utmost importance, but that requires deciphering the cellular, molecular, and systemic processes that accrue over the lifespan.

Ageing is initiated at the basic level of biological organisation, underpinned by 12 interconnected hallmarks[2]. Alterations to the epigenome are a primary hallmark of ageing, and a focus of this PhD[2]. Specifically, the age-related changes in DNAm, which accrues numerous, widespread changes. Of importance are two distinct linear changes: *differential* and *variable* patterns of DNAm. Age-associated *differentially methylated positions* (DMPs) capture age-related DNAm changes that are *shared* between individuals, whereas age-associated *variably methylated positions* (VMPs) capture DNAm changes that *diverge* between people over the lifespan[3].

While all humans experience similar ageing symptoms with chronological time, the degree and speed at which these changes occur varies between individuals, leading to inter-individual differences in the time of onset and severity of age-associated disease and disability (i.e. individuals who are the same chronological age will differ in their 'biological' age)[1]. It is therefore plausible that at the epigenetic level, two individuals with identical chronological ages (and patterns of DMPs) may display divergent patterns across VMPs. Albeit DMPs and VMPs are not mutually exclusive (i.e. a CpG can be considered both a DMP and a VMP), both features contribute to age-related DNAm

changes, but in fundamentally different ways, and the vast majority of studies so far, including epigenetic clock studies, have solely investigated DMPs. Focusing only on patterns of DMPs is limiting when trying to understand aspects of biological ageing, particularly when making sense of why individuals of the same age display vastly different ageing rates. However, we still do not know the extent to which the methylome is differentially and variability methylated in humans, and whether differences exist between different tissues. Between-tissue comparisons may offer insights into the fundamental molecular and cellular mechanisms that give rise to DMPs and VMPs, as tissues vary widely in their degree of cellular heterogeneity, proliferative capacity, and metabolic rates. **The first and second aims of this PhD are to explore the differential (aim 1) and variable (aim 2) patterns of DNAm that accrue over the lifespan in different tissues.**

Insights into epigenetic ageing can also be drawn from quantifying the changes at the whole methylome level using single measurements of *entropy.* Entropy is a probability formula that estimates the amount of information in a set of CpGs[3]. It has been shown in blood that the ageing methylome displays increasing *entropy* or 'methylation disorder' over time, implying that the methylome loses information with age[3]. This is because it is 'easier' to predict the methylation state across all CpGs in young vs older individuals. However, it remains to be seen whether tissues and cells display varied rates of entropic decay. **The third aim of this PhD is to explore the age-associated entropy dynamics across different tissues.**

Thus, the overarching objective of this thesis was to uncover the tissue-specific DNAm signatures of ageing and build a comprehensive multi-tissue map of epigenetic ageing in humans. Specifically, we aimed to quantify the age-related DNAm changes that are *shared* between all individuals (DMPs) (aim 1) and those that *diverge* (VMPs) (aim 2) over the life course. In addition, we sought to understand how entropy

changes in different tissues, and how DMPs and VMPs contribute to these entropy dynamics (aim 3).

Commencing with the Literature Review (**Chapter 2)**, we provide a comprehensive review of the various ways we can quantify ageing using DNAm data, including differential methylation, variable methylation, epigenetic clocks, entropy, and correlation networks. We also investigate the evidence linking DNAm to ageing phenotypes, and the longevity interventions that alter DNAm and / or it's machinery to extend healthspan and lifespan. We conclude this chapter with theories on the causes of epigenetic ageing, and the future of this research.

Chapter 3 describes the methodological approach undertaken to answer each of the project aims. Briefly, we describe our data collection process, which involved collecting and pre-processing the methylomes of 40,830 human samples from 113 datasets across 6 human tissues to build an exhaustive DNAm database. We then describe the bioinformatics and statistical tools used to quantify the features of ageing in each tissue, and our meta-analysis approach to identify robust DNAm changes with age. The DNAm database was the first major output of the study, and laid the foundation to power the large-scale, cross-sectional, multi-tissue EWAS meta-analysis of age.

Chapter 4 details the results of the analyses in whole blood, the largest tissue cohort in our study, which is followed by **Chapter 5,** and includes the results from all the remaining tissues, including brain, buccal, adipose, muscle and skin. The final discussion (**Chapter 6**) is the interpretation of the findings, the limitations of the study, and the future directions of this work. Notably, this study forms part of a larger research project, which is to expand the results in as many human tissues as possible and publish the largest quantitative map of epigenetic ageing in humans as an online webtool for use by the broader scientific community.

Chapter 2: Literature Review

This chapter is based on the following publication:

Abstract

The DNAm landscape accrues significant damage during ageing that leads to a cascade of undesirable consequences. This review provides a comprehensive overview of the various DNAm changes that have been described during ageing, including differential methylation, variable methylation, entropy, and correlation networks, as well as the statistical tools that can be used to quantify them. We also detail the evidence linking DNAm to ageing phenotypes, and the longevity strategies that alter both DNAm patterns and machinery to extend healthspan and lifespan. Lastly, we discuss theories on the mechanistic causes of epigenetic ageing.

Introduction

The world's rapidly ageing population has become one of society's greatest challenges[4]. By 2050, it is projected that in many parts of the world 25% of the population will be aged >65 years[5], but steady increases in life expectancy (lifespan) are not concomitant with an increase in 'healthspan' (disease-free, healthy lifespan). Instead, the ageing population exhibits rising morbidity rates and a decline in quality of life[1,6], which comes at social and economic costs[4,7].

Ageing is the time-dependent decline in functional capacity across the lifespan, characterised by the accumulation of molecular damage, which stems from a progressive reduction in the capacity to repair this damage[8–10]. This includes changes that erode the structure and function of all tissues with chronological time (i.e. *primary ageing*), as well as additional deleterious changes that are aggravated by environmental perturbation and disease (i.e. *secondary ageing*)[11–13]. Theories of ageing can be grouped into two main schools of thought[14]: i) ageing is a tightly regulated, programmed process, and the pathological consequences are an extension of biological processes such as growth and development; ii) ageing is a consequence of damage and stochastic errors that impair the capacity for maintenance. This pan-tissue deterioration is underpinned by a common set of biological defects, considered the 'hallmarks of ageing,' which are grouped into three categories: the primary hallmarks that cause the damage; the antagonistic hallmarks that are compensatory responses to the damage; and the integrative hallmarks responsible for ageing phenotypes[9]. Alterations to the epigenome are considered primary hallmarks of ageing[9].

The epigenome is a dynamic maintenance system operating via a range of chemical modifications that control chromatin organisation and regulate gene activity without changing the DNA sequence[15,16]. The best characterised epigenetic mark is DNAm, which is the covalent attachment of a methyl group to the 5th carbon of the cytosine residue (5mC). In mammals, DNAm usually occurs at cytosine-phosphate-guanine dinucleotides ('CpGs')[8] and carries out distinct functions in different genomic regions[17]. Patterns of DNAm (the 'methylome') are laid down early during embryonic development and are maintained through cell divisions to ensure cell identity. Therefore, DNAm patterns strongly differ between cell types within the same tissue, and between tissues[18,19]. However, DNAm can also be dynamically added by DNA methyl transferases (DNMTs) or removed by ten-eleven translocation (TET) enzymes[16]. Because the methylome operates at the interface between the genome and the environment[8], it can 'shift' in

response to environmental stimuli[20] (such as exercise[21–23], diet[24,25], smoking[26–28] or pollutants[29,30]). Moreover, the integrity of the methylome is tightly associated with healthy ageing and altered DNAm patterns have been associated with a broad range of age-related diseases, including Alzheimer's disease (AD)[31–35], cardiovascular disease (CVD)[36–38], and cancer[39–42].

However, many questions remain unresolved. What are the differences in DNAm associated with primary and secondary ageing? Does the methylome age differently in different tissues? What are the biological mechanisms causing DNAm changes throughout the lifespan? This review aims to answer these questions by painting a comprehensive picture of the ageing human methylome. First, we review the different types of changes that have been observed with age, namely *differential* and *variable* DNAm, *clocks*, *entropy,* and *correlation networks,* and describe the strength of evidence linking these changes to age-related diseases. Then, we propose molecular mechanisms that explain these changes and show how environmental factors that are known to accelerate or decelerate age-related DNAm changes can help us understand these mechanisms. Finally, we propose future directions for research in the field.

1. The ageing human methylome

Our understanding of DNAm changes that accrue over time is bound by the statistical and computational ways used to quantify those changes (Table 1). Common ways to quantify age-related DNAm changes are to look for CpG sites displaying differences in *average* DNAm levels between young and old individuals, or CpG sites displaying differences in DNAm *variance* between young and old individuals[19]. While these are the most widely used measurements, they are one-dimensional (i.e. they focus on individual CpGs). Changes at the whole methylome level can also be quantified using single measurements

(i.e. entropy) or changes in coordinated DNAm levels at multiple CpGs (i.e. correlation networks)[43–45].

1.1. Changes in average DNA methylation

Early studies investigating global changes in DNAm, using various chromatography techniques, were the first to report a global loss of DNAm occurs with age in some human and rodent tissues[46–49]. Studies using advanced sequencing technologies, such as whole-genome bisulphite sequencing (WGBS), have since reported conflicting evidence. While a global decrease in DNAm with ageing has been observed in human T cells[50], studies in human (e.g. brain, epidermis, muscle, heart, liver) and rodent tissues (e.g. liver, hippocampus) observed no statistically significant shifts in global DNAm levels during ageing[51–56]. This suggests the net effect of global DNAm may heavily depend on the methods used, which have vastly different detection techniques, or the tissue of interest.

These technological advancements (e.g. microarray, reduced representation bisulphite sequencing (RRBS) and WGBS) have since prompted the development of epigenome-wide association studies (EWAS), which have revealed predictable and consistent shifts in DNAm at specific CpG sites across the lifespan, i.e. differentially methylated positions (DMPs)[29,42,57–60]. DMPs exhibit an average 'shift' in their mean methylation level as humans age[19] (Figure 1). DNAm is, therefore, either increased (hypermethylated at the CpG site) or decreased (hypomethylated at the CpG site) in a consistent manner[61,62]. Age-related DMPs can be identified with linear models, implemented in packages such as *limma*[63] (Table 1). Differential methylation may also occur over a whole genomic region, as CpG sites ~500 bp apart are typically highly correlated; these form age-related changes in mean DNAm levels over multiple, contiguous CpGs, i.e. differentially methylated regions (DMRs)[19]. Various statistical algorithms exist for detecting DMRs, such as *DMRcate*[64], *BumpHunter*[65], *comb-p*[66], *blockFinder*[67], or *Probe Lasso*[68]. Since DMPs and

DMRs closely correlated with *chronological age,* these sites capture the age-associated DNAm changes *shared* between individuals over the lifespan[69].

1.1.1. Epigenetic clocks: predictors of chronological age

Subsets of DMPs have been used to build both multi-tissue and tissue-specific 'epigenetic clocks' capable of predicting the chronological age of a sample with high accuracy[60,68–71]. The first epigenetic clocks, including the saliva clock by Bocklandt *et al.*[71], Horvath's pan-tissue clock[60] and Hannum's whole blood clock[41], were developed using a machine learning algorithm trained to predict chronological age. By computationally distilling the widespread DMPs shared by all individuals across the lifespan, the algorithm selects several CpG sites that predict chronological age with high accuracy, a parameter known as 'DNAm age' or 'epigenetic age'[41,60,67,72]. The majority of epigenetic clocks are built using elastic net regression (Table 1)[67,73,74]. Typically, chronological age (or a transformed version of age or mortality risk) is regressed on a set of CpG sites, and the algorithm selects the most informative CpG sites from a pool of tens of thousands of potential sites for the age prediction[72,75].

A multitude of specialised clocks have been developed, including tissue-specific clocks[70,76–78], clocks for different animal species[79–90], and even multi-tissue, multi-species clocks[56]. For example, a multi-tissue epigenetic clock captures the time-dependent DNAm changes that are *intrinsic* to the methylome (reflected by the age-related DNAm patterns that validate consistently across many tissues and cell types)[55,60,76,91–93]. Conversely, tissue-specific clocks capture intrinsic changes as well as *extrinsic* or within-tissue DNAm changes, including age-related changes in tissue or cell composition[72,92,93]. In recent developments, a novel statistical framework for profiling epigenetic age at the single-cell resolution, 'scAge', has been introduced, providing novel insights into the heterogeneity in epigenetic ageing of individual cell types[94].

Notwithstanding the extraordinary ability of epigenetic clocks to predict chronological age, it is important to remember the clock is simply a multivariate age predictor generated from a subset of CpGs and does not capture the entirety of the ageing methylome[95]. This is owing to the fact that the specific CpG sites selected by the different clocks do not overlap, and different epigenetic clocks capture different biological signals[67]. As such,
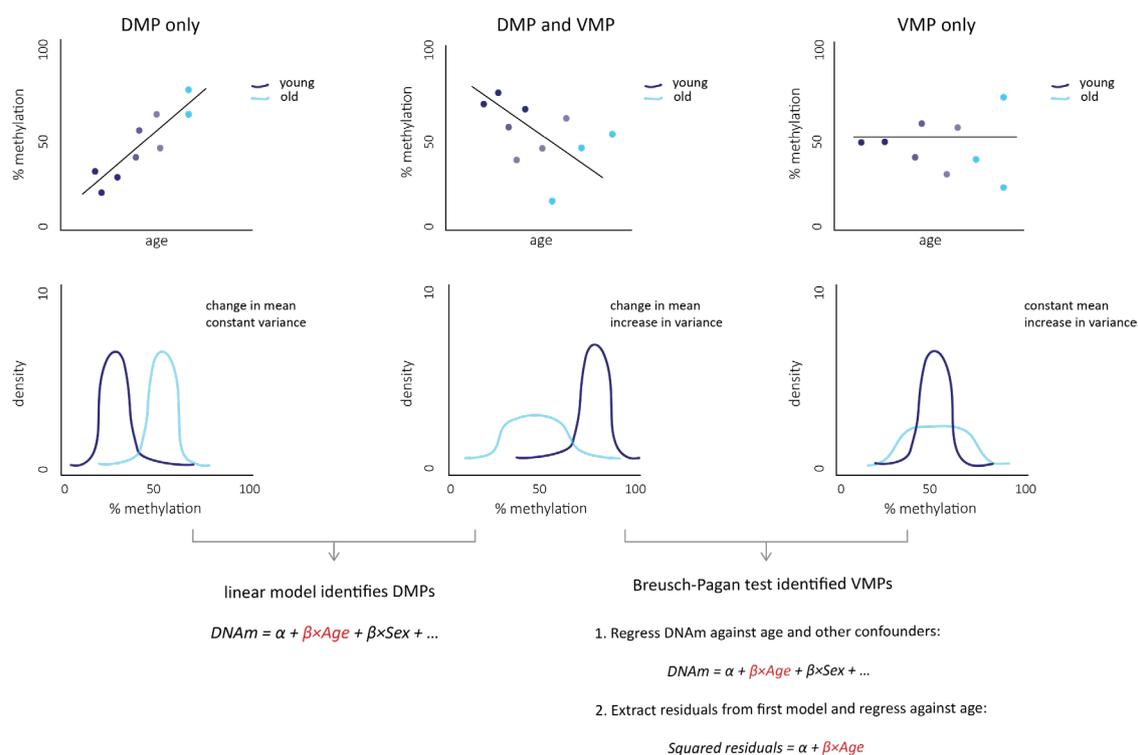
Figure 1 Linear models classify age-associated changes in DNAm.

Linear plots (top) and corresponding density plots (bottom) represent chronological changes at individual cytosine–guanine dinucleotides (CpGs) during ageing. CpGs that change in mean methylation are classified as differentially methylated positions (DMPs). A young individual with high methylation fraction at a particular DMP can be distinguished from an older person with low methylation at the same DMP. DMPs are identified using a linear model and can be homoscedastic (no variance in residuals) or heteroscedastic (variance of residuals increases with age). Variably methylated positions (VMPs) are classified according to the relative changes in variability with age. All VMPs are heteroscedastic. For example, older individuals will show striking variability in methylation status at a particular VMP, compared with young individuals. VMPs can be identified using the Breusch–Pagan statistical test, which is a two-way linear regression formula. First linear model regresses DNAm (DNAm) against age and other confounders, second linear model regresses the squared residuals of the first model against age. α, y intercept; β, regression coefficient for each predictor variable.

9

the clock concept should not be confused with global methylation signatures of ageing (i.e. EWAS of DMPs and VMPs, correlation networks and entropy).

1.2. Changes in DNA methylation variance

Some CpG sites exhibit increased variability with age and are known as age-associated variably methylated positions (VMPs)[43,62,70–75]. This mechanism of 'epigenetic drift' in ageing was first discovered in twin studies, whereby the methylomes of monozygotic twins diverge as they get older[71,76–78]. VMPs therefore capture the *stochastic* changes in DNAm that accumulate with age[62], and are largely driven by differences in environmental factors that accumulate throughout the lifespan[29,71,79,80]. Unlike DMPs, VMPs do not necessarily display shifts in their mean methylation over time, but instead show increases in DNAm variance, as a result of aberrant and unpredictable changes (Figure 1)[19,43,62,71,72]. Although some overlap between DMPs and VMPs does exist, characterised by CpG sites that display both a change in mean DNAm and variance with age, a significant proportion of VMPs are independent of differential DNAm shifts, and represent their own class of age-related DNAm changes[62]. The overwhelming majority of VMPs increase in variance with age, however several studies have reported that a small proportion of VMPs decrease in variance with age[43,71–73,75], with a tendency of these sites to approach fully methylated or unmethylated states (e.g. methylation fraction of 0 or 1)[75]. Although we could not find evidence in the current literature, we cannot rule out the possibility that variability may be related to technical issues of measurement technologies, which can show more error at intermediate DNAm level.

Various statistical tests have been used to identify VMPs, which all detect heteroscedasticity (i.e. change in variance) (Table 1). To test for heteroscedasticity in DNAm with age as a continuous variable, the Breusch-Pagan[43,62,74,81,82], or the White test can be used[76]. To test for heteroscedasticity in DNAm between discrete groups (i.e.

newborns vs centenarians), Bartlett's test, Levene's test, or Brown-Forsythe test can be used[19,71].The R package *DiffVar* has been developed to detect VMPs in microarray data modelled off Levene's test[73]. Variably methylated regions (VMRs) can be detected using packages such as *minfi* and *DMRcate*[19,64,67].

1.3. Increase in DNA methylation 'chaos' (entropy)

DNAm is binary in nature, however, the DNAm fraction of a particular CpG site is measured over a population of cells and is represented as a gradient (0 – 100%). In the mammalian methylome, CpG sites are typically highly methylated or unmethylated, with few sites showing intermediate levels of methylation. At many CpG sites, methylation levels shift over time from states of high or low methylation to an intermediate fraction of approximately 50%, representing a 'smoothening' of the epigenetic landscape[43,69] (Figure 2). CpG sites that are hyper- or hypomethylated in youth change to become less ordered and predictable at older ages[43,69]. Ageing is therefore associated with a reduction in the stringency of epigenetic maintenance systems[8,83], where small perturbations at individual CpG sites (i.e. DMPs and VMPs) cumulatively result in the inability of youthful DNAm patterns to be maintained throughout the lifespan. This has been described as epigenetic chaos, or loss of information, which increases with age[84].

Studies in blood have quantified these methylome-wide DNAm changes as a single measure of **'entropy,'** or methylation disorder (Table 1)[43,62,74,84–86]. Entropy is a scientific concept, as well as a quantifiable measure of randomness, uncertainty, or disorder. Specifically, Claude Shannon introduced the concept of 'Shannon entropy,' or 'information entropy', which measures the amount of information in a variable (i.e. a set of CpG sites)[87]. Entropy is low if it is easy to predict the information stored in said variable, because there is less surprise or uncertainty. Entropy increases if a parameter is difficult to predict, as it takes on many possible values, and thus there is more uncertainty

with an increasing number of possible outcomes[87]. For clarity, the *uncertainty* of a single CpG associated with entropy is distinct from the *stochasticity* of a VMP. While VMPs capture a variable change with age, entropy is linked to the ability to predict the methylation status of a CpG for any given cell (Figure 2). For instance, estimating the CpG status by sampling a pool of cells, where some cells are methylated and some cells are unmethylated, would yield a high entropy value because there is a 50% chance you could accurately estimate the methylation fraction.

In a single measure, Shannon entropy can therefore estimate the amount of information for a set of CpG sites, quantifying the total accumulation of differential age-related changes in DNAm at all CpG sites[43,84] (Figure 2). Shannon entropy increasing with age implies that the ageing methylome shifts to an epigenetic state of high disorder (i.e. tends towards a methylation fraction of 50%), displaying increased 'chaos' or information loss over time[43,62,84,88–90]. Shannon entropy is measured using a probability formula, which has been adapted to handle DNAm data[91].

Our current understanding of epigenetic entropy and age is limited by studies in blood. Considering tissues age at different rates[92], future investigations should explore whether this is underpinned by differing rates of entropic decay. Moreover, our understanding of the mechanisms underlying entropy are limited due to studies performed using bulk tissue. Since cell-to-cell heterogeneities increase with age[93], it is plausible that increases in entropy may simply reflect DNAm variability between cells within a tissue. Novel approaches to estimate the epigenetic age of single cells suggests that individual cell types do not age at the same rate[94]. As such, measuring entropy at the single-cell resolution may reveal interesting entropy dynamics, compared to whole tissue analyses. For example, if cells display attenuated epigenetic ageing rates, is this reflected by a slower rate of entropic decay?
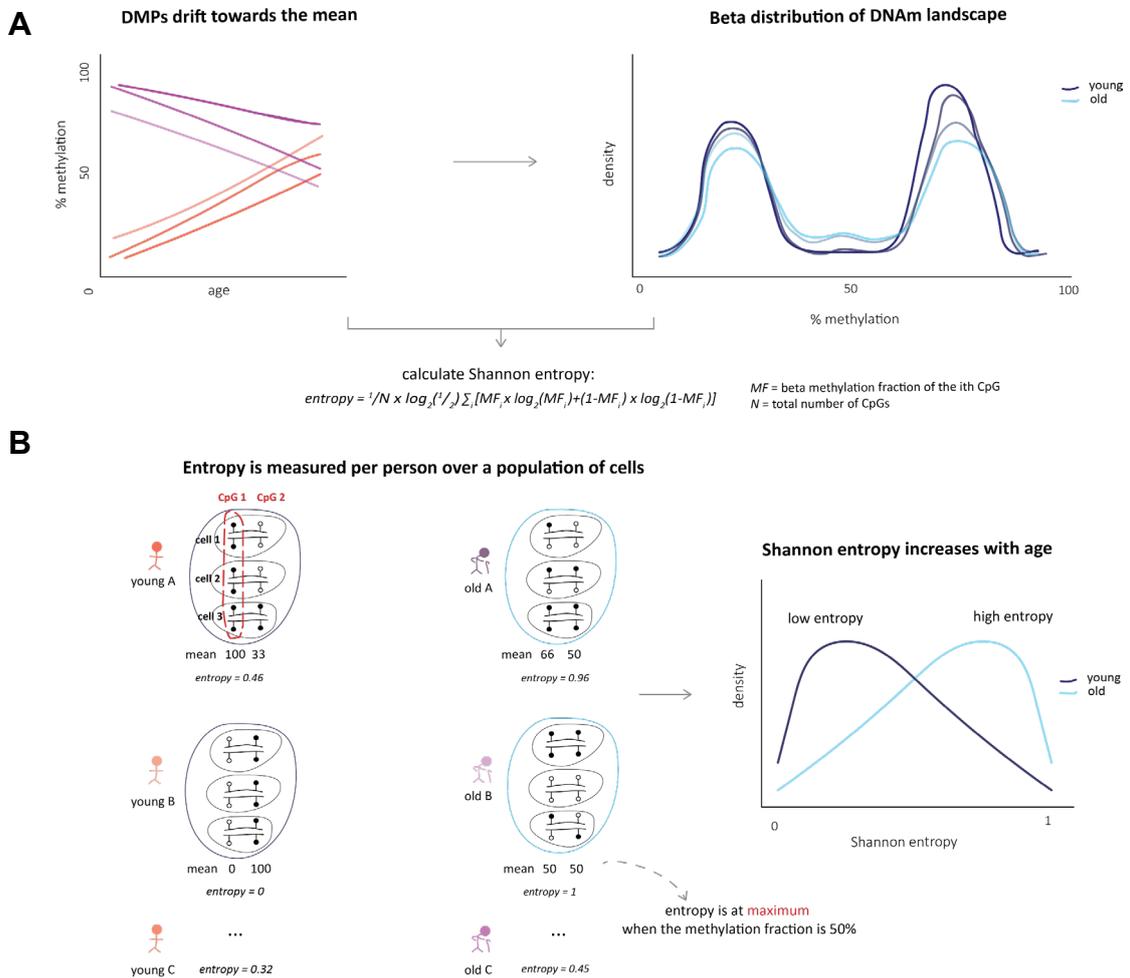
Figure 2 Entropy measures chaos in the ageing methylome.

A) Pattern of differentially methylated positions (DMPs) that drift towards a methylation fraction (MF) of 50% with increasing age (left). These data can also be shown as a beta distribution, whereby all cytosine–guanine dinucleotides (CpGs) in the methylome in young individuals show a bimodal distribution with two distinct peaks, which tend to shift to the middle with age (right). To quantify total accumulation of these age-related changes in DNAm (DNAm) across all CpGs, a Shannon entropy probability formula can be used. B) To calculate entropy for each individual, mean MF of each CpG is first measured over a population of cells and then input into the entropy formula. If a CpG takes on one or only a few values, it is easy to predict its value and entropy is considered low ('young B'). By contrast, if a CpG takes on numerous values, it is more difficult to predict and entropy is high. Entropy is at a maximum if mean methylation of a CpG is 50% ('older B'). As the methylome accumulates changes at multiple CpGs with age, Shannon entropy increases. i, single CpG; N, total number of CpGs.

1.4. Change in DNA methylation at coordinated CpGs (correlation networks)

The identification of DMPs and VMPs provides valuable insights into the shift of DNAm patterns with age at individual CpGs. However, the methylome forms a complex network of coordinated CpG sites that show similar methylation status (i.e. co-methylation)[56,60,95–97]. This inter-correlation between CpGs can have a physical explanation, as CpGs tend to influence the methylation status of other nearby CpGs, owing to DNMT and TET enzymes that maintain co-methylation dynamics and ensure epigenetic faithfulness during cell divisions[96,98]. Long-range co-methylation relationships also exist[95,99,100], as distal CpGs can be brought into spatial proximity due to chromatin folding[97]. Local and long-range CpG correlations reflect the three-dimensional architecture of DNA, and such coordinated DNAm patterns play an important role in regulating cellular activities[95,97,100]. As such, co-methylation between individual CpGs can reflect biological pathways, as the chain reactions that regulate cellular functions depend on the coordinated activation/silencing of multiple genes[19,101].

DNAm analysis may therefore benefit from adopting a 'systems' biology approach that encompasses the interconnectedness of the entire methylome[44,102]. Focusing on interconnected CpGs narrows the focus on a reduced set of entities, which has the benefit of alleviating the multiple testing burden (typical DNAm data contains hundreds of thousands of CpGs) and enhancing biological signal. Weighted correlation network analysis, also known as weighted gene co-expression analysis (WGCNA) has been used to identify clusters of co-methylated CpG sites (**modules**) that are associated with ageing in humans, both in single tissues, such as saliva[103], and across tissues such as brain and blood[45]. WGCNA constructs co-methylation networks by measuring the pairwise correlations between CpG sites (Figure 3), and then transforming this correlation into a measure of *proximity* (i.e. network interconnectedness) (Table 1)[44,104]. Highly interconnected CpGs are then clustered into modules that typically contain hundreds or

thousands of CpGs. To represent a sample's profile at each of these modules, DNAm levels at the CpGs contained within each module are 'summarized' using a data reduction technique (e.g. singular value decomposition). For a given module, each sample's profile is then represented by the module eigengene (ME), which is not a gene, but a 'summary' of the DNAm levels at the CpGs within that module. Module membership, which is measured by the correlation between a CpG and the ME, determines the connectivity of the CpGs in the module. CpGs with high module membership are considered highly connected 'hub' genes[44,104,105]. Modules that are present in multiple datasets represent common and robust CpG relationships that reflect true underlying biology and not technical noise. The interesting questions to ask are then: do older individuals display distinct DNAm patterns at those modules (i.e. highly correlated CpGs)[106]? What biological pathways do these modules reflect?

While WGCNA reveals whether interconnected CpGs become simultaneously hypo- or hypermethylated with age, it remains to be seen whether there are CpGs that become increasingly disconnected with age. This would translate to CpGs that are highly correlated in young, but poorly correlated in the elderly (Figure 3). This loss in connectivity may have important functional consequences on cellular function, as it would indicate a loss of coordinated gene expression and, therefore, potentially less efficient pathway activation. Few studies have assessed methylome connectivity using alternate methodologies[50,107]; in one such study comparing the correlation patterns of neighbouring CpGs, it was noted that there are reduced correlations in the methylomes of centenarians compared to neonates[50]. Since the interdependence between neighbouring CpGs ensures epigenetic fidelity during mitosis[96,98], neighbouring CpGs that lose their correlation with age could gradually introduce noise that is propagated during subsequent cell divisions. Coordinated DNAm changes can also arise from differential transcription factor (TF) binding during ageing[108], due to altered DNAm at TF binding sites compromising binding of TFs[108,109].

15

A

Raw DNAm matrix

	young$_1$	young$_2$	young$_3$	old$_1$	old$_2$	old$_3$
CpG$_1$	0.7	0.65	0.75	0.9	0.82	0.81
CpG$_2$	0.2	0.25	0.33	0.5	0.43	0.43
...	...	...	...	...	...	...
CpG$_m$	0.51	0.4	0.75	0.55	0.35	0.44

CpG$_1$ and CpG$_2$ are highly correlated

Positive correlation between CpG$_1$ and CpG$_2$

% methylation at CpG$_2$

% methylation at CpG$_1$

young
old

cor = 0.8
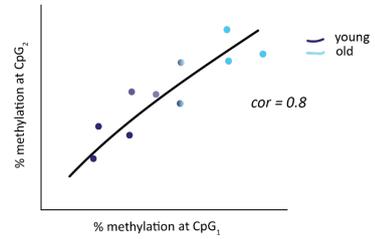
similarity matrix $s_{ij} = (0.5 + 0.5 \times |cor(MF_i, MF_j)|)$

soft-thresholding $a_{ij} = s_{ij}^{\beta}$

TOM

hierarchical clustering

	CpG$_1$	CpG$_{152}$	...	CpG$_m$
CpG$_1$				
CpG$_{152}$				
...				
CpG$_m$				

TOM

dynamic tree cut

module identification

highly connected "hub" CpG

	CpG$_1$	CpG$_{152}$	...	CpG$_m$
CpG$_1$				
CpG$_{152}$				
...				
CpG$_m$				
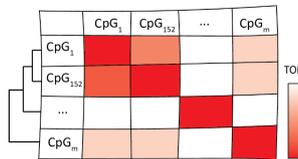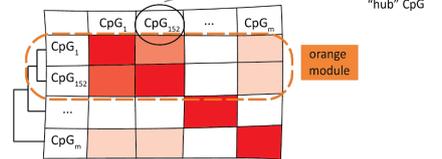
orange module

B

Raw DNAm matrix of CpGs in orange module

dimensionality reduction (e.g. SVD)

	young$_1$	young$_2$	young$_3$	old$_1$	old$_2$	old$_3$
PC$_1$	0.3	1.2	0.8	3.0	1.8	2.1
PC$_2$	-1.1	2.5	0.3	-5	-4	4
...	...	...	...	...	...	...

eigengene of orange module

Correlation between module eigengene and age

orange module eigengene

age

young
old

cor = 0.75

C

young

% methylation at CpG$_2$

% methylation at CpG$_1$

cor = 0.8

old

% methylation at CpG$_2$

% methylation at CpG$_1$

cor = -0.01

16
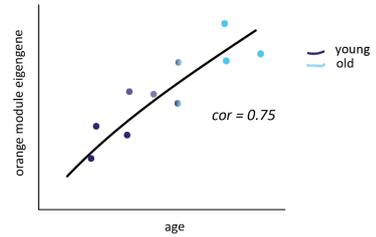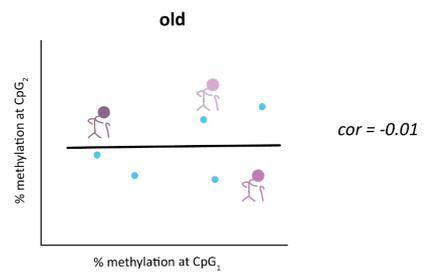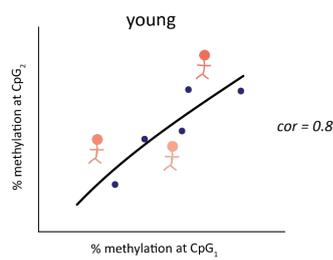
Figure 3 Correlation networks reveal connectivity in the ageing methylome.

A) Highly correlated cytosine–guanine dinucleotides (CpGs) that exhibit coordinated methylation changes with age (top panel; raw DNAm (DNAm) matrix and graph) can be summarized into highly informative modules (bottom right) using weighted gene correlation network analysis (WGCNA). The WGCNA package measures strength of correlation between CpGs, taking into account the methylation fraction (MF) for each sample at each CpG. First step involves constructing a similarity matrix from the raw DNAm matrix, a matrix of absolute values of correlation coefficients between CpGs. Note that s is similarity measure of MFs for ith and jth CpGs. This is followed by constructing an adjacency matrix, that uses a soft-thresholding parameter to measure strength of connection, whilst preserving underlying correlation relationship. Adjacency measure (a) for ith and jth CpGs calculated by raising the similarity measure to the power of β, the soft-thresholding parameter. For module detection, a topological overlap measure (TOM) is used to measure interconnectedness (proximity) of CpGs and is combined with unsupervised hierarchical clustering to organize CpGs with similar co-methylation dynamics. Gradient of TOM represents degree of interconnectedness, whereby white denotes low TOM (or low interconnectedness) and red denotes higher TOM (or higher interconnectedness). Modules then defined by 'cutting' branches of identified clusters, using methods such as Dynamic Tree Cut, which cuts branches of a module based on their shape. B) Module eigengenes are a mathematical construct used to summarize module connections into a single value using a dimensionality reduction technique such as singular value decomposition (SVD) (left). This is useful because modules often contain hundreds or thousands of CpGs. Module eigengenes can be correlated to specific traits of interest, such as age (right). Module eigengene that is positively correlated with age implies that all CpGs within that module become similarly hypermethylated with age. PC refers to the principal component; the output of the dimensionality reduction technique. The first PC is the module eigengene. C) Future work could investigate co-methylation relationships that become disconnected with age.

Table 1 Statistical tests and software for analysing age-associated changes in DNAm

Feature	Statistical test	Software packages	Advantages	Disadvantages	Ref
DMP (CpG site that changes in average DNAm with age)	Linear model	lmFit (limma)	Provides a genome-wide view of methylome shifts shared by individuals over time Useful for age-prediction algorithms Building blocks of epigenetic clocks Detected in relatively small sample sizes	Analysis excludes informative CpGs that change in variance as a function of age but not in average methylation	62,63,110

DMR (region of multiple, contiguous DMPs, such as of CpGs)	Different software use difference statistical approaches	minfi (DMRcate, blockFinder), bumphunter, comb-p, ChAMP (Probe Lasso)	DNAm up to ~500 bp is typically highly correlated DMR analysis reduces spatial redundancy DNAm altered over a region may offer better functional relevance (that is, directly linking to gene expression changes)	Isolated CpG sites that may be informative are discarded Methylation arrays offer unequal coverage, potentially limiting the number of important regions discovered	19,65–68,111–113
VMP (CpG site with a change in variability in DNAm with age)	Breusch-pagan test, White test to identify VMPs with age as continuous variable Bartlett's test, Levene's test, Brown-Forsythe test to identify VMPs between discrete groups	Lmtest (bptest), DiffVar (modelled off Levene's test)	Individual CpGs that change in variability with age inform of DNAm changes that differ between individuals over time	Large sample sizes across a broad age range required for sufficient statistical power Sparsity of large datasets in tissues other than blood could hamper the detection of VMPs across different tissues and cell types	19,62,73
VMR (genomic region of CpGs displaying variable DNAm	Different software use difference statistical approaches	minfi (DMRcate)	Genomic regions that are rich in VMPs may offer better functional relevance	Isolated CpGs that may be informative are discarded Methylation arrays have limited coverage, which	62,64,67

changes with age)				could exclude regions of importance	
Entropy (a single quantifiable measure of the methylome-wide DNAm changes for a sample at a point in time)	A Shannon entropy probability formula adapted for DNAm data	None currently available	A single entropy value provides a snapshot of the amount of epigenetic 'noise' or information loss for a single sample at a particular age Can be calculated for a specific set of CpGs to identify regions of high versus low methylation disorder associated with ageing.	Sensitive to batch effects, i.e., two samples of the same age from two different batches are difficult to compare	43,62,84, 85
Correlation networks (clusters of co-methylated CpGs for 'modules' that change with age)	Pairwise correlations are used to construct co-methylation networks; modules are identified using hierarchical clustering	WGCNA	Adopts a systems biology approach to explore the interconnectedness of the entire methylome, alleviating the multiple testing burden of individual CpGs Multiple datasets can be simultaneously analysed as a network-based meta-analysis technique	Technique does not assess modules that become disconnected with age	44,45,1 04
Epigenetic age (output of an epigenetic clock, which	Machine learning algorithms, such as elastic net regression	glmnet	Easy method of obtaining the epigenetic of a single sample at a point in time	A narrow measure of the methylome Age estimation depends on the datasets used to build the clock, as well as	106,114

estimates age from a subset of CpGs correlated with chronological age and age-related phenotypes)			Clocks that capture biological ageing parameters can predict healthspan and mortality risk Clocks used to assess the success of longevity interventions and rejuvenation experiments	the parameters that the clock was trained against.	

DMP, differentially methylated position; DMR, differentially methylated region; VMP, variably methylated position; VMR, variably methylated region; CpG, cytosine-guanine dinucleotide; DNAm, DNAm; WGCNA, weighted gene correlation network analysis

2. DNAm changes as hallmarks of ageing

Despite all individuals having identical rates of chronological ageing, there are marked disparities in individual rates of biological ageing[1,115,116]. As such, biological age represents the tissue and organismal functional status and age-associated risk of disease and disability[1,10,69,117], influenced by intrinsic factors such as sex and genetics, as well as the cumulative, lifelong exposure to environmental stimuli[79].

Here we review the evidence that DNAm is a hallmark of ageing against the following criteria outlined in the original publication[9]: i) changes arise during normal ageing; ii) the experimental acceleration of epigenetic ageing should lead to an acceleration of ageing, or iii) the experimental deceleration of epigenetic ageing should slow down ageing.

2.1. Age-related changes in DNA methylation and disease

DNAm is tightly linked to age-related diseases. Altered DNAm patterns have been observed in the left ventricles and blood of patients with CVD[36–38]. Individuals with

atherosclerosis show aberrant DNAm patterns in blood, endothelial and vascular smooth muscle cells (e.g. aorta and arteries)[118–121]. Changes in blood DNAm patterns are associated with hypertension[122], and blood pressure (BP) regulation[123]. DNAm alterations are observed in adipose, liver, and pancreatic islets in people with type 2 diabetes (T2D)[124–127]. There are DNAm changes in cartilage in people with osteoarthritis[128,129], and bone in people with osteoporosis and osteoarthritis (OA)[130]. Changes in DNAm in multiple brain regions have been reported in people with AD[31–35], with DNAm likely playing a functional role in AD pathogenesis[33]. DNAm changes in lens epithelium can cause the gene expression alterations associated with the development of cataracts[131]. Aberrant DNAm is a feature across multiple cancerous types[132–136].

Moreover, many age-associated DMPs overlap with the DNAm changes altered in disease. This has been reported in AD[34], T2D[124], and cancer[42,135–137]. There is also evidence to suggest a potential causal role of age-associated DNAm in endometrial cancer pathogenesis[138]. Epigenetic age acceleration (i.e. the difference between chronological age and epigenetic age estimated by epigenetic clocks) is associated with AD[139,140], dementia[139], BP[141], cancer[142,143], CVD[117], frailty[117,144], insulin[117,141], OA[145,146], and Parkinson disease[147]. Similarly, age-associated VMPs have been reported in cancer[62,148]. Higher entropy has been associated with chronic lymphocytic leukaemia (a cancer affecting the elderly)[149]. Moreover, an age-related co-methylation module present in blood and brain tissue contains CpGs associated with early AD[45].

Considering the integrity of the methylome is of paramount importance to the health of the individual, it is plausible that ageing phenotypes are the downstream consequences of disrupted DNAm patterns[150,151]. However, the precise mechanisms by which DMPs and VMPs individually contribute to ageing is not well defined. Two individuals with identical chronological ages (and therefore similar patterns at DMPs) may display divergent patterns across VMPs[62] (Figure 4). VMPs may reflect the introduction of

additional noise at key genomic regions that become naturally dysregulated with age. As such, identifying DMPs may pinpoint sites, genes, and pathways related to primary ageing, whereas VMPs may pinpoint sites, genes, and pathways related to secondary ageing[8]. If this holds true, the stochasticity introduced by VMPs may reflect mosaicism in ageing cells and tissues, reflecting the interindividual variation in risk of disease, in addition to the age-related changes that more or less track chronological age[69]. In support of this theory, age-associated epigenetic heterogeneity (or epigenetic 'chaos') is associated with an increase in age-related noisy gene expression (e.g. transcriptional variation during ageing)[152]. Moreover, a discernible relationship between VMPs and gene expression exists (i.e. VMPs associate with the expression of genes in *trans*)[62].

Heterogeneity in *biological* ageing, and therefore age-associated risk of disease, is mirrored at the epigenetic level. An individual whose biological age deviates from their chronological age, due to either positive (i.e. exercise, healthy diet) or negative (i.e. pollution) environmental influences, may therefore display patterns of VMPs that more closely resemble the methylome of a younger or older person (Figure 4). We propose that homoscedastic DMPs represent the DNAm changes that precisely track the chronological ageing process (i.e. primary ageing), whereas VMPs represent the DNAm changes that track the influence of intrinsic and environmental factors (i.e. biological or secondary ageing)[10,43,62,79,153]. This, however, remains a hypothesis that is yet to be tested. Although primary and secondary ageing are considered two distinct processes, at the DNAm level, the overlap that exists between DMPs and VMPs could, in theory, reflect biological processes part of "normal" primary ageing that are susceptible to erroneous, secondary changes. As such, these processes may not be completely independent of one another.

Evidence to support this hypothesis stems from the link between age-associated DNAm variability and cancer, whereby VMPs that undergo age-associated changes overlap with VMPs in healthy tissue that develops cancer in the future[148]. Ageing and cancer may share

a common origin[154], which hints at a causal role for DNAm in cancer initiation that may extend to other diseases[80]. For example, DNAm variability correlates with type 1 diabetes disease predisposition[155].

Despite numerous studies linking ageing with widespread changes to the methylome, some studies investigating the associations of DNAm with age-related disease have not yielded the same level of success, identifying only a handful of disease-related CpGs[38,122–124]. One possibility is that disease-affected tissues (e.g. pancreatic islets, heart) may be difficult to obtain in large sample numbers, and blood may not serve as a useful surrogate. Cell-type heterogeneity that is unaccounted for may also obscure findings[156,157]. Moreover, the bedrock of EWAS studies has been largely limited to identifying DMPs. Shifting focus to VMPs, entropy, and co-methylation networks in the context of human disease will help answer important questions about the biology of ageing, for example, whether age-related diseases display different rates of entropic decay; the co-methylation networks that become disrupted with diseases of ageing, and the underlying mechanisms; and whether VMPs display different patterns across cell types and tissues.

The lack of causality is another significant challenge in the interpretation of DNAm studies. The relationship between DNAm and gene expression is complex and improving our knowledge of how DNAm functions in different genomic contexts is necessary to accurately interpret how these DNAm changes may affect ageing and disease[158]. Confounding variables, including cell-type heterogeneity, genetic variation and reverse causation (i.e. DNAm may be the consequence of the phenotype and not the cause) can challenge the understanding of DNAm alterations[93,157]. While causal inference can be improved using multi-omics[19], there is scarcity in the availability of DNAm datasets with matched gene expression and other 'omics' data. Moreover, functional studies are not straightforward, since DNAm changes that arise during ageing are spread throughout the

methylome. Whether DNAm changes alter a phenotype at a single locus or multiple loci is therefore difficult to determine.
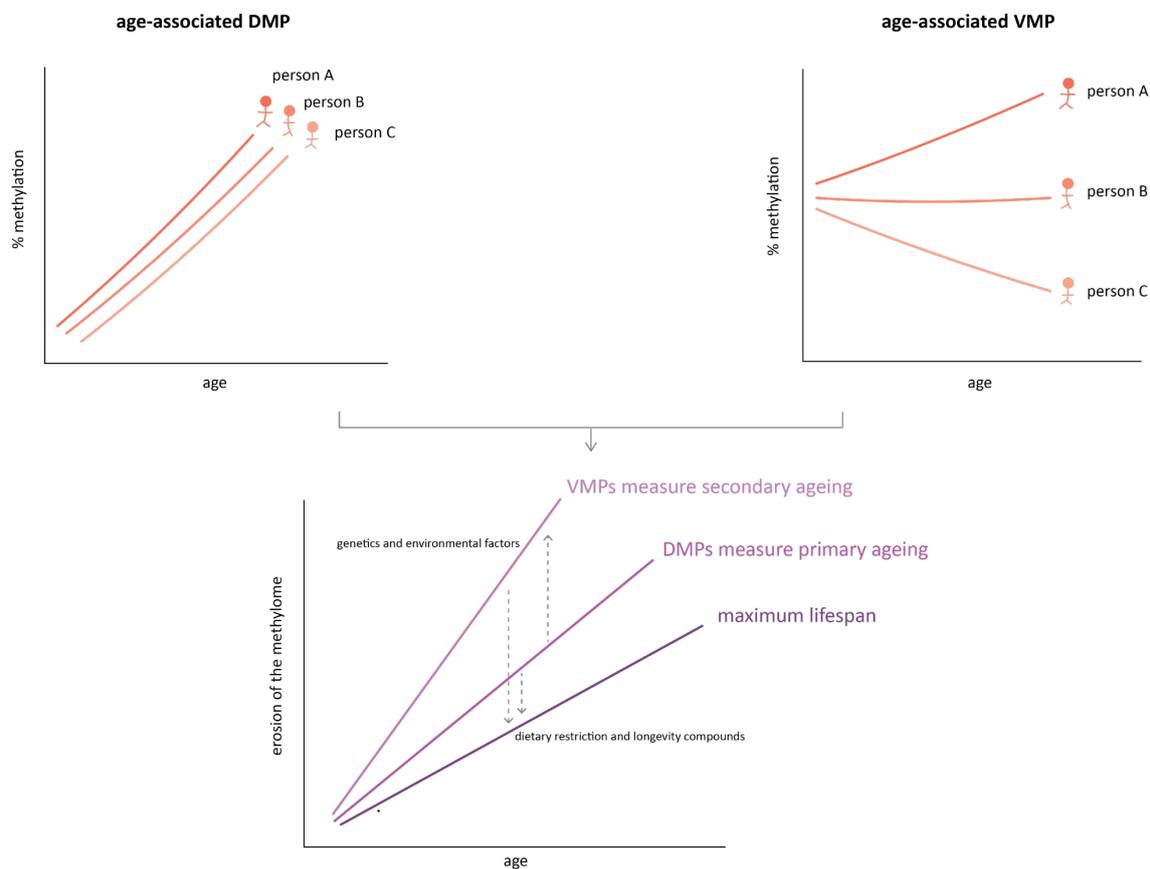
Figure 4 DMPs and VMPs reflect primary and secondary ageing processes.

Differentially methylated position (DMPs) represent changes in DNA methylation (DNAm) that are shared across the lifespan (top left). Variably methylated positions (VMPs) represent DNAm changes that differ across the lifespan (top right). Two individuals of same chronological age may have similar patterns of DMPs but display divergent patterns across VMPs. DMPs likely track a primary ageing process (intrinsic age-related deterioration occurring with chronological time), whereas VMPs likely track a secondary ageing process (additional and heterogeneous age-related changes accelerated or decelerated by environmental influence, genetics or disease). VMPs may therefore represent biological ageing at the epigenetic level. 'Maximum lifespan' indicates that longevity interventions delay DNAm changes associated with primary and secondary ageing. Dashed arrows indicate direction of effect on the methylome. For example, dietary restriction and longevity compounds attenuate age-associated erosion of the methylome. Conversely, unfavourable environmental influences and disease can accelerate erosion of the methylome.

2.2. Longevity-promoting interventions

Should the erosion of the methylome play a causative role in age-related decline, then such damage accumulation should, in theory, be alleviated or reversed by longevity-promoting interventions. This concept underlies certain dietary, drug, and reprogramming strategies.

Dietary restriction

Dietary restriction (DR) entails limiting the quantity of food or the timing of food intake. Calorie restriction (CR) is the chronic reduction of calories below standard without malnutrition and is a well-known longevity-enhancing strategy[159–161]. In humans, CR reduces the risk for age-associated diseases, including T2D, cancer, and CVD, and CR has consistently been shown to extend both healthspan and lifespan in mammals[159,160,162]. Intermittent fasting or time-restricted feeding seem to be equally important to promote healthy ageing[159]. Because CR and fasting are often combined during DR, there is confusion about which is necessary to reap benefits.

Several animal studies have investigated the effect of CR on the DNAm signatures of ageing, reporting that CR shifts the methylome to a younger profile by attenuating the age-related DNAm alterations in several tissues, such as blood[89,162], liver[55,163], kidney[164] and the hippocampus[165]. For instance, a study in liver tissue of female mice reported that CR increased methylation at DMPs that become hypomethylated with age, and decreased methylation at DMPs that become hypermethylated with age[55]. Another study in the blood of rhesus monkeys and mice reported that CR counteracts epigenetic drift, shifting the methylation patterns of CR old mice and monkeys to resemble their younger counterparts[162]. Moreover, CR decelerates epigenetic ageing by directly affecting the DNAm changes that underpin the epigenetic clock[90,162]. Co-methylation networks exploring the effects of CR on conserved ageing modules in multiple species and tissues

have also reported that CR led to DNAm changes in the opposite direction to age (i.e. the CpGs in the module that become hypermethylated with age, decrease in DNAm following CR)[166]. The most recent data in humans from the Comprehensive Assessment of Long-term Effects of Reducing Intake of Energy (CALERIE) trial, has reported that humans who have who restricted their caloric intake by 25% for a 2 year period had a lower epigenetic age measured by the DunedinPACE DNAm clock compared to ad libitum fed controls, but it did not lead to significant changes in epigenetic ageing measured by other biological clocks, such as PhenoAge and GrimAge[167]. Albeit modest changes in epigenetic ageing parameters, long-term follow-up on these participants is needed to establish the effects of the CR intervention epigenetic ageing.

Although the evidence outlined is limited, it is suggestive that calorie restriction promotes longevity at least in part, by slowing down the epigenetic changes associated with primary ageing. Furthermore, more research is required to determine whether the influence of calorie restriction on DNAm is what underpins calorie restriction's healthspan- and lifespan-extending properties.

Longevity drugs

There are promising longevity drugs to slow ageing. For an extensive overview of these compounds and their influence on all the hallmarks of ageing, please refer to a detailed review[168]. Here, we focus on drugs that have been shown to act directly upon the methylome and / or its machinery.

Rapamycin is a lifespan-extending compound that inhibits the mammalian target of rapamycin complex 1 (mTORC1), a master regulator of metabolism and cell growth[168]. In mice, repression of mTOR under rapamycin treatment slows ageing by extending lifespan[169,170], as well as healthspan, by protecting against age-related diseases, such as AD[171], cancer[172] or T2D[173]. One mouse study reported that while rapamycin rescues many

ageing traits, some ageing phenotypes remained unaltered or worsened[174]. Moreover, not all tissues respond equally to treatment[170]. In other mammals, rapamycin improved healthspan parameters in dogs, including cardiac function[175], and did not have marked metabolic consequences following long-term treatment in adipose and liver tissue of nonhuman primates[176]. Notably, rapamycin offers geroprotective properties such as improved immune function in elderly humans[177]. At the epigenetic level, rapamycin slows the epigenetic ageing of human keratinocytes in culture[178]. Age-associated DNAm changes are supressed in mice livers after treatment with rapamycin. Interestingly, rapamycin seems to be less effective than CR[55,90]. Rapamycin has no statistically significant effect on the epigenetic age of marmosets according to a marmoset-specific epigenetic clock[179].

Metformin is a drug that exerts anti-ageing effects by activating nutrient sensors, such as AMP-dependent kinase (AMPK), and inhibiting mTORC1. Metformin is widely used to treat T2D; however, it seems to protect against ageing phenotypes, such as cancer and inflammation. Animal studies support metformin as a promising drug for extending healthspan and lifespan[180]. The Targeting Ageing with Metformin (TAME) Trial has been initiated to investigate the effect of metformin in delaying age-related diseases[168]. A study in humans that was designed to reverse aspects of ageing in thymus tissue reported a cocktail of drugs, including growth hormone, metformin and dehydroepiandrosterone (DHEA), that reversed epigenetic ageing and increased predicted human lifespan[181]. However, the results from this study are inconclusive since no control or placebo was used, and the sample size was limited (n=9). Preliminary evidence from another small study reported metformin slows down epigenetic ageing in patients with T2D[182].

Nicotinamide adenine dinucleotide (NAD+) plays a central role in metabolism, acting as an essential co-enzyme for redox reactions[183]. In addition, NAD+ is a co-factor for NAD+-dependent enzymes, such as sirtuins, CD38, and poly(ADP-ribose) polymerases

(PARP). In humans, NAD+ levels have been shown to decline with age in the skin[184], brain[185], liver[186] and blood[187]. Dietary supplementation with precursor NAD+ molecules, nicotinamide riboside (NR) and nicotinamide mononucleotide (NMN), has emerged as a potential therapeutic strategy for ameliorating age-associated diseases, and extending healthy lifespan[168,188].

NAD+ is used by other molecules to coordinate epigenetic modifications, including DNAm. For example, NAD+ is required for the activity of epigenetic regulators, such as SIRT1, and a decline in NAD+ causes changes to histone modifications, altering chromatin structure and gene expression[183,189]. Moreover, SIRT1 affects DNAm at regions that become specifically altered with age[190]. To our knowledge, no studies to date have explored the influence of NAD+ enhancers on the global DNAm signatures of ageing.

Alpha-ketoglutarate (α-KG) is a key metabolite in the Krebs cycle, but also assists in demethylation as a co-factor for TET enzymes[191]. In mice, supplementation with α-KG decreases the severity of ageing phenotypes, such as osteoporosis[192], delays the age-related decline in fertility[193], and extends both healthspan and lifespan[194]. In mammals, altered levels of α-KG during ageing may alter the activity of TET enzymes required for stable maintenance processes, and supplementation with α-KG may increase its availability to act as a co-factor for TET enzymes[69,195]. A recent study in humans reported that a cocktail containing α-KG and vitamins, known as Rejuvant®, taken for four to ten months decreased biological age, measured by the TruAge clock, by an average of 8 years[196]. More robust evidence is needed to corroborate these results.

Spermidine is a naturally occurring polyamine that plays an essential role in metabolism[168,197]. Spermidine synthesis declines during ageing in both humans and mice[198]. In mice, spermidine supplementation extends lifespan and healthspan, offering cardioprotective benefits[199] and preventing against liver fibrosis and hepatocellular

carcinoma[200]. In humans, spermidine intake is correlated with lowered blood pressure and incidence of heart disease[199], and is linked to lower mortality risk by up to 5.7 years[201].

One mechanism by which spermidine may promote healthy ageing is through alterations in DNAm[198]. In mice, lifelong consumption of a polyamine-rich diet inhibits aberrant age-associated DNAm[202]. Increases in polyamine metabolism increase the availability of essential substrate S-adenosylmethionine (SAM), which favourably alters the activity of the DNMT enzymes that maintain patterns of DNAm[202–204]. These effects seem to be driven by increases in spermine levels (a derivative of spermidine), and more research is required to determine the influence of spermidine supplementation, particularly in humans, on the DNAm signatures of ageing.

Epigenetic rejuvenation

DNAm is at the core of epigenetic reprogramming experiments. Rejuvenation experiments 'reset' epigenetic patterns to youthful states by reprogramming the age of a cell. The overexpression of Yamanaka factors (OCT4, SOX2, KLF4 and MYC (OSKM)) is the most common strategy for cellular reprogramming and epigenetic rejuvenation[205]. By inducing these four factors, somatic cells can regain pluripotency and reset their DNAm age to zero, as measured by the pan-tissue epigenetic clock[15]. However, reprogramming a somatic cell to a pluripotent cell also leads to a loss of original cell identity. To overcome this, transient reprogramming experiments (i.e. *partial reprogramming*), have been introduced to achieve rejuvenation without loss of somatic identity[205].

In mice retinas, *in vivo* overexpression of just three Yamanaka factors (OSK) demonstrated that the global DNAm signatures of ageing that arise both in normal ageing and following injury, can be reversed to the point where cells do not lose their identity, and youthful DNAm signatures can be recovered[206,207]. Moreover, this yields a younger

transcriptome and restores youthful vision in old, vision-impaired mice. Interestingly, TET demethylating enzymes were necessary for the reprogramming to occur[206]. Similarly in human dermal fibroblasts, *in vitro* transient reprogramming using OSKM induces a significant reduction in DNAm age of ~30 years, measured by the pan-tissue epigenetic clock, and the transient reprogramming also rescues the transcriptional and morphological features of a youthful fibroblast[208]. These experiments highlight a very important aspect of epigenetic ageing, that is, youthful epigenetic information can be recovered, and maintenance methylation enzymes are required to recover this information. *How* the 'lost' information is recovered to reprogram the methylome is not completely understood, but may involve the persistent epigenetic memory at enhancers that allows cells to rescue their initial identity[208].

Exercise

Another promising healthspan-promoting strategy is exercise. The effect of exercise on the global DNAm signatures of ageing in humans is largely understudied, despite the plethora of healthy ageing benefits exercise has on offer[209]. One pioneering study in a large cohort of human skeletal muscle samples has recently shown that individuals with a higher level of baseline fitness display younger epigenetic and transcriptomic profiles, and that exercise training the shifts the methylome and the transcriptome to a more youthful molecular profile[210]. This work included well controlled exercise cohorts such as the Gene SMART (Skeletal Muscle Adaptive Response to Training)[211], with a large biobank of muscle and blood epigenetic data across the lifespan for healthy males and females. Future work should elucidate the type and duration of exercise training that would confer the greatest longevity benefit, and whether the rejuvenating effects are seen in tissues other than muscle.

In summary, DNAm alterations meet, to some degree, the criteria to be considered a hallmark of ageing: i) DNAm changes arise during normal ageing in arguably every

tissue, cell and species; ii) DNAm ageing can be accelerated experimentally using model organisms, and is associated with many age-related phenotypes in humans and animal models alike; iii) longevity interventions rescue age-associated DNAm changes in model organisms; and iv) resetting the methylome is necessary to reverse the age of a cell in mouse models and human cells. To strengthen the case that DNAm is a true hallmark of ageing in humans, more evidence to support longevity interventions (such as exercise training and certain diets) or unfavourable lifestyle changes, as modifiers of the pace of DNAm ageing is warranted. Importantly, conclusions can only be derived from the available evidence, which is sparse in experiments designed to uncover the mechanisms of ageing. Thus, it is premature to say that DNAm alterations are a primary hallmark, and not caused by another feature of ageing (i.e. molecular damage). Until such a time that this evidence becomes available, it remains inconclusive whether DNAm alterations are a cause or a consequence of ageing.

3. Origins of age-related DNA methylation changes

3.1. Genomic location of age-related methylome changes

EWAS have identified *hyper-* or *hypo*-DMPs associated with ageing in multiple human tissues and cell types, such as whole blood, monocytes, mesenchymal stem cells, buccal, brain, kidney, cervix, lung, liver, adipose, pancreas, placenta, small intestine, bladder, breast, thyroid, skin, and skeletal muscle[29,42,57,59,124,135,212–215]. Although tissues and cells have unique DNAm ageing signatures, there are conserved DNAm changes across cell types during ageing[61,80,216].

Gains in methylation with age accrue more frequently in CpG-rich regions, as these regions tend to be naturally unmethylated. Specifically, hyper-DMPs occur in promoters of key developmental genes harbouring both active and inactive histone marks (known

as bivalent chromatin domains) in embryonic stem cells (ESCs), as well as in regions actively repressed by Polycomb complexes[42,57,59,213]. Polycomb group proteins form Polycomb Repressive Complexes (PRCs) that associate with DNA and chromatin to control developmental regulators[217,218]. In ESCs, PRCs maintain pluripotency by repressing developmental genes that are preferentially activated upon cellular differentiation[217]. Hypermethylation may therefore be associated with decreased plasticity due to the permanent silencing of genes required for differentiation[59].

In contrast, hypomethylation happens preferentially in CpG regions of low density located at introns and intergenic regions[135]. Hypo-DMPs generally harbour active histone marks that are associated with enhancers[71,135]. Compared with gains in methylation, losses of methylation are less conserved across tissues, and functionally enriched for disparate pathways[57,135]. For example, in a study comparing blood, brain, kidney, and muscle tissue, only kidney and blood hypo-DMPs were enriched for a similar pathway related to immune response, and skeletal muscle hypo-DMPs were strongly enriched for muscle-specific pathways[57]. This is possibly owing to the role of enhancers in tissue-specific gene expression[17,57].

Convincing evidence recently published from a large meta-analysis of age across multiple species and tissues corroborates results from individual EWASs[58]. This study confirmed that hypermethylation in CpG islands is a conserved phenomenon across tissues, and the overwhelming majority of universal hyper-DMPs are proximal to genes encoding TFs that bind to Homeobox domains. These TFs are involved in central developmental processes, and many of the hyper-DMPs are located in regions targeted by the Polycomb machinery and in bivalent chromatin domains[58]. Moreover, there was a greater enrichment of the number of hyper-DMPs across tissues, compared to the number of hypo-DMPs. This echoes previous findings that hypo-DMPs may reflect tissue-specific

operations. For example, hypo-DMPs in brain and cortex tissues were enriched for circadian rhythm pathways, but not in other tissues such as skin or blood[58].

Although VMPs are less well characterized, a noteworthy study in blood reported increases in DNAm variability with age at bivalent regions and sites residing in Polycomb repressed regions[62]. In alignment with hyper-DMPs, VMPs that increase in both mean methylation and variance were strongly enriched for CpG islands (CGIs), and VMPs that decrease in mean methylation but increase in variability were enriched for non-CGIs[62]. In another smaller study of mesenchymal stem cells, increases in DNAm variability with age are preferentially located at non-CGIs and intergenic regions[71]. VMPs do not appear to be driven by changes in cell composition during ageing[62]. Interestingly, a significant proportion of VMPs are associated with gene expression changes in *cis*, which relate to pathways such as neuron differentiation and neuron development, and VMPs associated with gene expression in *trans* (i.e. the CpG and gene are located on different chromosomes or the same chromosome but more than 5 Mb apart) correspond to pathways such as metabolism, apoptosis, and the DNA damage response[62]. While there is evidence from cancer studies that 'epigenetic drift' occurs in other healthy tissues, such as the colon, these studies were not focused on identifying and characterising age-related VMPs[219,220]. Clearly, more studies using large sample sizes across a broad age range are needed to characterise VMPs in other tissues.

Co-methylation network analysis corroborates the results for the hyper-DMPs from EWAS of age in mammals[58,166]. WGCNA was recently used to cluster co-methylated CpGs across multiple tissues and in 331 eutherian species, including humans. Pooling DNAm data from ~15,000 samples, 55 co-methylation modules were identified, several of which were associated with biological traits, such as chronological age, sex, and maximum lifespan[166]. To harmonise chronological ageing between species, *relative* age was used as the ratio between age of organism and maximum lifespan of the species (e.g.

the relative age of a 40-year-old human is 0.33 because maximum human lifespan is 122.5 years). One conserved module in mammalian tissues was positively correlated with the relative age of all mammalian species. This means that the CpG sites in this module are correlated with each other and are collectively hypermethylated with age across all tissues and species. This module was enriched for pathways such as embryonic stem cell regulation, axonal fasciculation, angiogenesis, and diabetes-related processes. An earlier study of co-methylation networks in blood and brain tissue of humans had reported similar findings[45]. In both studies, the top "hub" CpGs (the most highly connected CpGs in the module occupying central network positions) reside in genomic regions adjacent to Polycomb targets and repressive histone marks. This is indicative of a conserved ageing phenomenon across tissues and species, whereby a subset of highly correlated CpGs that are unmethylated in young, become methylated with age at distinct regions that control development[58,166].

3.2. Mechanisms of age-related methylome changes

Pin-pointing the precise upstream mechanisms that cause the methylome to age (and how fast) is challenging. Several tenable theories have been put forward.

3.2.1. *The epigenetic clock theory of ageing*

The methylome seems to reflect an innate ageing process that is intricately linked with development and differentiation[221]. In support of this hypothesis, epigenetic clocks estimate with remarkable accuracy chronological age in arguably every species, tissue and cell type, from prenatal mammalian tissue to tissues of the oldest living mammals on earth[58,61]. Furthermore, the epigenetic clock can provide an accurate estimation of gestational age, which involves a highly coordinated developmental process with little noise perturbing the system[222], and it can even measure rejuvenation events that occur

during embryogenesis, marking the beginning of organismal ageing (ground zero) that continues uninterrupted throughout life (Figure 5)[94,223]. As such, the epigenetic clock reflects an ageing process that is not monotonous, but instead aligns with the nonlinear periods of growth and development[221,224,225]. For example, the epigenetic clock is accelerated during the first few years of life, and slows after puberty, paralleling the human developmental process[226]. To this end, the fundamental processes established in early life, are the same processes that ultimately cause an organism to age[221,226].

A tenable hypothesis to explain this phenomenon is presented in a recent review[226]. Albeit limited in mechanistic evidence, there is a convincing argument to support a link between stemness and epigenetic ageing, whereby the 'ticking' of the epigenetic clock is proposed as a measure of asymmetric stem cell or progenitor cell divisions (i.e. when tissue stem cells differentiate into non-stem cell cells) (Figure 5), or the increase or decrease in stem cell numbers in different tissues[226,227]. Epigenome-wide investigations substantiate these findings, evidencing the dysregulation of developmental genes that govern cell identity as a conserved feature of mammalian ageing[42,57–59,213]. Simply, the methylome is precisely altered in the genomic locations that function to preserve stem cell identity and function. In support of this hypothesis is the recent evidence from profiling epigenetic age in murine single cells, whereby epigenetic ageing is precisely tracked in hepatocytes (i.e. epigenetic age increases in old versus young hepatocytes), whereas muscle stem cells display minimal changes in epigenetic age[94], suggesting that the epigenetic clock ticks when stem cells are stimulated to divide[93].

In a recent study that analysed the methylome and transcriptome of CpGs from four epigenetic clocks, it was reported that the DNAm of some of the clock CpGs associate with gene expression in *trans,* and the genes involved play a role in T cell processes[228]. Upon further investigation, it was proposed that the differences in DNAm between immune cells, namely naïve and activated T and NK cells, may be driving the epigenetic

clock progression. However, these conclusions were drawn from only a subset of clock CpGs, and it remains unknown what percentage of the clock's predictive capability is attributable to T and NK cells.

While the maintenance of DNAm patterns established during development is key to maintaining youthful epigenetic states and robust cell identity, this is challenging due to the plasticity of DNAm[229]. DNAm dynamically responds to environmental cues, repairing DNA, and participating in transcription and replication. Consequently, failures in DNAm maintenance and tissue homeostasis can have detrimental effects on the organism, leading to ageing and age-associated disease[227,229]. Molecular damage, metabolism or activated developmental programmes, for example, can alter the methylome of adult stem cells, leading to stem cell dysfunction, and the subsequent decline in tissue and organ function[227]. The clock theory of ageing therefore proposes that the widespread decay of the methylome reflects an 'epigenetic maintenance system' that is operating to support development, cell differentiation, and maintain cell identity[61,221].

TET and DNMT enzymes are essential to maintain DNAm patterns, and their importance is highlighted by genetic disorders harbouring mutations in their genes. For example, mutations in *DNMT3A* and *TET2* are implicated in the early-onset of haematological malignancies in the elderly[230]. Unlike other Mendelian disorders, conditions associated with *DNMT1* mutations uniquely display a gradual and progressive onset of symptoms, such as hearing loss and dementia, that are absent in youth but manifest in adulthood[231,232]. Conversely, supplementation with longevity drugs, such as α-KG and polyamines, promote methylation maintenance by altering the activity DNMT and TET enzyme substrates[195,203]. Evidence from epigenetic rejuvenation experiments also suggests that TET and DNMT enzymes are necessary for reprogramming an aged cell to a youthful epigenetic state[206].

In summary, the epigenetic maintenance system responsible for primary ageing is also susceptible to the gradual accumulation of errors or biological 'noise,' leading secondary ageing changes. In consequence, biological ageing is proposed to be the unintended consequence of both developmental and maintenance programmes (i.e. the ability of the stem cell niche to maintain tissue homeostasis)[221,226]. However, key questions remain: how do epigenetic enzymes lose their ability to perform their function with advancing age? What mechanistic understanding can be determined from single-cell epigenetic age analyses? And perhaps more importantly, is there a single upstream mechanism driving these changes or is it multifaceted?

3.2.2. Metabolic signalling and chrono-epigenetic ageing

The circadian system is an autonomous internal oscillator that provides rhythmic coordination to physiological, behavioural, and metabolic processes, synchronizing the external environment with internal processes to maintain organismal health[233]. Ageing is accompanied by the loss of robust circadian oscillations and the desynchronisation of these processes and has been linked with metabolic disorders and multiple ageing pathologies in humans and in mice[234,235]. On the contrary, interventions that restore circadian rhythms in rodents are associated with longevity[234].

'Chrono-epigenetics' is the umbrella term used to describe the circadian dynamics of the epigenome, which affect histone modifications, chromatin architecture, and DNAm[236]. Specifically, CpGs exhibit circadian behaviour, which is facilitated by the rhythmic action of DNMT and TET enzymes[236]. Evidence from mouse studies has shown that the light entrainment of the circadian clock is dependent on DNAm, supporting the role of DNAm as a mediator between the external environment and internal rhythms[237]. Moreover, experiments in liver and lung tissue of mice, as well as human neutrophils, have shown that CpGs exhibit rhythmic oscillations that overlap significantly with CpGs that are differentially methylated with age[238,239] (Figure 5). One caveat to mention is that

even if DNAm oscillations are detected in a "purified cell type", the overlap with age-associated DNAm could reflect the subtle shifts in cell subtypes if the adjustment for cell-type heterogeneity is imperfect due to unaccounted cellular heterogeneity. Enrichment of oscillating CpG was observed at distal regions and enhancers of both highly expressed and circadian genes. Notably, oscillating CpGs were shown to *precede* age-dependent CpG changes, and the amplitudes of the oscillating CpG correlated with the magnitude of the linear age-dependent change[238,239] (Figure 5). The authors characterised CpGs that oscillate consistently between individuals as DMPs, and those under more lenient control (and more susceptible to environmental perturbation) as VMPs[236]. In support of this, there are linear age-dependent DNAm changes in the *CLOCK* gene, which is one of the 353 CpGs in Horvath's pan-tissue clock[61].

The role of enhancers in coordinating rhythmic expression may shed light on shared mechanisms between circadian disruption and epigenetic ageing. In mammals, a "core clock" involving four key factors (transcription factors CLOCK and BMAL1, and genes cryptochrome (*Cry*) and period (*Per*)) regulate the 24-hour cycles via a transcription-translation feedback loop (TTFL)[240,241]. Cell-autonomous peripheral clocks drive tissue-specific rhythmic gene expression, involving the coordinated activity of cell-specific enhancers[242–244]. Age-associated hypomethylation at enhancers of highly expressed genes may therefore involve the reduced precision of methylation enzymes DNMT and TET that is linked to circadian disruption, culminating in downstream consequences of altered gene-enhancer interactions (potentially disturbing co-methylation networks) and gene regulation[69].

If disruption of the chrono-epigenome is a proximate cause of ageing, it is not clear what is driving these decreases in circadian oscillations in such a precise manner that parallels the 'ticking' of the epigenetic clock. We hypothesise that metabolic processes are part of this equation.

The circadian clock and metabolism operate in a reciprocal relationship (i.e. the rhythmicity of metabolic processes are an output of the clock, and metabolic signals and states feed back onto the clock)[240]. Many metabolic genes that oscillate in tissue-specific rhythms participate in the same metabolic pathways implicated in both ageing and longevity[234,240] (Figure 5). For example, DR extends lifespan through beneficial effects on nutrient-sensors that are under circadian control, including the inhibition of pro-ageing factors insulin-like growth factor-1 (IGF-1), and the activation of longevity factors AMPK, sirtuins, nicotinamide phosphoribosyltransferase (NAMPT), and forkhead transcription factors (FOXOs)[159,245]. Similarly, "longevity drugs" (e.g. rapamycin, metformin, α-KG, spermidine) mimic the longevity effects of DR by targeting the same metabolic pathways[234]. Animal studies have demonstrated that CR can restore the tissue-specific circadian rhythmicity of key metabolic genes, suggesting that longevity benefits involve the restoration of healthy circadian cycles[234]. Polyamines also show circadian rhythmicity and in turn, regulate the circadian period. In mice, supplementation with spermidine counteracts the age-associated decline in circadian cycles by regulating the interaction between the core clock factors[246].

A vital link between ageing, metabolism, and circadian rhythms is NAD+: the central catalyst of metabolism. A reciprocal relationship between NAD+ and circadian rhythm exists, which is meditated by the rate-limiting enzyme NAMPT[183]. In mice livers, NAD+ reprograms metabolic and stress-response pathways through the restoration of circadian function that declines with age[247]. NAD+ acts through SIRT1, an NAD+ dependent enzyme, which interacts with the core clock components to restore robust circadian oscillations[247] (Figure 5). Although it was not demonstrated in this experiment, PARP1 is another NAD+ dependent enzyme that has been linked to entrainment of the circadian clock[248,249]. Whether PARP1 also participates in the reprogramming of circadian rhythms is unknown. Beyond supplementation, NAD+ is increased by DR, exercise, dietary interventions, and healthy circadian cycles[183]. Why NAD+ declines with age is not clear,

but likely involves multiple pathways, such as altered metabolic activity and inflammatory processes. PARP enzymes also have a dual role in assisting with DNA repair processes, likely contributing to the depletion of the NAD+ pool (and subsequent SIRT1 activity) during ageing[183,250].

Considering the above evidence, is it possible that daily metabolic stress may contribute to gradual deterioration of circadian function that explains, in part, the age-dependent DNAm changes associated with primary ageing? This would lend itself to the assumption that these changes would then be exaggerated in situations of more severe metabolic and circadian disruption, driving variability in the rates of epigenetic ageing. Future work should determine how different stimuli (i.e. light, nutrients, or DNA damage) may trigger the responses of temporal methylation and demethylation events through DNMT and TET activity, which could explain the role of maintenance enzymes in modulating this process. Moreover, whether longevity interventions (i.e. calorie restriction, exercise, NAD+ enhancers) can restore CpG oscillations to youthful states would contribute significantly to our understanding of ageing from a chrono-epigenetic-metabolic perspective.

3.2.3. *DNA damage and the relocalisation of chromatin modifiers*

Overlapping with the above theories is the emerging evidence that DNA damage, particularly in the form of double-stranded breaks (DSBs), may be driving the ageing process[207,251]. DNA damage is responsible for a variety of DNA lesions, arising from both exogenous (e.g. ultraviolet (UV), chemicals, x-rays) and endogenous sources (e.g. oxidative stress, metabolic stress, replication errors, spontaneous hydrolytic reactions)[9,251]. The accumulation of DNA damage leads to a collection of molecular consequences, such as genomic instability and epigenetic alterations, that underpin a spectrum of ageing phenotypes[251].

DSBs are a particularly lethal form of DNA damage that trigger the sophisticated DNA damage and repair (DDR) machinery[251]. DSB signals recruit epigenetic modifiers, such as sirtuins and PARP enzymes, from their native loci to repair sites and remodel the epigenetic landscape[252,253] (Figure 5). Epigenome integrity is restored after DNA repair, a crucial operation that preserves cell identity and function[229]. However, according to the relocalisation of chromatin modifiers hypothesis, during ageing, the incomplete return of these epigenetic modifiers to their original positions introduces noise into the epigenome in predictable ways, such as at key developmental regions that govern cell identity, and further increases susceptibility to more DSBs[207]. Taken together, in addition to their role in NAD+ metabolism, sirtuins and PARP1 also form part of the DDR machinery, suggesting the regulation of DNA repair and cellular metabolism are coordinated[254] (Figure 5). In worms and mice, PARP1 is chronically activated during ageing, potentially due to overactivation of DNA repair enzymes. This is important because the increased requirements for DNA repair that activate PARP1, deplete NAD+ pools (and inhibit sirtuin activity)[250], which are required for maintaining healthy metabolic and circadian processes.

The role of sirtuins in the relocalisation of chromatin modifiers process has been described in yeast and mice[255]. More recently it has been experimentally shown in mice to cause the loss of cell identity and accelerate the epigenetic clock[207]. Notwithstanding the need for more research to determine whether the relocalisation of chromatin modifiers contribute to mammalian ageing, this is compelling evidence, which could explain how 'random' damage can induce a precise and predictable pattern of DNAm changes.

The specific mechanism that causes DSBs to accelerate the epigenetic clock is unclear, but it may involve the relocalisation of methylation enzymes to sites of DNA repair[207] (Figure 5). DNMTs, including DNMT1 and DNMT3B, along with other chromatin modifiers, SIRT1, PARP1, and Polycomb group proteins, are recruited to DSBs and sites

of oxidative damage[256,257]. The localisation of these repressive proteins may inhibit transcription at damage sites to prevent interference with repair. Most DNAm alterations that occur during repair are likely transient and can be restored through demethylation, however, chronic DNA damage (i.e. during ageing) may lead to DNAm modifications that cumulate with age[258]. Specifically, promoter regions are susceptible to persistent repressive DNAm[256]. It has been postulated that transcription protects promoter regions from silencing, and even transient transcription inhibition would lend promoters more vulnerable to more stable silencing events[258]. Interestingly, CpG-rich regions are preferentially targeted by the damage-induced complex, where they are translocated away from CpG-poor regions, which may explain the age-associated hypermethylation at CGI promoters, and conversely hypomethylation at CpG-poor regions[257,259]. It has also been hypothesised that Polycomb group target genes are susceptible to hypermethylation due to age-related degradation of Polycomb machinery. This may lead to PRCs being unable to recognise and target unmethylated CpG-rich regions. Unmethylated CpG-rich regions ordinarily protected by PRCs become increasingly accessible to DNMT3A and DNMT3B *de novo* methylating enzymes, facilitating increased methylation at these sites[260]. There is also altered patterning of Polycomb histone marks during ageing, such as H3K27me3, which changes in a context-dependent manner. How Polycomb histone marks play a role in this process may depend on specific loci and cells and remains an interesting avenue for exploration.

A

predicted age

accelerated epigenetic ageing
in early years

old age

puberty

rejuvenation event
during early embryogenesis

birth

chronological age

'ground zero'

Exit from embryonic stem cell stage (ageing begins)
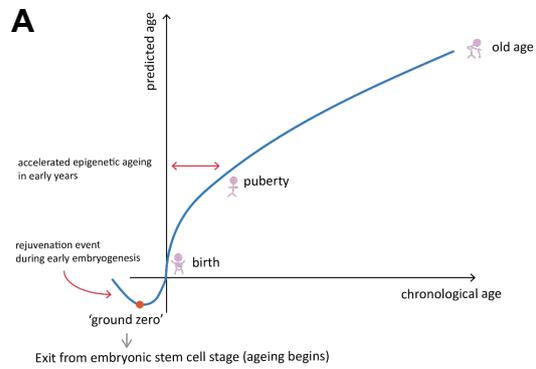
Epigenetic clock 'ticks' during asymmetric stem cell divisions

tissue stem cell

asymmetric
division

stem cell

transit-amplifying cell
(non-stem cells)

differentiated cells

**DEVELOPMENT &
MAINTENANCE**

**CIRCADIAN
RHYTHM**

**DAMAGE &
REPAIR**

METABOLISM

B

young
robust CpG oscillations

% meth

ageing

old
dampened CpG oscillations

NAD+ enhancers?
calorie restriction?
longevity drugs?
exercise?

Oscillating CpGs precede age-dependent CpG changes

% methylation

age

C

birth

ageing

↑ **DNA damage**

old age

sirtuins
PARP1

↓ NAD+

Transcriptional silencing

DSB break repair

DNMT1

DNMT3B

ageing

SIRT1

PARP1

PRC

relocalisation of
chromatin modifiers

D

young

mTOR

NAMPT

AMPK

IGF-1

ageing

old

calorie restriction
NAD+

NAD+ regulates circadian and metabolic processes

↑ NAD+

NAD+ restores circadian function via SIRT1

SIRT1

CLOCK   BMAL1

promoter

NAD+ declines during ageing

NAD+

age

increasing NAD+ extends lifespan

43

Figure 5 Proposed mechanisms of epigenetic ageing.

Epigenome maintains development and maintenance programmes, circadian rhythms, DNA repair and metabolic fitness to maintain health of an organism, and we propose that, during ageing, it contributes to decline of these coordinated processes. A) The epigenetic clock, stem cells and the developmental process are thought to be linked, with the clock tracking human growth and development from embryogenesis (during which the methylome is at the lowest biological age) into older age. Ageing thus begins soon after the embryonic stem cell state and continues throughout the lifespan until death. The epigenetic clock is proposed to 'tick' when tissue stem cells undergo asymmetric division and differentiate into non-stem cells (trans-amplifying cells), and rate of tissue turnover, which is measured as the rate of epigenetic ageing by the clock. B) Chrono-epigenetic theory proposes that cytosine-guanine dinucleotides (CpGs) that exhibit circadian oscillations are robust in young but dampen with age, possibly due to altered activity of ten–eleven translocation (TET) and DNA methyltransferase (DNMT) maintenance enzymes. Age-related changes in the amplitudes of the oscillations precede the linear DNAm changes and may predict age-dependent linear outcomes. Whether nicotinamide adenine dinucleotide (NAD+) enhancers, calorie restriction, longevity compounds, or exercise restores CpG rhythmicity that declines with age is unknown. C) DNA damage, which is present from birth until death, might also be an upstream cause of ageing. The cellular pool of NAD+ is utilised by DNA repair processes, suggesting that cell metabolism and DNA repair activities are interdependent. Upon DNA damage, NAD+-dependent enzymes, sirtuin 1 (SIRT1) and poly(ADP-ribose) polymerase 1 (PARP1), are recruited to double-stranded breaks (DSBs) to assist in repair. Other molecules, including DNMT1, DNMT3B and the Polycomb Repressive Complex (PRC) are also recruited to damage sites to silence transcription. CpG-rich promoter regions preferentially targeted by the repair machinery are susceptible to stable silencing events, possibly explaining why hypermethylation at Polycomb target genes is conserved in ageing. As DNA damage increases with age, this targeting puts strain on DNA repair machinery and reduces the pool of NAD+ available for cellular metabolism. D) The circadian clock controls rhythmic expression of metabolic genes, which lose robust oscillations with age. NAD+ regulates circadian and metabolic processes, and increasing levels of NAD+ (possibly via caloric restriction) restores metabolic rhythms by increasing SIRT1 activity; SIRT1 interacts directly with the core clock components to reprogramme healthy circadian function. NAD+ declines with age, and repleting NAD+ extends healthy lifespan in model organisms. AMPK, AMP-activated kinase; IGF1, insulin-like growth factor 1; NAMPT, nicotinamide phosphoribosyltransferase.

The DNA damage and relocalisation of chromatin modifiers hypothesis as a driver of ageing neatly positions itself in the ageing puzzle, alongside the developmental, metabolic, and chrono-epigenetic theories. Sources of DNA damage are ubiquitous in daily life, and even arise as early as development[261]. In fact, DSBs occur at a rate of 10 – 50 per cell per day[262]. Repair processes may therefore begin early during development, with the insidious DNA damage ensuing throughout life. An overly simplistic explanation is therefore that certain enzymes and proteins juggle DNA repair with other crucial processes, including epigenetic maintenance, metabolic regulation, and circadian control. The hyperactivity of the repair machinery during ageing is, unfortunately, at the cost of these processes. From this perspective, the DNAm changes that arise during ageing are the response to damage, signalling to the cell to hunker down and suppress instability.

But this has the unintended consequence of introducing epigenetic noise, compromising cell identity, impairing transcription, and ultimately causing ageing.

4. Pioneering epigenetic research

There is much that we still do not understand about the upstream causes of mammalian ageing, despite ageing itself driving the progression of most chronic diseases[4]. While we know that the methylome is extensively remodelled over the lifespan, the full extent to which chronological and biological ageing is quantified at the DNAm level is far from complete. Much attention in recent years in this field has been placed on building epigenetic clocks. However, we propose that instead of *building* new clocks, the focus should be shifted to *breaking down* the entire ageing methylome into its individual parts to first understand the mechanistic processes underling the extraordinary epigenetic ageing phenomenon. Although attempts have been made to break down the clocks into their various components[142,143,263], we propose that teasing apart the global changes in DNAm, by measuring DMPs, VMPs, co-methylation networks and entropy, in multiple tissues, is needed to obtain a better understanding of the ageing methylome in its entirety. The separation of these factors, down to the specific set of CpGs, could bring new mechanistic insights into the chronological and biological ageing process. For example, we still do not precisely understand what mechanisms are responsible for hypomethylation at enhancer regions.

At the epigenome-wide level, focusing only on patterns of DMPs is limiting when trying to understand aspects of biological ageing, particularly when making sense of why individuals of the same age may display vastly different ageing rates. However, there is a lack of research focusing specifically on VMPs in different tissues. Importantly, the classification of CpGs as DMPs and VMPs largely depends on the specifics of the cohort. For example, smoking exposure changes the *mean* methylation status at certain CpGs (i.e.

DMPs) that can be used to predict smoke exposure[132,264], however in the context of ageing, it is plausible that these CpGs might be VMPs. Important questions to consider are how much variable methylation is underpinning biological ageing, and does this differ in different tissues? Since very large sample sizes distributed across a broad age range are required to detect VMPs, this gap in knowledge is likely owing to the difficulty in sampling tissues other than blood. Nonetheless, uncovering VMPs presents a promising avenue to explore for identifying markers of biological age[43,62]. Future experiments could then investigate if slowing down ageing at DMPs will lengthen lifespan, and if slowing down ageing at VMPs will lengthen healthspan.

A major flaw in the available evidence is that many conclusions are drawn from experiments conducted in laboratory animals (e.g. mechanistic studies and longevity interventions), and there are limitations in extrapolating these findings to humans[265]. It is possible that this limitation can be alleviated by the use of third generation dual species epigenetic clocks (i.e. clocks built for humans and model organisms such as rats, mice, pigs, sheep, and primates)[58,179,266–270], which are based on CpGs that are highly conserved across mammals[271]. Nonetheless, experiments in humans that can measure the effect of interventions to suppress or slow ageing at the DNAm level would be an invaluable contribution to this field of research. Furthermore, exploration of whether age and sex matter when adopting different longevity protocols (i.e. intermittent fasting) is largely unknown. This brings to the forefront another significant challenge in epigenetic research, that is, we know surprisingly little about sex-specific differences in ageing. There is consistent evidence that females tend to outlive males[272], suggesting there is a robust feature of biology at play. Multi-tissue epigenetic clocks have also shown that males exhibit accelerated epigenetic ageing when compared to females[43,273]. Less understood, however, are the global patterns of DNAm that diverge between sexes across the lifetime.

Not addressed extensively in this review is cell-type heterogeneity, which remains a significant confounder in solid tissues, as well as in blood, and a major challenge in epigenetic research[19,274]. Considering cell types change with age, tackling the issue of cellular heterogeneity would be of great benefit to the ageing field. Recent advancements in single-cell DNAm analysis stems from the development of a novel computation tool capable of delineating the differences in cell type-specific epigenetic ageing[93,94]. This technology has the potential to accelerate our understanding of the functional and mechanistic consequences of DNAm changes during ageing[93], an exciting prospect for future explorations into biological ageing.

The potential confounding effect of 5-hydroxymethylcytosine (5hmC) in DNAm studies is also not covered in this review. Some platforms, such as the widely used Illumina HumanMethylation arrays, rely on bisulphite sequencing, which cannot distinguish between 5mC and 5hmC[275,276]. This has been shown in brain tissue[277], highlighting the possibility that age-related changes in DNAm attributed to 5mC could be due to changes in 5hmC.

None of the above can be achieved without access to large amounts of data from multiple human tissues. Epigenetic datasets of sufficient size in healthy, non-diseased human tissues across a broad age range are in short supply, and even more so in tissues other than blood. Nonetheless, existing datasets present a valuable resource for 'omics' research, which relies on large sample sizes to detect small effect sizes. Large-scale meta-analyses overcome many limitations from small study designs, and are a valuable tool in epigenetic research[278]. Lastly, DNAm does not act in isolation, but is simply a cog in a very large epigenetic machine. Future work should consider the entire epigenetic network, such as histone marks and chromatin changes, which may become disrupted with age.

5. Conclusions

The goal of ageing research is to target biological processes that not only make us live longer, but also those which help us to do so in a more youthful state. An important step towards achieving this goal is to identify the epigenetic processes that 'unravel' across the lifespan.

Although the manifestations of ageing are a feature of later life, age-associated alterations to the methylome are evidence that the underlying cellular and molecular changes begin earlier and even during development[92,279]. Arguably the most astounding feature of the ageing methylome is the consistency at which DNAm changes universally track chronological ageing[61], hinting there is a molecular clock ticking inside our cells. However, the picture of biological ageing is far from complete.

Making sense of the ageing methylome is not an easy feat. It requires the application of computational tools that accurately analyse and interpret the versatile DNAm marks that change across the lifespan. Notwithstanding the challenges, we are at the precipice of major geroscience discoveries, but extensive collaborate efforts from researchers across multiple fields, sharing ideas and data, are needed to collectively move forward ageing research.

Chapter 3: Methodology

This chapter will detail the analytical framework for this project, including the overall study design, data collection methods, and the statistical techniques used to answer each of the research aims.

1. Overall study design

This research was conducted as a large-scale, multi-tissue EWAS meta-analysis of age. Bioinformatics techniques were applied to analyse and interpret large amounts of existing epigenomic data. By exploiting the power of meta-analysis, we overcome many limitations of 'omics' research. Specifically, very large sample sizes are required to detect changes with small effect sizes, which is the case of age-related changes in DNAm profiles[280]. Our approach was therefore robust for identifying subtle, yet highly reproducible shifts in DNAm that accrue over chronological time in a wide variety of populations (e.g. males/females, conditioned/healthy individuals, etc.).

Briefly, the first step in the analytical framework was to gather existing DNAm datasets from our laboratory and our collaborators', in conjunction with public repositories, to assemble an exhaustive database of DNAm profiles across six human tissues (Figure 6). Using a published statistical pipeline, the second step was to utilise a variety of statistical techniques to quantity the various features of ageing in each dataset[3], including DMPs, VMPs, and entropy. The third step was to meta-analyse results within individual tissues (i.e. blood) to identify tissue-specific DMPs, VMPs and entropy. Finally, we performed various downstream analyses in blood to interpret the results in the context of chronological and biological ageing.

The specific aims of this research were as follows:

1. What are the tissue-specific patterns of *differential* DNAm (DMPs) that accrue with age in humans?

2. What are the tissue-specific patterns of *variable* DNAm (VMPs) that accrue with age in humans?

3. What are the age-associated *entropy* dynamics across different human tissues?

1.1. Data collection

To create the largest database of DNAm datasets to date, we have carried out a comprehensive data mining enterprise collecting the methylomes of 40,830 human samples from 113 datasets, across 6 tissues, profiled on the Illumina Methylation array platforms (27K, 450K and EPIC) (Figure 7, Supplementary Tables 1 - 6). This includes 1 dataset from our own laboratory (Gene SMART (GSE151407 & GSE171140))[211], 5 datasets from our collaborators (Finnish Twin Cohort (FTC)[281], Biological Atlas of Severe Obesity (ABOS), Limb Immobilisation and Transcriptional/Epigenetic Responses (LITER), Epigenetica & Kracht (EPIK) and EXACT), 97 open-access datasets from Gene Expression Omnibus (GEO) repository, 5 from ArrayExpress, 5 datasets from the controlled access database of Genotypes and Phenotypes (dbGaP), 1 dataset from the controlled access European Genome-Phenome Archive (EGA), and 1 dataset from a controlled access independent repository (Supplementary Tables 1 - 6). The bulk of the datasets are sourced from whole blood, totalling 56 datasets (n = 32,136) (Figure 7). The remaining samples are split across 27 datasets from brain tissue (n = 5,094), 6 datasets from skin tissue (n = 668), 4 datasets from buccal tissue (n = 473), 8 datasets from adipose tissue (n = 1,313), and 12 datasets from muscle tissue (n = 1,146). Datasets with fewer than 30 samples or an age standard deviation < 5 were excluded, as low age variability and low sample size severely impairs the ability of the linear models to detect age-related patterns reliably. Samples with a cancer diagnosis were also excluded, as cancer samples show highly unusual DNAm patterns that would likely skew the analysis[61].

1.2. Pre-processing

Datasets with raw DNAm data available were pre-processed, normalised, and filtered using the R statistical software (www.r-project.org), and following the *ChAMP*[282] pipeline. Methylated and unmethylated signals or IDAT files were used for the pre-processing. In accordance with the default parameters of the *champ.load* function, any sample with >10% of probes with a detection *p*-value > 0.01 was removed[282]. All probes with missing β-values (a detection *p*-value > 0.01), with a bead count < 3 in at least 5% of samples, non-CG probes and probes aligning to multiple locations were filtered out. Probes located on the X and Y chromosomes were also filtered out in datasets containing both males and females, as well as probes mapping to single nucleotide polymorphisms (SNPs)[282]. Additional cross-hybridising probes identified by Pidsley *et al.*[283] were also filtered out[7].

The methylation β-values obtained were calculated as the ratio of the methylated probe intensity and the overall intensity, as follows:

$$Beta\ value = \frac{intensity\ of\ methylated\ allele}{(intensity\ of\ unmethylated\ allele + intensity\ of\ methylated\ allele + 100)}$$

The Type I and Type II probe designs that are generated from the 450K and EPIC Illumina arrays were normalised using the *champ.norm* function[282]. We explored the technical and biological sources of variation in each dataset using a singular value decomposition method provided by the *champ.SVD* function[282]. The *ComBat* function from the *sva* package was used to adjust for technical variation from the slide and position on the slide if this information was available[284]. We could not perform batch correction if this information was unavailable. Finally, samples whose annotated sex was discordant with predicted sex (according to the *getSex* function from the *minfi* package), were removed[67]. Any missing information required for pre-processing, including raw IDAT files, batch information, detection *p*-values or the age of the samples, was requested from the corresponding authors at the time of pre-processing (Supplementary Tables 1 – 6).

Outline of project methodology

1. Data mining and preprocessing

ArrayExpress   EGA
GEO   dbGaP
Independent labs

40,830 DNAm profiles
blood, brain, adipose, skin, buccal and muscle
113 datasets
27K, 450K, EPIC Illumina arrayS

2. Statistical analysis

aim 1 — **DMPs**

$DNAm \sim age + sex ...$

MF %

old
young

age

aim 2 — **VMPs**

$DNAm \sim age + sex ...$
$resid \sim age$

MF %

old
young

age

aim 3 — **Entropy**

$1/N \times log_2(^1/_2) \sum_i [MF_i \times log_2(MF_i) + (1-MF_i) \times log_2(1-MF_i)]$

entropy

old
young

age

3. Meta-analyse results

Tissue-specific meta-analysis of age

	%DNAm change per year
blood dataset 1	0.15
blood dataset 2	0.21
blood dataset 3	0.19
blood dataset 4	0.17
blood dataset 5	0.15
...	...
...	...
blood dataset 27	-0.08
meta-analysis	0.14

0

Tissue-specific meta-analysis of age

	%DNAm change per year
brain dataset 1	0.15
brain dataset 2	0.21
brain dataset 3	0.19
brain dataset 4	0.17
brain dataset 5	0.15
...	...
...	...
brain dataset 27	-0.08
meta-analysis	0.14

0

4. Biological interpretation in blood

CpGs in chromatin states

active TSS
enhancers
repressed PRC
ZNF genes
Quiescent

Pathway enrichment

pathway 1
pathway 2
pathway 3
pathway 4
pathway 5
...
...
...
...

Figure 6 Framework of the project methodology.

Raw DNA methylation (DNAm) profiles from the 27K, 450K and EPIC Illumina Array platforms were sourced from open-access and controlled access databases, including ArrayExpress, Gene Expression Omnibus (GEO), European Genome-Phenome Archive (EGA), Database of Genotypes and Phenotypes (dbGaP) and independent labs. The DNAm profiles of 40,830 samples were collected and pre-processed, for both males and females, from 113 datasets in blood, brain, muscle, skin, adipose and buccal tissue (top panel). For each project aim, the relationship between age and average DNAm, or the relationship between age and DNAm variance, or the relationship between age and entropy was estimated in each dataset independently (second panel). For each CpG and for entropy, summary statistics were then meta-analysed across datasets to identify DMPs, VMPs and to determine whether entropy was significantly associated with age. We performed a tissue-specific meta-analysis of age for DMPs, VMPs and entropy, by pooling the results for a single tissue (i.e. brain, blood) (third panel). We then performed a functional analysis of the DMPs and VMPs to interpret the findings in blood, including pathway analysis, and enrichment in chromatin states (fourth panel).

Overview of tissue datasets
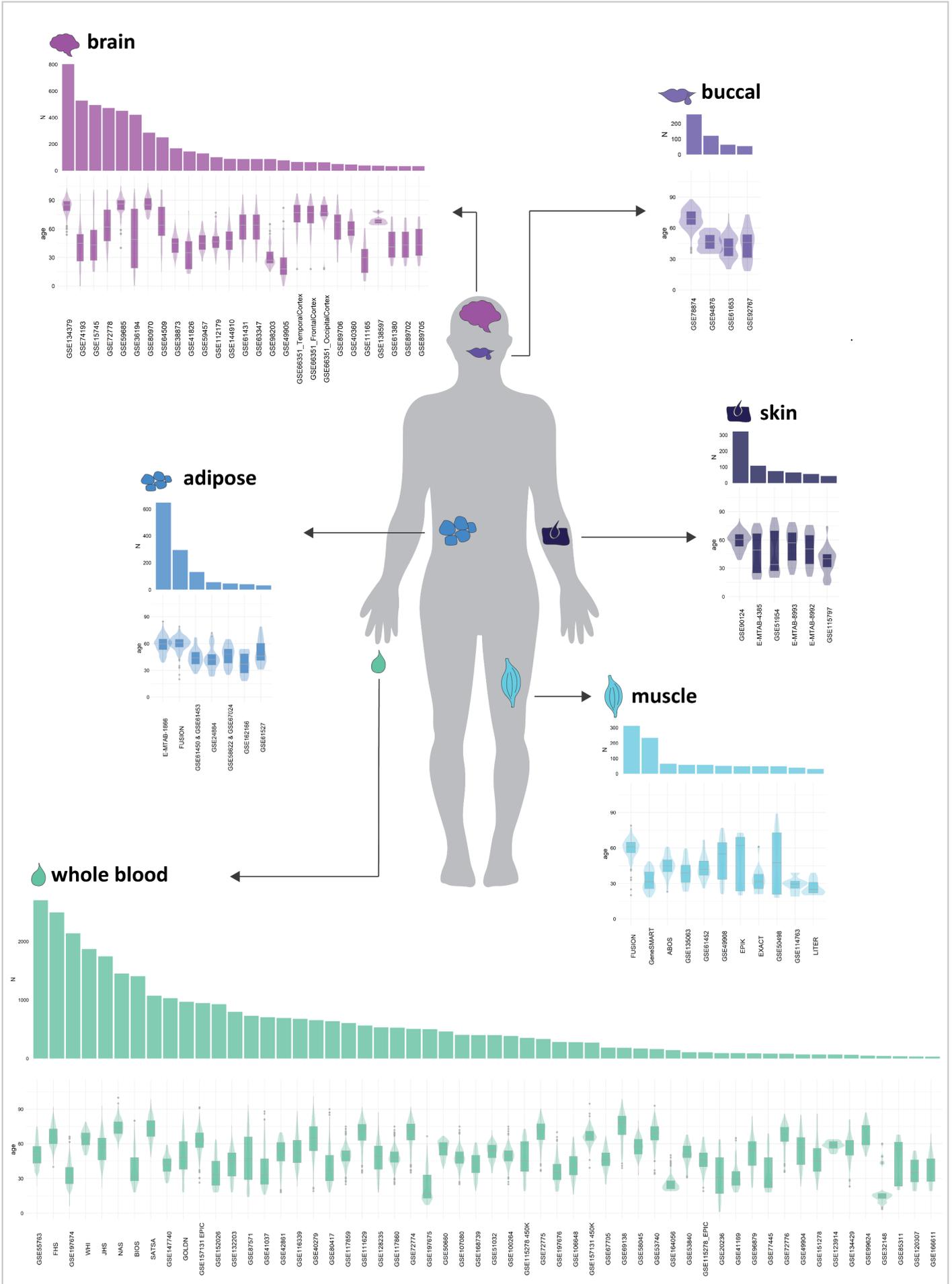
Figure 7 Overview figure of each dataset in each human tissue.

This diagram is an overview of each dataset collected and processed for each human tissue, including whole blood, skin, adipose, brain, buccal and muscle. For each tissue, the graph in the bottom panel illustrates the median age and interquartile range using boxplots, underlaid by violin plots that illustrate the distribution of samples across the age range. The bar plots at the top panel display the number of samples in each dataset (N), ordered from largest (left) to smallest (right). Datasets without access to phenotype information include FTC in muscle and GSE68336 in adipose and have been excluded from this figure.

2. Statistical framework

A variety of statistical tests were used to quantify the age-related changes in the methylome, which are extensively described in Chapter 2 of this thesis.

Here, we outline the methodology used to quantify the DMPs, VMPs, and entropy in each dataset, followed by the meta-analysis approach used to each answer each project aim.

2.1. DNA methylation features

2.1.1. Differentially methylated positions (DMPs)

DMPs are classified as CpGs that exhibit an average change in methylation as a function of age[3]. To identify DMPs, we used linear regression models and moderated Bayesian statistics, as implemented in the *limma* package[63,110]. A linear model was fitted for each dataset independently. DNAm was regressed against age and other dataset-specific covariates (i.e., sex, batch, body mass index (BMI)), when this information was available (Supplementary Tables 1 – 6), for each CpG probe. If a dataset included repeated measures of the same individual or related individuals (e.g., twins), we added a random effect to the model (using the *duplicateCorrelation* function of the *limma* package). Since beta values represent a fraction that ranges between 0 and 1, linear models were performed using M-values (a logit transformation of beta values), which have more favourable statistical properties for differential analysis[285].

Here are two examples of linear models used to identify DMPs:

A. Linear model for a dataset without repeated measures:

$$DNAm \sim \beta_0 + \beta_1 * age + \beta_2 * sex$$

54

Where DNAm corresponds to the M-values (i.e. the log-transformed beta value) for each CpG probe, as equal to the sum of a constant β_0 (intercept) and each explanatory variable, age and sex, scaled by their respective regression coefficients (β_1 and β_2).

B. Linear model for a dataset with repeated measures:

$$DNAm \sim \beta_0 + \beta_1 * age + \beta_2 * sex + (1|unique\ sample\ ID)$$

Where DNAm corresponds to the M-values (i.e. the log-transformed beta value) for each CpG probe, as equal to the sum of a constant β_0 (intercept) and each explanatory variable, age and sex, scaled by their respective regression coefficients (β_1 and β_2). The last variable in the equation (1|*unique sample ID)* represents the duplicate correlation for a replicate sample.

To then compare the effect of cell type heterogeneity on age-related DNAm changes, we repeated all linear models adjusting for cell type proportions in two tissues, whole blood and buccal. In each whole blood dataset, we applied the *champ.refbase* method, as implemented in the *ChAMP* package, to estimate the cell type proportions for granulocytes, monocytes, Natural Killer cells (NK), CD4T cells, CD8T cells and B cells, for each sample[282]. We then repeated the linear model for each blood dataset including the estimated cell type proportions for the 5 largest cell types to remove the confounding effect of cell type proportion on DMPs. In buccal tissue, we used the *epidish* function from the *EpiDISH* package to estimate the proportion of immune cells, fibroblasts, and epithelial cells for each sample[286]. Once again, we repeated each linear model and adjusted for the two largest cell type proportions in buccal tissue.

Two examples of the linear models with cell type adjustments.

C. Linear model for with cell type adjustments in whole blood:

$$DNAm \sim \beta_0 + \beta_1 * age + \beta_2 * sex + \beta_3 * CD8T + \beta_4 * CD4T + \beta_5 * NK + \beta_6 * Bcell + \beta_7 * Gran$$

Where DNAm corresponds to the M-values (i.e. the log-transformed beta value) for each CpG probe, as equal to the sum of a constant β_0 (intercept) and each explanatory

variable, age, sex, CD8T, CD4T, NK, Bcell, and Gran, scaled by their respective regression coefficients ($\beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \beta_6, \beta_7$).

D. Linear model for with cell type adjustments in buccal tissue:

$$DNAm \sim \beta_0 + \beta_1 * age + \beta_2 * sex + \beta_3 * fibroblasts + \beta_4 * epithelial$$

Where DNAm corresponds to the M-values (i.e. the log-transformed beta value) for each CpG probe, as equal to the sum of a constant β_0 (intercept) and each explanatory variable, age, sex, fibroblasts and epithelial, scaled by their respective regression coefficients ($\beta_1, \beta_2, \beta_3, \beta_4$).

We limited cell type adjustments to two tissues as repeating the analysis for all six tissues is beyond the scope of this thesis, however, this will be completed in future work.

2.1.2. Variably methylated positions (VMPs)

Age-related VMPs were identified using the Breusch-Pagan test for heteroscedasticity, which is a two-way regression that models the change in DNAm variance as a function of age (i.e. it tests if the variance in DNAm levels (adjusted for covariates) is dependent on age)[3,81]. DNAm was first regressed against age and other confounders for each dataset (i.e. the linear model to identify DMPs, Section 2.1.1.) to obtain residuals (Supplementary Tables 1 – 6). The residuals were then extracted from this model and squared. We ran the Shapiro-Wilk test to remove CpGs where residuals strongly deviated from normality (i.e. DNAm sites that are associated with SNPs not filtered out during pre-processing)[43], and subsequently regressed the squared residuals of the remaining markers against age.

The second regression to identify VMPs was as follows:

$$squared\ residuals \sim \beta_0 + \beta_1 * age$$

Where the residuals are those extracted from the linear regression used to identify DMPs, as equal to the sum of a constant β_0 (intercept) and the explanatory variable, age, scaled by the regression coefficient β_1.

All analyses were repeated in blood and buccal datasets that were corrected for cell types to compare the effect of cell type heterogeneity on age-related VMPs.

2.1.3. Shannon entropy

Shannon entropy is a single quantifiable measure of the methylome-wide DNAm changes for a single sample at a point in time[3]. Shannon entropy ranges between 0 and 1, taking its maximum value when the methylation fraction in a given set of CpGs, measured over a population of cells, is 50%[85]. Shannon entropy was calculated for each sample in each dataset, using a probability formula adapted to handle DNAm data[3,43,62].

To calculate the genome-wide Shannon entropy, a linear model is fitted for each dataset, adjusting for dataset covariates where appropriate (e.g. sex), as follows:

$$DNAm \sim \beta_0 + \beta_1 * sex$$

Where DNAm corresponds to the M-values (i.e. the log-transformed beta value) for each CpG probe, as equal to the sum of a constant β_0 (intercept) and the explanatory variable, sex, scaled by the regression coefficient (β_1).

Age was not included in this model, as we did not want to remove the effect of age on the beta values. The *mean* M-values from the original matrix were added to the residuals from the above linear model, and then transformed back to obtain *adjusted* beta values, as the Shannon entropy formula has been adapted specifically to handle beta values. Using the beta values adjusted for covariates, genome-wide Shannon entropy was computed for each sample in each dataset according to the formula[3,43]:

$$Entropy = \frac{1}{N * log_2 \frac{1}{2}} \sum_i [(DNAm_i * log_2 DNAm_i) + (1 - DNAm_i) * log_2 (1 - DNAm_i)]$$

Where *DNAm_i* is the methylation fraction (e.g. beta value between 0 and 1) for the ith CpG probe and *N* is the total number of CpGs.

To calculate the effect of age on Shannon entropy, a linear model was fitted for each dataset as follows:

$$Shannon\ entropy \sim \beta_0 + \beta_1 * age$$

Where Shannon entropy is equal to the sum of a constant *β_0* (intercept) and the explanatory variable, age, scaled by the regression coefficient (*β_1*).

Since genome-wide entropy captures the complexity of the entire system in a single measure, we sought to determine the contribution of the various features of ageing that may be driving genome-wide changes in entropy. To do so, we then repeated the analysis by calculating entropy for each sample, in each dataset, in each tissue, on all the *age-related CpGs* identified from the meta-analysis in aims 1 and 2 (i.e. a complete list of both DMPs and VMPs in each tissue), as well as the *non-age-related CpGs* (i.e. the complete list of non-DMPs and non-VMPs in each tissue). In blood and buccal tissue, we also calculated entropy on β-values corrected for cell type heterogeneity, to compare the effect of cell type composition on age-related changes in entropy.

2.2. Adjusting for age-related diseases

Age is the leading risk factor for age-associated disease and disability[287,288]. As such, we took careful consideration when adjusting for disease-specific covariates in our linear models to ensure that we did not unnecessarily remove the effect of age on DNAm.

While no official classification of 'age-related' disease exists, 92 conditions have been classified as 'age-related' based upon an *exponential* increase in incidence with age using a two-step mathematical modelling technique[288]. To have a direct view of the relationship between age and each disease/condition present in our datasets (Supplementary Tables 1 - 6), we used the GDB data exploration tool (https://vizhub.healthdata.org/gbd-compare/).

Based on their classification and the GDB tool, we *included* the following conditions/phenotypes as covariates that show no relationship with age, including depression, anaemia, inflammatory bowel disease / Crohn's diseases, lupus, non-alcohol steatohepatitis (NASH) / non-alcoholic fatty liver disease (NAFLD) and asthma, and *excluded* conditions/phenotypes that show a relationship with age (either increase or decrease in prevalence with age), including progressive supranuclear palsy, Alzheimer's disease/dementia, cardiovascular disease, ischemic stroke, Parkinson's disease, COVID-19, cardiomyopathy, chronic obstructive pulmonary disease (COPD), schizophrenia, anxiety disorders, osteoarthritis, rheumatoid arthritis, multiple sclerosis, psoriasis, type 2 diabetes and cirrhosis.

Importantly, some diseases did show a relationship with age, but cannot be considered age-related as they don't depend on age-related functional decline of the body, but on age-related differences in behaviour (e.g. HIV infection, drug use disorder), and were included as covariates. In addition, factors that are not age-related but accelerate or slow down ageing (e.g. smoking, BMI, hypertension, exercise training and bariatric surgery) were also included as covariates (Supplementary Tables 1 – 6).

2.3. Meta-analyses

Primary DNAm studies are limited in their ability to detect the significance of an age-related DNAm change with a small effect, largely due to limited sample sizes or a narrow age range of samples. To obtain the needed statistical power to verify that such an effect exists, we utilise meta-analysis techniques to detect robust age-related DNAm changes[278]. Furthermore, as this meta-analysis pools summary statistics from a large range of heterogeneous cohorts (i.e. containing both males and females, different ethnicities, with a wide range of health statuses), our results are valid across all kinds of free-living human beings (e.g. they are not limited to healthy individuals, or to males only).

Here, we outline how each project aim will be answered using a meta-analysis approach.

Aim 1: Tissue-specific EWAS meta-analysis of differential methylation and age

The first step was to perform a *tissue-specific* meta-analysis of age to identify DMPs in each tissue separately. As described above, an EWAS of age was performed in each dataset independently. Each EWAS was adjusted for bias and inflation using the empirical null distribution as implemented in *bacon*[280]. The results from the independent EWAS for a **single tissue** (i.e. blood) were then pooled, using an inverse variance based fixed-effects meta-analysis implemented in METAL[278]. This approach computes a weighted average of the results, using the individual effect size estimates and standard errors extracted from each independent EWAS. The overlap in CpGs between datasets was imperfect (not all CpGs were present in all datasets), as we used three different Illumina array platforms, and since different CpGs are filtered out during the pre-processing of individual datasets. We restricted our analysis to CpGs that were present in at least three datasets. Age-related DMPs were then identified using a stringent meta-analysis false discovery rate (FDR) < 0.005. Where appropriate, the meta-analyses were repeated with datasets adjusted cell types.

Aim 2: Tissue-specific EWAS meta-analysis of variable methylation and age

Following the same process as above, we used METAL to pool results from the independent EWAS in each independent tissue (i.e. blood), but we followed a *sample size*-based fixed effects meta-analysis (instead of an inverse-variance method)[278]. This approach was more appropriate to meta-analyse the χ^2 test statistic that is the output of the Breusch-Pagan test; it relies on the *sample size* of each dataset and the *p*-value at each CpG. We restricted our analysis to CpGs that were present in at least 15% of the samples in each tissue. Age-related VMPs were identified using a stringent meta-analysis FDR <0.005. Where appropriate, the meta-analyses were repeated with datasets adjusted cell types.

Aim 3: Shannon entropy meta-analysis of age

To identify the change in Shannon entropy with age in each tissue, the summary statistics (i.e effect size and standard error) extracted from the independent Shannon entropy regressions (Supplementary Tables 7 – 12), as described above, were pooled using a fixed effects meta-analysis using the R package *metafor*[289]. We meta-analysed the summary statistics for the genome-wide entropy results, as well as for the age-related CpGs, and the non-age-related CpGs.

All meta-analyses in blood were repeated for the cell-type corrected analyses, to compare the effect of cell heterogeneity on entropy.

2.4. Biological interpretation of meta-analyses

To aid in the biological interpretation of the identified age-related DNAm sites, we have performed downstream analysis in blood only, to determine the functional relevance of DMPs and VMPs in the context of chronological and biological ageing. First, we tested whether DMPs and VMPs showed any enrichment in chromatin states. This was done by comparing the distribution of the blood-specific VMPs and DMPs with that of non-VMPs and non-DMPs, respectively, in the different chromatin states profiled in peripheral blood mononuclear cells (PBMCs) from the Roadmap Epigenomics Project with a Fischer's exact test[290]. To gain insights into the cellular and phenotype consequences of ageing on the blood methylome, we tested whether genes belonging to gene ontology (GO) terms (GO gene set in MsigDB), human phenotype ontologies (HPO gene set in MsigDB), canonical pathways (CP gene set in MsigDB) and expression signatures of genetic and chemical perturbations (CGP gene set in MsigDB) were enriched among the VMPs and DMPs using the *gsameth* function from the *missMethyl* package[291]. An improved adaptation of Zhou et al.'s comprehensive annotation was used to assign one or more genes to each VMP and DMP[292] All GO, HPO, CP and CGO terms were deemed significant at an FDR < 0.005[293,294].

Chapter 4: Signatures of differential and variable DNA methylation, entropy, and ageing in blood

In chapter 4, we first characterise the unique signatures of age-related DMPs and VMPs, as well as entropy in **whole blood**. Then, we perform functional genomic enrichment tests to provide a functional interpretation of our findings.

1. Ageing is characterised by small, widespread changes in DNA methylation mean and variance in blood

We first conducted a meta-analysis of age-related differential methylation in 56 whole blood datasets, reaching a total sample size of 32,136 blood methylomes (Supplementary Table 1). Nearly half of the tested sites (333,300 CpGs) were DMPs (48%) at FDR < 0.005. DMPs changed in *average* DNAm levels with older age (Figure 8A), and two-thirds of them (66%) decreased in DNAm levels ('hypoDMPs'). HyperDMPs increase by an average of 0.027% per year of age and hypoDMPs decreased by an average of -0.034% per year of age.

Importantly, there was an inverse correlation between the overall methylation fraction of a CpG and the direction of change during ageing: DMPs whose DNAm levels were usually high (> 75% on average), were overwhelmingly hypoDMPs, while DMPs whose DNAm levels were usually low (< 25% on average), were overwhelmingly hyperDMPs (Supplementary Figure 1). In contrast, DMPs with intermediate DNAm levels trend equally frequently towards high and low DNAm levels. For example, in the BIOS blood dataset, 21% of DMPs were considered 'intermediately' methylated, of which 40% gained methylation with age, and 60% lost methylation with age (Supplementary Figure 1).

The meta-analysis identified DMPs as those CpGs with highly consistent effects across datasets. For example, cg16867657 was estimated to gain 0.45% DNAm per year of age across the different datasets (Figure 8B), and the effect size barely varied between

datasets (from 0.25% to 0.81%). This CpG is in the promoter of *ELOVL2*, which has been associated with ageing in a plethora of studies[295–298].
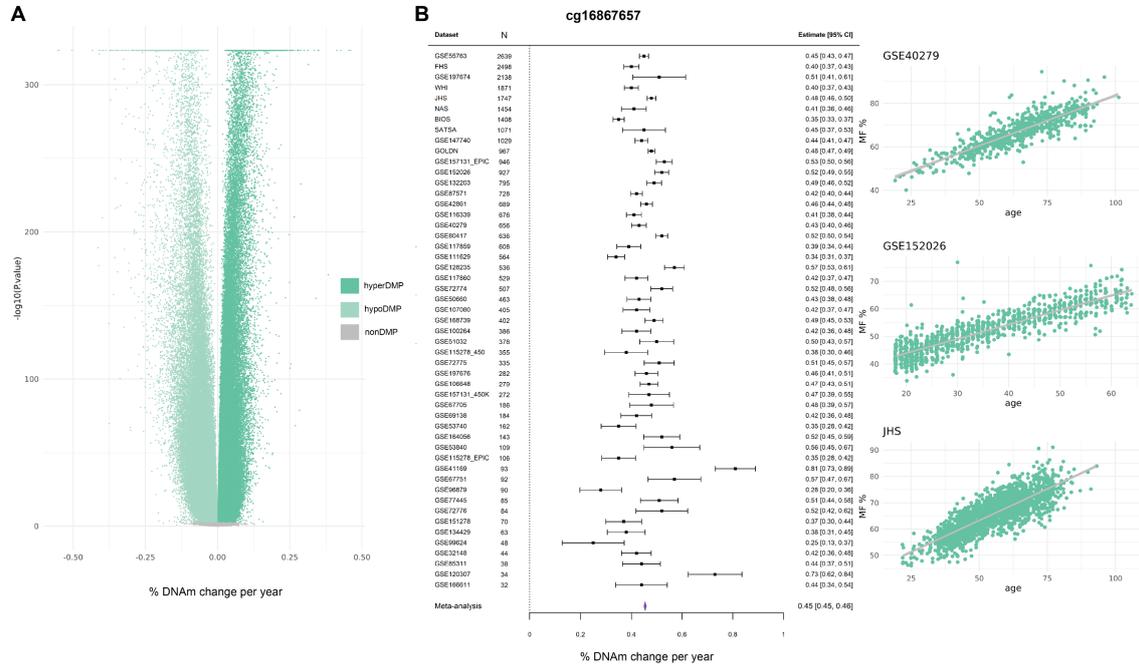
Figure 8 Differential meta-analysis of age in blood.

A) A volcano plot displaying the meta-analysis effect size (x-axis) and significance (y-axis) for the 696,228 tested CpGs in the differential methylation meta-analysis of age in blood. The strongest associations have the smallest *p*-values and will be the highest points on the plot. Hypomethylated (hypoDMP), hypermethylated (hyperDMP) and non-DMPs are represented by the colours in the legend. DMPs are classified at a false discovery rate (FDR) < 0.005. B) Forest plot of a highly significant CpG (cg16867657). The dataset and sample size are on the left side of the plot, with the corresponding effect size and errors represented by the point and error bars. The meta-analysis effect size is represented by the purple polygon. The methylation plots for this CpG from three independent blood cohorts, including GSE40279, GSE152026 and Jackson Heart Study (JHS) are displayed on the far right, with age on the x-axis and methylation fraction (MF) as a percentage on the y-axis. Each point on the plot represents a single sample.

These results remain largely unchanged after the meta-analysis was adjusted for blood cell type proportions (Pearson's correlation of meta-analyses effect sizes = 0.94, *p* value < 2.2e-16) (Supplementary Figure 2). Only 14% of DMPs were no longer significant after accounting for the confounding effect of blood cell type composition, with only a small percentage of those at the threshold of significance (~8%). It is therefore plausible that at those DMPs, the age-related change in DNAm is likely to reflect the change in the % of different blood cell types that comes with age.

We then meta-analysed the same 56 whole blood datasets to identify changes in methylation variability (VMPs) during ageing (see Methods section 2.3 for meta-analysis rationale). We identified 243,958 VMPs (37% of tested CpGs) at FDR < 0.005 (Figure

9A), nearly all of which *increased* in variance (99% of VMPs). These results are robust as we only included CpGs assayed in ≥ 5,000 samples (15% of total sample). The magnitude of the age-related changes in variance is small, for example, the average increase in variance across all datasets for the most significant VMP, cg21899500, is 0.01% per year of age (Figure 9B).
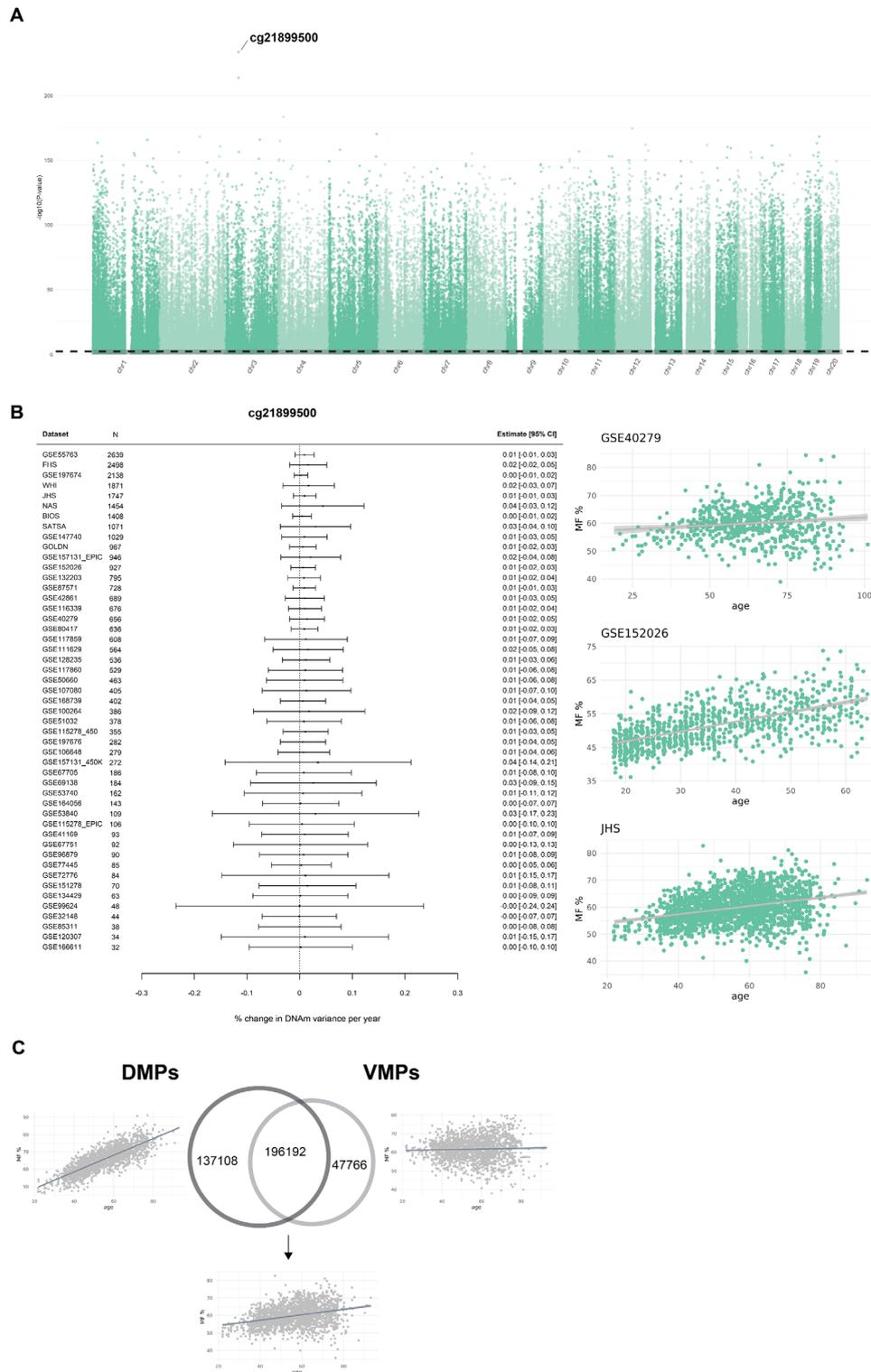
Figure 9 Variable meta-analysis of age in blood.

A) Manhattan plot displaying the genome-wide distribution of variably methylated positions (VMPs) in blood. The autosomal chromosomes are displayed along the x-axis, and the negative logarithm of the *p*-values on the y-axis for each of the 691,302 CpGs included in the meta-analysis. The strongest associations have the smallest *p*-values and will be the highest points on the plot. The dashed line is the FDR threshold that separates the significant VMPs above the line, and the non-significant CpGs below the line in the grey shaded area. B) Forest plot of the top VMP (cg21899500) that is hypermethylated with age and increases in variance with age, and the three adjacent methylation plots from three independent datasets, GSE40279, GSE152026 and Jackson Heart Study (JHS). C) Venn diagram of the overlap between DMPs and VMPs.

There was a significant overlap between DMPs and VMPs (i.e. a CpG site whose average DNAm level changed during ageing was also more likely to see its variance increase with age; Fischer's exact test *p* value < 2.2 x 10^{-16}). We identified 196,192 CpGs that were classified as both DMPs and VMPs (Figure 9C). Among those, 73,357 (37%) increased in both average methylation and variance, and 122,835 (63%) decreased in average methylation but increased in variance. This hypo/hypermethylation distribution pattern among VMPs suggests that VMPs are not particularly biased towards an increase or a decrease in average DNAm.

We repeated the VMP meta-analysis after adjusting for blood cell type composition and as for the DMP analysis, results remained largely unchanged (Pearson's correlation of meta-analyses Zscore = 0.93, *p*-value < 2.2e-16) (Supplementary Figure 3). However, VMPs seemed to be more sensitive than DMPs to confounding by cell type proportions, as more than a third of VMPs (37%) were *only* significant in the meta-analysis **not** adjusted for cell types. With that said, an additional 5,913 CpGs were classified as VMPs (4%) only after we adjusted for cell types. We identified 159,166 VMPs (22% of tested CpGs) after correcting for cell type composition.

2. Variable methylation introduces noise at key regions that become differentially methylated with age in blood

We then investigated whether DMPs and VMPs are enriched in functional regions of the blood genome. To do this, we compared the distributions of DMPs and VMPs to that of non-DMPs and non-VMPs, in different chromatin states profiled in PBMCs[290], using a comprehensive annotation of the Illumina Methylation arrays[292]. HypoDMPs were over-represented in quiescent chromatin states and those weakly repressed by Polycomb complexes, while hyperDMPs were over-represented in bivalent regions and regions repressed by Polycomb complexes (χ^2 test p-value $< 2.2e-16$) (Figure 10A). Similarly, hyperVMPs are over-represented in bivalent enhancers and regions repressed by the Polycomb complex (Figure 10B). ConstantVMPs display similar distributions to hyperVMPs, over-represented in weak repressed Polycomb and quiescent regions. All classes of VMPs are depleted in active transcription start sites (χ^2 test p-value $< 2.2e-16$).
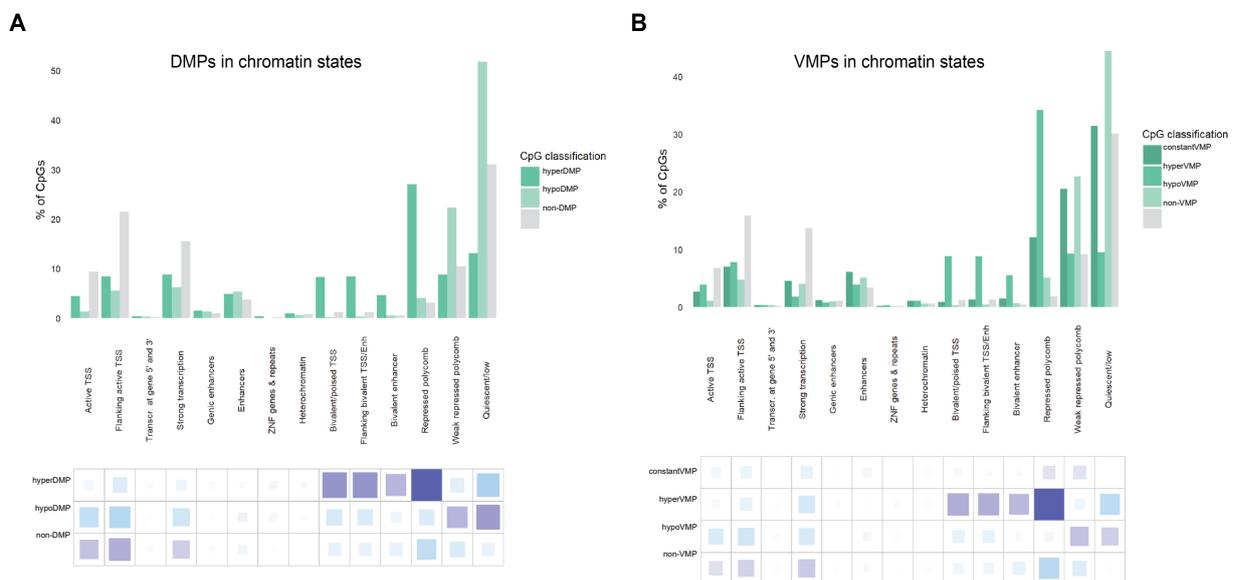
Figure 10 Chromatin state enrichment of age-related CpGs in blood.

A) Distribution of hypomethylated and hypermethylated differentially methylated positions (DMPs) and non-DMPs in chromatin states from peripheral blood mononuclear cells (PBMCs); B) Distribution of hypermethylated, constant, hypomethylated variably methylated positions (VMPs) and non-VMPs in chromatin states from PBMCs. The grids under the graph represent the residuals from the χ^2 test, with the size of the blocks in the grid being proportional to the cell's contribution. Purple indicates over-representation or enrichment in the chromatin state, and blue represents under-representation or depletion in the chromatin state.

3. Key pathways relating to development, metabolism, signalling, and homeostatic processes become differentially and variably methylated with age

We next sought to understand whether the change in average DNAm, and the change in DNAm variance during ageing, are associated with specific sets of genes involved in interesting biological pathways. As DMPs and VMPs showed a significant overlap in the first place, it was unsurprising to see that many pathways were enriched in both differentially methylated genes (DMGs) and variably methylated genes (VMGs). We identified 790 overlapping *gene ontology* terms that were overrepresented (FDR < 0.005) in both the DMG and VMG analysis (Figure 11A), the most significant of which relating to development, neurological processes, cell signalling, differentiation and homeostatic processes. Less significant were the 170 GO terms unique to DMGs (Figure 11B), such as mRNA metabolic processes, catalytic complex, histone binding and methylation, and the 158 GO terms unique to VMGs (Figure 11B), such as G-protein-coupled receptor pathways, muscle cell development and metabolic processes. Considering the significant overlap between the DMGs and the VMGs, we noted a positive correlation between the GO terms identified in each analysis (Spearmen's correlation of p values = 0.58, p value < 2.2e-16) (Supplementary Figure 4).
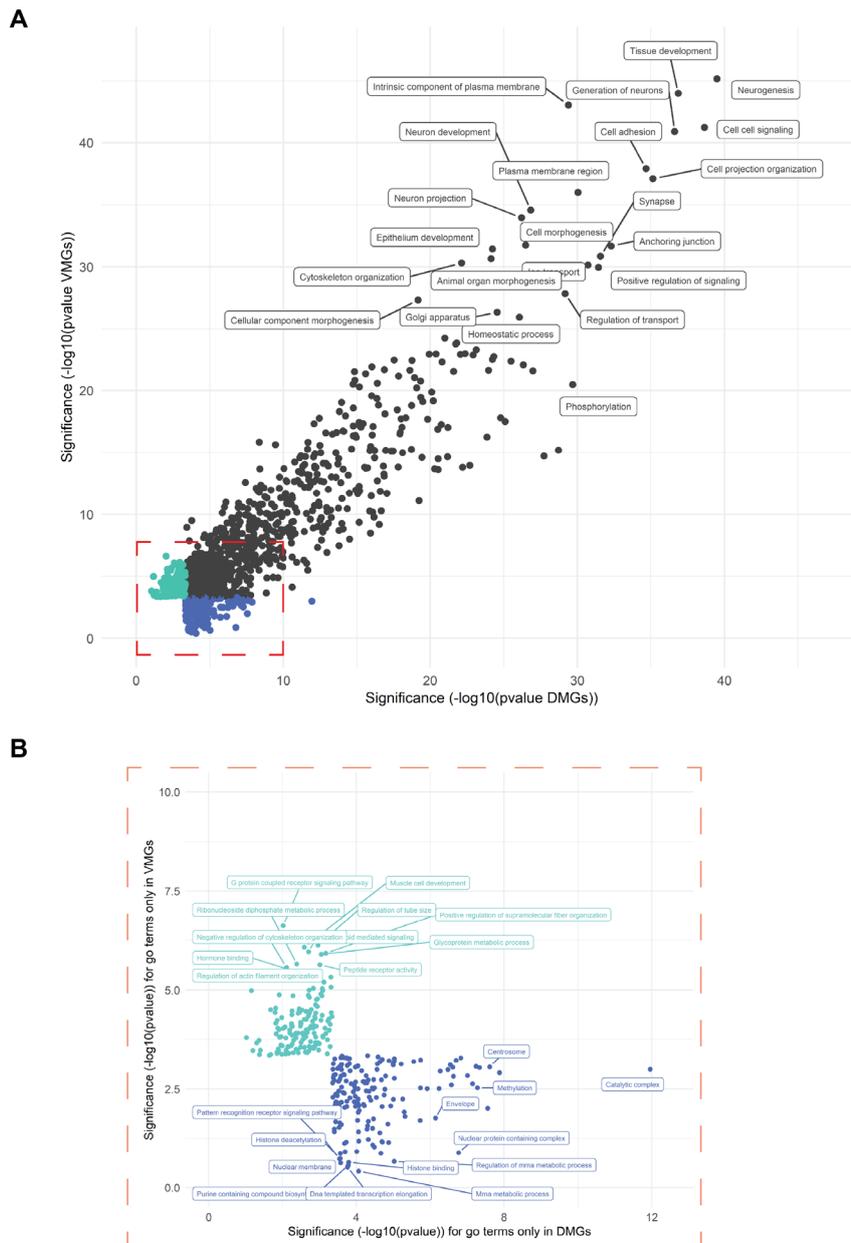
Figure 11 Overrepresentation analysis of the age-related CpGs for Gene Ontologies in blood.

Overrepresentation analysis of the age-related differentially methylated genes (DMGs) assigned to the differentially methylated positions (DMPs), as well as the age-related variably methylated genes (VMGs) assigned to the variably methylated positions (VMPs) for Gene Ontology from MSigDB. A) The significance of the DMGs is on the x-axis and the significance of the VMGs is on the y-axis; all pathways are significant at an FDR < 0.005. Pathways in turquoise represent those present only in the variable methylation analysis, pathways in royal blue represent those present only in the differential methylation analysis, and pathways in dark grey are significant in both. B) A magnification of the pathways that are unique to the differential (royal blue) and variable (turquoise) analysis.

We also identified 403 overlapping *human phenotype ontology* (HPO) terms that were overrepresented (FDR < 0.005) in both the DMG and VMG analysis (Figure 12A), as

well as 114 HPO terms that were identified only in the DMG analysis, and 80 HPO terms that were identified only in the VMG analysis (Figure 12B). Analogous to GO terms, the most significant pathways in each analysis are overlapping, including onset (i.e. the age group in which disease manifestations appear), constitutional symptoms, various abnormal organ systems, abnormal inflammatory response and gait disturbance (Figure 12A). The most significant HPO terms unique to the differential analysis include peripheral neuropathy, optic atrophy, abnormalities in brain and cranial features and hypertrophic cardiomyopathy (Figure 12B). In the variable analysis, the most significant,
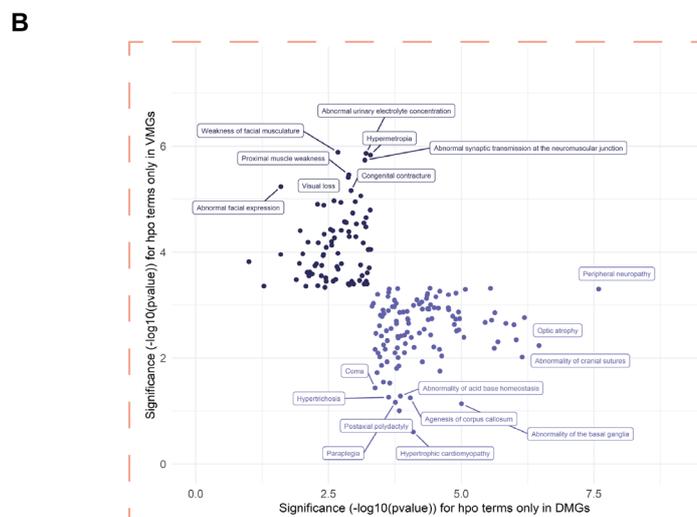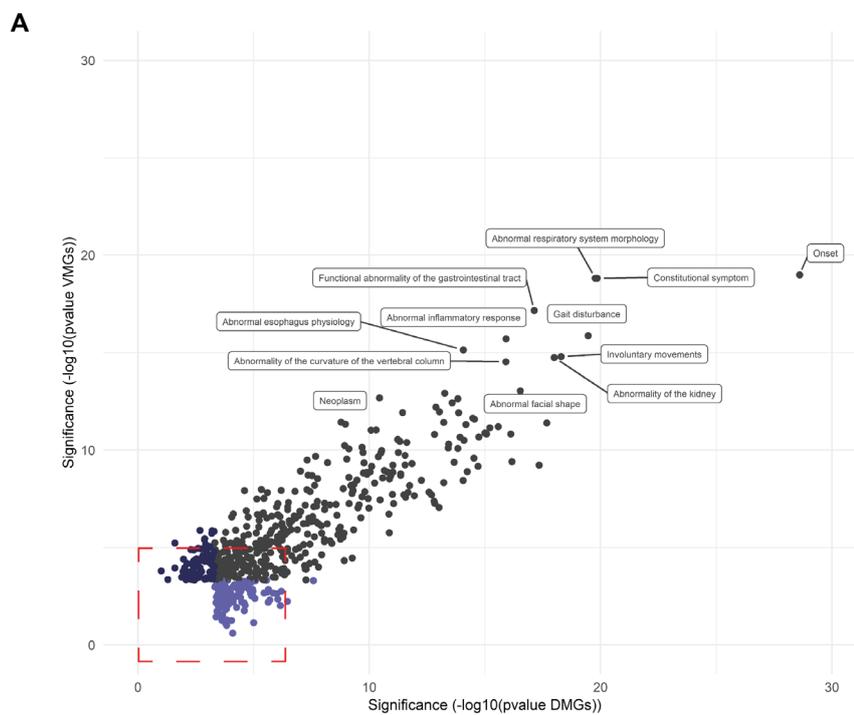
Figure 12 Overrepresentation analysis of the age-related CpGs for Human Phenotype Ontologies in blood.

Overrepresentation analysis of the age-related differentially methylated genes (DMGs) assigned to the differentially methylated positions (DMPs), as well as the age-related variably methylated genes (VMGs) assigned to the variably methylated positions (VMPs) for Human Phenotype Ontology (HPO) from MSigDB. A) The significance of the DMGs is on the x-axis and the significance of the VMGs is on the y-axis; all pathways are significant at an FDR < 0.005. Pathways in dark purple represent those present only in the variable methylation analysis, pathways in light purple represent those present only in the differential methylation analysis, and pathways in dark grey are significant in both. B) A magnification of the pathways that are unique to the differential (light purple) and variable (dark purple) analysis.

unique HPO terms include abnormalities to facial structure, musculoskeletal abnormalities, and visual changes (Figure 12B).

Once again, the most significant *canonical pathways* (CP) overrepresented (FDR < 0.005) in both the DMG and VMG analysis were overlapping (Figure 13A). We identified 58 overlapping pathways, the most significant relating to the proteins in the extracellular matrix (ECM), the neuronal system, transport of small molecules, lipid metabolism and signalling pathways. Albeit less significant than the overlapping pathways, unique to the differential analysis were pathways relating to transcription, the ciliary landscape and endoderm differentiation, and unique to the variable analysis were pathways relating to ECM regulation, signalling pathways, GPCR ligand binding and endoderm differentiation (Figure 13B).

We repeated the enrichment on the *expression signatures of genetic and chemical perturbations* (CGP), and once again the most significant pathways overrepresented (FDR < 0.005) in both the DMG and VMG analysis were overlapping (Figure 14A). There were 401 significant overlapping CGP terms, the most significant relating to H327Kme3 embryonic stem cells, PRC subunits, H3K4me3 and H3K27Kme3 in the brain, and various cancer signatures (Figure 14A). CGP signatures unique to the differential analysis include to various cancer signatures and FOXO3 targets, and CGP

signatures unique to the variable analysis include HDAC2 targets, lung cancer and stem cells in adipose and bone (Figure 14B).
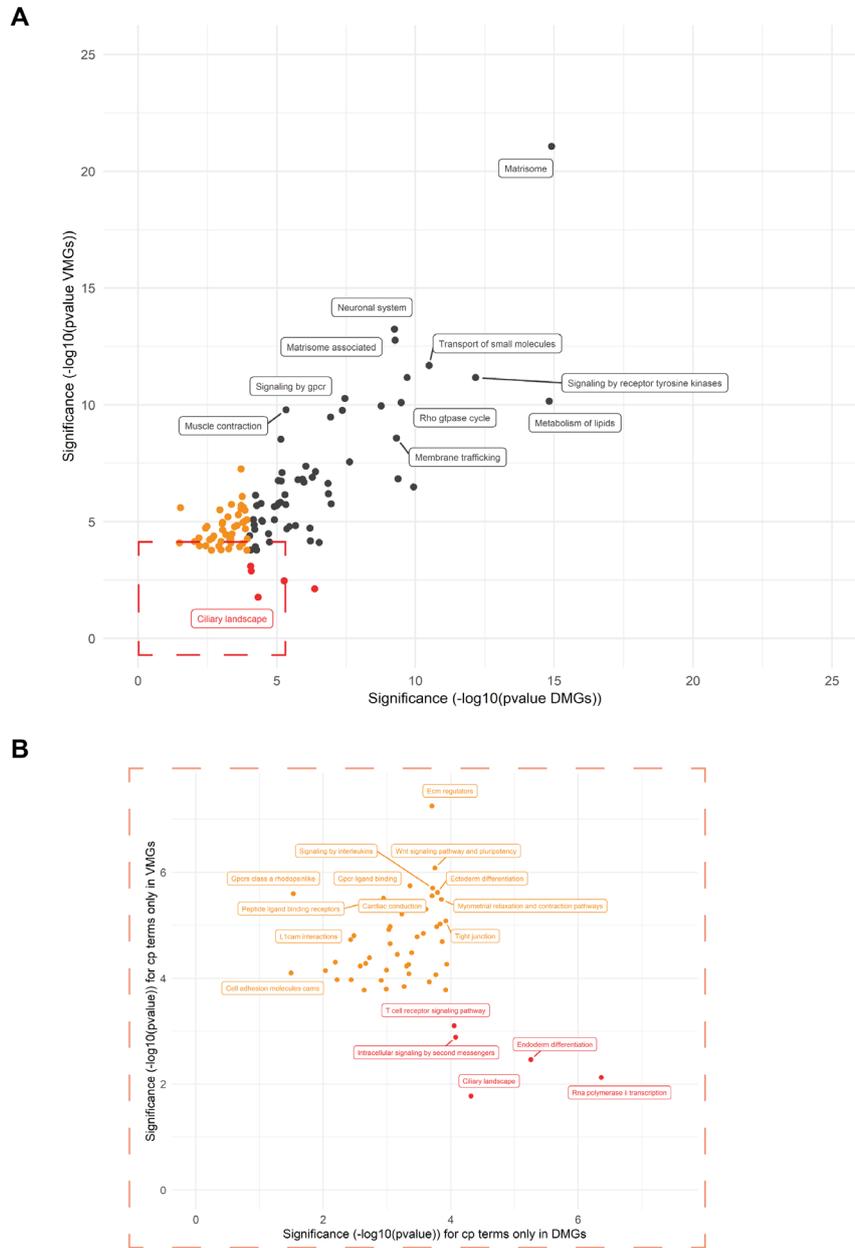
Figure 13 Overrepresentation analysis of the age-related CpGs for Canonical Pathways in blood.

Overrepresentation analysis of the age-related differentially methylated genes (DMGs) assigned to the differentially methylated positions (DMPs), as well as the age-related variably methylated genes (VMGs) assigned to the variably methylated positions (VMPs) for Canonical Pathways (CP) from MSigDB. A) The significance of the DMGs is on the x-axis and the significance of the VMGs is on the y-axis; all pathways are significant at an FDR < 0.005. Pathways in orange represent those present only in the variable methylation analysis, pathways in red represent those present only in the differential methylation analysis, and pathways in dark grey are significant in both. B) A magnification of the pathways that are unique to the differential (red) and variable (orange) analysis.
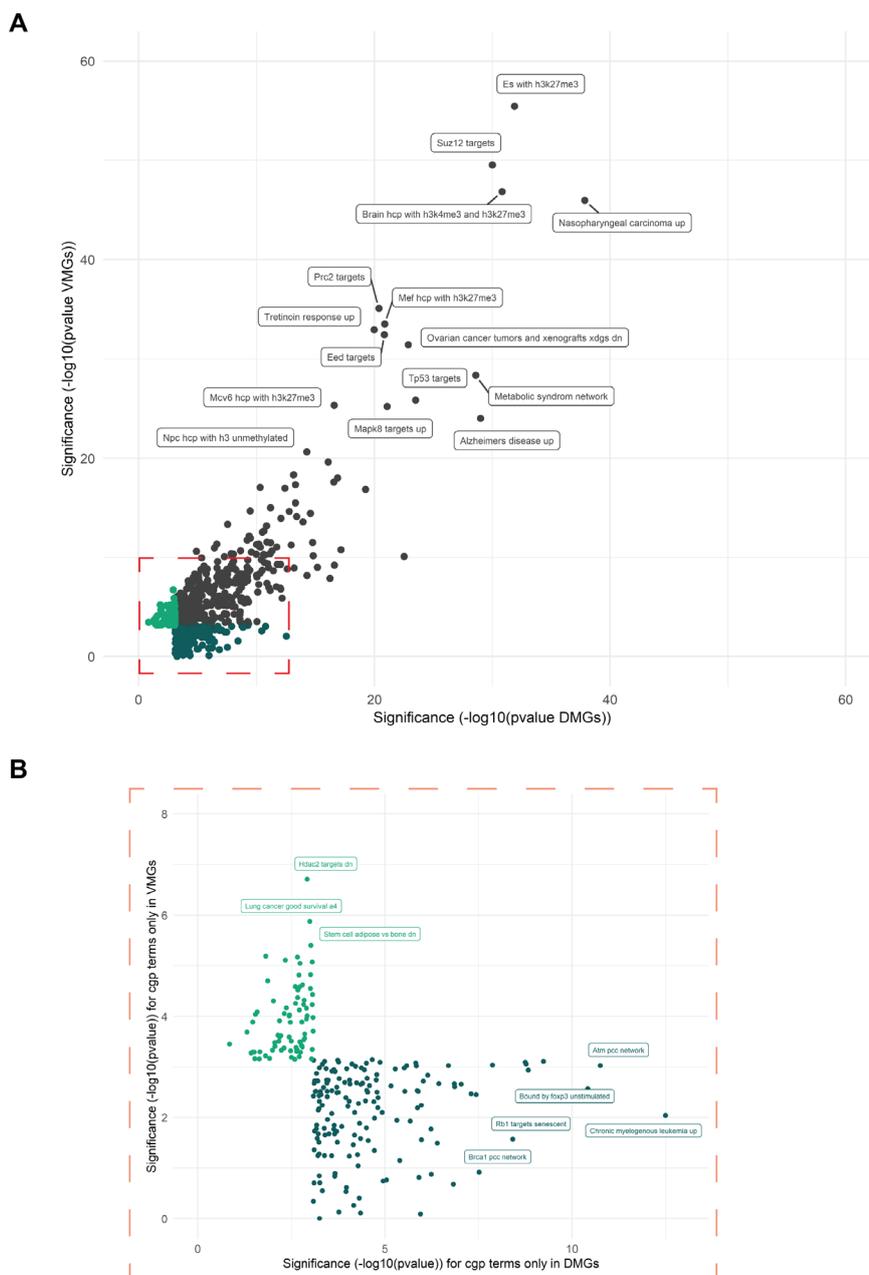
Figure 14 Overrepresentation analysis of the age-related CpGs for Expression Signatures of Genetic and Chemical Perturbations in blood.

Overrepresentation analysis of the age-related differentially methylated genes (DMGs) assigned to the differentially methylated positions (DMPs), as well as the age-related variably methylated genes (VMGs) assigned to the variably methylated positions (VMPs) for Expression Signatures of Genetic and Chemical Perturbations (CGP) from MSigDB. A) The significance of the DMGs is on the x-axis and the significance of the VMGs is on the y-axis; all pathways are significant at an FDR < 0.005. Pathways in green represent those present only in the variable methylation analysis, pathways in emerald represent those present only in the differential methylation analysis, and pathways in dark grey are significant in both. B) A magnification of the pathways that are unique to the differential (emerald) and variable (green) analysis.

4. Entropy increases in the ageing blood methylome, driven by the cumulative changes in differential methylation

Lastly, we determined whether the ageing blood methylome increases in entropy ('chaos') with age, and what is driving these changes. We took the same statistical approach as with the DMP and VMP analyses described above: first, we estimated the strength of the association between age and entropy in each independent cohort; then, we pooled these effect sizes across the different cohorts using a fixed-effects meta-analysis (Supplementary Table 7). Entropy captures the amount of information encoded by the epigenome: CpGs are deemed highly 'predictable', if they have either high (~100%) or low (~0%) DNAm levels; conversely, CpGs whose methylation fraction is closer to 50% are deemed 'unpredictable'. As the methylation state of genes determines cellular identity and therefore cellular function, entropy (i.e. 'chaos') increases when multiple CpGs throughout the genome drift towards a methylation fraction of 50%. An entropy of 0 means that every CpG is either methylated at 0% or 100%, and an entropy of 1 means that every CpG is methylated at exactly 50%. In these two opposite scenarios, the methylome of a cell is either entirely predictable, or entirely unpredictable.

When taking *all* CpGs into account (DMPs and non-DMPs, VMPs and non-VMPs), we observed a very small but significant increase in entropy of 0.0005 per decade of age (p-value < 0.0001). Furthermore, the increase in entropy was highly variable between cohorts ($I^2 = 88\%$), with some datasets showing a change of -0.002 per decade, and others a change of 0.008 per decade (Supplementary Figure 5).

As an increase in entropy with age reflects a drift towards a methylation fraction of 50% at multiple CpGs, we hypothesised that this increase in entropy would be driven by sites that change with age (DMPs and / or VMPs). To answer this, we re-calculated entropy in each sample from each dataset, but only taking into account the methylation levels at DMPs and VMPs (i.e. the age-related CpGs) (Supplementary Table 7). As a 'control' comparison, we also re-calculated entropy in each sample from each dataset, but only taking into account the methylation levels of CpGs that were neither DMPs nor VMPs (i.e. non-age-related CpGs).

Figure 15 Meta-analysis of entropy and age in blood.

A) A forest plot of the two meta-analyses comparing the changes in entropy between the non-age-related CpGs (left) and the age-related CpGs (right) identified in blood. The meta-analyses effect sizes are represented by the orange and light green polygons. On the x-axes is the change in entropy per decade of age, and on the y-axes are the effect sizes and standard error measurements from the independent EWAS. The dataset name and sample size (N) are to the left of the forest plot. B) Three graphs with entropy (y-axis) plotted against age (x-axis) for the age-related CpGs and non-age-related CpGs from three independent blood datasets, GOLDN, GSE87571 and GSE197674.

In line with our hypothesis, we found that non-age-related CpGs do *not* contribute to the global increase in entropy with age, with a meta-analysis effect size of -0.0003 change in entropy per 10 years of age (*p*-value <0.0001, *I2* statistic 46%) (Figure 15A). In contrast, age-related CpGs increase in entropy by 0.002 per decade of age (*p*-value <0.0001, I^2 statistic 85%) (Figure 15A). This suggests that the global increase in entropy stems from the linear changes occurring at DMPs and VMPs. Moreover, the baseline entropy (i.e. the entropy value at the youngest age in a particular dataset) for the non-age-related sites is lower than the baseline entropy for the age-related sites (Figure 15B). This is because there is a higher proportion of intermediately methylated CpGs in the age-related sites (~20%), compared to the non-age-related sites (~1%).

To determine the contribution of differential or variable methylation to changes in entropy, we calculated entropy for two of the largest datasets (BIOS and FHS) on CpGs that were either DMPs or VMPs. Further, we wanted to determine if constant VMPs (i.e. VMPs that are *not* DMPs) would contribute to the changes in entropy (Supplementary Figure 6). We observed that both DMPs and VMPs show a similar increase in entropy during ageing, which is unsurprising considering that many DMPs are VMPs, and vice-versa. We observed that the *overall* entropy (i.e. regardless of its association with age) is lower for DMPs than for VMPs, which reflects the *type* of CpG affected by differential methylation or by a change in variance. While both DMPs and VMPs affect CpGs whose methylation levels start at high or low levels, VMPs have a greater proportion of CpGs with *intermediate* DNAm levels at baseline (~28% of VMPs are intermediately methylated vs ~20% of DMPs). While the overall entropy at constant VMPs is high, we did not observe an increase in entropy at those sites, suggesting that it is the differential shifts in DNAm towards the mean that contribute to the overall increases in entropy with age.

To further our investigation into the contribution of DMPs to changes in entropy, we used the BIOS blood dataset, which has a large sample size and distribution of samples across a large age range (Figure 7). We observed that although the majority of DMPs (~73%) *converge* to the mean with age, one third of DMPs (~27%) *diverge* away from the mean towards high and low methylation fractions (Figure 16A). To determine this effect on entropy, we then recalculated entropy on the converging and diverging DMPs, respectively. Remarkably, we found a highly significant **increase** in entropy in the converging sites of 0.005 increase in entropy per decade of age (p-value < 2.2e-16), and a stark contrast with the diverging sites, which significantly **decrease** entropy with age, and could be considered "anti-entropic" since they become *more* predictable with age (Figure 16B).
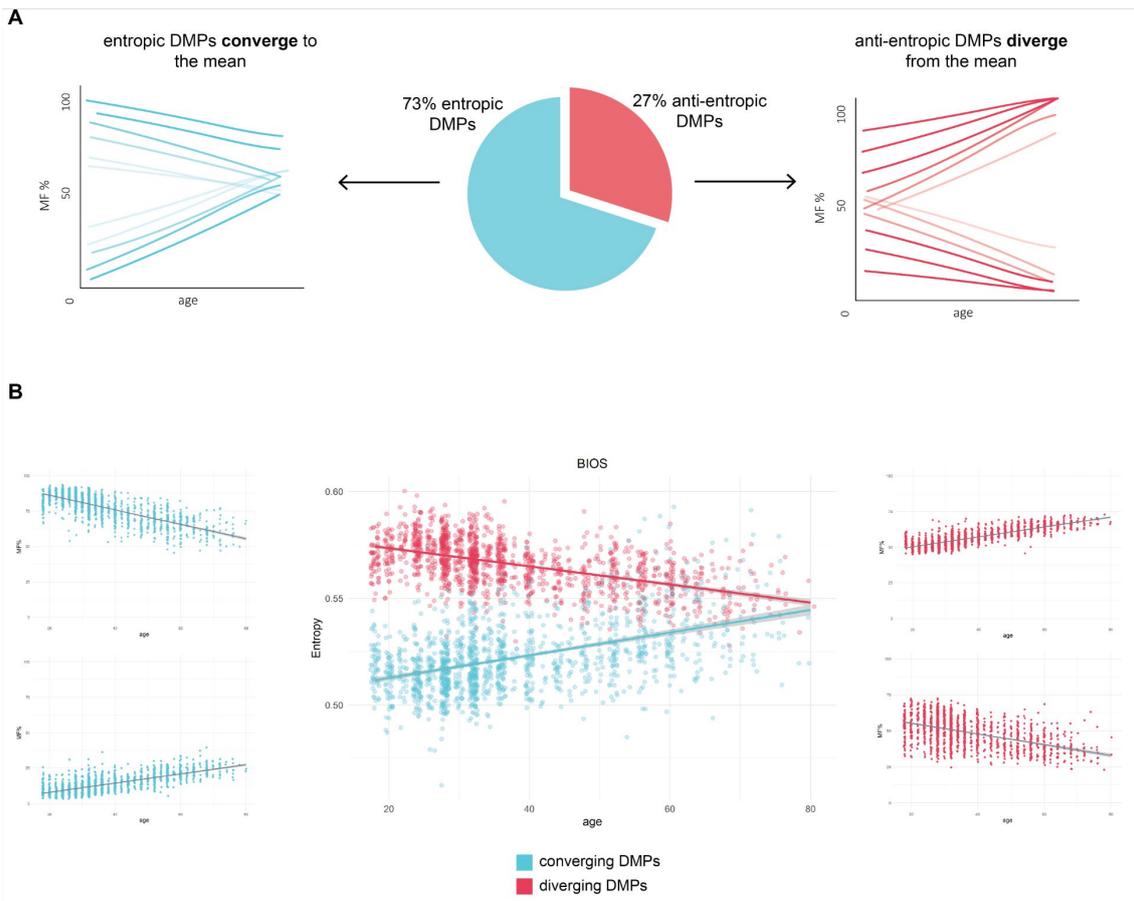
Figure 16 Contribution of differential methylation to changes in entropy with age.

A) A pie chart of the proportion of differentially methylated positions (DMPs) that are entropic and converge to the mean (blue) and the proportion of DMPs that are anti-entropic and diverge from the mean (red) for the BIOS dataset in blood. The arrows point to hypothetical graphs that illustrate the aggregate regression lines for the CpGs that converge (left) and diverge (right) with age. B) In the far left (blue) and far right (red) panels are the methylation plots for two CpG sites, respectively, that are highly or lowly methylated in young (blue) and converge to the mean with age, and two methylation plots that are intermediately methylated in young (red) and diverge from the mean with age.

To determine the influence of cell type heterogeneity on the entropy dynamics in blood, we repeated the analyses after adjusting for blood cell types. There was a moderate correlation of 0.53 (*p*-value = 2.9e-5) between the effect sizes before vs after adjustment for cell type proportions (Supplementary Figure 7A), yet the overall meta-analysis effect size remained unchanged (0.0005 change in entropy per decade of age) (Supplementary Figure 7B). The moderate correlation likely reflects the large uncertainty around the entropy estimates for individual datasets (see the wide error bar on forest plots).

Chapter 5: Signatures of differential and variable DNAm, entropy and ageing in brain, buccal, adipose, skin, and muscle tissue

In this chapter, we repeated the analyses performed in blood (Chapter 4) in five other tissues: brain, buccal, adipose, skin, and muscle tissue.

1. Brain confers a unique signature of variable methylation and age

First, we conducted a meta-analysis of differential methylation and age in 27 brain datasets sourced from different brain regions, reaching a total sample size of 5,094 methylomes. We identified 65,416 DMPs (9% of tested CpGs) of which the *average* DNAm levels change with older age (FDR < 0.005) (Figure 17A), 46% of which decreased in DNAm levels ('hypoDMPs').

Figure 17 Differential meta-analysis of age in brain.

A) A volcano plot displaying the meta-analysis effect size (x-axis) and meta-analysis significance (y-axis) for the 694,945 tested CpGs in the differential methylation meta-analysis of age in brain tissue. The strongest associations have the smallest *p*-values and will be the highest points on the plot. Hypomethylated (hypoDMP), hypermethylated (hyperDMP) and non-DMPs are represented by the colours in the legend. DMPs are classified at a false discovery rate (FDR) < 0.005. B) A forest plot of a highly significant CpG (cg08342886). The dataset and sample size are on the left side of the plot, with the corresponding effect size and errors represented by the point and error bars. The meta-analysis effect size is represented by the purple polygon. The methylation plots for this CpG from two independent brain cohorts, including GSE72778 and GSE64509, are displayed on the far right, with age on the x-axis and methylation fraction (MF) on the y-axis. Each point on the plot represents a single sample.

We then meta-analysed the same 27 brain datasets to detect changes in methylation variability during ageing (see Methods section 2.3 for meta-analysis rationale). We identified 506 VMPs (0.1% of tested CpGs) at FDR < 0.005 (Figure 18A). Contrary to blood, we found roughly equal numbers of VMPs that *increase* (52% of VMPs) and *decrease* (48% of VMPs) in variance with age. These results are robust as we only included CpGs present in ≥ 750 samples (15% of brain samples). Furthermore, the paucity of identified VMPs is surprising, considering the relatively large sample size in this tissue (for reference, ~34% of tested CpGs were VMPs in blood). As in the case for blood, the
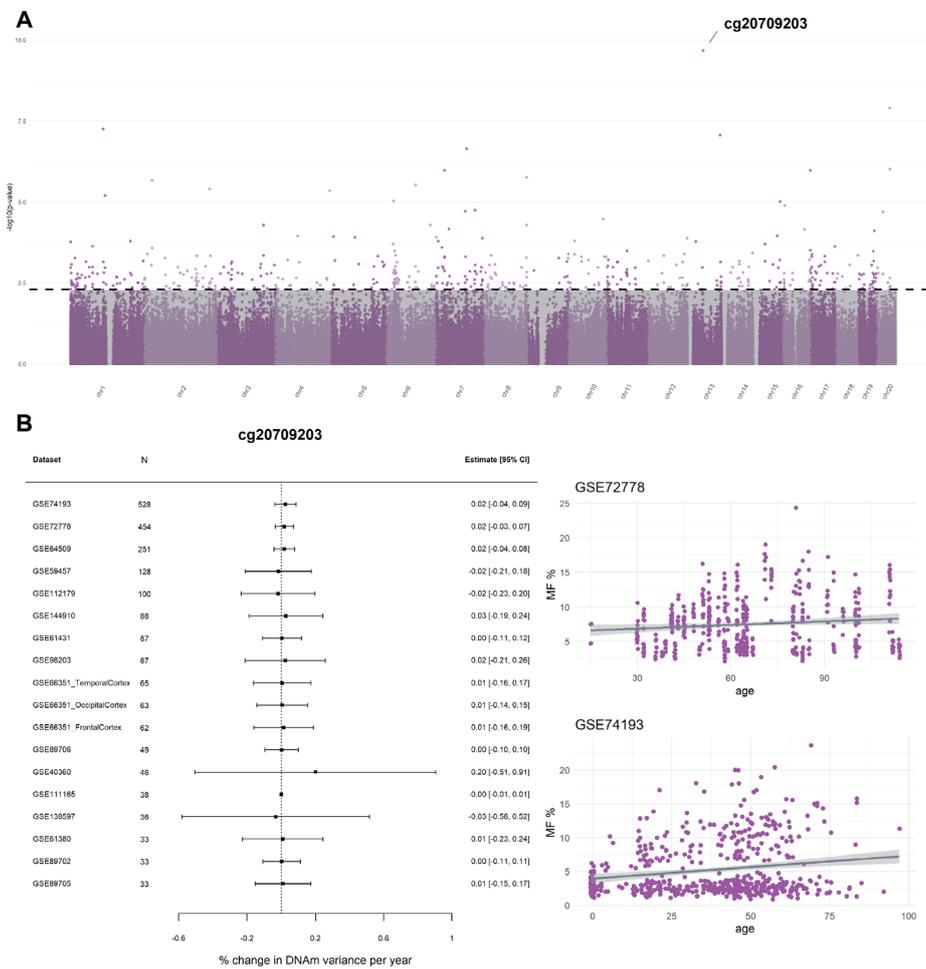
Figure 18 Variable meta-analysis of age in brain.

A) A Manhattan plot to display the genome-wide distribution of significant VMPs in brain. The autosomal chromosomes are displayed along the x-axis, and the negative logarithm of the *p*-values on the y-axis for each of the 372,762 CpGs included in the meta-analysis. The strongest associations have the smallest *p*-values and will be the highest points on the plot. The dashed line is the FDR threshold that separates the significant VMPs above the line, and the non-significant CpGs below the line in the grey shaded area. B) Forest plot of most significant VMP (cg20709203) that is hypermethylated with age and increases in variance with age, and the two adjacent methylation plots from two independent datasets, GSE72778 and GSE74193.

shifts in variance with age were very small, as evidenced by the VMP cg20709203 (Figure 18B). There is an overlap of 264 VMPs that are also DMPs, 70% of which decrease in variance with age.

Lastly, we determined whether the ageing brain methylome increases in entropy with age. When taking all CpGs into account, we did not detect a significant change in entropy (meta-analysis p-value = 0.13) (Supplementary Figure 8). However, if we only calculated entropy on the age-related CpGs in brain (i.e. all DMPs and VMPs), we found a tiny (when compared to tissues such as blood), yet highly significant increase in entropy of 0.0007 per decade of age (p-value <0.0001) (Figure 19A). Unsurprisingly, entropy calculated on non-age-related sites showed no significant increase during ageing (p-value = 0.15, Figure 19A). The increase in entropy in brain was much smaller than for blood, but the heterogeneity between datasets was just as stark ($I^2 = 81\%$). Since majority of the age-related CpGs are DMPs, the increase in entropy in brain is largely driven by differential shifts (Figure 19B).
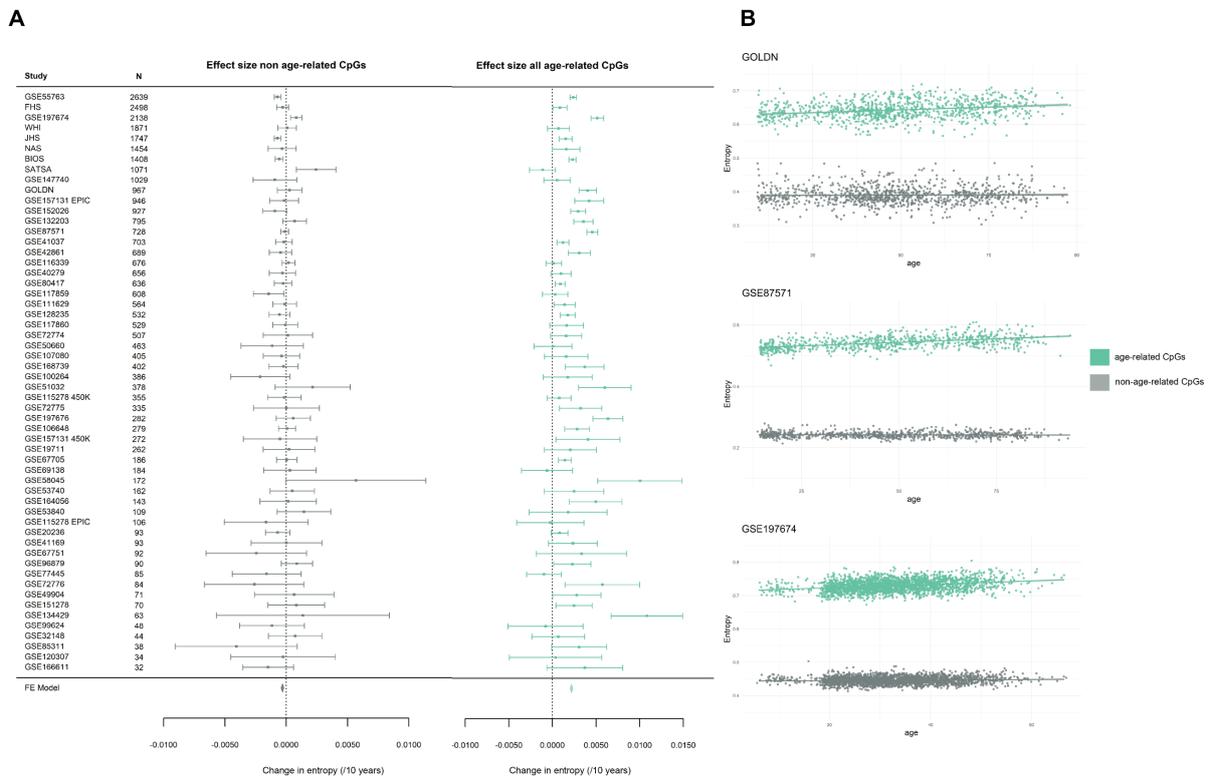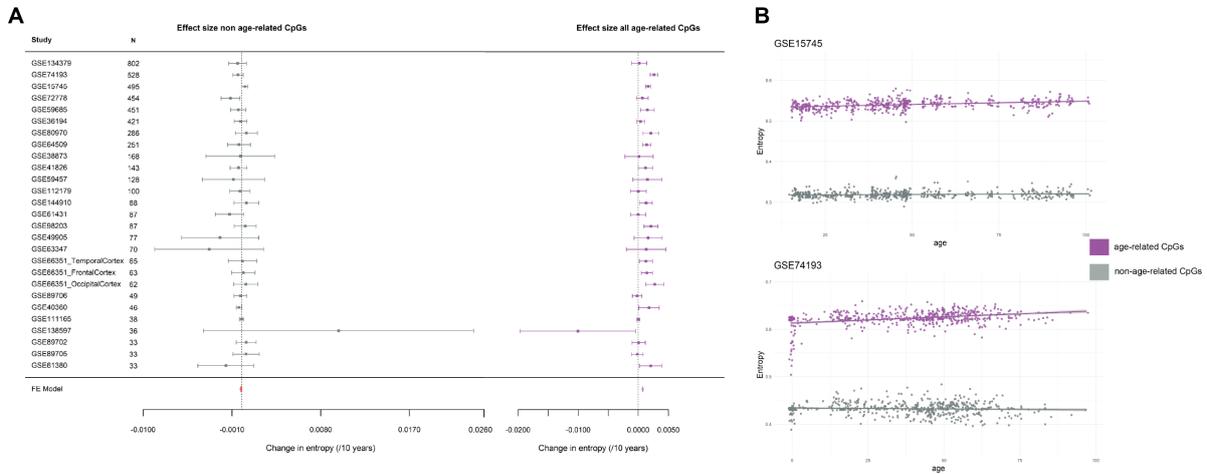

Figure 19 Meta-analysis of entropy and age in brain.

A) A forest plot of the two meta-analyses comparing the changes in entropy between the non-age-related CpGs (left) and the age-related CpGs (right) identified in brain. The meta-analyses effect sizes are represented by the orange and lilac polygons. On the x-axes is the change in entropy per decade of age, and on the y-axes are the effect sizes and standard error measurements from the independent EWAS. The dataset name and sample size (N) are to the left of the forest plot. B) Two graphs with entropy (y-axis) plotted against age (x-axis) for the age-related CpGs and non-age-related CpGs for two independent brain datasets, GSE15745 and GSE74193.

2. Buccal tissue exhibits meaningful shifts in the methylome despite the limited sample size of this tissue

We reached a total sample size of 473 from 4 datasets in buccal tissue. We identified 763 DMPs (0.2 % of tested CpGs) (FDR < 0.005) (Figure 20A), the majority of which (75%) were hypermethylated. In this tissue, we could adjust for cell type proportions (immune cells, epithelial cells, and fibroblasts) and identified 5,552 DMPs after correction, of which 761 overlapped with the results from the unadjusted meta-analysis (Fischer's exact test p value < 2.2×10^{-16}). There was a strong correlation between the effect sizes for the cell type adjusted meta-analysis and the unadjusted meta-analysis (Pearson's correlation = 0.73, p value < 2.2×10^{-16}) (Supplementary Figure 9). We therefore identified an additional 4,791 DMPs whose change in DNAm with age was only visible after taking cellular heterogeneity into account.

Figure 20 Differential meta-analysis of age in buccal.

A) A volcano plot displaying the meta-analysis effect size (x-axis) and meta-analysis significance (y-axis) for the 425,823 tested CpGs in the differential methylation meta-analysis of age in buccal tissue. The strongest associations have the smallest p-values and will be the highest points on the plot. Hypomethylated (hypoDMP), hypermethylated (hyperDMP) and non-DMPs are represented by the colours in the legend. DMPs are classified at a false discovery rate (FDR) < 0.005. B) A forest plot of a highly significant CpG (cg10501210). The dataset and sample size are on the left side of the plot, with the corresponding effect size and errors represented by the point and error bars. The meta-analysis effect size is represented by the purple polygon. The methylation plots for this CpG from two independent buccal cohorts, including GSE94876 and GSE92767, are displayed on the far right, with age on the x-axis and methylation fraction (MF) on the y-axis. Each point on the plot represents a single sample.

Despite the smaller sample size for this tissue, we identified 1,875 VMPs (0.5% of tested CpGs) at FDR < 0.005, all of which *increased* in variance with age (Figure 21A). For robustness, we only included CpGs that were present in >=70 samples (15% of buccal samples). After adjusting for cell type proportions, we did not identify any VMPs and the effect sizes were only moderately consistent between the two analyses (Pearson's correlation of meta-analyses Zscores = 0.72, *p*-value < 2.2e-16) (Supplementary Figure 10A). This suggests that the change in variability in DNAm levels with age in buccal are explained by changes in the relative proportions of different cell types. However, this does not preclude the possibility of VMPs in buccal tissue that change with age independently of cell type proportion, we were likely underpowered to detect VMPs in this tissue (see inflation of p-value histogram, Supplementary Figure 10B).

Figure 21 Variable meta-analysis of age in buccal.

A) A Manhattan plot to display the genome-wide distribution of significant VMPs in buccal. The autosomal chromosomes are displayed along the x-axis, and the negative logarithm of the adjusted *p*-values on the y-axis for each of the 365,932 CpGs included in the meta-analysis. The strongest associations have the smallest *p*-values and will be the highest points on the plot. The dashed line is the FDR threshold that separates the significant VMPs above the line, and the non-significant CpGs below the line in the grey shaded area. B) Forest plot of most significant VMP (cg11074814) that is hypermethylated with age and increases in variance with age, and the two adjacent methylation plots from two independent datasets, GSE78874 and GSE94876.

To determine whether the buccal methylome changes in entropy with age, we conducted a meta-analysis of entropy and age. When taking into account all CpGs, we failed to detect any significant change in entropy (*p*-value = 0.81) (Supplementary Figure 11A), even when adjusting for the two most abundant cell types in this tissue (*p*-value = 0.57) (Supplementary Figure 11B).

In agreement with previous results in blood and brain, we detected an increase in entropy when restricting the entropy calculations to age-related CpGs (DMPs and VMPs) (Figure 22A). The effect size in buccal tissue was similar to that of blood, with an increase in entropy of 0.004 per decade of age, and with substantial heterogeneity between datasets (*p*-value < 0.0001, I^2 statistic 82%). Unsurprisingly, entropy did not show any increase with age when calculated on non-age-related CpGs (*p*-value = 0.87) (Figure 22A). This corroborated the results from the other tissues and is in accordance with the hypothesis, namely that the changes in entropy are driven by the collective changes occurring at the individual age-related CpGs, majority of which are DMPs. We also noted that baseline entropy for the non-age-related CpGs was much lower than for the age-related sites (Figure 22B).

Figure 22 Meta-analysis of entropy and age in buccal.

A) A forest plot of the two meta-analyses comparing the changes in entropy between the non-age-related CpGs (left) and the age-related CpGs (right) identified in buccal. The meta-analyses effect sizes are represented by the orange and purple polygons. On the x-axes is the change in entropy per decade of age, and on the y-axes are the effect sizes and standard error measurements from the independent EWAS. The dataset name and sample size (N) are to the left of the

forest plot. B) A graph with entropy (y-axis) plotted against age (x-axis) for the age-related CpGs and non-age-related CpGs for an independent buccal dataset, GSE94876.

3. Adipose tissue displays the greatest increase in entropy with age compared to all analysed tissues

With a total sample size of 1,313 samples from 8 datasets, we identified 13,093 DMPs (2% of tested CpGs) at FDR < 0.005 (Figure 23A), 74% of which showed an *increase* in mean DNAm levels with age.

We then conducted a *sample size* based meta-analysis of variable methylation and age in 7 of the 8 adipose datasets with a total sample size of 1,243. GSE68336 was not included in the VMP analysis, as we do not have the phenotype data available for this dataset. We did not identify any VMPs (FDR < 0.005), but we were likely underpowered in this tissue, as we detected a subtle inflection in the *p*-value histogram (Supplementary Figure 12).
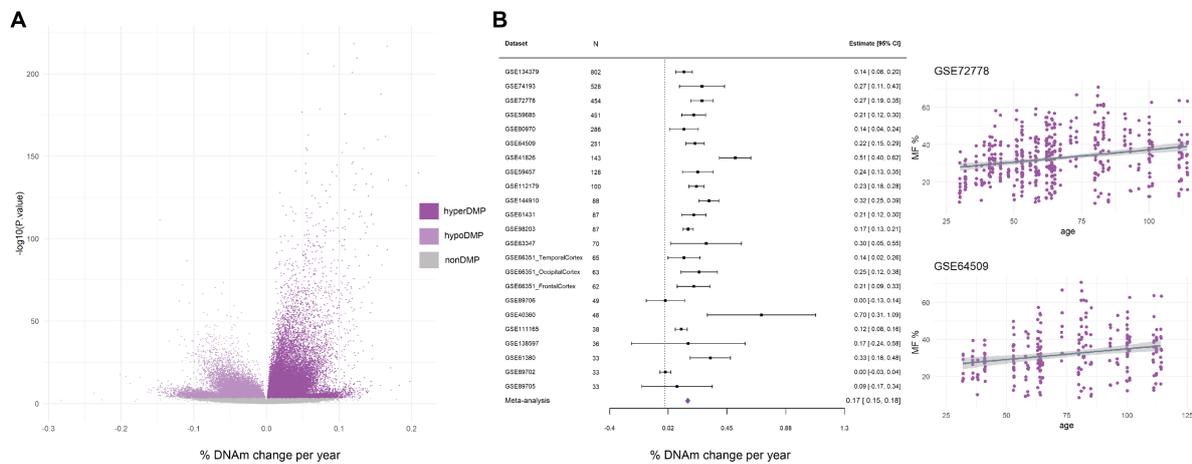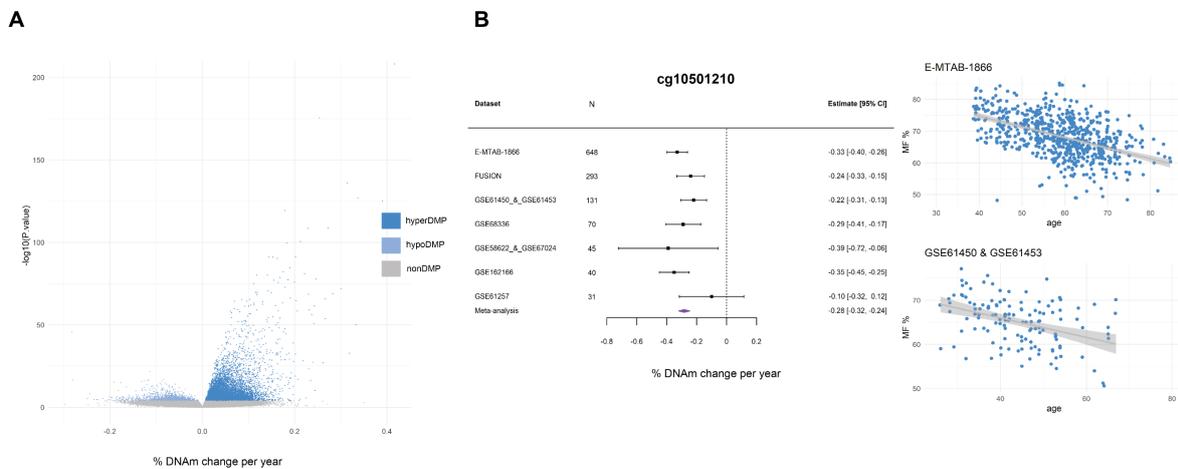

Figure 23 Differential meta-analysis of age in adipose.

A) A volcano plot displaying the meta-analysis effect size (x-axis) and meta-analysis significance (y-axis) for the 658,080 tested CpGs in the differential methylation meta-analysis of age in adipose tissue. The strongest associations have the smallest *p*-values and will be the highest points on the plot. Hypomethylated (hypoDMP), hypermethylated (hyperDMP) and non-DMPs are represented by the colours in the legend. DMPs are classified at a false discovery rate (FDR) < 0.005. B) A forest plot of a highly significant CpG (cg10501210). The dataset and sample size are on the left side of the plot, with the corresponding effect size and errors represented by the point and error bars. The meta-analysis effect size is represented by the purple polygon. The methylation plots for this CpG from two independent adipose cohorts, including E-MTAB-1866 and GSE61450 & GSE61453, are displayed on the far right, with age on the x-axis and methylation fraction (MF) on the y-axis. Each point on the plot represents a single sample.

Once again, we then conducted a fixed-effects meta-analysis of entropy and age in adipose tissue with a combined sample size of 1,243 from 7 adipose datasets. As above, GSE68336 was not included in the entropy analysis. Results in adipose were entirely consistent with other tissues: entropy showed no significant change when calculated over all tested CpGs (p-value = 0.016) (Supplementary Figure 13), or when restricted to the non-age-related CpGs (p-value = 0.064) (Figure 24A). In contrast, entropy calculations on age-related CpGs yielded a 0.008 increase in entropy per decade of age (p-value <0.0001, I^2 statistic 70%) (Figure 24A), which was the largest increase in entropy detected, of all analysed tissues.
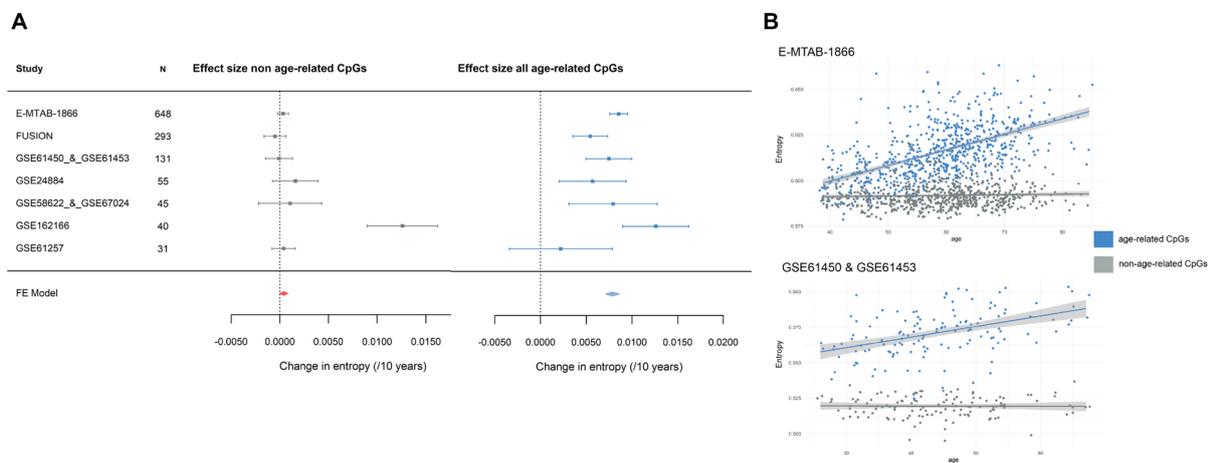
Figure 24 Meta-analysis of entropy and age in adipose.

A) A forest plot of the two meta-analyses comparing the changes in entropy between the non-age-related CpGs (left) and the age-related CpGs (right) identified in adipose. The meta-analyses effect sizes are represented by the orange and royal blue polygons. On the x-axes is the change in entropy per decade of age, and on the y-axes are the effect sizes and standard error measurements from the independent EWAS. The dataset name and sample size (N) are to the left of the forest plot. B) A graph with entropy (y-axis) plotted against age (x-axis) for the age-related CpGs and non-age-related CpGs for two adipose datasets, GSE61450 & GSE61453 and E-MTAB-1866.

4. Muscle is the only tissue that did not display an increase in entropy across the lifespan

With a combined sample size of 1,146 samples from 12 datasets, we identified 68,777 DMPs (9% of tested CpGs) at FDR < 0.005 (Figure 25A), with even numbers of hyper- and hypoDMPs (53 and 47% of DMPs, respectively). Ten of the datasets used in this meta-analysis have been previously meta-analysed[213].
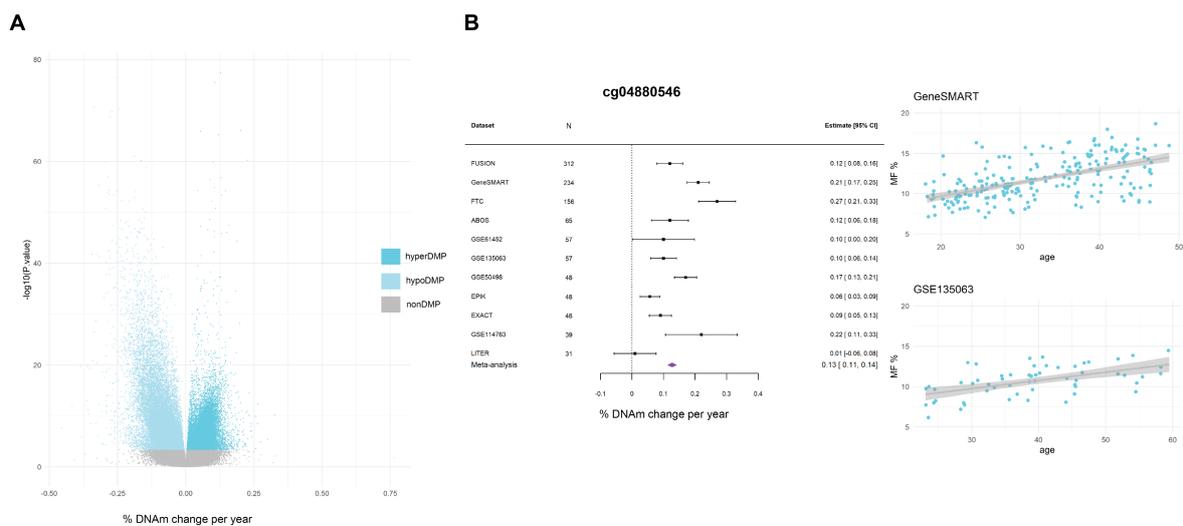
Figure 25 Differential meta-analysis of age in muscle.

A) A volcano plot displaying the meta-analysis effect size (x-axis) and meta-analysis significance (y-axis) for the 780,032 tested CpGs in the differential methylation meta-analysis of age in muscle tissue. The strongest associations have the smallest *p*-values and will be the highest points on the plot. Hypomethylated (hypoDMP), hypermethylated (hyperDMP) and non-DMPs are represented by the colours in the legend. DMPs are classified at a false discovery rate (FDR) < 0.005. B) A forest plot of a highly significant CpG (cg04880546). The dataset and sample size are on the left side of the plot, with the corresponding effect size and errors represented by the point and error bars. The meta-analysis effect size is represented by the purple polygon. The methylation plots for this CpG from two independent muscle cohorts, including GeneSMART and GSE135063, are displayed on the far right, with age on the x-axis and methylation fraction (MF) on the y-axis. Each point on the plot represents a single sample.

We then conducted a *sample size* based meta-analysis of variable methylation and age in 11 muscle datasets with a total sample size of 991. The FTC cohort was not included in this analysis, as we do not have the raw data and phenotypes are not available. We did not identify any VMPs at FDR < 0.005. The muscle datasets are small and show a limited distribution of samples across the age range, which makes it challenging to detect VMPs. As suggested by the *p*-value histogram, we may have simply been underpowered to detect VMPs (Supplementary Figure 14).

In muscle, we failed to detect any significant change in entropy, whether we considered all tested CpGs (*p*-value = 0.35) (Supplementary Figure 15), and non-age-related CpGs (*p*-value = 0.29) (Figure 26A). To our surprise, we also failed to detect any change in entropy when considering age-related CpGs (*p*-value = 0.88) (Figure 26A). Muscle is, therefore, the only tissue that displayed no significant increase in entropy across the lifespan. It is unclear why muscle was an exception in this regard, but we have suggested a hypothesis in the discussion (See Chapter 6). As for other tissues, however, the baseline

entropy for the age-related CpGs was much higher than for the non-age-related CpGs (Figure 26A). A limitation in muscle tissue is that the datasets are small in size, and datasets with larger sample sizes such as FUSION do not have sufficient spread of samples across a broad age range, which may be why we are unable to detect significant changes in this tissue. Although some datasets do display increases in entropy with age, such as GeneSMART, the results are highly heterogenous across the different muscle datasets.
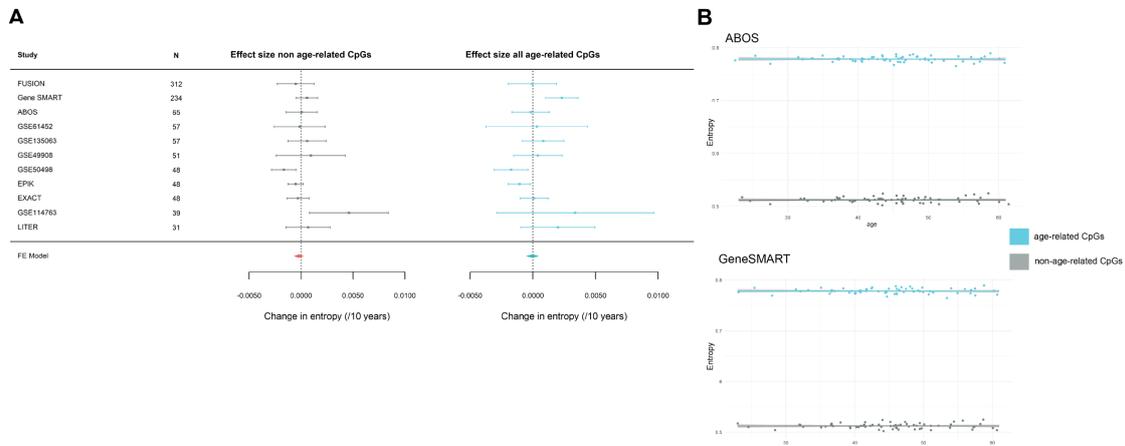
Figure 26 Meta-analysis of entropy and age in muscle.

A) A forest plot of the two meta-analyses comparing the changes in entropy between the non-age-related CpGs (left) and the age-related CpGs (right) identified in muscle. The meta-analyses effect sizes are represented by the orange and light blue polygons. On the x-axes is the change in entropy per decade of age, and on the y-axes are the effect sizes and standard error measurements from the independent EWAS. The dataset name and sample size (N) are to the left of the forest plot. B) A graph with entropy (y-axis) plotted against age (x-axis) for the age-related CpGs and non-age-related CpGs for two muscle datasets, ABOS and GeneSMART.

5. Skin tissue exhibits large shifts in differential methylation and entropy relative to other tissues

With a total sample size of 668 from 6 datasets, we identified 69,956 DMPs (19% of tested CpGs) in skin (FDR < 0.005) (Figure 27A), the majority of which were hypermethylated with age (63% of DMPs were 'hyperDMPs').
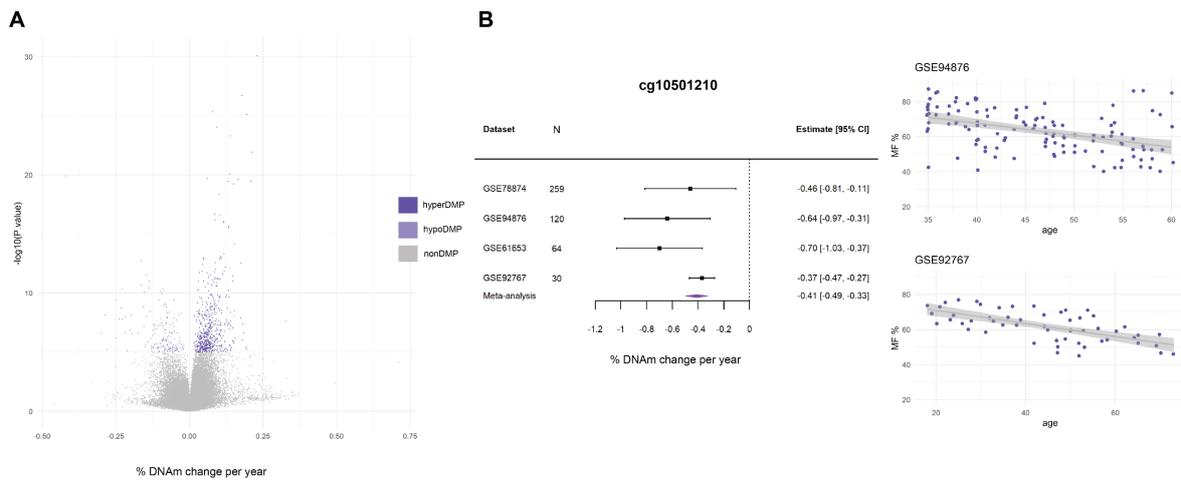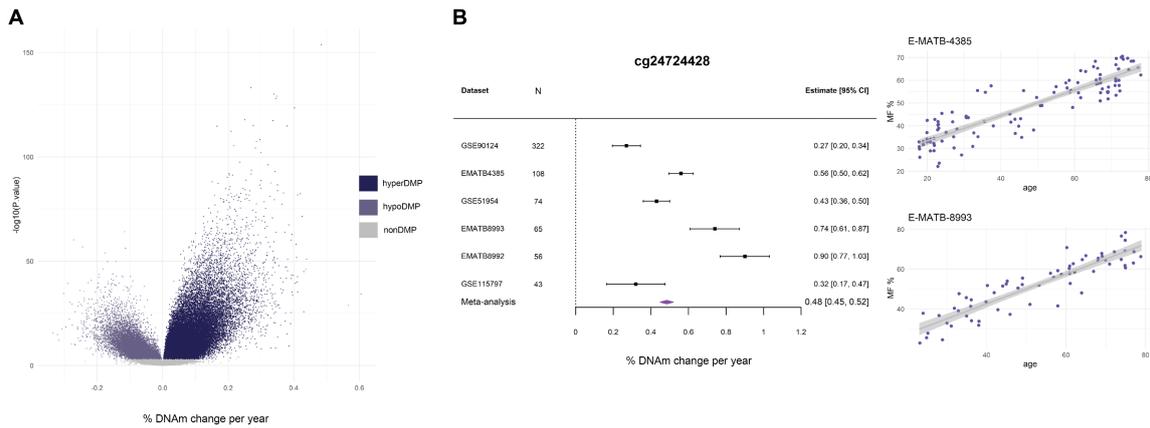

Figure 27 Differential meta-analysis of age in skin.

A) A volcano plot displaying the meta-analysis effect size (x-axis) and meta-analysis significance (y-axis) for the 371,112 tested CpGs in the differential methylation meta-analysis of age in skin tissue. The strongest associations have the smallest *p*-values and will be the highest points on the plot. Hypomethylated (hypoDMP), hypermethylated (hyperDMP) and non-DMPs are represented by the colours in the legend. DMPs are classified at a false discovery rate (FDR) < 0.005. B) A forest plot of a highly significant CpG (cg24724428). The dataset and sample size are on the left side of the plot, with the corresponding effect size and errors represented by the point and error bars. The meta-analysis effect size is represented by the purple polygon. The methylation plots for this CpG from two independent skin cohorts, including E-MATB-4385 and E-MATB-8993, are displayed on the far right, with age on the x-axis and methylation skin (MF) on the y-axis. Each point on the plot represents a single sample.

Despite the relatively small sample size in this tissue, we identified 1,267 VMPs (0.3% of tested CpGs) at FDR < 0.005 (Figure 28A), the vast majority of which *increase* in variance with age (97% of VMPs).
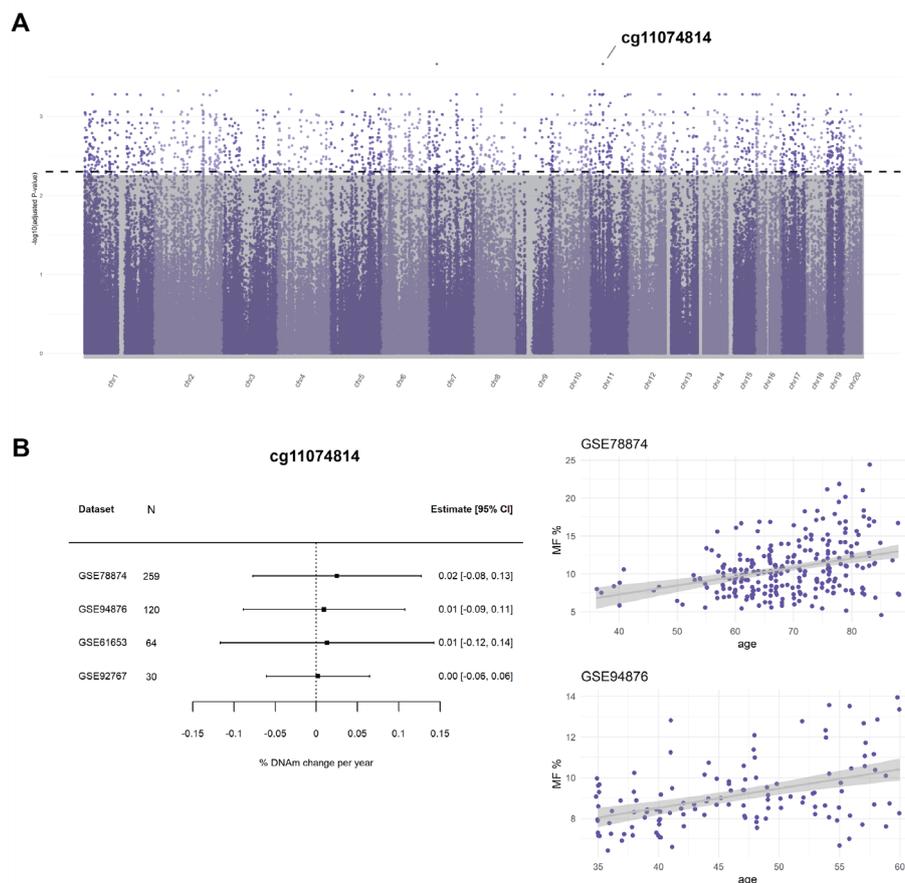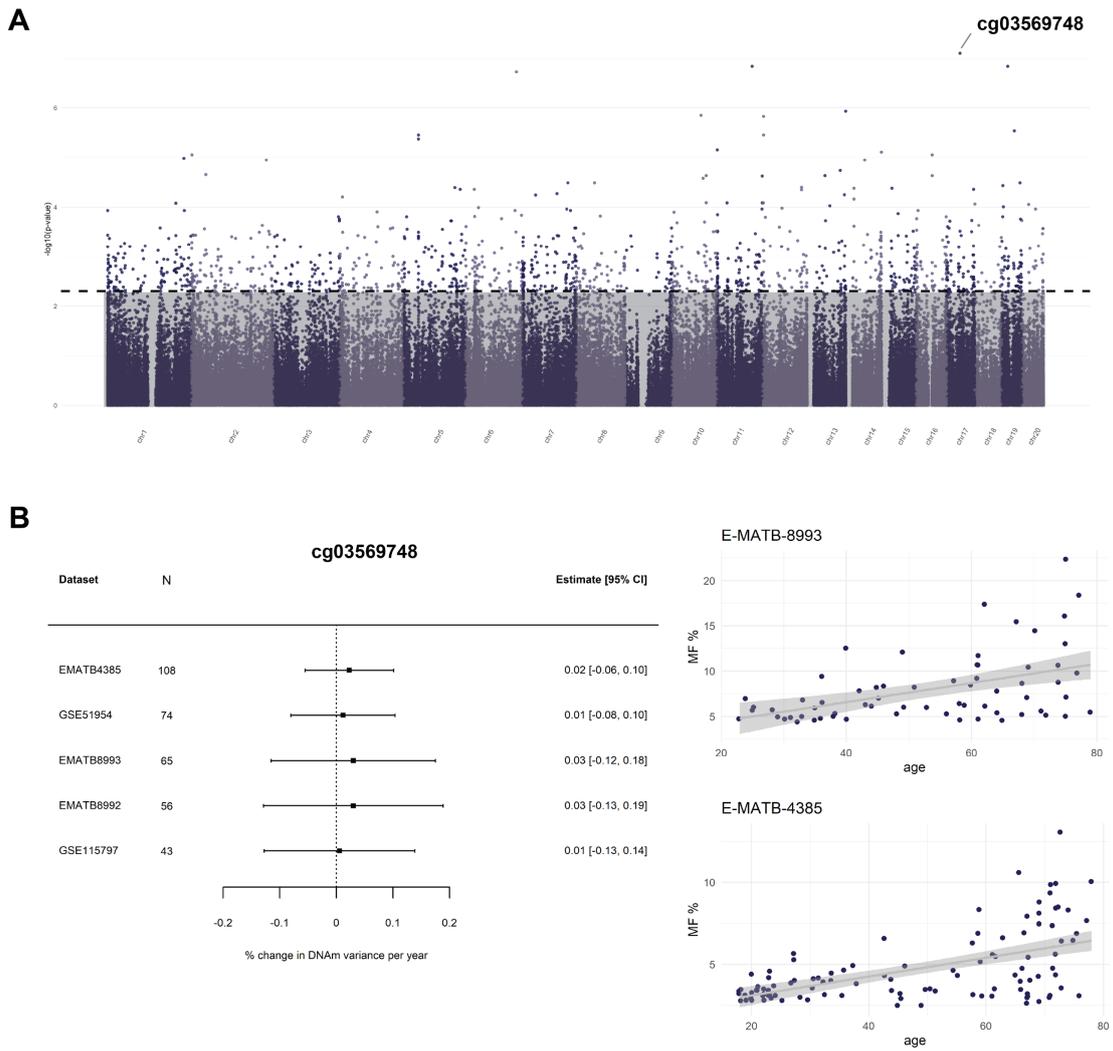
Figure 28 Variable meta-analysis of age in skin.

A) A Manhattan plot to display the genome-wide distribution of significant VMPs in skin. The autosomal chromosomes are displayed along the x-axis, and the negative logarithm of the adjusted *p*-values on the y-axis for each of the 370,802 CpGs included in the meta-analysis. The strongest associations have the smallest *p*-values and will be the highest points on the plot. The dashed line is the FDR threshold that separates the significant VMPs above the line, and the non-significant CpGs below the line in the grey shaded area. B) Forest plot of most significant VMP (cg03569748) that is hypermethylated with age and increases in variance with age, and the two adjacent methylation plots from two independent datasets, E-MATB-8993 and E-MATB-4385.

As in the other tissues, we then sought to determine whether entropy increases in the skin methylome. In the genome-wide meta-analysis of entropy and age we found the skin methylome significantly increases in entropy by 0.001 per decade of age (*p*-value < 0.0001) (Supplementary Figure 16). As with all other tissues, entropy calculated on non-age-related sites did not show any significant increase in entropy (*p*-value = 0.64) (Figure 29A). However, entropy calculated only on the age-related sites yielded a significant

0.005 increase in entropy per decade of age, with substantial heterogeneity between datasets (*p*-value < 0.0001, I^2 statistic 88%) (Figure 29A). This constitutes the second largest change in entropy behind adipose tissue.
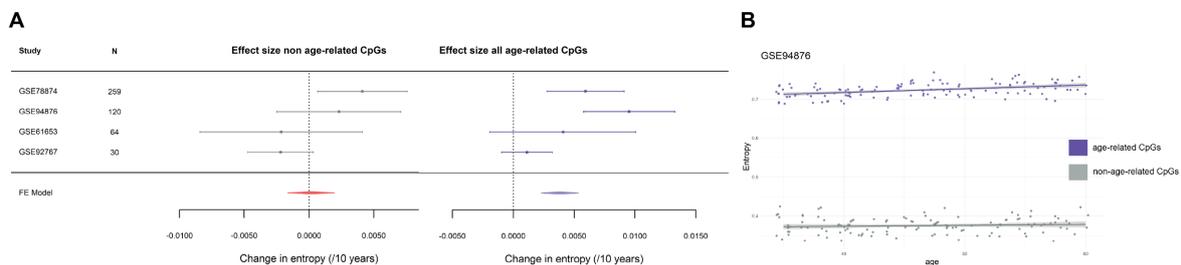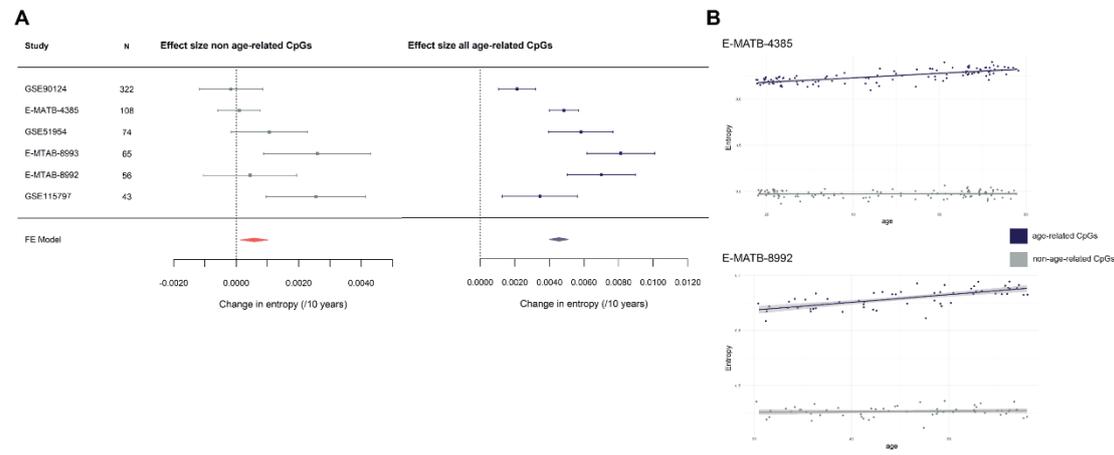

Figure 29 Meta-analysis of entropy and age in skin.

A) A forest plot of the two meta-analyses comparing the changes in entropy between the non-age-related CpGs (left) and the age-related CpGs (right) identified in skin. The meta-analyses effect sizes are represented by the orange and dark purple polygons. On the x-axes is the change in entropy per decade of age, and on the y-axes are the effect sizes and standard error measurements from the independent EWAS. The dataset name and sample size (N) are to the left of the forest plot. B) A graph with entropy (y-axis) plotted against age (x-axis) for the age-related CpGs and non-age-related CpGs for two skin datasets, E-MATB-4385 and E-MATB-8992.

In summary, tissues exhibit highly interesting patterns in their respective ageing methylomes (Figure 30). This leaves us with a myriad of questions about why tissues age differently, which we will attempt to answer in the next chapter of this thesis.

brain
9% DMPs
46% hypo
54% hyper

0.1% VMPs
52% increase
48% decrease

0.0007 change in entropy / decade

buccal
0.2% DMPs
25% hypo
75% hyper

0.5% VMPs
100% increase
0% decrease

0.004 change in entropy / decade

adipose
2% DMPs
36% hypo
74% hyper

0% VMPs

0.008 change in entropy / decade

skin
19% DMPs
37% hypo
63% hyper

0.3% VMPs
97% increase
3% decrease

0.005 change in entropy / decade

whole blood
48% DMPs
66% hypo
34% hyper

37% VMPs
99% increase
1% decrease

0.002 change in entropy / decade

muscle
9% DMPs
47% hypo
53% hyper

0% VMPs

no change in entropy / decade

Figure 30 Summary figure of the results from the differential, variable and entropy analysis in each human tissue, including whole blood, skin, adipose, brain, buccal and muscle.

For each tissue, we detail the proportion of differentially methylated positions (DMPs), variably methylated positions (VMPs) and change in entropy at the age-related CpGs that were identified for each tissue, respectively.

Chapter 6: Discussion

This chapter will discuss the results from Chapter 4 and 5, addressing our current knowledge, where this work has contributed to the field of epigenetic ageing, and the limitations thereof. We will also explore ideas for future research.

1. Establishing the multi-tissue database of the human methylome

We have collected and pre-processed 40,830 human methylomes, from 113 datasets, spanning 6 different tissues. To our knowledge, this is the largest database of human methylomes, and is the first major output of this study. Whole blood was by far the tissue best represented, comprising the bulk (79%) of the samples, followed by brain tissue (12%), adipose (3%), muscle (3%), skin (2%) and buccal tissue (1%).

We noted that our ability to detect all kinds of age-related changes relied first and foremost on the *sample size* of the tissue cohort and on the *age distribution within individual datasets*. This is particularly relevant for the Breusch-Pagan test for heteroscedasticity used to detect VMPs, as it requires large sample sizes and a broad age range, and possibly the presence of individuals at the 'upper limit' of old age. A recent preprint revealed that the variance of DNAm in mouse blood did not increase linearly with age but was instead only visible at very old ages [299]. Therefore, low sample size, narrow age range, and the scarcity of very old individuals are probably the reason for only few human studies investigating the changes in variance or stochasticity with age. In comparison, the standard linear model used to detect DMPs is less sensitive to the characteristics of the dataset substructure. It should be noted that datasets varied widely in sex distribution, disease status and other factors that could *modulate* the effect of age on the methylome (e.g. age-related changes may be more pronounced in different ethnicities, or in different sexes, which introduces some 'noise'/variability in observed effect size). We see this heterogeneity as a strength more than a weakness when you leverage a large sample size, because it implies that the DMPs, VMPs and entropy we detected stand true for a broad range of individuals and can be considered "universal". One caveat to note is that the ethnic representation of the samples was overwhelmingly white, clearly highlighting the need to obtain a better representation of ethnicities in DNAm profiles.

Whole blood contains very large individual datasets, 43% of which have >500 samples. Moreover, samples within these datasets are widely distributed across the age range (i.e. samples are spread from very young to very old ages). Blood is therefore the most statistically powered tissue to identify age-related changes in the methylome. Brain is the second largest tissue cohort, however only 7% (2 out of 27) of the datasets contained > 500 samples. Many of the brain datasets have a skewed distribution of samples towards older ages. For example, GSE134379 has the largest sample size, but the majority of the samples lie between 80 – 100 years. There is a greater proportion of males in this tissue (63%), and datasets came from diverse brain regions with potentially vastly different ageing patterns. While the substructure of the individual datasets is inferior to blood, the large sample size was sufficient to reach high statistical power in this tissue. Adipose tissue comprised only one dataset with a large sample size >500, with majority of the datasets with <70 samples. The distribution of samples across the age range for adipose was consistent across all datasets, with no obvious skewing of samples to very young or very old ages. There were, however, datasets with samples sourced from different adipose tissue (subcutaneous or visceral), which could introduce noise in the detected effect sizes. Importantly, adipose tissue is predominantly female (77%). The remaining tissues, muscle, skin and buccal, had no datasets with very large sample numbers (>500). In muscle, the samples are predominantly male (71%), and are skewed towards younger ages and 'healthy' phenotypes. Skin samples are also predominantly female (89%), but similarly to muscle, majority of samples are from 'healthy' tissue. The skin datasets span a large age range, despite the smaller sample sizes of individual datasets. Buccal tissue is the smallest tissue cohort that is slightly more male (68%). Two out of the four buccal datasets had samples with different ethnicities, which was accounted for in the analysis.

In summary, datasets with substructures skewed to a larger proportion of males, include brain, muscle and buccal, whereas adipose and skin have predominantly females, and blood is evenly proportioned across sexes. Muscle and skin have predominantly 'healthy' phenotypes. Blood and buccal tissue have more ethnic diversity in their population substructure. Brain tissue has a very old cohort, and muscle has a young cohort. When looking at the specifics of each tissue, these substructures give insight into our findings.

2. Building a comprehensive picture of the ageing blood methylome

Ageing in the blood methylome is characterised by widespread differential shifts (48% of tested CpGs are DMPs) that accumulate slowly and are small in magnitude, with a greater proportion of DMPs that lose methylation with age, yet many hypermethylated and hypomethylated DMPs trend towards intermediate levels over time (~70% of DMPs). This implies that a substantial fraction of DMPs are intermediately methylated in young, and diverge from the mean and trend towards high and low DNAm fractions over time (~30% of DMPs). While we know from previous studies that differential methylation is a feature of ageing in blood[43,50,57,59,298,300], our findings are unparalleled, revealing the sheer magnitude and omnipresence of these changes that may have been previously underappreciated. There is also a significant stochastic or variable component to the blood methylome, with 37% of tested CpGs changing in variance with age. This study is 38 times larger than the largest known study of VMPs in blood[62]. Relative to DMPs, the magnitude of change in VMPs is small, and the majority of VMPs *increase* in variance, or 'stochasticity,' with age. Unlike previously reported[62], we noted a significant overlap between DMPs and VMPs (i.e. the CpGs that change in average methylation and variance with age), but we also identified a notable proportion of CpGs that change *only* in the average methylation (i.e. 41% of DMPs are homoscedastic), and conversely there are age-related changes at CpGs independent of differential shifts (i.e. 20% of VMPs are constant VMPs). Therefore, the largest class of age-related linear changes in blood are those that change in both average methylation *and* variance (i.e. CpGs classified as both DMPs and VMPs), followed by CpGs that change *only* in average methylation (i.e. DMPs only), and then CpGs that change *only* in variance (i.e. constant VMPs) (see Chapter 2, "Literature Review").

We then explored these age-related changes in the context of entropy, a well-described feature of epigenetic ageing that captures the state of disorder or 'chaos' in a set of CpGs [43,62,207,299]. By comparing the change in entropy on the subset of age-related CpGs (i.e. a complete list of the CpGs that are DMPs, VMPs or both) and non-age-related CpGs identified in blood, we reported that entropy *increases* only at the age-related sites and remains *unchanged* at the non-age-related sites. This result validates our initial findings that the sites we have classified as DMPs and VMPs are true age-related CpGs, which acquire higher levels of disorder over time. Conversely, non-age-related CpGs maintain lower levels of methylation disorder and do not contribute to overall increases in entropy with age. We also show that the **increase** in entropy is driven by the differential shifts that trend from high and low methylation fractions in young,

to intermediate methylation states at older ages; also referred to as the 'smoothening of the epigenetic landscape,' however, intermediately methylated CpGs that drift *away* from the mean towards fully methylated and unmethylated states **decrease** entropy, exhibiting 'anti-entropic' properties. It has been proposed that these 'anti-entropic' CpGs could represent genes that act to control the increase in 'noise' with age and hence become increasingly regulatd[301]. Since the majority of DMPs are entropic, the net global effect is an increase in entropy with age. Entropy is useful in that it captures the entire state of the methylome in a *single quantitative measure.*

We then explored the biological significance of these age-related changes by investigating their distribution in chromatin states and their enrichment in biological pathways. Our results replicate and substantiate the evidence that gains in average methylation (hyperDMPs) *and* variance (hyperVMPs) accrue at key developmental regions, such as those harbouring active and inactive histones and regions actively repressed by the Polycomb complex[3]. In contrast, CpGs that lose methylation (hypoDMPs) but increase in variance (hypoVMPs) accrue in regions weakly repressed by the Polycomb complex, quiescent states, and enhancers. Interestingly, constant VMPs are enriched in quiescent states and regions weakly repressed by the Polycomb complex, but also in active transcription sites. To further substantiate the above findings were the overlapping biological pathways enriched with VMGs and DMGs. The results from the GO pathway enrichment identified key processes related to development, differentiation and cell-signalling are altered during ageing[3]. Moreover, DMGs and VMGs were highly enriched for CGP expression signatures related to the PRC subunits, such as SUZ12 and EED, as well as repressive and activating histone marks, H3K27me3 and H3K4me3, which co-occupy bivalent promoter regions that prepare key developmental genes to be switched on during differentiation in specific cells, such as embryonic stem cells[302].

Although considerably addressed in the literature review (see Chapter 2, "Origins of age-related DNAm changes"), it is reiterated that a plausible consequence for gains in methylation could be that hypermethylation diminishes the plasticity of genes that are required for differentiation by altering the precise genomic locations that preserve stem cell identity and function[3]. We hypothesise this inhibits the capacity of cells to robustly maintain their identity into old age; a process that is *shared* among individuals during chronological ageing (as evidenced by gains in methylation at DMPs). Age-related loss of cellular identity has been previously reported in gene expression signatures in different somatic tissues of mice, whereby the transcription profiles *diverge* during development and *converge* during ageing[303]. A similar

idea is described in a recent study as "ex-differentiation," whereby aged cells lose their ability to maintain their identity due to epigenetic changes at developmental genes[207].

Less understood is the mechanism of hypomethylation, but it is proposed that the loss of methylation at enhancer regions could reflect tissue-specific operations[3]. If so, comparing our results across other tissues would be necessary to confirm this. It has also been hypothesised that hypomethylation is immunogenic and contributes to age-associated inflammatory processes[304]. Further evidence to support this comes from the pathway enrichment of the HPO gene set, which revealed DMGs and VMGs to be enriched for abnormal inflammatory cascades, and multiple organ system abnormalities, an unsurprising finding given that ageing affects every human tissue. Chronic inflammation has recently been included as a hallmark of ageing, which manifests systemically but is also associated with various ageing phenotypes, such as arteriosclerosis, disc degeneration and neuroinflammation[2], thus is not isolated to a specific tissue.

Lastly, the overlapping CP gene sets that were enriched for DMGs and VMGs largely related to components of the extracellular matrix (ECM). During ageing, the integrity of the ECM declines, leading to a loss of cellular and tissue support that drives cellular ageing processes and ultimately leads to disease progression[305].

Enrichment and pathway analysis give insight into the roles that differential and variable methylation could be playing in chronological and 'biological' ageing. The cumulative erosion of the methylome appears to underpin a cascade of altered biological systems. Differential shifts from high and low methylation states initialised during development in young lose and gain methylation, a mechanism of ageing *shared* by all individuals over the lifespan[299]. Since the plasticity of DNAm means the methylome is susceptible to environmental perturbation[3], stochasticity or variability is introduced at many of the DMPs, shifting the methylation to higher or lower states and contributing to the *differences* in organismal fitness over the lifespan, as detected as VMPs. Daily metabolic stress and the resultant damage accumulation may play a fundamental role in *driving* these changes[306]. A primary role of the metabolism is to increase and maintain organismal fitness, however, the myriad of environmental insults challenges the capacity to maintain this homeostasis, leading to gradual and continuous damage accumulation[306]. The stochasticity or variability is therefore a good indicator of cumulative damage of the environment or non-specific damage, whereby the epigenome is adapting to changing environmental cues[299].

In a nutshell, the core instruction manual that coordinates how the genome must be differently executed in different cells to achieve specific cellular functions becomes disrupted with primary ageing (i.e. DMPs), likely due to daily metabolic stressors, but is also susceptible to additional damage accumulation via secondary ageing mechanisms (i.e. VMPs). Moreover, we hypothesise that additional damage can accumulate at regions that are *unaffected* by primary ageing (i.e. constant VMPs) but can take additional "hits" from environmental insults. Collectively, the gains and losses (DMPs) and stochastic changes (VMPs) contribute to continuous entropic decay, characterised by a loss of cell identity (i.e. reflected by the smoothening of the epigenetic landscape), and ageing.

To build on our knowledge of epigenetic ageing in blood, we then explored these features of ageing, including DMPs, VMPs and entropy, in other somatic tissues, including brain, skin, muscle, buccal and adipose.

3. Quantifying ageing in the brain, adipose, skin, muscle and buccal methylomes

All other investigated tissues, including brain, adipose, buccal, skin and muscle, exhibit *differential* shifts with age to varying degrees. DMPs comprise 19% of the skin methylome, 9% of the brain methylome, 9% of the muscle methylome, 2% of the adipose methylome and only 0.2% of the buccal methylome. Two standout tissues include skin and brain. Skin has a small sample size (n = 668) and majority of the skin samples possess 'normal' phenotypes, yet almost one fifth of the skin methylome is differentially methylated, whereas brain tissue has a substantially larger sample size with many samples coming from individuals with age-related diseases, yet less than one tenth of the brain methylome is differentially methylated with age. The gains and losses in methylation also differed between the tissues. Muscle and brain were the only tissues to have equal proportions of hypermethylated and hypomethylated DMPs. Buccal, skin, and adipose tissue all had a greater proportion of hypermethylated DMPs (75%, 63% and 74%, respectively), and blood was the *only* tissue to exhibit more hypomethylation (66%) with age. It is uncertain why these uneven proportions in gains and losses of methylation exist, but hypermethylation at specific developmental regions such as those marked by the PRC, as in our results in blood, appears to be a conserved feature of ageing across tissues[58,213,214]. Albeit beyond the scope of this project, enrichment of hyper- and hypomethylated DMPs in tissue-specific chromatin states is necessary to verify this.

Aside from blood, *variable* shifts in DNAm were detected only in some of the tissues, and to a much lesser extent. VMPs comprise 0.1% of the brain methylome, 0.5% of the buccal methylome, and 0.3% of the skin methylome. One interesting finding was buccal tissue, whereby the number of VMPs exceeds the number of DMPs in this tissue. We were unable to detect VMPs in adipose tissue and muscle, but this is likely due to the lack of statistical power in these tissues. Brain is particularly interesting in the context of VMPs, as it is the *only* tissue with equal numbers of VMPs that increase and decrease in variance with age. All the remaining tissues where VMPs were detected have > 97% of VMPs *increasing* in variance with age.

To capture the entirety of the age-related changes in the methylome of each tissue, we then compared the change in entropy with age at the age-related CpGs (which were predominantly DMPs) to the non-age-related CpGs. Excluding muscle, the increase in entropy only at the *age-related CpGs* in each tissue validated the results of the independent meta-analyses used to classify DMPs and VMPs in each tissue. It was interesting to note that adipose tissue displays the largest change in entropy with age, followed by skin and then buccal. Comparatively, brain displayed a much smaller increase in entropy with age, and muscle showed no change in entropy with age.

Building on the evidence in blood, the differences in the age-related DNAm changes may reflect the turnover rates of tissues (i.e. the ability of tissue stem cells to maintain tissue homeostasis by balancing cell proliferation with cell death) by accumulating in regions that function to maintain stem cell identity and functionality[3]. In support of this, stem cells and differentiated cells (i.e. non stem cells) isolated from colon tissue in mice display highly correlated age-related shifts (i.e. DMPs) in DNAm, which occur predominantly in CpG islands that become hypermethylated with age [307]. Similar results were also reported in upper and lower intestinal tissue in mice[307,] reiterating that DNAm changes with age are a stem cell phenomenon.

Mammalian tissues have differing proliferative capacities and therefore rates of turnover (i.e. different stem cell division rates)[308]. Tissues such as the brain and heart have lower or very specific regenerative capacities, whereas muscle, adipose and liver regenerate upon injury, and tissues such as the skin, buccal and bone marrow have high proliferative capacities[308]. This has been shown to be reflected by the degree to which entropy increases with age in different mice tissues, whereby tissues with low turnover rates, such as the heart or kidneys, have lower rates of entropy with age, than tissues with greater proliferative capacities, such as the intestines[307].

Our results partially support this hypothesis. Highly proliferative tissues, such as buccal, skin and blood[307,308] displayed very large increases in entropy with age driven by the gains and losses of methylation at DMPs. Furthermore, these highly proliferative tissues also had the largest proportion of VMPs with majority increasing in variability with age. Comparatively, brain tissue, which has a lower or 'specific' capacity to regenerate[308], had a lower overall increase in entropy. Furthermore, the brain methylome exhibited unusual changes in variability with half of the brain VMPs *decreasing* in variability with age, a unique feature of this tissue. This finding was similarly reported in a recent preprint, noting that human brain tissue displayed decreasing variance with age when compared to blood[301]. Akin to anti-entropic DMPs, many of the sites that decrease in variance with age trend to polar beta values (i.e. 0 and 100%) and likely capture genes that are regulated as a response to or consequence of age [301]. Adipose was an outlier in this context. Although it is considered to proliferate only in response to injury (i.e. 'intermediate' proliferative capacity)[308], it displayed the largest increase in entropy with age. Muscle was the only tissue to exhibit *no* increase in entropy with age. Albeit a tissue with lower regenerative capacity[308], the muscle samples used in this analysis were skewed to younger, healthy phenotypes.

As discussed in blood and in Chapter 2, the "Epigenetic Clock Theory of Ageing" proposes that DMPs (which drive the changes in entropy) measure the asymmetric stem cell divisions (i.e. when tissue stem cells differentiate into non-stem cell cells) in different tissues over chronological time. Considering stem cells decline in replicative function (i.e. the ability to produce differentiated cells) with age in many mammalian tissues[309], worsened by insults such as chronic inflammation or injury, their ability to maintain homeostasis and repair and differentiation becomes impaired over the lifespan[308]. If this hypothesis is true, it would make sense that variability introduced by VMPs reflect the accumulation of damage from environmental insults (i.e. sedentary behaviour, sun exposure, toxin exposure, inflammation, injury) that *speed up* the proliferation of the stem cell pool to replace damaged tissue, thereby increasing a person's biological age.

4. Cellular heterogeneity in blood and buccal tissue

Since bulk tissue analyses reflect the aggregate cell type DNAm changes as well as age-related changes in cellular composition, cellular heterogeneity presents a major confounder in DNAm studies. We explored this confounding effect on the methylation signatures of ageing in two

tissues, blood and buccal. In blood, we found a high overlap of significant CpGs in both the adjusted and unadjusted analyses of DMPs and to a lesser extent, VMPs. These represent age-related DNAm changes that are either present in all underlying subtypes, or only in a predominant subtype[310], a finding that overlaps with previous studies[62,214]. While majority of the age-related changes therefore occur *independently* of changes in cell composition, a portion of VMPs (one third) were significant *only* in the unadjusted analyses and are likely measuring variability that *is* related to changes in cell composition with age. It has been suggested that the increase in 'transcriptional noise' associated with the loss of cell identity lacks sufficient evidence and may instead be linked to age-related changes in cell composition[311]. Thus, VMPs may be capturing an important phenotype of ageing in blood (and potentially other tissues), such as age-related adaptations in the immune system[312], which warrants further exploration.

In buccal tissue, the effects of cellular heterogeneity (i.e. the proportions of immune cells, epithelial cells, and fibroblasts) are more pronounced than in blood, however, we used very stringent FDR thresholds in a small sample size that possibly underpowered the analysis to detect the full extent of age-related DNAm changes. As such, we interpret these results with scepticism. Nonetheless, the additional DMPs identified in the *adjusted* analysis suggest changes are occurring in *at least* one of the cell subtypes and are independent of shifts in cellular composition. Whereas the VMPs identified only in the *unadjusted* analysis likely capture the relative cellular proportions in this tissue and not the age-related increases in variance in one or more cell subtypes. However, whether there are changes in variability independent of cell type composition can only be fully determined with a tissue cohort with sufficient statistical power.

In a nutshell, our preliminary results into cellular heterogeneity highlight the confounding effect of cell type composition on DNAm, an area of interest earmarked for future work. Importantly, performing cell type correction is recommended *in addition* to an unadjusted analysis[310], particularly in ageing studies, since increasing cellular heterogeneity could reflect an important biological process or age-related phenotype. Future work should also Endeavor to identify age-related changes *within* individual cell types[313], to understand the relative contribution of different cellular compartments in age-related diseases.

5. Limitations

Identifying quantitative markers of ageing strongly depends on the available resources to implement the best study design. To overcome the limitation of underpowered cross-sectional studies, particularly in the case of non-blood tissues, we used a large-scale meta-analysis approach. Nonetheless, we were still limited in the ability to quantify the full extent of the age-related changes (particularly in the context of VMPs) in some tissues due to insufficient sample numbers and skewed age distributions. We attempted to overcome this limitation by only including datasets with an age standard deviation >5 and a sample size of >30, yet the age range (i.e. reaching to very old ages), sex ratios and disease status were highly skewed between tissues. As a result, the unevenness in statistical power between tissues makes a tissue comparison imperfect. We therefore cannot negate the possibility that the sample size of the skin, adipose, muscle and buccal datasets impacted the ability to detect age-related CpGs with smaller effect sizes (such is the case in blood) and hence the age-related DMPs in these tissues would have much larger effect sizes and inflate the entropy values. Furthermore, the heterogeneity of the brain regions and cell types of the samples in the datasets of this tissue cohort could also explain the unusual changes variability. While our database is, to our knowledge, one of the largest collections of human methylomes across a multitude of tissues, even larger sample numbers across a broad age range is needed to gain confidence in our results.

Another limitation in our study is that we do not know the extent to which age-related changes in the relative proportions of different cell subtypes confounded the signals detected in brain, adipose, skin and muscle, and which cell types are driving the age-related DNAm changes. As mentioned in the above section, this is simply beyond the scope of this PhD, but is an area of interest for future work.

Finally, a limitation that exists not only in this work but is applicable to broader epigenetic research, is the incompleteness of the data that is made available through online databases. While databases such as the GEO, ArrayExpress and dbGaP provide an invaluable tool for epigenetic research, there are inferior quality control standards that ensure completeness and accuracy of the available data. *Many* of the datasets on these platforms are missing crucial information, raw data, or both, with minimal information regarding preprocessing steps. This presents a major challenge with reproducibility. While we made every attempt to gather the

missing information and preprocess all datasets ourselves, this was an impossibility in some cases.

6. Future work

This research forms part of a larger research endeavour, which is to build and publish the largest quantitative map of human ageing as an online webtool for use by the broader scientific community. As such, we envision the quantitative map of the human methylome to be expanded to as many human tissues as possible. Thus far, our database comprises ~56,000 samples from 186 datasets in 14 human tissues, and it continues to expand as new datasets become available. We intend to quantify the features of ageing in each tissue, including *DMPs, VMPs* and *entropy,* as well as the downstream biological interpretation of the age-related changes. We also aim to identify the *pan-tissue* signature of epigenetic ageing by pooling the results across all tissues to distinguish the DNAm changes that are unique to tissues from those that are shared between tissues.

The future of this project will also involve the identification of the tissue-specific and pan-tissue coordinated DNAm changes at multiple CpGs, known as *correlation networks* (see Chapter 2)*.* By adopting a 'system's biology' approach to DNAm analysis, we can focus on the interconnectedness of the entire methylome by building correlation networks using the co-methylation relationships of multiple CpGs. WGCNA has been used to identify clusters of co-methylated CpG sites (modules) that are associated with ageing and ageing phenotypes[3]. Modules that are present in multiple datasets represent common and robust CpG relationships that reflect the true underlying biology and not technical noise. Considering the sheer number of individual CpGs that change with age, correlation networks may prove a valuable data reduction technique to simplify findings and alleviate the multiple testing burden.

We also intend on determining the influence of cellular heterogeneity on these age-related signatures by performing adjusted and unadjusted analyses in each tissue, but also determining *which cell types* these age-related changes are occurring in. To do this, we will use statistical tools such as CellDMC to identify cell-type specific DMPs and VMPs[313].

Once the above is completed, we aim to publish all the results as an open-access online resource to ensure that the utility of the project is unlimited. Our ambition is for this webtool to become a highly utilised platform that can be used to assess the effect of longevity interventions, such

as exercise, nutrition, supplementation, etc., on these age-related signatures *in vivo* humans. Our online webtool will secure the longevity of this project, foster open science and accelerate the outcomes for future research. Finally, all code will be made available to ensure reproducibility of our work.

Appendices

Supplementary Table 1 Description of blood datasets.

Dataset ID	Database	Origin tissue	N	N after pre-processing	Age (mean ± SD)	Age range (min - max)	Array	% Male	Phenotype	Access / info required	Check for sex	Pre-processed ourselves	Covariates
GOLDN	dbGaP	Blood	967	967	48.9 ± 16.4	18 - 88	450K	47.4	Normal	Yes. Access through dbGaP.	Yes	Yes	age + sex
WHI	dbGaP	Blood	1890	1871	64.2 ± 7.0	50 - 79	450K	0.0	Postmenopausal women	Yes. Access through dbGaP.	N/A. All samples are female.	Yes	age
GSE55763	GEO	Blood	2707	2639	51 ± 10.1	23 - 75	450K	67.7	Normal		Yes. 4 samples with incorrect sex removed.	No. R keeps crashing when trying to pre-process.	age + sex
GSE128235	GEO	Blood	537	532	47.7 ± 13.4	18-87	450K	42.7	Depression / healthy control		Yes	Yes. 1 sample corrupted IDAT file. 4 samples without a diagnosis removed.	age + sex + diagnosis
GSE99624	GEO	Blood	48	48	67.2 ± 9.9	49 - 87	450K	18.7	Osteoporosis / healthy controls		Yes	Yes	age + sex
GSE115278_450	GEO	Blood	366	355	46.9 ± 15.4	19 - 73	450K	35.8	Normal		Yes. 11 samples removed with incorrect sex.	Yes	age + sex + study
GSE115278_EPIC	GEO	Blood	108	106	46.1 ± 9.6	19 - 66	EPIC	36.8	Normal		Yes. 2 samples removed with incorrect sexes	Yes	age + sex + study
GSE87571	GEO	Blood	732	728	47.4 ± 21.0	14 - 94	450K	46.7	Normal		Yes. 3 samples removed with incorrect sex.	Yes. 1 sample missing age information.	age + sex
GSE53740	GEO	Blood	384	162	68.1 ± 10.3	34 - 93	450K	48.4	Progressive supranuclear palsy / frontotemporal dementia / healthy controls		Yes. Remove batch 2 all samples have been scrambled. Remove 63 samples from batch 1 with mismatch sex or undefined sex.	Yes	age + sex + race
GSE58045	GEO	Blood	172	172	57.2 ± 8.2	32 - 80	27K	0.0	Normal twins		N/A. All samples female.	No	age + (1|family number)
NAS	dbGaP	Blood	1454	1454	74.5 ± 7	55 - 100	450K	100.0	Normal / chronic disease	Yes. Access through dbGaP.	N/A. All samples are male.	Partly. Raw files not available but ran normalisation and batch correction.	age + year + (1|ID)
BIOS	EGA	Blood	4386	1408	37.5 ± 13.9	18 - 80	450K	34.2	Unknown	Yes. Application sent to authors.	Not all sex info supplied. 3 samples with sex did not match predicted sex.	Yes. Samples without ages remove, samples with corrupt IDATs removed.	age + sex
GSE77445	GEO	Blood	85	85	33.8 ± 15.9	18 - 69	450K	50.5	Normal		No. Only signal intensities for a reduced set of CpGs are provided.	No	age + sex
JHS	Independent	Blood	1747	1747	55.7 ± 12.3	22 - 93	EPIC	37.0	Cardiometabolic conditions in African population	Yes. Independent application.	Yes.	No	age + sex + education + BMI + smoker + alcohol
FHS	dbGaP	Blood	2562	2498	66.3 ± 8.9	40 - 92	450K	45.8	Normal / cardiovascular disease	Yes. Access through dbGaP.	Yes. 19 samples removed	Yes. 40 Samples had corrupt IDATS removed. 1 sample failed QC.	age + sex
GSE49904	GEO	Blood	71	71	55 ± 14.5	23 - 83	27K	31.0	Iron deficiency anaemia / Anaemia / Healthy		N/A	No	age + sex + ethnicity + diagnosis + smoking history
GSE80417	GEO	Blood	675	636	40.4 ± 15.0	18 - 90	450K	59.3	Schizophrenia / normal		Yes.	No. 39 samples with no age / incorrect age.	age + sex
SATSA	ArrayExpress	Blood	1072	1071	73.1 ± 9.7	48 - 98	450K	39.7	Normal		Yes. 1 sample removed	Yes	age + sex + (1|twin pair)
GSE42861	GEO	Blood	689	689	51.9 ± 11.8	18 - 70	450K	28.6	Rheumatoid arthritis / healthy controls		Yes	Yes	age + sex + disease status
GSE51032	GEO	Blood	845	378	53.2 ± 7.2	34-72	450K	20.4	Only normal samples		Yes. 10 samples incorrect sex removed.	No. IDATS corrupted. Removed 424 samples with a cancer diagnosis. 9 samples failed QC	age + sex
GSE67705	GEO	Blood	284	186	46.1 ± 8.8	25 - 67	450K	100.0	HIV positive / HIV negative		N/A. All samples male.	No. 98 Samples failed QC and were removed.	age + HIV status
GSE32148	GEO	Blood	48	44	19.7 ± 14.6	10-76	450K	47.9	Normal / Crohn's / Ulcerative Colitis		Yes. 4 samples removed with incorrect sex.	No	age + sex + disease status
GSE106648	GEO	Blood	279	279	41.2 ± 11.2	16 - 66	450K	27.6	Multiple sclerosis / healthy control		Yes	Party. No detection pvalue information.	age + sex + smoking status
GSE69138	GEO	Blood	589	184	74.8 ± 12.6	39 - 99	450K	51.1	Ischaemic stroke subtypes		Yes. 1 sample removed with incorrect sex	Yes	age + sex + stroke subtype
GSE50660	GEO	Blood	464	463	55.4 ± 6.6	38 - 67	450K	70.5	Smoker / non-smoker		Yes. 1 sample removed.	No. Sample IDs in raw files don't match the phenotype table.	age + sex + smoking status
GSE40279	GEO	Blood	656	656	64.0 ± 14.7	19 - 101	450K	48.5	Healthy	Emailed authors for raw IDATS or detection pvals.	No	No. Could not pre-process. Raw sample IDs did not match the phenotype table.	age + sex + ethnicity
GSE41037	GEO	Blood	717	703	37 ± 16	16 - 88	27K	61.9	Schizophrenia / healthy control		Yes. 5 samples don't match	Yes. 12 samples failed QC	age + sex
GSE41169	GEO	Blood	95	93	31.6 ± 10.4	18 - 65	450K	29.5	Schizophrenia / healthy control	Emailed authors for raw data	Yes. 2 samples removed with incorrect sexes	No	age + sex
GSE53840	GEO	Blood	120	109	52 ± 8	31 - 68	450K	100.0	HIV viral load		Yes	Yes. 9 samples no age info.	age + HIV viral load
GSE67751	GEO	Blood	92	92	49.5 ± 8.9	24 - 68	450K	45.7	HIV / control	Emailed authors for batch info	Yes	No	age + sex + HIV status
GSE72775	GEO	Blood	335	335	70.2 ± 10.3	36.5-90.5	450K	58.8	Caucasian/Hispanic	Emailed authors for batch info	Yes	Partly. No batch info.	age + sex + ethnicity
GSE111629	GEO	Blood	572	564	69.3 ± 11.3	35 - 92	450K	56.2	Parkinson's disease / control		Yes	Yes. 8 samples failed QC.	age + sex
GSE72774	GEO	Blood	508	507	69.6 ± 11.2	35.1 - 92	450K	55.4	Parkinson's disease	Emailed authors for batch info	Yes. 1 sample removed with incorrect sex prediction.	Partly. No batch info.	age + sex + number of years in school + smoking status
GSE72776	GEO	Blood	84	84	66.4 ± 11.7	34 - 50	450K	59.5	Parkinson's disease	Emailed authors for batch info	Yes	Partly. No batch info.	age + sex
GSE166611	GEO	Blood	32	32	38.5 ± 12.4	19 - 69	450K	0.0	Normal weight / Obese		Yes	Yes	age + BMI
GSE164056	GEO	Blood	143	143	25.8 ± 6.4	19 - 50	EPIC	34.3	Social Anxiety Disorder		Yes	Yes	age + sex

Database	Dataset ID	Tissue / cell type	Origin tissue	N	N after pre-processing	Age (mean ± SD)	Age range (min - max)	Array	% Male	Phenotype	Access / info required	Check for Sex	Pre-processed ourselves	Covariates
GSE85311	GEO	Blood		38	38	47.4 ± 17.9	20 - 68	450K	71.0	Young / old sedentary /old exercise trained		Yes	Partly. No batch info.	age + sex + training status
GSE151278	GEO	Blood		70	70	47.1 ± 14.6	20-86	450K	61.4	Psoriasis		Yes	Yes	age + sex + drug response
GSE96879	GEO	Blood		90	90	51.6 ± 14.4	22 - 84	450K	0.0	Lupus / Healthy controls		N/A	Yes	age + ethnicity + disease status
GSE134429	GEO	Blood		64	63	56.0 ± 12.7	23 - 81	EPIC	11.1	Rheumatoid arthritis / healthy controls	Emailed authors for batch info	Yes. 1 sample with incorrect sex	Partly. No batch info.	age + sex + batch + patient cohort + donor
GSE120307	GEO	Blood		34	34	35.5 ± 11.0	19-54	450K	53.0	Healthy / psychiatric disorder (twin pairs)		Yes	No	age + sex + (1|twin pair)
GSE20236	GEO	Blood		93	93	32.0 ± 18.8	49 - 74	27K	0.0	Normal	Emailed authors for batch info	N/A	No. Sample IDs in the raw files are missing.	age
GSE19711	GEO	Blood		540	262	64.9 ± 6.7	52 - 78	27K	0.0	Normal samples only		N/A	Yes. Removed 266 cancer samples. 2 samples failed QC	age + ca125
GSE157131_EPIC	GEO	Blood		946	946	62.6 ± 9.7	26.4 - 91.6	EPIC	100.0	Hypertension / control	Emailed authors for batch info	All samples are male in description of study, and the sex prediction confirmed this. In the raw phenotype table, some of the samples are 'female'. Phenotypes were changed to be all male.	Partly. No batch info	age
GSE157131_450K	GEO	Blood		272	272	66.7 ± 6.5	39 - 94	450K	100.0	Hypertension / control	Emailed authors for batch info	All samples are male in description of study, and the sex prediction confirmed this. In the raw phenotype table, some of the samples are 'female'. Phenotypes were changed to be all male.	Partly. No batch info	age
GSE117859	GEO	Blood		608	608	49.4 ± 7.6	25 - 75	450K	100.0	Smoking + HIV		N/A	Yes	age + smoking
GSE117860	GEO	Blood		529	529	48.1 ± 7.8	25 - 75	450K	100.0	Smoking + HIV		N/A	Yes. More than half the CpGs were removed during pre-processing due to NA probes. After checking the probes this is a 450K dataset and not 850K as stipulated on GEO.	age + smoking
GSE147740	GEO	Blood		1129	1029	41.6 ± 7.8	26 - 59	EPIC	60.3	Normal		Yes. 3 Samples with incorrect sex were removed.	Yes. Removed samples without actual ages (samples with ages as a range)	age + sex
GSE152026	GEO	Blood		934	927	35.2 ± 12.8	18 - 64	EPIC	54.9	Psychosis / control		Yes. 1 Sample with incorrect sex removed	Yes	age + sex + disease status
GSE132203	GEO	Blood		795	795	42.3 ± 12.3	18 - 76	EPIC	28.2	Psychiatric disorders		Yes	Yes	age + sex + ethnicity + child abuse
GSE100264	GEO	Blood		386	386	49.5 ± 7.3	25 - 75	450K	100.0	Drug use + Hepatitis C + HIV+ / control		Yes	Yes	age + hepatitis C infection
GSE107080	GEO	Blood		405	405	47.9 ± 8.0	25 - 75	EPIC	100.0	Drug use + Hepatitis C + HIV+ / control		Yes	Yes	age + hepatitis C infection + race + smoking + artadherence
GSE116339	GEO	Blood		679	676	53.9 ± 12.9	23 - 88.5	EPIC	41.3	Polybrominated biphenyl exposure		Yes. 3 samples with incorrect sex removed	Yes	age + sex + PBB exposure
GSE168739	GEO	Blood		407	402	42.0 ± 10.3	19 - 61	EPIC	44.9	COVID-19		Yes. 5 samples with incorrect sex removed	Yes	age + sex
GSE197674	GEO	Blood		2138	2138	32.9 ± 10.0	6 - 66.4	EPIC	47.1	Survivors of childhood cancer		Yes. All samples clustered with the correct sex	Partly. Used their pre-processed matrix but filtered additional champ probes and pidsley cross-reactive probes. Performed combat for slide and array.	age + sex + abdominal pelvic rt + brain rt + chest rt + alkylating agent + anthracyclines + corticosteroids + epipodophyllotoxins + platinum + vincristine
GSE197676	GEO	Blood		282	282	35.8 ± 10.2	18.6 - 70.2	EPIC	45.6	Healthy / non-diseased		Yes	Yes	age + sex
				36954	32136									

GOLDN, Genetics of Lipid-Lowering Drugs and Diet Network Study; WHI, Women's Health Initiative; NAS, Normative Aging Study; BIOS, Biobank-based integrative omics study; JHS, Jackson Heart Study; FHS, Framingham Heart Study; SATSA, Swedish Adoption/Twin Study of Aging; GEO, Gene Expression Omnibus; dbGaP, Database of Genotypes and Phenotypes; N, Sample number; SD, Standard deviation; HIV, human immunodeficiency virus; COVID-19, coronavirus disease of 2019; QC, quality control

Supplementary Table 2 Description of brain datasets.

Database	Dataset ID	Tissue / cell type	Origin tissue	N	N after pre-processing	Age (mean ± SD)	Age range (min - max)	Array	% Male	Phenotype	Access / info required	Check for Sex	Pre-processed ourselves	Covariates
GEO	GSE64509	Multiple brain regions	Brain	260	251	74 ± 25	15 - 114	450K	48.4	Normal	Emailed authors for batch info	Yes. 9 samples with incorrect sex removed.	Partly. Missing batch info. Removed multihit probes from their pre-processed matrix.	age + sex + brain region
GEO	GSE80970	Prefrontal cortex and superior temporal gyrus	Brain	286	286	86 ± 7.8	70 - 108	450K	37.8	Alzheimer's / control		Yes	Yes	age + sex + tissue + (1|donor id)
GEO	GSE15745	Cerebellum, Frontal cortex, Pons and Temporal cortex	Brain	506	495	47 ± 24	15 - 101	27K	68.0	Normal	Emailed authors for batch info	Yes. 4 samples that did not cluster with their respective sex or tissue were removed.	Yes. 7 samples failed QC and were removed	age + sex + tissue + (1|ID)
GEO	GSE36194	Cerebellum and frontal cortex	Brain	724	421	51 ± 30	0 - 102	27K	65.0	Normal	Emailed authors for batch info	Yes. 40 samples that did not cluster with their respective sex or tissue were removed	Yes. 3 samples failed QC. An additional 254 samples from this dataset appear in GSE15745. I have removed them since they also appear in the other analysis. 4 samples did not have batch info and 2 samples no age info.	age + sex + tissue + (1|ID)
GEO	GSE112179	Neurons	Brain	100	100	48 ± 11	23 - 77	EPIC	75.0	Schizophrenia and bipolar disorder		Yes	Yes	age + sex
GEO	GSE66351_Occipital Cortex	Glia & neurons from occipital cortex	Brain	62	62	75 ± 17	18 - 94		45.2	Alzheimer's		Yes	Yes	age + sex + cell type
GEO	GSE66351_FrontalCortex	Frontal cortex	Brain	63	63	74 ± 15	18 - 97	450K	47.6			Yes	Yes	age + sex + (1|donor id)
GEO	GSE66351_TemporalCortex	Temporal cortex	Brain	65	65	74 ± 14	18 - 97		47.7			Yes	Yes	age + sex

Database	Dataset ID	Tissue / cell type	Origin tissue	N	N after pre-processing	Age (mean ± SD)	Age range (min - max)	Array	% Male	Phenotype	Access / info required	Check for Sex	Pre-processed ourselves	Covariates
GEO	GSE38873	Cerebellum	Brain	168	168	44 ± 9.9	20 - 70	27K	63.0	Schizophrenia / Bipolar / Control	Emailed authors for sex info	Cannot check as no sex info provided	Yes	age + sex + batch
GEO	GSE41826	Postmortem frontal cortex	Brain	145	143	34 ± 16	13 - 79	450K	52.4	Major depression / Control		Yes. 2 samples removed	Yes	age + sex + diagnosis
GEO	GSE89702	Cerebellum	Brain	33	33	45 ± 16	21 - 72	450K	84.8	Schizophrenia / controls		Cannot check as sex chromosome probes were removed from raw data	Yes	age + sex
GEO	GSE89705	Striatum putamen (prefrontal cortex?)	Brain	33	33	46 ± 17	21 - 73	450K	81.8	Schizophrenia / controls	Emailed authors for sex chr probes	Cannot check as sex chromosome probes were removed from raw data	Yes	age + sex
GEO	GSE89706	Striatum putamen	Brain	49	49	63 ± 17	25 - 96	450K	63.3	Schizophrenia / controls		Cannot check as sex chromosome probes were removed from raw data	Yes	age + sex
GEO	GSE61431	Cerebellum and frontal cortex	Brain	87	87	62 ± 17	25 - 96	450K	64.3	Schizophrenia / controls	Emailed authors for detection p-values	Yes	Partly. Missing batch info. Removed multihit probes from their pre-processed matrix.	age + sex + source tissue
GEO	GSE61380	Frontal cortex	Brain	33	33	44 ± 16	21 - 73	450K	84.8	Schizophrenia / controls		Yes	Yes	age + sex
GEO	GSE74193	Prefrontal cortex	Brain	675	528	38 ± 22	-0.5 - 97	450K	65.0	Schizophrenia / controls		Yes. 8 samples removed	Yes. 139 technical replicates were averaged	age + sex
GEO	GSE40360	Frontal lobe white matter	Brain	47	46	60 ± 11	35 - 81	450K	47.8	Multiple sclerosis	Emailed authors for batch info	Yes	Partly. Missing batch info. Removed multihit probes from their pre-processed matrix.	age + sex
GEO	GSE59685	Cerebellum, entorhinal cortex, frontal cortex, superior temporal gyrus	Brain	531	451	84 ± 9.3	40 - 105	450K	41.5	Alzheimer's disease		Yes	Yes. 80 samples from blood were removed.	age + sex + tissue source + (1|subject id)
GEO	GSE49905	Cerebral cortex	Brain	78	77	23 ± 17	0 - 82	27K	72.7	Autism / control		Yes. 1 sample removed	Yes	age + sex + diagnosis
GEO	GSE59457	Multiple brain regions	Brain	130	128	46 ± 9.9	26 - 68	450K	73.4	HIV / control	Emailed authors for batch info	Yes. 1 sample removed	Partly. Missing batch info. Removed multihit probes from their pre-processed matrix.	age + predictedsex + tissue + HIV status
GEO	GSE63347	Multiple brain regions	Brain	71	70	54 ± 9	32 - 64	450K	52.8	Down's syndrome / control	Emailed authors for batch info	Yes. 1 sample removed	Partly. Missing batch info. Removed multihit probes from their pre-processed matrix.	age + sex + tissue + down syndrome status
GEO	GSE72778	Multiple brain regions	Brain	475	454	65 ± 23	15 - 114	450K	55.9	Huntington's disease	Emailed authors for batch info	Yes. 1 sample removed	Partly. Missing batch info. Removed multihit probes from their pre-processed matrix. 2 samples removed with entirely missing data.	age + sex + brain bank + brain region + postmortem interval + (1|personid)
GEO	GSE144910	Superior temporal gyrus	Brain	96	88	48 ± 14	17 - 83	EPIC	71.6	Schizophrenia / controls	Emailed authors for batch info	Yes	Yes. 8 technical replicates were averaged	age + sex + race
GEO	GSE138597	Multiple brain regions	Brain	36	36	70 ± 5.2	64 - 79	EPIC	100.0	Alzheimer's/Parkinson's/control		Yes	Yes	age at death + race
GEO	GSE134379	Middle temporal gyrus and cerebellum	Brain	808	802	84 ± 7.9	54 - 103	450K	50.4	Alzheimer's / control		Yes	Yes. 6 samples failed QC	age + sex + brain region
GEO	GSE111165	Multiple brain regions	Brain	101	38	29 ± 16	5 - 61	EPIC	68.4			Yes	Yes	age + sex + brain region + (1|subject)
GEO	GSE98203	Neuronal nuclei	Brain	88	87	31 ± 11	15 - 65	450K	80.5	heroine users/suicide/controls		Yes. 1 sample removed	Yes	age + gender + cohort
				5750	5094									

GEO, Gene Expression Omnibus; N, Sample number; SD, Standard deviation; HIV, human immunodeficiency virus; QC, quality control

Supplementary Table 3 Description of skin datasets.

Database	Dataset ID	Tissue / cell type	Origin tissue	N	N after pre-processing	Age (mean ± SD)	Age range (min - max)	Array	% Male	Phenotype	Access / info required	Check for Sex	Pre-processed ourselves	Covariates
GEO	GSE51954	Epidermal and dermal	Skin	78	74	48.6 ± 23.7	20 - 84	450K	44.9	Sun exposed skin / sun protected skin		Yes	Yes. 4 samples without ages.	age + sex + exposure + tissue
GEO	GSE90124	Skin	Skin	322	322	59.4 ± 9.2	38.7 - 83.1	450K	0	Normal twins		N/A. All samples are female.	Yes	age + total body naevus count + (1|relatedness id)
ArrayExpress	EMATB4385	Epidermis	Skin	108	108	47.5 ± 20.7	18 - 78	450K	0	Normal		N/A. All samples are female.	Yes	age
GEO	GSE115797	Skin	Skin	48	43	37.0 ± 14.2	12 - 74	450K	20.8	Psoriasis / normal		Yes. 4 samples with incorrect sex removed.	Yes. 1 sample removed during QC.	age + sex + disease state + (1|Unique ID)
ArrayExpress	EMATB8992	Epidermis	Skin	56	56	50.4 ± 16.9	21 - 76	450K	0	Normal	Emailed authors for phenotypes and batch info	N/A. All samples are female.	Partly. No batch info.	age
ArrayExpress	EMATB8993	Epidermis	Skin	65	65	52.9 ± 16.7	23 - 79	459K	0	Normal	Emailed authors for phenotypes and batch info	N/A. All samples are female.	Partly. No batch info.	age + batch
				677	668									

GEO, Gene Expression Omnibus; N, Sample number; SD, Standard deviation; QC, quality control

Supplementary Table 4 Description of buccal datasets.

Database	Dataset ID	Tissue / cell type	Origin tissue	N	N after pre-processing	Age (mean ± SD)	Age range (min – max)	Array	% Male	Phenotype	Access / info required	Check for Sex	Pre-processed ourselves	Covariates
GEO	GSE78874	Saliva	Buccal	259	259	69 ± 9.7	36 – 88	450K	56.3	Normal		Yes	Yes	age + sex + ethnicity
GEO	GSE61653	Saliva	Buccal	128	64	41 ± 12	20 – 74	450K	17.2	Normal	Emailed authors for age, sex and batch info	Yes	Yes	age + sex

Database	Dataset ID	Tissue / cell type	Origin tissue	N	N after pre-processing	Age (mean ± SD)	Age range (min - max)	Array	% Male	Phenotype	Access / info required	Check for Sex	Pre-processed ourselves	Covariates
GEO	GSE94876	Buccal cells	Buccal	120	120	46 ± 7.7	35 – 60	450K	100	Smoker / non-smoker		Yes	Yes	age + ethnicity + smoking type
GEO	GSE92767	Saliva	Buccal	54	30	43.9 ± 15.1	18 – 73	450K	100	Normal	Emailed authors for batch info	N/A. All samples are male.	Yes	age
				561	473									

GEO, Gene Expression Omnibus; N, Sample number; SD, Standard deviation; QC, quality control

Supplementary Table 5 Description of adipose datasets.

Database	Dataset ID	Tissue / cell type	Origin tissue	N	N after pre-processing	Age (mean ± SD)	Age range (min - max)	Array	% Male	Phenotype	Access / info required	Check for Sex	Pre-processed ourselves	Covariates
ArrayExpress	E-MTAB-1866	Adipose	Adipose	648	648	59 ± 9.4	39 - 85	450K	0.0	Healthy MZ and DZ twins	Emailed authors for age & batch info and to reupload raw data	N/A. All samples are females.	No	age
dbGaP	FUSION	Adipose	Adipose	295	293	60 ± 8.1	20 - 79	EPIC	55.6	Healthy/T2D	Yes	Yes. 2 samples removed	Yes	age + sex + experimenter
GEO	GSE61450_&_GSE61453	Subcutaneous and visceral adipose tissue	Adipose	142	131	44 ± 9.5	26 - 67	450K	29.0	Severely obese		Yes. 3 samples removed + 2 samples whose tissues don't match the right tissue	Yes. 6 samples failed QC	age + sex + tissue type
GEO	GSE68336	Adipose	Adipose	70	70	31 ± 4.4	23 - 36	450K	42.9	MZ twins discordant and concordant for BMI	Emailed authors for batch info; they performed the EWAS for us as we can't access the age of the twins	Yes	Yes	age + sex + BMI
GEO	GSE24884	Adipose	Adipose	56	55	43 ± 12	23 - 72	27K	0.0	Obese / Non-obese		Yes	Yes. 1 sample failed QC	age + sex + obesity status
GEO	GSE58622_&_GSE67024	Adipose	Adipose	45	45	46 ± 11	25 - 65	450K	0.0	Post obese / Control	Emailed authors for position on batch info	Yes	Partly (they had missing position on batch info) - adjusted for batch but not position on batch	age + obesity status
GEO	GSE61257	Adipose	Adipose	32	31	50 ± 14	31 - 79	450K	25.8	Normal / Healthy Obese / NASH / NAFLD	Emailed authors for batch info	Yes. 1 sample removed.	Partly (they had missing batch info) – removed multihit probes from their pre-processed matrix	age + sex + disease status
GEO	GSE162166	Subcutaneous and visceral adipose tissue	Adipose	40	40	38 ± 12	18 - 54	EPIC	32.5	T2D / controls	Emailed authors for age info	Yes	Yes	age + sex + tissue + BMI + (1|ID)
				1328	1313									

GEO, Gene Expression Omnibus dbGaP, Database of Genotypes and Phenotypes; FUSION, Finland-United States Investigation of NIDDM Genetics; N, Sample number; SD, Standard deviation; T2D, Type 2 Diabetes; NASH, Non-alcoholic steatohepatitis; Non-alcoholic fatty liver disease; MZ, Monozygotic; DZ, Dizygotic; BMI; Body mass index; QC, quality control

Supplementary Table 6 Description of muscle datasets.

Database	Dataset ID	Tissue / cell type	Origin tissue	N	N after pre-processing	Age (mean ± SD)	Age range (min - max)	Array	% Male	Phenotype	Access / info required	Check for Sex	Pre-processed ourselves	Covariates
dbGaP	FUSION	Muscle	Muscle	313	312	60 ± 8.0	20-79	EPIC	57.0	Healthy/T2D	Yes. Access through dbGaP.	Yes. 1 sample removed	Yes	age + sex + scientist
GEO	Gene SMART (GSE151407 & GSE171140)	Vastus lateralis	Muscle	234	234	32 ± 8.1	18-45	EPIC	80.0	Healthy		Yes	Yes	age + sex + timepoint + batch + (1|ID)
Collaborators	FTC	Vastus lateralis	Muscle	156	156	43 ± 16		EPIC	58.0	Healthy	Yes. Independent collaboration.	No	No	Analysis run by collaborators
Collaborators	ABOS	Rectus abdominis	Muscle	72	65	45 ± 8.2	23-61	450K	0.0	Lean/obese/obese with T2D	Yes. Independent collaboration.	Yes	Yes. Duplicates averaged	age + BMI + T2D
Collaborators	LITER	Vastus lateralis	Muscle	63	31	26 ± 5.9	20-39	EPIC	100.0	Healthy	Yes. Independent collaboration.	Yes	Yes. Many files had a wrong size suggesting missing probe data, so they had to be removed	age + BMI + sample group + (1|participant)
GEO	GSE61452	Rectus abdominis	Muscle	60	57	44 ± 9.5	26-67	450K	26.3	Morbidly obese who underwent bariatric surgery		Yes. 1 sample removed	Yes. 2 samples failed QC	age + BMI + sex
GEO	GSE135063	Vastus lateralis	Muscle	57	57	39 ± 10	23-60	EPIC	28.0	Healthy/obese	Emailed authors for age.	Yes: sex was predicted	Yes	age + sex + obesity + timepoint + (1|ID)
GEO	GSE49908	Vastus lateralis	Muscle	51	51	50 ± 17	21-77	27K	100.0	Healthy		Yes	Yes	age
GEO	GSE50498	Vastus lateralis	Muscle	48	48	47 ± 26	18-89	450K	100.0	Healthy		No	No	age + BMI + sex
Collaborators	EPIK	Vastus lateralis	Muscle	48	48	45 ± 22	20-71	EPIC	100.0	Healthy	Yes. Independent collaboration.	Yes	Yes	Analysis run by collaborators
Collaborators	EXACT	Vastus lateralis	Muscle	48	48	33 ± 10	21-61	EPIC	100.0	Healthy		Yes	Yes	age + VO2max + timepoint + (1|ID)
GEO	GSE114763	Vastus lateralis	Muscle	40	39	29 ± 6	19-39	EPIC	100.0	Healthy		Yes	Yes. Duplicates averaged	age + timepoint + (1|ID)
				1190	1146									

GEO, Gene Expression Omnibus; dbGaP, Database of Genotypes and Phenotypes; FTC, Finnish Twin Cohort; ABOS, Biological Atlas of Severe Obesity; FUSION, Finland-United States Investigation of NIDDM Genetics; LITER, Limb Immobilisation and Transcriptional/Epigenetic Responses; GeneSMART, Gene Skeletal Muscle Adaptive Response to Training; EPIK, Epigenetica & Kracht, N, Sample number; SD, Standard deviation; T2D, Type 2 Diabetes; BMI, Body mass index; QC, quality control
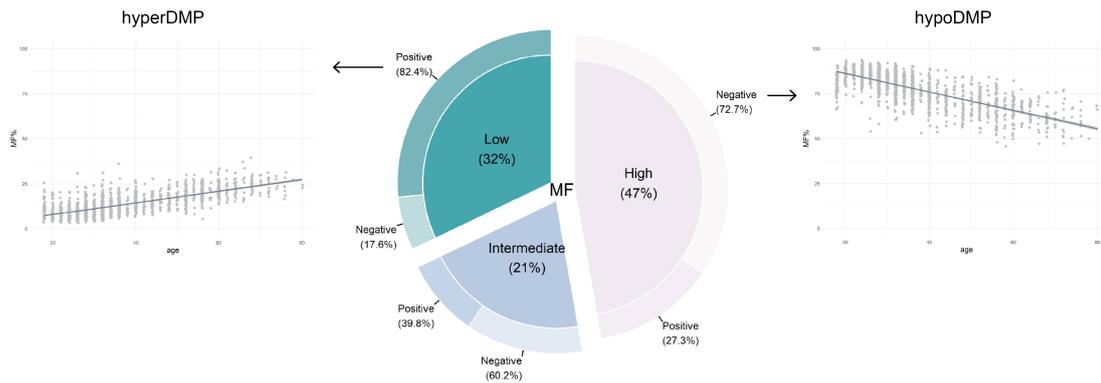
Supplementary Figure 1 Doughnut chart of the proportions of highly, intermediately, and lowly methylated DMPs in a single blood dataset.

The inner pie of the doughnut represents the proportions of age-related differentially methylated positions (DMPs) that are highly methylated (<75%), intermediately methylated (25 – 75%) and lowly methylated (<25%) at baseline in a single blood dataset (BIOS), with the direction of change in methylation with age in the outer circle. A positive direction implies that a DMP increases in methylation with age, and a negative direction implies a CpG loses methylation with age.
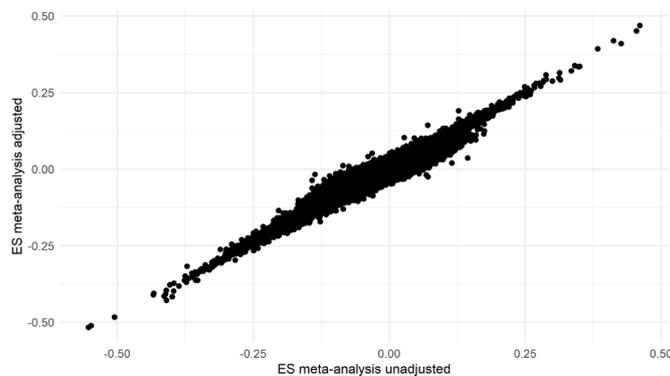
Supplementary Figure 2 A correlation plot of the effect sizes of the CpGs meta-analysed in the cell type adjusted meta-analysis versus the meta-analysis not adjusted for cell types of differential methylation and age in blood.

The effect sizes (ES) from the meta-analysis not adjusted for cell types on the x-axis and the ES for the meta-analysis adjusted for cell types on the y-axis. The Pearson's correlation coefficient = 0.94 with *p*-value < 2.2e-16.
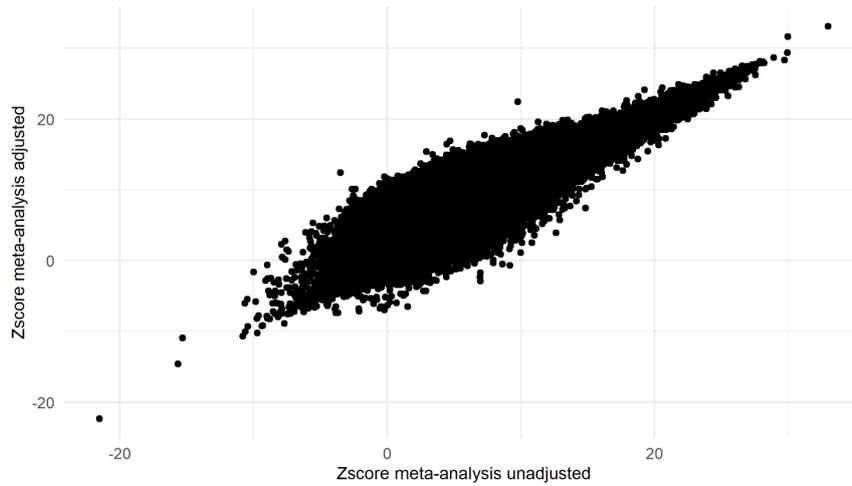
Supplementary Figure 3 A correlation plot of the Zscores of the CpGs meta-analysed in the cell type adjusted meta-analysis versus the meta-analysis not adjusted for cell types of variable methylation and age in blood.

The Zscores from the meta-analysis not unadjusted for cell types on the x-axis and the Zscore for the meta-analysis adjusted for cell types on the y-axis. The Pearson's correlation coefficient = 0.93 with p-value < 2.2e-16.
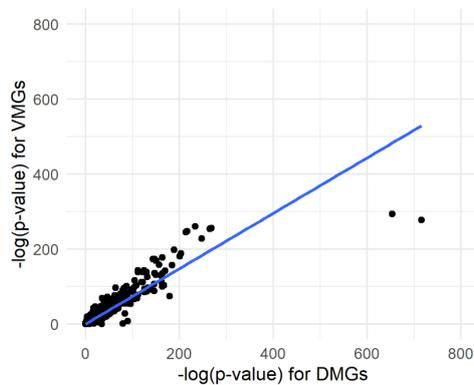
Supplementary Figure 4 A correlation plot of the -log(p-values) for the GO terms for DMGs and VMGs.

The -log(p-values) for the gene ontology (GO) terms in the differentially methylated genes (DMGs) (x-axis) and the VMGs (y-axis) analysis. The Spearmen's correlation coefficient = 0.84.

Study	N	Estimate [95% CI]
GSE55763	2639	0.0011 [0.0009, 0.0013]
FHS	2498	0.0004 [-0.0002, 0.0010]
GSE197674	2138	0.0032 [0.0026, 0.0038]
WHI	1871	0.0004 [-0.0004, 0.0013]
JHS	1747	0.0009 [0.0003, 0.0015]
NAS	1454	-0.0005 [-0.0017, 0.0007]
BIOS	1408	0.0011 [0.0008, 0.0014]
SATSA	1071	0.0002 [-0.0013, 0.0017]
GSE147740	1029	-0.0001 [-0.0016, 0.0013]
GOLDN	967	0.0025 [0.0016, 0.0033]
GSE157131 EPIC	946	0.0022 [0.0010, 0.0034]
GSE152026	927	0.0012 [0.0005, 0.0020]
GSE132203	795	0.0023 [0.0015, 0.0031]
GSE87571	728	0.0026 [0.0022, 0.0030]
GSE41037	703	0.0005 [-0.0002, 0.0011]
GSE42861	689	0.0016 [0.0006, 0.0026]
GSE116339	676	0.0002 [-0.0004, 0.0008]
GSE40279	656	0.0005 [-0.0005, 0.0015]
GSE80417	636	0.0004 [-0.0002, 0.0010]
GSE117859	608	0.0003 [-0.0011, 0.0018]
GSE111629	564	0.0008 [-0.0001, 0.0016]
GSE128235	532	0.0001 [-0.0005, 0.0007]
GSE117860	529	0.0009 [-0.0003, 0.0022]
GSE72774	507	0.0010 [-0.0007, 0.0026]
GSE50660	463	-0.0004 [-0.0026, 0.0017]
GSE107080	405	0.0008 [-0.0009, 0.0024]
GSE168739	402	0.0019 [0.0006, 0.0033]
GSE100264	386	-0.0001 [-0.0020, 0.0018]
GSE51032	378	0.0045 [0.0017, 0.0073]
GSE115278 450K	355	0.0004 [-0.0008, 0.0015]
GSE72775	335	0.0019 [-0.0003, 0.0040]
GSE197676	282	0.0038 [0.0024, 0.0051]
GSE106648	279	0.0017 [0.0007, 0.0026]
GSE157131 450K	272	0.0021 [-0.0007, 0.0049]
GSE19711	262	0.0011 [-0.0012, 0.0034]
GSE67705	186	0.0008 [0.0003, 0.0014]
GSE69138	184	-0.0002 [-0.0024, 0.0020]
GSE58045	172	0.0079 [0.0031, 0.0127]
GSE53740	162	0.0016 [-0.0007, 0.0040]
GSE164056	143	0.0028 [0.0005, 0.0051]
GSE53840	109	0.0016 [-0.0011, 0.0044]
GSE115278 EPIC	106	-0.0009 [-0.0039, 0.0021]
GSE20236	93	0.0001 [-0.0000, 0.0002]
GSE41169	93	0.0013 [-0.0011, 0.0038]
GSE67751	92	0.0009 [-0.0035, 0.0053]
GSE96879	90	0.0017 [0.0003, 0.0030]
GSE77445	85	-0.0012 [-0.0035, 0.0010]
GSE72776	84	0.0022 [-0.0015, 0.0058]
GSE49904	71	0.0016 [-0.0011, 0.0043]
GSE151278	70	0.0018 [-0.0000, 0.0036]
GSE134429	63	0.0065 [0.0015, 0.0116]
GSE99624	48	-0.0016 [-0.0044, 0.0012]
GSE32148	44	0.0007 [-0.0016, 0.0030]
GSE85311	38	0.0000 [-0.0036, 0.0037]
GSE120307	34	0.0001 [-0.0043, 0.0045]
GSE166611	32	0.0015 [-0.0016, 0.0046]
FE Model		0.0005 [0.0005, 0.0006]

Change in entropy (/10 years)

Supplementary Figure 5 Forest plot of the genome-wide meta-analysis of entropy and age in blood.

On the x-axis is the change in entropy per decade of age, with the effect size and standard errors from each independent epigenome-wide association study (EWAS) plotted on the y-axis. The dataset and sample size (N) are on the left side of the plot and the 95% confidence intervals (CI) are displayed on the right panel. The meta-analysis effect size is represented by the blue polygon.

Supplementary Figure 6 A comparison of the changes in entropy with age in DMPs, VMPs, non-age-related CpGs, constant VMPs and all age-related CpGs.

Entropy calculated on the genome-wide set of CpGs (purple), the age-related CpGs (dark purple), the constant VMPs (grey blue), the DMPs only (turquoise), the non-age-related sites (green) and the VMPs only (yellow), plotted against age (x-axis) for the two largest datasets; A) BIOS dataset and the B) FHS dataset in blood.
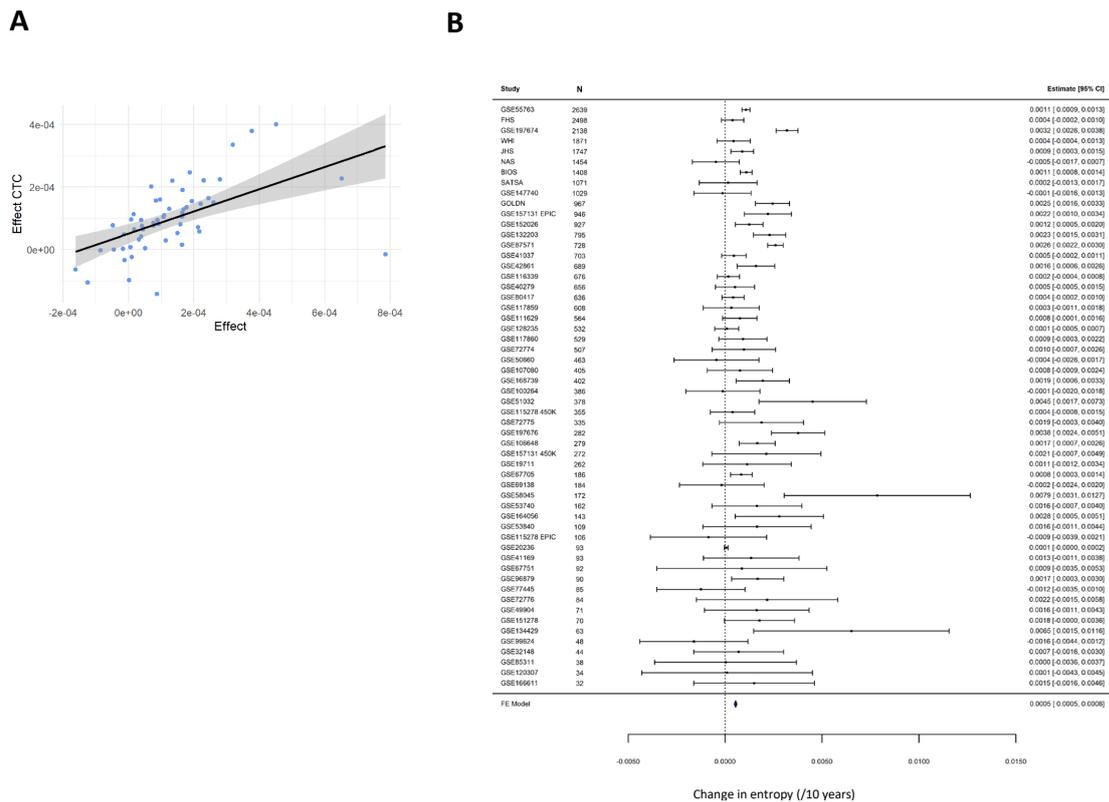
Supplementary Figure 7 Entropy changes in blood after correction for cell types.

A) The effect sizes from the independent epigenome-wide association studies (EWAS) of entropy and age for the analyses unadjusted for the 5 largest blood cell types (x-axis) and the effect size for the cell type corrected (CTC) analyses. Each point on the graph represents a single blood dataset. The Pearson's correlation = 0.54. B) A forest plot of the genome-wide meta-analysis of entropy and age in the cell type corrected datasets. The change in entropy per decade of age is located on the x-is, with the effect size and standard error measurements from the independent EWAS on the y-axis. The 95% confidence intervals (CI) are on the right pane. The meta-analysis effect size is represented by the navy polygon at the bottom of the plot.

Supplementary Table 7 Shannon entropy results in blood.

Dataset	Tissue	N	Age (mean ± SD)	Array	% Male	Entropy p-value	Effect	Stderr	Entropy p-value CTC	Effect CTC	Stderr CTC	ALL p-value	ALL effect	ALL stderr	NON p-value	NON effect	NON stderr
BIOS	Blood	1408	37.5 ± 13.9	450K	34.2	5.62461E-12	0.000108893	1.56711E-05	1.66797E-15	0.000109871	1.36385E-05	2.48E-28	0.000230373	2.04E-05	1.64E-05	-5.70E-05	1.64E-05
FHS	Blood	2498	66.3 ± 8.9	450K	45.8	0.1877697	3.86966E-05	2.93697E-05	2.78042E-05	9.45327E-05	2.25156E-05	0.036621	8.69E-05	4.16E-05	0.2459571	-2.85E-05	2.46E-05
GOLDN	Blood	967	48.9 ± 16.4	450K	47.4	5.824E-08	0.000245031	4.48189E-05	0.00002118	0.000164522	3.85012E-05	6.11E-15	0.000404495	5.10E-05	0.5808556	2.84E-05	5.14E-05
GSE100264	Blood	386	49.5 ± 7.3	450K	100	0.9039949	-1.177589e-05	9.756643e-05	0.7386943	-3.318863e-05	9.941942e-05	0.218594	0.000175854	0.000143	0.08593418	-0.000211998	0.000123135
GSE106648	Blood	279	41.2 ± 11.2	450K	27.6	0.000546446	0.000165539	4.73E-05	5.87E-08	0.000190881	3.42E-05	0.00012	0.000283737	7.27E-05	0.8445477	6.92E-06	3.52E-05
GSE107080	Blood	405	47.9 ± 8.0	EPIC	100	8.585942e-05	7.605117e-05	8.585942e-05	0.1803641	8.5567e-05	6.376264e-05	0.209893	0.000158405	0.000126	0.6157777	-3.81E-05	7.58E-05
GSE111629	Blood	564	69.3 ± 11.3	450K	56.2	0.09432215	7.60E-05	4.54E-05	0.02716773	7.68E-05	3.47E-05	0.023896	0.00014016	6.19E-05	0.8094899	-1.17E-05	4.87E-05
GSE115278 450K	Blood	355	46.9 ± 15.4	450K	35.8	0.5078759	3.8748E-05	0.000058459	0.4362365	4.20312E-05	4.20312E-05	0.264123	7.82E-05	6.99E-05	0.8314082	-1.49E-05	6.99E-05
GSE115278 EPIC	Blood	106	46.1 ± 9.6	EPIC	36.8	0.5765098	-8.57169E-05	0.000152996	0.9879491	-1.93584E-06	0.000127859	0.906673	-2.30E-05	0.000196	0.3511263	-0.000163515	0.000174581
GSE116339	Blood	676	53.9 ± 12.9	EPIC	41.3	0.5539667	1.750642e-05	2.956551e-05	2.050586e-05	6.457464e-05	2.050586e-05	0.731308	1.56E-05	4.55E-05	0.4511984	1.99E-05	2.64E-05
GSE117859	Blood	608	49.4 ± 7.6	450K	100	0.6654736	3.19082E-05	7.37624E-05	0.5534623	3.25011E-05	5.48163E-05	0.665474	3.19E-05	7.38E-05	0.02726577	-0.000142777	6.45E-05
GSE117860	Blood	529	48.1 ± 7.8	450K	100	0.1489717	9.180556e-05	6.352066e-05	0.07426908	8.187646e-05	4.57793e-05	0.095856	0.000163884	9.82E-05	0.8964397	-6.76E-06	5.19E-05
GSE120307	Blood	34	35.5 ± 11.0	450K	53	0.9649455	9.95E-06	0.000224609	0.897244	-2.31E-05	0.000177596	0.892103	3.69E-05	0.00027	0.9041536	-2.65E-05	0.000218004
GSE128235	Blood	532	47.7 ± 13.4	450K	42.7	0.7846098	8.58E-06	3.14E-05	0.00234564	9.61E-05	3.14E-05	6.86E-07	0.000175678	4.38E-05	0.2050711	-5.50E-05	4.33E-05
GSE132203	Blood	795	42.3 ± 12.3	EPIC	28.2	8.25274E-08	0.000229697	4.24405E-05	7.7381E-10	0.000221362	3.55529E-05	3.12E-10	0.000355602	5.58E-05	0.162107	6.88E-05	4.92E-05
GSE147740	Blood	1029	41.6 ± 7.8	EPIC	60.3	0.8578721	-1.34189E-05	7.49119E-05	0.4528796	4.86315E-05	6.47635E-05	0.485971	5.40E-05	7.74E-05	0.3093811	-9.26E-05	9.11E-05
GSE151278	Blood	70	47.1 ± 14.6	450K	61.4	0.05822633	0.0001772981	9.203258e-05	0.08005927	0.0001357616	7.640418e-05	0.021347	0.000247961	0.000105	0.4872273	8.21E-05	0.00011757
GSE152026	Blood	927	35.2 ± 12.8	EPIC	54.9	0.0006154573	0.0001244897	3.622545e-05	5.892645e-05	0.0001302386	3.227122e-05	2.08E-11	0.000294393	4.34E-05	0.06430396	-9.20E-05	4.96E-05
GSE157131 450K	Blood	272	66.7 ± 6.5	450K	100	0.1403107	0.0002125616	0.0001437215	0.5701558	7.182903e-05	0.0001263443	0.029841	0.000407517	0.000187	0.7411363	-5.07E-05	0.000153269
GSE157131 EPIC	Blood	946	62.6 ± 9.7	EPIC	100	0.0003969457	0.0002206612	6.206786e-05	0.002077936	0.000146279	4.74E-05	6.55E-07	0.000421581	8.42E-05	0.7718038	-1.75E-05	6.02E-05
GSE164056	Blood	143	25.8 ± 6.4	EPIC	34.3	0.00011632	0.000280024	0.00011632	0.01755237	0.000225162	9.36921E-05	0.001561	0.000495444	0.000154	0.891697	1.60E-05	0.000117325
GSE166611	Blood	32	38.5 ± 12.4	450K	0	0.3533241	0.0001497593	0.0001588485	0.6865235	5.329313e-05	0.000130776	0.104708	0.000371036	0.000222	0.1738786	-0.000147505	0.000105896
GSE168739	Blood	402	42.0 ± 10.3	EPIC	44.9	0.005971567	0.000194046	7.02016E-05	0.000830029	0.00015494	4.60011E-05	0.00116	0.000368954	0.000113	0.7208878	-2.16E-05	6.03E-05
GSE19711	Blood	262	64.9 ± 6.7	27K	0	0.3314212	0.000113503	0.000116644	0.7376015	2.90499E-05	8.66153E-05	0.180264	0.000204286	0.000152	0.8299482	2.31E-05	0.000107638
GSE197674	Blood	2138	32.9 ± 10.0	EPIC	47.1	4.40E-28	0.000319244	2.86E-05	1.72E-44	0.000335183	2.34E-05	3.18E-43	0.000513533	3.64E-05	0.000502173	8.19E-05	2.35E-05
GSE197676	Blood	282	35.8 ± 10.2	EPIC	45.6	7.01873E-05	0.000376987	7.01873E-05	4.1106E-09	0.000379745	6.25442E-05	8.73E-05	0.000635775	8.73E-05	0.4128983	5.81E-05	7.08E-05
GSE20236	Blood	93	32.0 ± 18.78	27K	0	0.2295936	5.89216E-06	4.87146E-06	0.03455883	8.24421E-06	3.8422E-06	0.094658	8.24E-05	4.88E-05	0.1785174	-6.90E-05	5.09E-05
GSE32148	Blood	44	19.7 ± 14.6	450K	47.9	0.558993	6.96E-05	0.000118187	0.03244751	0.000201963	9.13E-05	0.663959	6.74E-05	0.000154	0.5142416	7.29E-05	0.00011083
GSE40279	Blood	656	64.0 ± 14.7	450K	48.5	0.3204021	5.10554E-05	5.13436E-05	0.9272555	4.24542E-06	4.64825E-05	0.092281	9.91E-05	5.88E-05	0.5737251	-3.08E-05	5.47E-05
GSE41037	Blood	703	37 ± 16	27K	61.9	0.1592406	4.56E-05	3.23E-05	0.008152934	6.73E-05	2.53E-05	0.000591	0.000120773	3.50E-05	0.5985051	-1.81E-05	3.43E-05
GSE41169	Blood	93	31.6 ± 10.4	450K	29.5	0.2869219	0.000134718	0.000125766	0.06228045	0.000220851	0.000117007	0.103579	0.000233607	0.000142	0.9900006	1.85E-06	0.000147594
GSE42861	Blood	689	51.9 ± 11.8	450K	28.6	0.001366545	0.000159234	4.95317E-05	0.04421635	8.12344E-05	4.03001E-05	2.06E-06	0.000307344	6.42E-05	0.3254957	-4.59E-05	4.66E-05
GSE49904	Blood	71	55 ± 14.5	27K	31	0.2402566	0.000163367	0.000137914	0.8946044	1.60E-05	0.000120196	0.05103	0.000279712	0.000141	0.6982522	6.45E-05	0.000165601
GSE50660	Blood	463	55.4 ± 6.6	450K	70.5	0.6890406	-4.47802E-05	0.000111836	0.9971298	3.68951E-07	0.000102507	0.947926	7.25E-06	0.000111	0.3793442	-0.000114795	0.000130456
GSE51032	Blood	378	53.2 ± 7.2	450K	20.4	0.001526438	0.000451577	0.000141488	0.001150322	0.000400854	0.000122414	0.000101	0.000600497	0.000153	0.1682637	0.000215158	0.000155877
GSE53740	Blood	162	68.1 ± 10.3	450K	47.9	0.1677049	0.000164121	0.000118422	0.08266189	0.000105605	6.05E-05	0.156757	0.000247893	0.000174	0.5929959	4.94E-05	9.22E-05
GSE53840	Blood	109	52 ± 8	450K	100	0.2500448	0.000164896	0.00014258	0.1407098	0.000117615	7.92472E-05	0.429937	0.000180077	0.000227	0.19465	0.000145058	0.000111145
GSE55763	Blood	2639	51 ± 10.1	450K	67.7	3.15012E-24	0.000107633	1.04972E-05	3.25478E-35	0.000104553	8.32351E-06	4.77E-36	0.000238909	1.88E-05	2.13E-07	-7.01E-05	1.35E-05
GSE58045	Blood	172	57.2 ± 8.2	27K	0	0.001603156	0.00078535	0.000244892	0.9427179	-1.47415E-05	-1.47415E-05	7.35E-05	0.001002084	0.000247	0.05238299	0.000568455	0.00029097

112

Dataset	Tissue	N	Age (mean ± SD)	Array	% Male	Entropy p-value	Effect	Stderr	ALL p-value	ALL effect	ALL stderr	NON p-value	NON effect	NON stderr
GSE67705	Blood	186	46.1 ± 8.8	450K	100	0.00380295	8.38585E-05	2.86073E-05	0.001462808	0.000156706	2.61127E-05	0.000166	0.000143254	3.73E-05	0.9221143	4.15E-06	4.24E-05
GSE67751	Blood	92	49.5 ± 8.9	450K	45.7	0.6995618	8.66E-05	0.00022372	0.4221127	-0.000141879	0.000175932	0.209037	0.000333099	0.000263	0.246199	-0.000244177	0.000209193
GSE69138	Blood	184	74.8 ± 12.6	450K	51.1	0.8784567	-1.70775E-05	0.000111515	0.9707839	3.05626e-06	8.33324E-05	0.684125	-6.07E-05	0.000149	0.7793944	3.06E-05	0.000109034
GSE72774	Blood	507	69.6 ± 11.2	450K	55.4	0.2455719	9.69E-05	8.34E-05	0.02605122	0.0001603	7.18E-05	0.085205	0.00015757	9.14E-05	0.8923563	1.39E-05	0.000102413
GSE72775	Blood	335	70.2 ± 10.3	450K	58.8	0.09070863	0.000187568	0.000110556	0.00390439	0.000246771	8.49139e-05	0.009118	0.000323684	0.000123	0.9918559	1.40E-06	0.000137433
GSE72776	Blood	84	66.4 ± 11.7	450K	59.5	0.2467337	0.000216987	0.000185992	0.6720531	5.85E-05	0.00013777	0.010186	0.000572139	0.000218	0.2100965	-0.000261696	0.000207169
GSE134429	Blood	63	56.0 ± 12.7	EPIC	11.1	0.01163675	0.000651829	0.00025701	0.3016693	0.00022742	0.000219863	3.67E-07	0.001083024	0.000209	0.7060233	0.000135957	0.000360143
GSE77445	Blood	85	33.8 ± 15.9	450K	50.5	0.2884682	-0.000124334	0.000116111	0.3170136	-0.000104965	0.00010404	0.351017	-9.59E-05	1.02E-04	0.2711624	-0.00015937	1.44E-04
GSE80417	Blood	636	40.4 ± 15.0	450K	59.3	0.1578088	4.11E-05	2.91E-05	0.005250155	7.60E-05	2.71E-05	0.001604	9.04E-05	2.85E-05	0.4897902	-2.51E-05	3.64E-05
GSE85311	Blood	38	47.4 ± 17.9	450K	71	0.9905637	2.223144e-06	0.0001866718	0.549951	-9.716753e-05	0.000161003	0.06426	0.000306489	0.000161	0.1166512	-0.000407661	0.000253578
GSE87571	Blood	728	47.4 ± 21.0	450K	46.7	3.21171E-33	0.000260116	2.05873E-05	6.84639E-19	0.000150763	1.65182E-05	5.43E-42	0.000457339	3.16E-05	0.4415698	-1.27E-05	1.65E-05
GSE96879	Blood	90	51.6 ± 14.4	450K	0	0.01628075	0.0001682668	6.869033e-05	0.007405148	0.0001286493	4.692491e-05	0.038121	0.000228333	0.000108	0.1831673	8.75E-05	6.52E-05
GSE99624	Blood	48	67.2 ± 9.9	450K	18.7	0.2631376	-0.000161301	0.000142382	0.5391254	-6.37106E-05	0.000102965	0.723378	-7.80E-05	0.000219	0.00013508	-0.000116006	0.00013508
JHS	Blood	1747	55.7 ± 12.3	EPIC	37	0.00319957	8.835696e-05	2.993141e-05	4.047926e-05	9.463708e-05	2.299669e-05	4.56E-05	1.53E-04	3.73E-05	2.13E-07	-7.01E-05	1.35E-05
NAS	Blood	1454	74.5 ± 7	450K	100	0.4407512	-4.77873E-05	6.19699E-05	0.2034583	7.78616E-05	6.11959E-05	0.049396	0.00015854	8.06E-05	0.5675632	-3.34E-05	5.84E-05
SATSA	Blood	1071	73.1 ± 9.7	450K	39.7	0.8347775	1.59187E-05	7.62863E-05	0.08817062	0.000113456	6.64096E-05	0.138618	-0.000112596	7.60E-05	0.003085382	0.000243825	8.22E-05
WHI	Blood	1871	64.2 ± 7.0	450K	0	0.3131826	4.41529E-05	4.3766E-05	0.04760497	6.50322E-05	3.28084E-05	0.27251	6.98E-05	6.36E-05	0.8377477	7.85E-06	3.83E-05

GOLDN, Genetics of Lipid-Lowering Drugs and Diet Network Study; WHI, Women's Health Initiative; NAS, Normative Aging Study; BIOS, Biobank-based integrative omics study; JHS, Jackson Heart Study; FHS, Framingham Heart Study; SATSA, Swedish Adoption/Twin Study of Aging; N, Sample number; SD, Standard deviation; Stderr, Standard error; CTC, cell-type corrected; ALL, All age-related CpGs; NON, non-age-related CpGs

Supplementary Table 8 Shannon entropy results in brain.

Dataset	Tissue / cell type	N	Age (mean ± SD)	Array	% Male	Entropy p-value	Effect	Stderr	ALL p-value	ALL effect	ALL stderr	NON p-value	NON effect	NON stderr
GSE64509	Multiple brain regions	251	74 ± 25	450K	48.4	0.7156654	-2.01E-05	5.52E-05	6.08E-05	0.000140819	3.45E-05	0.6207545	-2.95E-05	5.96E-05
GSE80970	Prefrontal cortex and superior temporal gyrus	286	86 ± 7.8	450K	37.8	0.2109281	7.09E-05	5.65E-05	0.001862637	0.000208419	6.64E-05	0.4242368	4.56E-05	5.70E-05
GSE15745	Cerebellum, Frontal cortex, Pons and Temporal cortex	495	47 ± 24	27K	68.0	0.000109917	6.22E-05	1.60E-05	4.59E-14	0.000164921	2.12E-05	0.05285631	3.05E-05	1.57E-05
GSE36194	Cerebellum and frontal cortex	421	51 ± 30	27K	65.0	0.9659132	-1.33E-06	3.12E-05	0.2397821	3.89E-05	3.30E-05	0.6793406	-1.36E-05	3.29E-05
GSE112179	Neurons	100	48 ± 11	EPIC	75.0	0.7236421	-1.86E-05	5.25E-05	0.9890099	9.35E-07	6.77E-05	0.6956304	-2.06E-05	5.26E-05
GSE66351	Glia & neurons from occipital cortex	62	75 ± 17	450K	45.2	0.2205621	7.62E-05	6.16E-05	0.000628416	0.000273557	7.58E-05	0.5275455	4.02E-05	6.32E-05
	Frontal cortex	63	74 ± 15		47.6	0.5261179	3.50E-05	5.49E-05	0.003670915	0.000142449	4.71E-05	0.803568	1.51E-05	6.06E-05
	Temporal cortex	65	74 ± 14		47.7	0.7098155	2.55E-05	6.83E-05	0.02484509	0.000126534	5.50E-05	0.924959	6.95E-06	7.35E-05
GSE38873	Cerebellum	168	44 ± 9.9	27K	63.0	0.9538931	-9.40E-06	0.000162307	0.9224646	1.17E-05	0.000119808	0.930116	-1.56E-05	0.000178
GSE41826	Postmortem frontal cortex	143	34 ± 16	450K	52.4	0.8030961	-1.04E-05	4.17E-05	0.04901475	0.000119553	6.02E-05	0.4281818	-3.43E-05	4.31E-05
GSE89702	Cerebellum	33	45 ± 16	450K	84.8	0.4662771	3.79E-05	5.13E-05	0.9004415	7.05E-06	5.59E-05	0.4058369	4.35E-05	5.16E-05
GSE89705	Striatum putamen (prefrontal cortex?)	33	46 ± 17	450K	81.8	0.5822988	3.35E-05	6.03E-05	0.7510594	-1.52E-05	4.76E-05	0.5374964	4.25E-05	6.81E-05
GSE89706	Striatum putamen	49	63 ± 17	450K	63.3	0.6196349	-1.60E-05	3.19E-05	0.6284179	-1.96E-05	4.03E-05	0.6441831	-1.53E-05	3.29E-05
GSE61431	Cerebellum and frontal cortex	87	62 ± 17	450K	64.3	0.1424812	-0.000105933	7.16E-05	0.9731912	-2.17E-06	6.42E-05	0.09123636	-0.000125311	7.34E-05
GSE61380	Frontal cortex	33	44 ± 16	450K	84.8	0.4201309	-0.000105103	0.000128636	0.0396598	0.000206768	9.63E-05	0.2726959	-0.000162144	0.000145
GSE74193	Prefrontal cortex	528	38 ± 22	450K	65.0	0.6822943	1.07E-05	2.61E-05	1.34E-15	0.000260334	3.16E-05	0.1475972	-3.88E-05	2.68E-05
GSE40360	Frontal lobe white matter	46	60 ± 11	450K	47.8	0.9110029	2.16E-06	1.92E-05	0.04369901	0.000178997	8.62E-05	0.03059615	-3.03E-05	1.36E-05
GSE59685	Cerebellum, entorhinal cortex, frontal cortex, superior temporal gyrus	451	84 ± 9.3	450K	41.5	0.8600311	-6.92E-06	3.92E-05	0.008981589	0.000150417	5.73E-05	0.3658708	-3.60E-05	3.97E-05
GSE49905	Cerebral cortex	77	23 ± 17	27K	72.7	0.4631012	-0.000131709	0.000178582	0.1732988	0.000162676	0.00011833	0.2716595	-0.000221698	0.000200
GSE59457	7 brain regions	128	46 ± 9.9	450K	73.4	0.7466247	-4.88E-05	0.000150633	0.218566	0.000151824	0.000122783	0.5947207	-8.60E-05	0.000161
GSE63347	Multiple brain regions	70	54 ± 9	450K	52.8	0.3266036	-0.000257492	0.000260591	0.435731	0.000130991	0.000167069	0.2447171	-0.000331489	0.000282

GSE72778	Multiple brain regions	454	65 ± 23	450K	55.9	0.06135526	-8.81E-05	4.70E-05	0.1703127	6.84E-05	4.98E-05	0.01743013	-0.000115777	4.85E-05
GSE144910	Superior temporal gyrus	88	48 ± 14	EPIC	71.6	0.3914515	5.31E-05	6.16E-05	0.01536863	0.000129191	5.21E-05	0.4787681	4.53E-05	6.36E-05
GSE138597	Multiple brain regions	36	70 ± 5.2	EPIC	100.0	0.2051167	0.000796853	0.000616824	0.04789221	-0.001006104	0.000490207	0.1696086	0.000980975	0.000699
GSE134379	Middle temporal gyrus and cerebellum	802	84 ± 7.9	450K	50.4	0.4687694	-3.30E-05	4.55E-05	0.8073313	1.57E-05	6.45E-05	0.351789	-4.28E-05	4.59E-05
GSE111165	Multiple brain regions	38	29 ± 16	EPIC	68.4	0.8417552	-2.20E-06	1.09E-05	0.8949735	1.60E-06	1.21E-05	0.8133227	-2.59E-06	1.09E-05
GSE98203	Neuronal nuclei	87	31 ± 11	450K	80.5	0.2203116	6.13E-05	4.97E-05	0.000602332	0.000210549	5.91E-05	0.5559209	3.36E-05	5.68E-05

N, Sample number; SD, Standard deviation; Stderr, Standard error; ALL, All age-related CpGs; NON, non-age-related CpGs

Supplementary Table 9 Shannon entropy results in buccal.

Dataset	Tissue	N	Age (mean ± SD)	Array	% Male	Entropy p-value	Effect	Stderr	Entropy p-value CTC	Effect CTC	Stderr CTC	ALL p-value	ALL effect	ALL stderr	NON p-value	NON effect	NON stderr
GSE78874	Buccal	259	69 ± 9.7	450K	56.3	0.01694848	0.0004215	0.0001754	0.03959663	0.000260033	0.00012571	0.000314347	0.00059306	0.000162358	0.02016652	0.000412237	0.000176337
GSE61653	Buccal	64	41 ± 12	450K	17.2	0.4092444	-0.000265	0.000318	0.008830213	0.000516793	0.0001911	0.1864148	0.00040812	0.000305471	0.5037349	-0.000215192	0.000319966
GSE94876	Buccal	120	46 ± 7.7	450K	100	0.2521155	0.000278	0.000242	0.9674149	6.43E-06	0.0001572	2.30E-06	0.00095258	0.000191697	0.3488984	0.000229826	0.000244374
GSE92767	Buccal	30	43.9 ± 15.1	450K	100	0.09642474	-0.000217	0.000128	0.3689671	-6.39E-05	7.05E-05	0.2958806	0.00011209	0.00010616	0.09458917	-0.000219449	0.00012888

N, Sample number; SD, Standard deviation; Stderr, Standard error; CTC, cell-type corrected; ALL, All age-related CpGs; NON, non-age-related CpGs.

Supplementary Table 10 Shannon entropy results in skin.

Dataset	N	Tissue	Age (mean ± SD)	Array	Entropy p-value	Effect	Stderr	ALL p-value	ALL effect	ALL stderr	NON p-value	NON effect	NON stderr
GSE90124	322	Skin	59.4 ± 9.2	450K	0.5599066	2.71E-05	4.64E-05	0.000122304	0.000212381	5.46E-05	0.7436627	-1.69E-05	5.18E-05
E-MATB-4385	108	Skin	47.5 ± 20.7	450K	0.004132492	9.935388e-05	3.389139e-05	1.32E-19	0.000483658	4.33E-05	0.7858886	9.49E-06	3.49E-05
E-MTAB-8992	56	Skin	50.4 ± 16.9	450K	0.03778073	0.000162048	7.609368e-05	4.35E-09	0.000700475	0.000100277	0.5629364	4.42E-05	7.59E-05
E-MTAB-8993	65	Skin	52.9 ± 16.7	450K	0.9499687	-5.442095e-06	8.638716e-05	1.88E-11	0.000812316	9.96E-05	0.004327886	0.000259659	8.77E-05
GSE51954	74	Skin	48.6 ± 23.7	450K	0.003729102	0.000196294	6.55E-05	5.01E-08	0.000581679	9.55E-05	0.0916882	0.00010636	6.22E-05
GSE115797	43	Skin	37.0 ± 14.2	450K	0.002091673	0.000272542	8.30E-05	0.003390085	0.0003442	0.000110646	0.00298965	0.000255342	8.09E-05

N, Sample number; SD, Standard deviation; Stderr, Standard error; ALL, All age-related CpGs; NON, non-age-related CpGs

Supplementary Table 11 Shannon entropy results in muscle.

Dataset	Tissue	N	Age (mean ± SD)	Array	% Male	Entropy p-value	Effect	Stderr	ALL p-value	ALL effect	ALL stderr	NON p-value	NON effect	NON stderr
FUSION	Muscle	312	60 ± 8.0	EPIC	57.0	0.602056	-4.58E-05	8.78E-05	0.9625795	-4.61E-06	9.82E-05	0.576371	-5.06E-05	9.06E-05
Gene SMART	Muscle	234	32 ± 8.1	EPIC	80.0	0.1519661	7.44E-05	5.18E-05	0.000605533	0.000230957	6.63E-05	0.2874738	5.67E-05	5.31E-05
Finnish Twin Cohort (FTC)	Muscle	156	43 ± 16	EPIC	58.0									
ABOS	Muscle	65	45 ± 8.2	450K	0.0	0.9571792	3.97E-06	7.36E-05	0.8269691	-1.69E-05	7.71E-05	0.9394379	5.86E-06	7.68E-05
LITER	Muscle	31	26 ± 5.9	EPIC	100.0	0.4720054	8.03E-05	0.000110194	0.1995586	0.000198938	0.000151542	0.5396265	6.76E-05	0.000108843
GSE61452	Muscle	57	44 ± 9.5	450K	26.3	0.9447985	-8.80E-06	0.000126567	0.8791115	3.16E-05	0.000206825	0.9181429	-1.29E-05	0.000124506
GSE135063	Muscle	57	39 ± 10	EPIC	28.0	0.4971413	6.05E-05	8.85E-05	0.3414824	8.13E-05	8.47E-05	0.5368168	5.80E-05	9.33E-05
GSE49908	Muscle	51	50 ± 17	27K	100.0	0.5833725	8.98E-05	0.00016269	0.6891093	3.97E-05	9.87E-05	0.583066	9.37E-05	0.000169544

GSE50498	Muscle	48	47 ± 26	450K	100.0	0.008037716	-0.000166626	6.01E-05	0.01558701	-0.00017414	6.93E-05	0.008184149	-0.00016603	6.01E-05
EPIK	Muscle	48	45 ± 22	EPIC	100.0	0.1251969	-5.73E-05	3.67E-05	0.02391557	-0.000107385	4.60E-05	0.173264	-5.18E-05	3.74E-05
EXACT	Muscle	48	33 ± 10	EPIC	100.0	0.6273395	-2.59E-05	5.29E-05	0.8592904	1.02E-05	5.74E-05	0.5880307	-2.97E-05	5.45E-05
GSE114763	Muscle	39	29 ± 6	EPIC	100.0	0.02186623	0.000448969	0.000187783	0.2964171	0.000339203	0.000320394	0.02298349	0.000460862	0.000194481

FTC, Finnish Twin Cohort; ABOS, Biological Atlas of Severe Obesity; FUSION, Finland-United States Investigation of NIDDM Genetics; LITER, Limb Immobilisation and Transcriptional/Epigenetic Responses; GeneSMART, Gene Skeletal Muscle Adaptive Response to Training; EPIK, Epigenetica & Kracht; N, Sample number; SD, Standard deviation; Stderr, Standard error; ALL, All age-related CpGs; NON, non-age-related CpGs

Supplementary Table 12 Shannon entropy results in adipose.

Dataset	Tissue	N	Age (mean ± SD)	Array	% Male	Entropy p-value	Effect	Stderr	ALL p-value	ALL effect	ALL stderr	NON p-value	NON effect	NON stderr
E-MTAB-1866	Adipose	648	59 ± 9.4	450K	0.0	0.03925426	6.03E-05	2.92E-05	2.12E-55	0.000853759	4.93E-05	0.2473112	3.33E-05	2.88E-05
FUSION	Adipose	293	60 ± 8.1	EPIC	55.6	0.5114692	-3.79E-05	5.76E-05	3.34E-08	0.000544908	9.60E-05	0.3831641	-5.05E-05	5.78E-05
GSE61450_&_GSE61453	Adipose	131	44 ± 9.5	450K	29.0	0.8015576	1.79E-05	7.09E-05	3.60E-08	0.000748553	0.000127689	0.8840291	-1.03E-05	7.06E-05
GSE68336	Adipose	70	31 ± 4.4	450K	42.9									
GSE24884	Adipose	55	43 ± 12	27K	0.0	0.1545539	0.000173284	0.000119982	0.003633812	0.000569234	0.000187027	0.1921208	0.000158545	0.000120004
GSE58622_&_GSE67024	Adipose	45	46 ± 11	450K	0.0	0.4512306	0.000127149	0.000167238	0.002365046	0.000794236	0.000245784	0.531651	0.000105066	0.000166618
GSE61257	Adipose	31	50 ± 14	450K	25.8	0.49665	4.41E-05	6.41E-05	0.4446858	0.000222513	0.00028716	0.5381725	3.82E-05	6.13E-05
GSE162166	Adipose	40	38 ± 12	EPIC	32.5	0.01179141	0.000364481	0.000137763	4.31E-08	0.001259691	0.000184694	4.31E-08	0.001259691	0.000184694

FUSION, Finland-Unites States Investigation of NIDDM Genetics; N, Sample number; SD, Standard deviation; Stderr, Standard error; ALL, All age-related CpGs; NON, non-age-related CpGs

Study	N		Estimate [95% CI]
GSE134379	802		-0.0003 [-0.0012, 0.0006]
GSE74193	528		0.0001 [-0.0004, 0.0006]
GSE15745	495		0.0006 [0.0003, 0.0009]
GSE72778	454		-0.0009 [-0.0018, 0.0000]
GSE59685	451		-0.0001 [-0.0008, 0.0007]
GSE36194	421		-0.0000 [-0.0006, 0.0006]
GSE80970	286		0.0007 [-0.0004, 0.0018]
GSE64509	251		-0.0002 [-0.0013, 0.0009]
GSE38873	168		-0.0001 [-0.0033, 0.0031]
GSE41826	143		-0.0001 [-0.0009, 0.0007]
GSE59457	128		-0.0005 [-0.0034, 0.0025]
GSE112179	100		-0.0002 [-0.0012, 0.0008]
GSE144910	88		0.0005 [-0.0007, 0.0017]
GSE61431	87		-0.0011 [-0.0025, 0.0003]
GSE98203	87		0.0006 [-0.0004, 0.0016]
GSE49905	77		-0.0013 [-0.0048, 0.0022]
GSE63347	70		-0.0026 [-0.0077, 0.0025]
GSE66351_TemporalCortex	65		0.0003 [-0.0011, 0.0016]
GSE66351_FrontalCortex	63		0.0003 [-0.0007, 0.0014]
GSE66351_OccipitalCortex	62		0.0008 [-0.0004, 0.0020]
GSE89706	49		-0.0002 [-0.0008, 0.0005]
GSE40360	46		0.0000 [-0.0004, 0.0004]
GSE111165	38		-0.0000 [-0.0002, 0.0002]
GSE138597	36		0.0080 [-0.0041, 0.0201]
GSE89702	33		0.0004 [-0.0006, 0.0014]
GSE89705	33		0.0003 [-0.0008, 0.0015]
GSE61380	33		-0.0011 [-0.0036, 0.0015]
FE Model			0.0001 [-0.0000, 0.0002]

x-axis: -0.0100 -0.0050 0.0000 0.0050 0.0100 0.0150 0.0200 0.0250

Change in entropy (/10 years)

Supplementary Figure 8 Forest plot of the genome-wide meta-analysis of entropy and age in brain.

On the x-axis is the change in entropy per decade of age, with the effect size and standard errors from each independent epigenome-wide association study (EWAS) plotted on the y-axis. The dataset and sample size (N) are on the left side of the plot and the 95% confidence intervals (CI) are displayed on the right panel. The meta-analysis effect size is represented by the blue polygon.
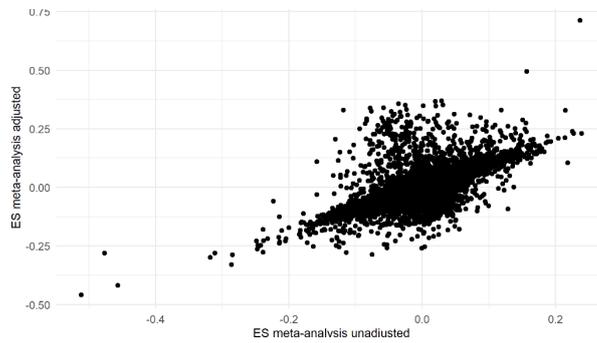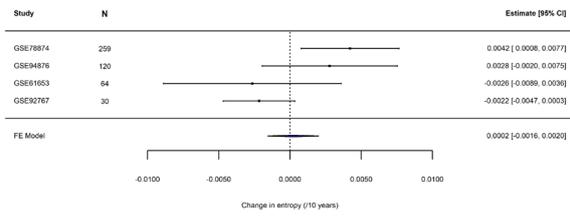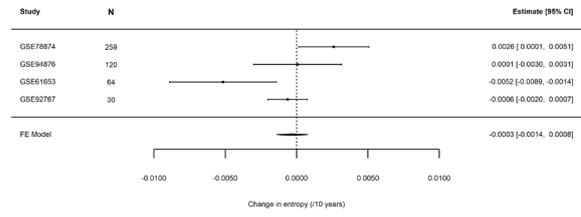
Supplementary Figure 9 A correlation plot of the effect sizes of the CpGs meta-analysed in the cell type adjusted meta-analysis versus the meta-analysis not adjusted for cell types of differential methylation and age in buccal tissue.

The effect sizes (ES) from the meta-analysis not adjusted for cell types on the x-axis and the ES for the meta-analysis adjusted for cell types on the y-axis. Pearson's correlation = 0.73, *p*-value < 2.2e-16. ES, effect size.

A

B
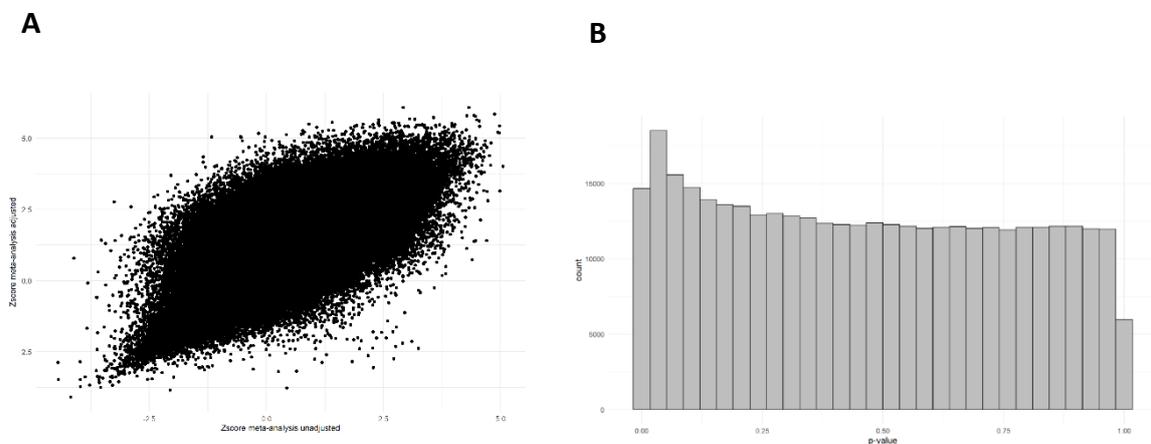
Supplementary Figure 10 Results of the meta-analysis of variable methylation and age in buccal tissue.

A) A correlation plot of the Zscores of the CpGs meta-analysed in the variable methylation and age meta-analyses in the cell type adjusted (y-axis) and the unadjusted meta-analysis (x-axis) in buccal tissue. Pearson's correlation coefficient = 0.72, *p*-value < 2.2e-26. B) A histogram of the *p*-values extracted from the meta-analysis of variable methylation and age in buccal tissue with cell type correction.

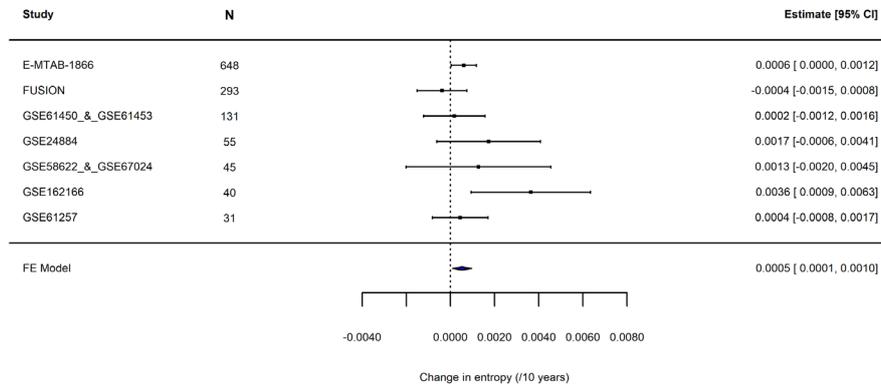**Supplementary Figure 11 Forest plots from the meta-analyses of entropy and age in buccal tissue before and after cell type correction.**

A) Forest plot of the genome-wide meta-analysis of entropy and age in buccal tissue that is not corrected for cell types. On the x-axis is the change in entropy per decade of age, with the effect size and standard errors from each independent epigenome-wide association study (EWAS) plotted on the y-axis. The dataset and sample size (N) are on the left side of the plot and the 95% confidence intervals (CI) are displayed on the right panel. The meta-analysis effect size is represented by the blue polygon. B) Forest plot of the genome-wide meta-analysis of entropy and age in buccal tissue that is corrected for cell types. On the x-axis is the change in entropy per decade of age, with the effect size and standard errors from each independent EWAS plotted on the y-axis. The dataset and N are on the left side of the plot and the 95% CI are displayed on the right panel. The meta-analysis effect size is represented by the blue polygon.
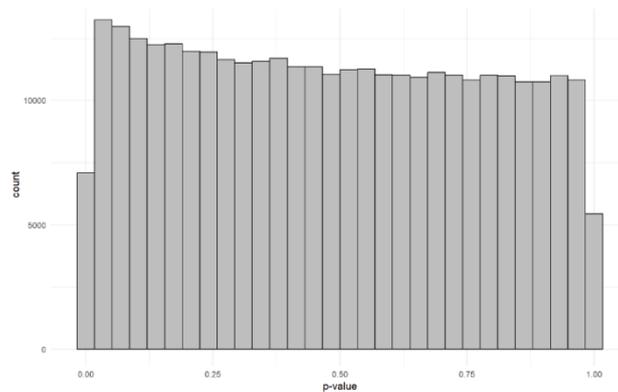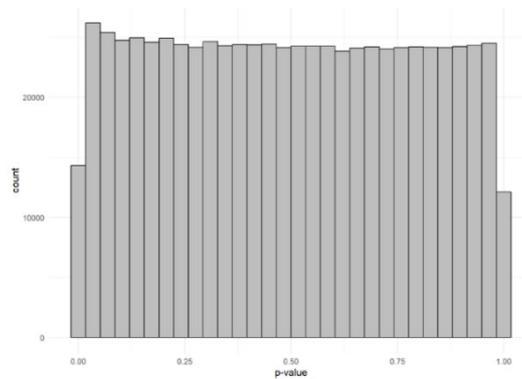

Supplementary Figure 12 A histogram of the *p*-values from the meta-analysis of variable methylation and age in adipose tissue.

Study	N		Estimate [95% CI]
E-MTAB-1866	648		0.0006 [0.0000, 0.0012]
FUSION	293		-0.0004 [-0.0015, 0.0008]
GSE61450_&_GSE61453	131		0.0002 [-0.0012, 0.0016]
GSE24884	55		0.0017 [-0.0006, 0.0041]
GSE58622_&_GSE67024	45		0.0013 [-0.0020, 0.0045]
GSE162166	40		0.0036 [0.0009, 0.0063]
GSE61257	31		0.0004 [-0.0008, 0.0017]
FE Model			0.0005 [0.0001, 0.0010]

-0.0040   0.0000 0.0020 0.0040 0.0060 0.0080

Change in entropy (/10 years)

Supplementary Figure 13 Forest plot of the genome-wide meta-analysis of entropy and age in adipose.

On the x-axis is the change in entropy per decade of age, with the effect size and standard errors from each independent epigenome-wide association study (EWAS) plotted on the y-axis. The dataset and sample size (N) are on the left side of the plot and the 95% confidence intervals (CI) are displayed on the right panel. The meta-analysis effect size is represented by the blue polygon.

Supplementary Figure 14 A histogram of the *p*-values from the meta-analysis of variable methylation and age in muscle tissue.
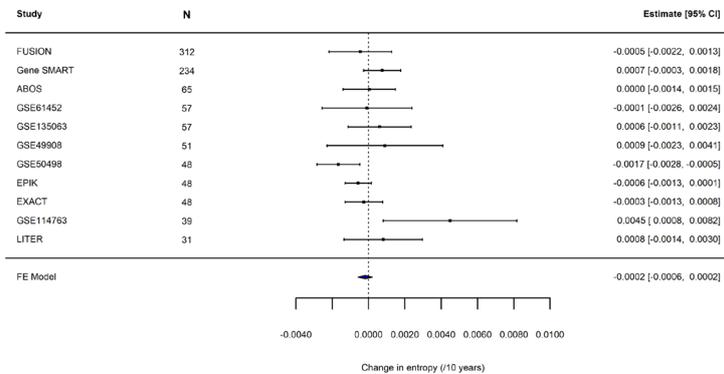
Supplementary Figure 15 Forest plot of the genome-wide meta-analysis of entropy and age in muscle.

Study	N	Estimate [95% CI]
FUSION	312	-0.0005 [-0.0022, 0.0013]
Gene SMART	234	0.0007 [-0.0003, 0.0018]
ABOS	65	0.0000 [-0.0014, 0.0015]
GSE61452	57	-0.0001 [-0.0026, 0.0024]
GSE135063	57	0.0006 [-0.0011, 0.0023]
GSE49908	51	0.0009 [-0.0023, 0.0041]
GSE50498	48	-0.0017 [-0.0028, -0.0005]
EPIK	48	-0.0006 [-0.0013, 0.0001]
EXACT	48	-0.0003 [-0.0013, 0.0008]
GSE114763	39	0.0045 [0.0008, 0.0082]
LITER	31	0.0008 [-0.0014, 0.0030]
FE Model		-0.0002 [-0.0006, 0.0002]

x-axis: Change in entropy (/10 years) [-0.0040, 0.0000, 0.0020, 0.0040, 0.0060, 0.0080, 0.0100]

On the x-axis is the change in entropy per decade of age, with the effect size and standard errors from each independent epigenome-wide association study (EWAS) plotted on the y-axis. The dataset and sample size (N) are on the left side of the plot and the 95% confidence intervals (CI) are displayed on the right panel. The meta-analysis effect size is represented by the blue polygon.
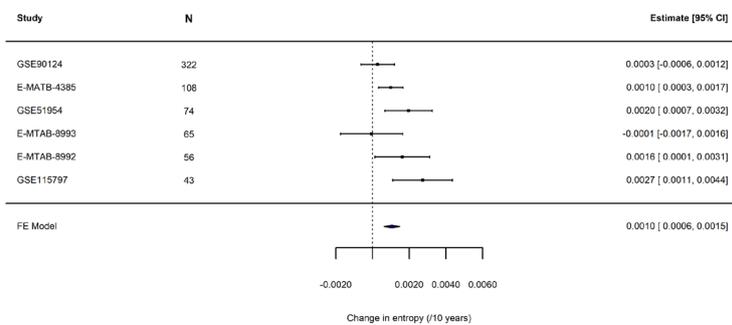
Supplementary Figure 16 Forest plot of the genome-wide meta-analysis of entropy and age in skin.

Study	N	Estimate [95% CI]
GSE90124	322	0.0003 [-0.0006, 0.0012]
E-MATB-4385	108	0.0010 [0.0003, 0.0017]
GSE51954	74	0.0020 [0.0007, 0.0032]
E-MTAB-8993	65	-0.0001 [-0.0017, 0.0016]
E-MTAB-8992	56	0.0016 [0.0001, 0.0031]
GSE115797	43	0.0027 [0.0011, 0.0044]
FE Model		0.0010 [0.0006, 0.0015]

x-axis: Change in entropy (/10 years) [-0.0020, 0.0020, 0.0040, 0.0060]

On the x-axis is the change in entropy per decade of age, with the effect size and standard errors from each independent epigenome-wide association study (EWAS) plotted on the y-axis. The dataset and sample size (N) are on the left side of the plot and the 95% confidence intervals (CI) are displayed on the right panel. The meta-analysis effect size is represented by the blue polygon.

References

1. Partridge L, Deelen J, Slagboom PE. Facing up to the global challenges of ageing. *Nature* 2018;**561**:45–56.

2. López-Otín C, Blasco MA, Partridge L, Serrano M, Kroemer G. Hallmarks of aging: An expanding universe. Cell. 2023;**186**:243–278.

3. Seale K, Horvath S, Teschendorff A, Eynon N, Voisin S. Making sense of the ageing methylome. Nat Rev Genet. 2022;**23**:585–605.

4. Seals DR, Justice JN, Larocca TJ. Physiological geroscience: Targeting function to increase healthspan and achieve optimal longevity. *Journal of Physiology* 2016;**594**:2001–2024.

5. Petsko GA. A seat at the table. *Genome Biol* 2008;**9**:113.

6. Crimmins EM. Lifespan and healthspan: Past, present, and promise. *Gerontologist* 2015;**55**:901–911.

7. Harper S. Economic and social implications of aging societies. *Science* 2014;**346**:587–591.

8. Jones MJ, Goodman SJ, Kobor MS. DNA methylation and healthy human aging. *Aging Cell* 2015;**14**:924–932.

9. López-Otín C, Blasco MA, Partridge L, Serrano M, Kroemer G. The Hallmarks of Aging. *Cell* 2013;**153**:1194–1217.

10. Gladyshev VN. Aging: progressive decline in fitness due to the rising deleteriome adjusted by genetic, environmental, and stochastic processes. *Aging Cell* 2016;**15**:594–602.

11. Cartee GD, Hepple RT, Bamman MM, Zierath JR. Exercise Promotes Healthy Aging of Skeletal Muscle. *Cell Metab* 2016;**23**:1034–1047.

12. Kolovou GD, Kolovou V, Mavrogeni S. We are ageing. *Biomed Res Int* 2014;**2014**.

13.   Holloszy JO. The Biology of Aging. *Mayo Clin Proc* 2000;**75**:S3–S9.

14.   Hägg S, Jylhävä J. Sex differences in biological aging with a focus on human studies. *Elife* 2021;**10**:1–27.

15.   Zhang W, Qu J, Liu G-H, Belmonte JCI. The ageing epigenome and its rejuvenation. *Nat Rev Mol Cell Biol* 2020;**21**:137–150.

16.   Kane AE, Sinclair DA. Epigenetic changes during aging and their reprogramming potential. *Crit Rev Biochem Mol Biol* 2019;**54**:61–83.

17.   Greenberg MVC, Bourc'his D. The diverse roles of DNA methylation in mammalian development and disease. *Nat Rev Mol Cell Biol* 2019;**20**:590–607.

18.   Bauer M. Cell-type-specific disturbance of DNA methylation pattern: A chance to get more benefit from and to minimize cohorts for epigenome-wide association studies. *Int J Epidemiol* 2018;**47**:917–927.

19.   Teschendorff AE, Relton CL. Statistical and integrative system-level analysis of DNA methylation data. *Nat Rev Genet* 2018;**19**:129–147.

20.   Feil R, Fraga MF. Epigenetics and the environment: Emerging patterns and implications. *Nat Rev Genet* 2012;**13**:97–109.

21.   Turner D, Gorski P, Maasar M, Seaborne R, Brown A, Kitchen M *et al.* DNA methylation across the genome in aged human skeletal muscle tissue and muscle stem cells: The role of HOX genes and physical activity. *Sci Rep* 2020;**10**:15360.

22.   Barrès R, Yan J, Egan B, Treebak JT, Rasmussen M, Fritz T *et al.* Acute exercise remodels promoter methylation in human skeletal muscle. *Cell Metab* 2012;**15**:405–411.

23.   Urdinguio RG, Tejedor JR, Fernández-Sanjurjo M, Pérez RF, Peñarroya A, Ferrero C *et al.* Physical exercise shapes the mouse brain epigenome. *Mol Metab* 2021;**54**:101398.

24.   Voisin S, Almén MS, Moschonis G, Chrousos GP, Manios Y, Schiöth HB. Dietary fat quality impacts genome-wide DNA methylation patterns in a cross-sectional study of Greek preadolescents. *European Journal of Human Genetics* 2015;**23**:654–662.

25. Pauwels S, Ghosh M, Duca RC, Bekaert B, Freson K, Huybrechts I *et al.* Maternal intake of methyl-group donors affects DNA methylation of metabolic genes in infants. *Clin Epigenetics* 2017;**9**:1–13.

26. Joubert BR, Felix JF, Yousefi P, Bakulski KM, Just AC, Breton C *et al.* DNA Methylation in Newborns and Maternal Smoking in Pregnancy: Genome-wide Consortium Meta-analysis. *Am J Hum Genet* 2016;**98**:680–696.

27. Elliott HR, Tillin T, McArdle WL, Ho K, Duggirala A, Frayling TM *et al.* Differences in smoking associated DNA methylation patterns in South Asians and Europeans. *Clin Epigenetics* 2014;**6**:1–10.

28. Tsaprouni LG, Yang TP, Bell J, Dick KJ, Kanoni S, Nisbet J *et al.* Cigarette smoking reduces DNA methylation levels at multiple genomic loci but the effect is partially reversible upon cessation. *Epigenetics* 2014;**9**:1382–1396.

29. Christensen BC, Houseman EA, Marsit CJ, Zheng S, Wrensch MR, Wiemels JL *et al.* Aging and environmental exposures alter tissue-specific DNA methylation dependent upon CpG island context. *PLoS Genet* 2009;**5**:e1000602.

30. Plusquin M, Guida F, Polidoro S, Vermeulen R, Raaschou-Nielsen O, Campanella G *et al.* DNA methylation and exposure to ambient air pollution in two prospective cohorts. *Environ Int* 2017;**108**:127–136.

31. Zhang L, Silva TC, Young JI, Gomez L, Schmidt MA, Hamilton-Nelson KL *et al.* Epigenome-wide meta-analysis of DNA methylation differences in prefrontal cortex implicates the immune processes in Alzheimer's disease. *Nat Commun* 2020;**11**.

32. Li P, Marshall L, Oh G, Jakubowski JL, Groot D, He Y *et al.* Epigenetic dysregulation of enhancers in neurons is associated with Alzheimer's disease pathology and cognitive symptoms. *Nat Commun* 2019;**10**:1–14.

33. Altuna M, Urdánoz-Casado A, Sánchez-Ruiz De Gordoa J, Zelaya M V., Labarga A, Lepesant JMJ *et al.* DNA methylation signature of human hippocampus in Alzheimer's disease is linked to neurogenesis. *Clin Epigenetics* 2019;**11**:1–16.

34. Pellegrini C, Pirazzini C, Sala C, Sambati L, Yusipov I, Kalyakulina A *et al.* A Meta-Analysis of Brain DNA Methylation Across Sex, Age, and Alzheimer's Disease Points for Accelerated Epigenetic Aging in Neurodegeneration. *Front Aging Neurosci* 2021;**13**:1–21.

35. Huo Z, Zhu Y, Yu L, Yang J, De Jager P, Bennett DA *et al.* DNA methylation variability in Alzheimer's disease. *Neurobiol Aging* 2019;**76**:35–44.

36. Palou-Márquez G, Subirana I, Nonell L, Fernández-Sanlés A, Elosua R. DNA methylation and gene expression integration in cardiovascular disease. *Clin Epigenetics* 2021;**13**:1–12.

37. Movassagh M, Choy MK, Goddard M, Bennett MR, Down TA, Foo RSY. Differential DNA methylation correlates with differential expression of angiogenic factors in human heart failure. *PLoS One* 2010;**5**:e8564.

38. Fernández-Sanlés A, Sayols-Baixeras S, Subirana I, Sentí M, Pérez-Fernández S, de Castro Moura M *et al.* DNA methylation biomarkers of myocardial infarction and cardiovascular disease. *Clin Epigenetics* 2021;**13**:1–11.

39. Klutstein M, Moss J, Kaplan T, Cedar H. Contribution of epigenetic mechanisms to variation in cancer risk among tissues. *Proc Natl Acad Sci U S A* 2017;**114**:2230–2234.

40. Issa JP. Aging and epigenetic drift: A vicious cycle. *Journal of Clinical Investigation* 2014;**124**:24–29.

41. Klutstein M, Nejman D, Greenfield R, Cedar H. DNA methylation in cancer and aging. *Cancer Res* 2016;**76**:3446–3450.

42. Teschendorff AE, Menon U, Gentry-Maharaj A, Ramus SJ, Weisenberger DJ, Shen H *et al.* Age-dependent DNA methylation of genes that are suppressed in stem cells is a hallmark of cancer. *Genome Res* 2010;**20**:440–446.

43. Hannum G, Guinney J, Zhao L, Zhang L, Hughes G, Sadda SV *et al.* Genome-wide Methylation Profiles Reveal Quantitative Views of Human Aging Rates. *Mol Cell* 2013;**49**:359–367.

44.   Langfelder P, Horvath S. WGCNA: An R package for weighted correlation network analysis. *BMC Bioinformatics* 2008;**9**:559.

45.   Horvath S, Zhang Y, Langfelder P, Kahn RS, Boks MPM, van Eijk K *et al.* Aging effects on DNA methylation modules in human brain and blood tissue. *Genome Biol* 2012;**13**:R97.

46.   Wilson VL, Smith RA, Ma S, Cutler RG. Genomic 5-methyldeoxycytidine decreases with age. *Journal of Biological Chemistry* 1987;**262**:9948–9951.

47.   Fuke C, Shimabukuro M, Petronis A, Sugimoto J, Oda T, Miura K *et al.* Age related changes in 5-methylcytosine content in human peripheral leukocytes and placentas: An HPLC-based study. *Ann Hum Genet* 2004;**68**:196–204.

48.   Vanyushin B, Nemirovsky L, Klimenko V, Vasiliev V, Belozersky A. The 5-Methylcytosine in DNA of Rats. *Gerotologia* 1973;**19**:138–152.

49.   Unnikrishnan A, Hadad N, Masser DR, Jackson J, Freeman WM, Richardson A. Revisiting the genomic hypomethylation hypothesis of aging. *Ann N Y Acad Sci* 2018;**1418**:69–79.

50.   Heyn H, Li N, Ferreira HJ, Moran S, Pisano DG, Gomez A *et al.* Distinct DNA methylomes of newborns and centenarians. *Proc Natl Acad Sci U S A* 2012;**109**:10522–10527.

51.   Unnikrishnan A, Freeman WM, Jackson J, Wren JD, Porter H, Richardson A. The role of DNA methylation in epigenetics of aging. *Pharmacol Ther* 2019;**195**:172–185.

52.   Lister R, Mukamel EA, Nery JR, Urich M, Puddifoot CA, Johnson ND *et al.* Global epigenomic reconfiguration during mammalian brain development. *Science* 2013;**341**.

53.   Raddatz G, Hagemann S, Aran D, Söhle J, Kulkarni PP, Kaderali L *et al.* Aging is associated with highly defined epigenetic changes in the human epidermis. *Epigenetics Chromatin* 2013;**6**:1–12.

54.   Hadad N, Masser DR, Logan S, Wronowski B, Mangold CA, Clark N *et al.* Absence of genomic hypomethylation or regulation of cytosine-modifying enzymes with aging in male and female mice. *Epigenetics Chromatin* 2016;**9**:1–20.

55. Cole JJ, Robertson NA, Rather MI, Thomson JP, McBryan T, Sproul D *et al.* Diverse interventions that extend mouse lifespan suppress shared age-associated epigenetic changes at critical gene regulatory regions. *Genome Biol* 2017;**18**:1–16.

56. Eckhardt F, Lewin J, Cortese R, Rakyan VK, Attwood J, Burger M *et al.* DNA methylation profiling of human chromosomes 6, 20 and 22. *Nat Genet* 2006;**38**:1378–1385.

57. Day K, Waite LL, Thalacker-Mercer A, West A, Bamman MM, Brooks JD *et al.* Differential DNA methylation with age displays both common and dynamic features across human tissues that are influenced by CpG landscape. *Genome Biol* 2013;**14**:1–19.

58. Lu AT, Fei Z, Haghani A, Robeck TR, Zoller JA, Li CZ *et al.* Universal DNA methylation age across mammalian tissues. *Nat Aging* 2023 doi:https://doi.org/10.1101/2021.01.18.426733.

59. Rakyan VK, Down TA, Maslau S, Andrew T, Yang TP, Beyan H *et al.* Human aging-associated DNA hypermethylation occurs preferentially at bivalent chromatin domains. *Genome Res* 2010;**20**:434–439.

60. Bell JT, Pai AA, Pickrell JK, Gaffney DJ, Pique-Regi R, Degner JF *et al.* DNA methylation patterns associate with genetic and gene expression variation in HapMap cell lines. *Genome Biol* 2011;**12**:R10.

61. Horvath S. DNA methylation age of human tissues and cell types. *Genome Biol* 2013;**14**:R115.

62. Slieker RC, van Iterson M, Luijk R, Beekman M, Zhernakova D V., Moed MH *et al.* Age-related accrual of methylomic variability is linked to fundamental ageing mechanisms. *Genome Biol* 2016;**17**:1–13.

63. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W *et al.* Limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* 2015;**43**:e47.

64. Peters TJ, Buckley MJ, Statham AL, Pidsley R, Samaras K, Lord RL *et al.* De novo identification of differentially methylated regions in the human genome. *Epigenetics Chromatin* 2015;**8**:1–16.

65. Jaffe AE, Murakami P, Lee H, Leek JT, Fallin MD, Feinberg AP *et al.* Bump hunting to identify differentially methylated regions in epigenetic epidemiology studies. *Int J Epidemiol* 2012;**41**:200–209.

66. Pedersen BS, Schwartz DA, Yang I V., Kechris KJ. Comb-p: Software for combining, analyzing, grouping and correcting spatially correlated P-values. *Bioinformatics* 2012;**28**:2986–2988.

67. Aryee MJ, Jaffe AE, Corrada-Bravo H, Ladd-Acosta C, Feinberg AP, Hansen KD *et al.* Minfi: A flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. *Bioinformatics* 2014;**30**:1363–1369.

68. Butcher LM, Beck S. Probe Lasso: A novel method to rope in differentially methylated regions with 450K DNA methylation data. *Methods* 2015;**72**:21–28.

69. Field AE, Robertson NA, Wang T, Havas A, Ideker T, Adams PD. DNA Methylation Clocks in Aging: Categories, Causes, and Consequences. *Mol Cell* 2018;**71**:882–895.

70. Wang Y, Pedersen NL, Hägg S. Implementing a method for studying longitudinal DNA methylation variability in association with age. *Epigenetics* 2018;**13**:866–874.

71. Fernández AF, Bayón GF, Urdinguio RG, Toraño EG, García MG, Carella A *et al.* H3K4me1 marks DNA regions hypomethylated during aging in human stem and differentiated cells. *Genome Res* 2015;**29**:27–40.

72. Oh G, Ebrahimi S, Wang SC, Cortese R, Kaminsky ZA, Gottesman II *et al.* Epigenetic assimilation in the aging human brain. *Genome Biol* 2016;**17**:1–11.

73. Phipson B, Oshlack A. DiffVar: a new method for detecting differential variability with application to methylation in cancer and aging. *Genome Biol* 2014;**15**:465.

74. Yusipov I, Bacalini MG, Kalyakulina A, Krivonosov M, Pirazzini C, Gensous N *et al.* Age-related DNA methylation changes are sex-specific: a comprehensive assessment. *Aging* 2020;**12**:24057–24080.

75. Vershinina O, Bacalini MG, Zaikin A, Franceschi C, Ivanchenko M. Disentangling age-dependent DNA methylation: deterministic, stochastic, and nonlinear. *Sci Rep* 2021;**11**:1–12.

76. Planterose Jiménez B, Liu F, Caliebe A, Montiel González D, Bell JT, Kayser M *et al.* Equivalent DNA methylation variation between monozygotic co-twins and unrelated individuals reveals universal epigenetic inter-individual dissimilarity. *Genome Biol* 2021;**22**:1–23.

77. Fraga MF, Ballestar E, Paz MF, Ropero S, Setien F, Ballestar ML *et al.* Epigenetic differences arise during the lifetime of monozygotic twins. *Proc Natl Acad Sci U S A* 2005;**102**:10604–10609.

78. Talens RP, Christensen K, Putter H, Willemsen G, Christiansen L, Kremer D *et al.* Epigenetic variation during the adult lifespan: Cross-sectional and longitudinal data on monozygotic twin pairs. *Aging Cell* 2012;**11**:694–703.

79. Van Dongen J, Nivard MG, Willemsen G, Hottenga JJ, Helmer Q, Dolan C V. *et al.* Genetic and environmental influences interact with age and sex in shaping the human methylome. *Nat Commun* 2016;**7**:1–13.

80. Teschendorff AE, West J, Beck S. Age-associated epigenetic drift: Implications, and a case of epigenetic thrift? *Hum Mol Genet* 2013;**22**:7–15.

81. Breusch TS, Pagan AR. A Simple Test for Heteroscedasticity and Random Coefficient Variation. *Econometrica* 1979;**47**:1287–1294.

82. Mur J, McCartney DL, Walker RM, Campbell A, Bermingham ML, Morris SW *et al.* DNA methylation in APOE: The relationship with Alzheimer's and with cardiovascular health. *Alzheimer's and Dementia: Translational Research and Clinical Interventions* 2020;**6**:1–9.

83. Hayflick L. Entropy explains aging, genetic determinism explains longevity, and undefined terminology explains misunderstanding both. *PLoS Genet* 2007;**3**:2351–2354.

84. Martin-Herranz D, Aref-Eshghi E, Bonder MJ, Stubbs T, Stegle O, Sadikovic B *et al.* Screening for genes that accelerate the epigenetic ageing clock in humans reveals a role for the H3K36 methyltransferase NSD1. *Genome Biol* 2019;**20**:1–19.

85. Jenkinson G, Pujadas E, Goutsias J, Feinberg AP. Potential energy landscapes identify the information-theoretic nature of the epigenome. *Nat Genet* 2017;**49**:719–729.

86. Yan Q, Paul KC, Lu AT, Kusters C, Binder AM, Horvath S *et al.* Epigenetic mutation load is weakly correlated with epigenetic age acceleration. *Aging* 2020;**12**:17863–17894.

87. Shannon CE. A Mathematical Theory of Communication. *Bell System Technical Journal* 1948;**27**:623–656.

88. Yan Q, Paul KC, Lu AT, Kusters C, Binder AM, Horvath S *et al.* Epigenetic mutation load is weakly correlated with epigenetic age acceleration. *Aging* 2020;**12**:17863–17894.

89. Sziráki A, Tyshkovskiy A, Gladyshev VN. Global remodeling of the mouse DNA methylome during aging and in response to calorie restriction. *Aging Cell* 2018;**17**:1–12.

90. Wang T, Tsui B, Kreisberg JF, Robertson NA, Gross AM, Yu MK *et al.* Epigenetic aging signatures in mice livers are slowed by dwarfism, calorie restriction and rapamycin treatment. *Genome Biol* 2017;**18**:1–11.

91. Mendelsohn AR, Larrick JW. The DNA methylome as a biomarker for epigenetic instability and human aging. *Rejuvenation Res* 2013;**16**:74–77.

92. Rando TA, Wyss-Coray T. Asynchronous, contagious and digital aging. *Nat Aging* 2021;**1**:29–35.

93. Rudolph KL. DNA-methylation aging at single-cell level. *Nat Aging* 2021;**1**:1086–1087.

94. Trapp A, Kerepesi C, Gladyshev VN. Profiling epigenetic age in single cells. *Nat Aging* 2021;**1**:1189–1201.

95.   Zhang L, Xie WJ, Liu S, Meng L, Gu C, Gao YQ. DNA Methylation Landscape Reflects the Spatial Organization of Chromatin in Different Cells. *Biophys J* 2017;**113**:1395–1404.

96.   Haerter JO, Lövkvist C, Dodd IB, Sneppen K. Collaboration between CpG sites is needed for stable somatic inheritance of DNA methylation states. *Nucleic Acids Res* 2014;**42**:2235–2244.

97.   Li G, Liu Y, Zhang Y, Kubo N, Yu M, Fang R *et al.* Joint profiling of DNA methylation and chromatin architecture in single cells. *Nat Methods* 2019;**16**:991–993.

98.   Lövkvist C, Dodd IB, Sneppen K, Haerter JO. DNA methylation in human epigenomes depends on local topology of CpG sites. *Nucleic Acids Res* 2016;**44**:5123–5132.

99.   Mallona I, Ausso S, Diez-Villanueva A, Moreno V, Peinado MA. DNA co-methylation networks outline the structure and remodeling dynamics of colorectal cancer epigenome. *bioRxiv* 2020;49–52.

100.   Fortin JP, Hansen KD. Reconstructing A/B compartments as revealed by Hi-C using long-range correlations in epigenetic data. *Genome Biol* 2015;**16**:1–23.

101.   West J, Widschwendter M, Teschendorff AE. Distinctive topology of age-associated epigenetic drift in the human interactome. *Proc Natl Acad Sci U S A* 2013;**110**:14138–14143.

102.   Willis CRG, Ames RM, Deane CS, Phillips BE, Boereboom CL, Abdulla H *et al.* Network analysis of human muscle adaptation to aging and contraction. *Aging* 2020;**12**:740–755.

103.   Bocklandt S, Lin W, Sehl ME, Sanchez FJ, Sinsheimer JS, Horvath S *et al.* Epigenetic Predictor of Age. *PLoS One* 2011;**6**:e14821.

104.   Zhang B, Horvath S. A general framework for weighted gene co-expression network analysis. *Stat Appl Genet Mol Biol* 2005;**4**.

105.   Yip AM, Horvath S. Gene network interconnectedness and the generalized topological overlap measure. *BMC Bioinformatics* 2007;**8**:1–14.

106. Bell CG, Lowe R, Adams PD, Baccarelli AA, Beck S, Bell JT *et al.* DNA methylation aging clocks: Challenges and recommendations. *Genome Biol* 2019;**20**:1–24.

107. West J, Widschwendter M, Teschendorff AE. Distinctive topology of age-associated epigenetic drift in the human interactome. *Proc Natl Acad Sci U S A* 2013;**110**:14138–14143.

108. Yuan T, Jiao Y, de Jong S, Ophoff RA, Beck S, Teschendorff AE. An Integrative Multi-scale Analysis of the Dynamic DNA Methylation Landscape in Aging. *PLoS Genet* 2015;**11**:1–21.

109. Bonder MJ, Luijk R, Zhernakova D V., Moed M, Deelen P, Vermaat M *et al.* Disease variants alter transcription factor levels and methylation of their binding sites. *Nat Genet* 2017;**49**:131–138.

110. Smyth G. limma: Linear Models for Microarray Data. 2005https://doi.org/10.1007/0-387-29362-0_23.

111. Guo S, Diep D, Plongthongkum N, Fung HL, Zhang K, Zhang K. Identification of methylation haplotype blocks aids in deconvolution of heterogeneous tissue samples and tumor tissue-of-origin mapping from plasma DNA. *Nat Genet* 2017;**49**:635–642.

112. Vanderkraats ND, Hiken JF, Decker KF, Edwards JR. Discovering high-resolution patterns of differential DNA methylation that correlate with gene expression changes. *Nucleic Acids Res* 2013;**41**:6816–6827.

113. Schlosberg CE, VanderKraats ND, Edwards JR. Modeling complex patterns of differential DNA methylation that associate with gene expression changes. *Nucleic Acids Res* 2017;**45**:5100–5111.

114. Friedman J, Hastie T, Tibshirani R. Regularization paths for generalized linear models via coordinate descent. *J Stat Softw* 2010;**33**:1–22.

115. Belsky DW, Caspi A, Houts R, Cohen HJ, Corcoran DL, Danese A *et al.* Quantification of biological aging in young adults. *Proc Natl Acad Sci U S A* 2015;**112**:E4104–E4110.

116. Jylhävä J, Pedersen NL, Hägg S. Biological Age Predictors. *EBioMedicine* 2017;**21**:29–36.

117. Levine ME, Lu AT, Quach A, Chen BH, Assimes TL, Bandinelli S *et al.* An epigenetic biomarker of aging for lifespan and healthspan. *Aging* 2018;**10**:573–591.

118. Lacey M, Baribault C, Ehrlich KC, Ehrlich M. Atherosclerosis-associated differentially methylated regions can reflect the disease phenotype and are often at enhancers. *Atherosclerosis* 2019;**280**:183–191.

119. Xue Y, Guo Y, Luo S, Zhou W, Xiang J, Zhu Y *et al.* Aberrantly Methylated-Differentially Expressed Genes Identify Novel Atherosclerosis Risk Subtypes. *Front Genet* 2020;**11**:1–12.

120. Bakshi C, Vijayvergiya R, Dhawan V. Aberrant DNA methylation of M1-macrophage genes in coronary artery disease. *Sci Rep* 2019;**9**:1–22.

121. Kim JY, Choi BG, Jelinek J, Kim DH, Lee SH, Cho K *et al.* Promoter methylation changes in ALOX12 and AIRE1: Novel epigenetic markers for atherosclerosis. *Clin Epigenetics* 2020;**12**:1–13.

122. Kazmi N, Elliott HR, Burrows K, Tillin T, Hughes AD, Chaturvedi N *et al.* Associations between high blood pressure and DNA methylation. *PLoS One* 2020;**15**:1–16.

123. Richard MA, Huan T, Ligthart S, Gondalia R, Jhun MA, Brody JA *et al.* DNA Methylation Analysis Identifies Loci for Blood Pressure Regulation. *Am J Hum Genet* 2017;**101**:888–902.

124. Bacos K, Gillberg L, Volkov P, Olsson AH, Hansen T, Pedersen O *et al.* Blood-based biomarkers of age-associated epigenetic changes in human islets associate with insulin secretion and diabetes. *Nat Commun* 2016;**7**:1–13.

125. Dayeh T, Volkov P, Salö S, Hall E, Nilsson E, Olsson AH *et al.* Genome-Wide DNA Methylation Analysis of Human Pancreatic Islets from Type 2 Diabetic and Non-Diabetic Donors Identifies Candidate Genes That Influence Insulin Secretion. *PLoS Genet* 2014;**10**:e1004160.

126. Volkov P, Bacos K, Ofori JK, Esguerra JLS, Eliasson L, Rönn T *et al.* Whole-Genome bisulfite sequencing of human pancreatic islets reveals novel differentially methylated regions in type 2 diabetes pathogenesis. *Diabetes* 2017;**66**:1074–1085.

127. Barajas-Olmos F, Centeno-Cruz F, Zerrweck C, Imaz-Rosshandler I, Martínez-Hernández A, Cordova EJ *et al.* Altered DNA methylation in liver and adipose tissues derived from individuals with obesity and type 2 diabetes. *BMC Med Genet* 2018;**19**:1–8.

128. Fernández-Tajes J, Soto-Hermida A, Vázquez-Mosquera ME, Cortés-Pereira E, Mosquera A, Fernández-Moreno M *et al.* Genome-wide DNA methylation analysis of articular chondrocytes reveals a cluster of osteoarthritic patients. *Ann Rheum Dis* 2014;**73**:668–677.

129. Aref-Eshghi E, Zhang Y, Liu M, Harper PE, Martin G, Furey A *et al.* Genome-wide DNA methylation study of hip and knee cartilage reveals embryonic organ and skeletal system morphogenesis as major pathways involved in osteoarthritis. *BMC Musculoskelet Disord* 2015;**16**:1–10.

130. Delgado-Calle J, Fernández AF, Sainz J, Zarrabeitia MT, Sañudo C, García-Renedo R *et al.* Genome-wide profiling of bone reveals differentially methylated regions in osteoporosis and osteoarthritis. *Arthritis Rheum* 2013;**65**:197–205.

131. Wang Y, Li F, Zhang G, Kang L, Guan H. Ultraviolet-B induces ERCC6 repression in lens epithelium cells of age-related nuclear cataract through coordinated DNA hypermethylation and histone deacetylation. *Clin Epigenetics* 2016;**8**:1–15.

132. Teschendorff AE, Yang Z, Wong A, Pipinikas CP, Jiao Y, Jones A *et al.* Correlation of smoking-associated DNA methylation changes in buccal cells with DNA methylation changes in epithelial cancer. *JAMA Oncol* 2015;**1**:476–485.

133. Teschendorff AE, Gao Y, Jones A, Ruebner M, Beckmann MW, Wachter DL *et al.* DNA methylation outliers in normal breast tissue identify field defects that are enriched in cancer. *Nat Commun* 2016;**7**:1–11.

134. Hansen KD, Timp W, Bravo HC, Sabunciyan S, Langmead B, McDonald OG *et al.* Increased methylation variation in epigenetic domains across cancer types. *Nat Genet* 2011;**43**:768–775.

135. Pérez RF, Tejedor JR, Bayón GF, Fernández AF, Fraga MF. Distinct chromatin signatures of DNA hypomethylation in aging and cancer. *Aging Cell* 2018;**17**:1–16.

136. Chatsirisupachai K, Lesluyes T, Paraoan L, Van Loo P, de Magalhães JP. An integrative analysis of the age-associated multi-omic landscape across cancers. *Nat Commun* 2021;**12**:1–17.

137. Levine ME, Hosgood HD, Chen B, Absher D, Assimes T, Horvath S. DNA methylation age of blood predicts future onset of lung cancer in the women's health initiative. *Aging* 2015;**7**:690–700.

138. Jones A, Teschendorff AE, Li Q, Hayward JD, Kannan A, Mould T *et al.* Role of DNA Methylation and Epigenetic Silencing of HAND2 in Endometrial Cancer Development. *PLoS Med* 2013;**10**:e1001551.

139. Lu AT, Hannon E, Levine ME, Crimmins EM, Lunnon K, Mill J *et al.* Genetic architecture of epigenetic and neuronal ageing rates in human brain regions. *Nat Commun* 2017;**8**:1–14.

140. Levine ME, Lu AT, Bennett DA, Horvath S. Epigenetic age of the pre-frontal cortex is associated with neuritic plaques, amyloid load, and Alzheimer's disease related cognitive functioning. *Aging* 2015;**7**:1198–1211.

141. Quach A, Levine ME, Tanaka T, Lu AT, Chen BH, Ritz B *et al.* Epigenetic clock analysis of diet, exercise, education, and lifestyle factors. *Aging* 2017;**9**:419–437.

142. Yang Z, Wong A, Kuh D, Paul DS, Rakyan VK, Leslie RD *et al.* Correlation of an epigenetic mitotic clock with cancer risk. *Genome Biol* 2016;**17**:1–18.

143. Teschendorff AE. A comparison of epigenetic mitotic-like clocks for cancer risk prediction. *Genome Med* 2020;**12**:1–17.

144. Breitling LP, Saum KU, Perna L, Schöttker B, Holleczek B, Brenner H. Frailty is associated with the epigenetic clock but not with telomere length in a German cohort. *Clin Epigenetics* 2016;**8**:1–8.

145. Vidal L, Lopez-Golan Y, Rego-Perez I, Horvath S, Blanco FJ, Riancho JA *et al.* Specific increase of methylation age in osteoarthritis cartilage. *Osteoarthritis Cartilage* 2016;**24**:S63–S534.

146. Vidal-Bralo L, Lopez-Golan Y, Mera-Varela A, Rego-Perez I, Horvath S, Zhang Y *et al.* Specific premature epigenetic aging of cartilage in osteoarthritis. *Aging* 2016;**8**:2222–2231.

147. Horvath S, Ritz BR. Increased epigenetic age and granulocyte counts in the blood of Parkinson's disease patients. *Aging* 2015;**7**:1130–1142.

148. Teschendorff AE, Jones A, Fiegl H, Sargent A, Zhuang JJ, Kitchener HC *et al.* Epigenetic variability in cells of normal cytology is associated with the risk of future morphological transformation. *Genome Med* 2012;**4**:1–14.

149. Landau DA, Clement K, Ziller MJ, Boyle P, Fan J, Gu H *et al.* Locally Disordered Methylation Forms the Basis of Intratumor Methylome Variation in Chronic Lymphocytic Leukemia. *Cancer Cell* 2014;**26**:813–825.

150. Li E, Zhang Y. DNA methylation in mammals. *Cold Spring Harb Perspect Biol* 2014;**6**.

151. Salameh Y, Bejaoui Y, El Hajj N. DNA Methylation Biomarkers in Aging and Age-Related Diseases. *Front Genet* 2020;**11**:1–11.

152. Zhang W, Qu J, Liu G-H, Belmonte JCI. The ageing epigenome and its rejuvenation. *Nat Rev Mol Cell Biol* 2020;**21**:137–150.

153. Deelen J, Beekman M, Capri M, Franceschi C, Slagboom PE. Identifying the genomic determinants of aging and longevity in human population studies: Progress and challenges. *BioEssays* 2013;**35**:386–396.

154. Capp JP, Thomas F. Tissue-disruption-induced cellular stochasticity and epigenetic drift: Common origins of aging and cancer? *BioEssays* 2021;**43**:1–10.

155. Rakyan VK, Beyan H, Down TA, Hawa MI, Maslau S, Aden D *et al.* Identification of type 1 Diabetes-associated DNA methylation variable positions that precede disease diagnosis. *PLoS Genet* 2011;**7**:1–9.

156. Houseman EA, Kile ML, Christiani DC, Ince TA, Kelsey KT, Marsit CJ. Reference-free deconvolution of DNA methylation data and mediation by cell composition effects. *BMC Bioinformatics* 2016;**17**:1–15.

157. Birney E, Smith GD, Greally JM. Epigenome-wide Association Studies and the Interpretation of Disease -Omics. *PLoS Genet* 2016;**12**:e1006105.

158. Jones PA. Functions of DNA methylation: Islands, start sites, gene bodies and beyond. *Nat Rev Genet* 2012;**13**:484–492.

159. Longo VD, Di Tano M, Mattson MP, Guidi N. Intermittent and periodic fasting, longevity and disease. *Nat Aging* 2021;**1**:47–59.

160. Flanagan EW, Most J, Mey JT, Redman LM. Calorie Restriction and Aging in Humans. *Annu Rev Nutr* 2020;**40**:105–133.

161. Gensous N, Franceschi C, Santoro A, Milazzo M, Garagnani P, Bacalini MG. The Impact of Caloric Restriction on the Epigenetic Signatures of Aging. *Int J Mol Sci* 2019;**20**:1–14.

162. Maegawa S, Lu Y, Tahara T, Lee JT, Madzo J, Liang S *et al.* Caloric restriction delays age-related methylation drift. *Nat Commun* 2017;**8**:1–11.

163. Hahn O, Grönke S, Stubbs TM, Ficz G, Hendrich O, Krueger F *et al.* Dietary restriction protects from age-associated DNA methylation and induces epigenetic reprogramming of lipid metabolism. *Genome Biol* 2017;**18**:1–18.

164. Kim CH, Lee EK, Choi YJ, An HJ, Jeong HO, Park D *et al.* Short-term calorie restriction ameliorates genomewide, age-related alterations in DNA methylation. *Aging Cell* 2016;**15**:1074–1081.

165. Hadad N, Unnikrishnan A, Jackson JA, Masser DR, Otalora L, Stanford DR *et al.* Caloric restriction mitigates age-associated hippocampal differential CG and non-CG methylation. *Neurobiol Aging* 2018;**67**:53–66.

166. Haghani A, Li CZ, Robeck TR, Zhang J, Lu AT, Ablaeva J *et al.* DNA Methylation Networks Underlying Mammalian Traits. *Science* 2023;**381**:647.

167. Waziry R, Ryan CP, Corcoran DL, Huffman KM, Kobor MS, Kothari M *et al.* Effect of long-term caloric restriction on DNA methylation measures of biological aging in healthy adults from the CALERIE trial. *Nat Aging* 2023 doi:10.1038/s43587-022-00357-y.

168. Partridge L, Fuentealba M, Kennedy BK. The quest to slow ageing through drug discovery. *Nat Rev Drug Discov* 2020;**19**:513–532.

169. Harrison DE, Strong R, Sharp ZD, Nelson JF, Astle CM, Flurkey K *et al.* Rapamycin fed late in life extends lifespan in genetically heterogeneous mice. *Nature* 2009;**460**:392–395.

170. Wilkinson JE, Burmeister L, Brooks S V., Chan CC, Friedline S, Harrison DE *et al.* Rapamycin slows aging in mice. *Aging Cell* 2012;**11**:675–682.

171. Spilman P, Podlutskaya N, Hart MJ, Debnath J, Gorostiza O, Bredesen D *et al.* Inhibition of mTOR by rapamycin abolishes cognitive deficits and reduces amyloid-β levels in a mouse model of alzheimer's disease. *PLoS One* 2010;**5**:1–8.

172. Popovich IG, Anisimov VN, Zabezhinski MA, Semenchenko A V., Tyndyk ML, Yurova MN *et al.* Lifespan extension and cancer prevention in HER-2/neu transgenic mice treated with low intermittent doses of rapamycin. *Cancer Biol Ther* 2014;**15**:586–592.

173. Reifsnyder PC, Flurkey K, Te A, Harrison DE. Rapamycin treatment benefits glucose metabolism in mouse models of type 2 diabetes. *Aging* 2016;**8**:3120–3130.

174. Neff F, Flores-Dominguez D, Ryan DP, Horsch M, Schröder S, Adler T *et al.* Rapamycin extends murine lifespan but has limited effects on aging. *Journal of Clinical Investigation* 2013;**123**:3272–3291.

175. Urfer SR, Kaeberlein TL, Mailheau S, Bergman PJ, Creevy KE, Promislow DEL *et al.* A randomized controlled trial to establish effects of short-term rapamycin treatment in 24 middle-aged companion dogs. *Geroscience* 2017;**39**:117–127.

176. Ross C, Salmon A, Strong R, Fernandez E, Javors M, Richardson A *et al.* Metabolic consequences of long-term rapamycin exposure on common marmoset monkeys (Callithrix jacchus). *Aging* 2015;**7**:964–973.

177. Mannick JB, Del Giudice G, Lattanzi M, Valiante NM, Praestgaard J, Huang B *et al.* mTOR inhibition improves immune function in the elderly. *Sci Transl Med* 2014;**6**:268ra179.

178. Horvath S, Lu AT, Cohen H, Raj K. Rapamycin retards epigenetic ageing of keratinocytes independently of its effects on replicative senescence, proliferation and differentiation. *Aging* 2019;**11**.

179. Horvath S, Zoller JA, Haghani A, Lu AT, Raj K, Jasinska AJ *et al.* DNA methylation age analysis of rapamycin in common marmosets. *Geroscience* 2021 doi:10.1007/s11357-021-00438-7.

180. Novelle MG, Ali A, Diéguez C, Bernier M, de Cabo R. Metformin: A hopeful promise in aging research. *Cold Spring Harb Perspect Med* 2016;**6**:1–12.

181. Fahy GM, Brooke RT, Watson JP, Good Z, Vasanawala SS, Maecker H *et al.* Reversal of epigenetic aging and immunosenescent trends in humans. *Aging Cell* 2019;**18**:1–12.

182. Li M, Duan L, Cai Y, Hao B, Chen J, Liu H. Effect of Metformin on the Epigenetic Age of Peripheral Blood in Patients With Diabetes Mellitus (DM). *Preprint* 2020;1–18.

183. Covarrubias AJ, Perrone R, Grozio A, Verdin E. NAD+ metabolism and its roles in cellular processes during ageing. *Nat Rev Mol Cell Biol* 2021;**22**:119–141.

184. Massudi H, Grant R, Braidy N, Guest J, Farnsworth B, Guillemin GJ. Age-associated changes in oxidative stress and NAD+ metabolism in human tissue. *PLoS One* 2012;**7**:1–9.

185. Zhu XH, Lu M, Lee BY, Ugurbil K, Chen W. In vivo NAD assay reveals the intracellular NAD contents and redox state in healthy human brain and their age dependences. *Proc Natl Acad Sci U S A* 2015;**112**:2876–2881.

186. Zhou CC, Yang X, Hua X, Liu J, Fan MB, Li GQ *et al.* Hepatic NAD+deficiency as a therapeutic target for non-alcoholic fatty liver disease in ageing. *Br J Pharmacol* 2016;2352–2368.

187. Clement J, Wong M, Poljak A, Sachdev P, Braidy N. The Plasma NAD + Metabolome Is Dysregulated in 'normal' Aging. *Rejuvenation Res* 2019;**22**:121–130.

188. Katsyuba E, Romani M, Hofer D, Auwerx J. NAD+ homeostasis in health and disease. *Nat Metab* 2020;**2**:9–31.

189. Zhang T, Kraus WL. SIRT1-dependent regulation of chromatin and transcription: Linking NAD+ metabolism and signaling to the control of cellular functions. *Biochim Biophys Acta Proteins Proteom* 2010;**1804**:1666–1675.

190. Wakeling LA, Ions LJ, Escolme SM, Cockell SJ, Su T, Dey M *et al.* SIRT1 affects DNA methylation of polycomb group protein target genes, a hotspot of the epigenetic shift observed in ageing. *Hum Genomics* 2015;**9**:14.

191. Salminen A, Kaarniranta K, Hiltunen M, Kauppinen A. Krebs cycle dysfunction shapes epigenetic landscape of chromatin: Novel insights into mitochondrial regulation of aging process. *Cell Signal* 2014;**26**:1598–1603.

192. Wang Y, Deng P, Liu Y, Wu Y, Chen Y, Guo Y *et al.* Alpha-ketoglutarate ameliorates age-related osteoporosis via regulating histone methylations. *Nat Commun* 2020;**11**:1–14.

193. Zhang Z, He C, Gao Y, Zhang L, Song Y, Zhu T *et al.* A-Ketoglutarate Delays Age-Related Fertility Decline in Mammals. *Aging Cell* 2021;**20**:1–15.

194. Asadi Shahmirzadi A, Edgar D, Liao CY, Hsu YM, Lucanic M, Asadi Shahmirzadi A *et al.* Alpha-Ketoglutarate, an Endogenous Metabolite, Extends Lifespan and Compresses Morbidity in Aging Mice. *Cell Metab* 2020;**32**:447–456.

195. Amenyah SD, Ward M, Strain JJ, McNulty H, Hughes CF, Dollin C *et al.* Nutritional epigenomics and age-related disease. *Curr Dev Nutr* 2020;**4**:1–16.

196. Demidenko O, Barardo D, Budovskii V, Finnemore R, Palmer FR, Kennedy BK *et al.* Rejuvant®, a potential life-extending compound formulation with alpha-ketoglutarate and vitamins, conferred an average 8 year reduction in biological aging, after an average of 7 months of use, in the TruAge DNA methylation test. *Aging* 2021;**13**:24485–24499.

197. Madeo F, Eisenberg T, Pietrocola F, Kroemer G. Spermidine in health and disease. *Science* 2018;**359**:1–10.

198. Soda K. Spermine and gene methylation: a mechanism of lifespan extension induced by polyamine-rich diet. *Amino Acids* 2020;**52**:213–224.

199. Eisenberg T, Abdellatif M, Schroeder S, Primessnig U, Stekovic S, Pendl T *et al.* Cardioprotection and lifespan extension by the natural polyamine spermidine. *Nat Med* 2016;**22**:1428–1438.

200. Yue F, Li W, Zou J, Jiang X, Xu G, Huang H *et al.* Spermidine prolongs lifespan and prevents liver fibrosis and hepatocellular carcinoma by activating MAP1S-mediated autophagy. *Cancer Res* 2017;**77**:2938–2951.

201. Kiechl S, Pechlaner R, Willeit P, Notdurfter M, Paulweber B, Willeit K *et al.* Higher spermidine intake is linked to lower mortality: A prospective population-based study. *American Journal of Clinical Nutrition* 2018;**108**:371–380.

202. Soda K, Kano Y, Chiba F, Koizumi K, Miyaki Y. Increased Polyamine Intake Inhibits Age-Associated Alteration in Global DNA Methylation and 1,2-Dimethylhydrazine-Induced Tumorigenesis. *PLoS One* 2013;**8**:1–7.

203. Fukui T, Soda K, Takao K, Rikiyama T. Extracellular spermine activates DNA methyltransferase 3A and 3B. *Int J Mol Sci* 2019;**20**:1–17.

204. Soda K. Polyamine metabolism and gene methylation in conjunction with one-carbon metabolism. *Int J Mol Sci* 2018;**19**.

205. Simpson DJ, Olova NN, Chandra T. Cellular reprogramming and epigenetic rejuvenation. *Clin Epigenetics* 2021;**13**:1–10.

206. Lu Y, Brommer B, Tian X, Krishnan A, Meer M, Wang C *et al.* Reprogramming to recover youthful epigenetic information and restore vision. *Nature* 2020;**588**:124–129.

207. Yang JH, Hayano M, Griffin PT, Amorim JA, Bonkowski MS, Apostolides JK *et al.* Loss of epigenetic information as a cause of mammalian aging. *Cell* 2023;**186**:305-326.e27.

208. Gill D, Parry A, Santos F, Okkenhaug H, Todd CD, Hernando-Herraez I *et al.* Multi-omic rejuvenation of human cells by maturation phase transient reprogramming. *Elife* 2022;**ee**:e71624.

209. Taylor D. Physical activity is medicine for older adults. *Postgrad Med J* 2014;**90**:26–32.

210. Voisin S, Seale K, Jacques M, Landen S, Harvey NR, Haupt LM *et al.* Exercise is associated with younger methylome and transcriptome profiles in human skeletal muscle. *Aging Cell* 2023 doi:10.1111/acel.13859.

211. Yan X, Eynon N, Papadimitriou ID, Kuang J, Munson F, Tirosh O *et al.* The gene SMART study: Method, study design, and preliminary findings. *BMC Genomics* 2017;**18**:15–28.

212. Bell JT, Tsai PC, Yang TP, Pidsley R, Nisbet J, Glass D *et al.* Epigenome-wide scans identify differentially methylated regions for age and age-related phenotypes in a healthy ageing population. *PLoS Genet* 2012;**8**.

213. Voisin S, Jacques M, Landen S, Harvey NR, Haupt LM, Griffiths LR *et al.* Meta-analysis of genome-wide DNA methylation and integrative OMICs in human skeletal muscle. *J Cachexia Sarcopenia Muscle* 2021;**12**:1064–1078.

214. Zhu T, Zheng SC, Paul DS, Horvath S, Teschendorff AE. Cell and tissue type independent age-associated DNA methylation changes are not rare but common. *Aging* 2018;**10**:3541–3557.

215. Thompson RF, Atzmon G, Gheorghe C, Liang HQ, Lowes C, Greally JM *et al.* Tissue-specific dysregulation of DNA methylation in aging. *Aging Cell* 2010;**9**:506–518.

216. Zhu T, Zheng SC, Paul DS, Horvath S, Teschendorff AE. Cell and tissue type independent age-associated DNA methylation changes are not rare but common. *Aging* 2018;**10**:3541–3557.

217. Lee TI, Jenner RG, Boyer LA, Guenther MG, Levine SS, Kumar RM *et al.* Control of Developmental Regulators by Polycomb in Human Embryonic Stem Cells. *Cell* 2006;**125**:301–313.

218. Bernstein BE, Mikkelsen TS, Xie X, Kamal M, Huebert DJ, Cuff J *et al.* A Bivalent Chromatin Structure Marks Key Developmental Genes in Embryonic Stem Cells. *Cell* 2006;**125**:315–326.

219. Yu M, Hazelton WD, Luebeck GE, Grady WM. Epigenetic aging: More than just a clock when it comes to cancer. *Cancer Res* 2020;**80**:367–374.

220. Luebeck GE, Hazelton WD, Curtius K, Maden SK, Yu M, Carter KT *et al.* Implications of epigenetic drift in colorectal neoplasia. *Cancer Res* 2019;**79**:495–504.

221. Horvath S, Raj K. DNA methylation-based biomarkers and the epigenetic clock theory of ageing. *Nat Rev Genet* 2018;**19**:371–384.

222. Knight AK, Craig JM, Theda C, Bækvad-Hansen M, Bybjerg-Grauholm J, Hansen CS *et al.* An epigenetic clock for gestational age at birth based on blood methylation data. *Genome Biol* 2016;**17**:1–11.

223. Kerepesi C, Zhang B, Lee S-G, Trapp A, Gladyshev VN. Epigenetic clocks reveal a rejuvenation event during embryogenesis followed by aging. *Sci Adv* 2021;**7**:4–11.

224. Raj K. The Epigenetic Clock and Aging. In: *Epigenetics of Aging and Longevity*. Elsevier Inc.; 2018. pp. 95–118.

225. Chiavellini P, Canatelli-Mallat M, Lehmann M, Gallardo MD, Herenu CB, Cordeiro JL *et al.* Aging and rejuvenation - a modular epigenome model. *Aging* 2021;**13**:4734–4746.

226. Raj K, Horvath S. Current perspectives on the cellular and molecular features of epigenetic ageing. *Exp Biol Med* 2020;**245**:1532–1542.

227. Ermolaeva M, Neri F, Ori A, Rudolph KL. Cellular and epigenetic drivers of stem cell ageing. *Nat Rev Mol Cell Biol* 2018;**19**:594–610.

228. Jonkman TH, Dekkers KF, Slieker RC, Grant CD, Ikram MA, van Greevenbroek MMJ *et al.* Functional genomics analysis identifies T and NK cell activation as a driver of epigenetic clock progression. *Genome Biol* 2022;**23**:24.

229. Dabin J, Fortuny A, Polo SE. Epigenome Maintenance in Response to DNA Damage. *Mol Cell* 2016;**62**:712–727.

230. Jaiswal S, Ebert BL. Clonal hematopoiesis in human aging and disease. *Science* 2019;**366**.

231. Klein CJ, Botuyan MV, Wu Y, Ward CJ, Nicholson GA, Hammans S *et al.* Mutations in DNMT1 cause hereditary sensory neuropathy with dementia and hearing loss. *Nat Genet* 2011;**43**:595–600.

232. Baets J, Duan X, Wu Y, Smith G, Seeley WW, Mademan I *et al.* Defects of mutant DNMT1 are linked to a spectrum of neurological disorders. *Brain* 2015;**138**:845–861.

233. Bell-Pedersen D, Cassone VM, Earnest DJ, Golden SS, Hardin PE, Thomas TL *et al.* Circadian Rhythms From Multiple Oscillators: Lessons From Diverse Organisms. *Nat Rev Drug Discov* 2005;**4**:1–13.

234. Acosta-Rodríguez VA, Rijo-Ferreira F, Green CB, Takahashi JS. Importance of circadian timing for aging and longevity. *Nat Commun* 2021;**12**:1–12.

235. Etchegaray JP, Mostoslavsky R. Interplay between Metabolism and Epigenetics: A Nuclear Adaptation to Environmental Changes. *Mol Cell* 2016;**62**:695–711.

236. Oh ES, Petronis A. Origins of human disease: the chrono-epigenetic perspective. *Nat Rev Genet* 2021;**22**:533–546.

237. Azzi A, Dallmann R, Casserly A, Rehrauer H, Patrignani A, Maier B *et al.* Circadian behavior is light-reprogrammed by plastic DNA methylation. *Nat Neurosci* 2014;**17**:377–382.

238. Oh G, Koncevičius K, Ebrahimi S, Carlucci M, Groot DE, Nair A *et al.* Circadian oscillations of cytosine modification in humans contribute to epigenetic variability, aging, and complex disease. *Genome Biol* 2019;**20**:1–14.

239. Oh G, Ebrahimi S, Carlucci M, Zhang A, Nair A, Groot DE *et al.* Cytosine modifications exhibit circadian oscillations that are involved in epigenetic diversity and aging. *Nat Commun* 2018;**9**:1–11.

240. Reinke H, Asher G. Crosstalk between metabolism and circadian clocks. *Nat Rev Mol Cell Biol* 2019;**20**:227–241.

241. Takahashi JS. Transcriptional architecture of the mammalian circadian clock. *Nat Rev Genet* 2017;**18**:164–179.

242. Yeung J, Naef F. Rhythms of the Genome: Circadian Dynamics from Chromatin Topology, Tissue-Specific Gene Expression, to Behavior. *Trends in Genetics* 2018;**34**:915–926.

243. Mure LS, Le HD, Benegiamo G, Chang MW, Rios L, Jillani N *et al.* Diurnal transcriptome atlas of a primate across major neural and peripheral tissues. *Science* 2018;**359**:1–9.

244. Ruben MD, Wu G, Smith DF, Schmidt RE, Francey LJ, Lee YY *et al.* A database of tissue-specific rhythmically expressed human genes has potential applications in circadian medicine. *Sci Transl Med* 2018;**10**:1–8.

245. López-Otín C, Galluzzi L, Freije JMP, Madeo F, Kroemer G. Metabolic Control of Longevity. *Cell* 2016;**166**:802–821.

246. Zwighaft Z, Aviram R, Shalev M, Rousso-Noori L, Kraut-Cohen J, Golik M *et al.* Circadian Clock Control by Polyamine Levels through a Mechanism that Declines with Age. *Cell Metab* 2015;**22**:874–885.

247. Levine DC, Hong H, Weidemann BJ, Ramsey KM, Affinati AH, Schmidt MS *et al.* NAD+ Controls Circadian Reprogramming through PER2 Nuclear Translocation to Counter Aging. *Mol Cell* 2020;**78**:835-849.e7.

248. Asher G, Reinke H, Altmeyer M, Gutierrez-Arcelus M, Hottiger MO, Schibler U. Poly(ADP-Ribose) Polymerase 1 Participates in the Phase Entrainment of Circadian Clocks to Feeding. *Cell* 2010;**142**:943–953.

249. Kumar V, Takahashi JS. PARP around the Clock. *Cell* 2010;**142**:841–843.

250. Imai S ichiro, Guarente L. NAD+ and sirtuins in aging and disease. *Trends Cell Biol* 2014;**24**:464–471.

251. Schumacher B, Pothof J, Vijg J, Hoeijmakers JHJ. The central role of DNA damage in the ageing process. *Nature* 2021;**592**:695–703.

252. Ciccarone F, Zampieri M, Caiafa P. PARP1 orchestrates epigenetic events setting up chromatin domains. *Semin Cell Dev Biol* 2017;**63**:123–134.

253. Oberdoerffer P, Michan S, McVay M, Mostoslavsky R, Vann J, Park SK *et al.* SIRT1 Redistribution on Chromatin Promotes Genomic Stability but Alters Gene Expression during Aging. *Cell* 2008;**135**:907–918.

254. Shimizu I, Yoshida Y, Suda M, Minamino T. DNA damage response and metabolic disease. *Cell Metab* 2014;**20**:967–977.

255. Jing H, Lin H. Sirtuins in epigenetic regulation. *Chem Rev* 2015;**115**:2350–2375.

256. O'Hagan HM, Mohammad HP, Baylin SB. Double strand breaks can initiate gene silencing and SIRT1-dependent onset of DNA methylation in an exogenous promoter CpG island. *PLoS Genet* 2008;**4**.

257. O'Hagan HM, Wang W, Sen S, DeStefano Shields C, Lee SS, Zhang YW *et al.* Oxidative Damage Targets Complexes Containing DNA Methyltransferases, SIRT1, and Polycomb Members to Promoter CpG Islands. *Cancer Cell* 2011;**20**:606–619.

258. O'Hagan HM. Chromatin modifications during repair of environmental exposure-induced DNA damage: A potential mechanism for stable epigenetic alterations. *Environ Mol Mutagen* 2014;**55**:278–291.

259. Ding N, Bonham EM, Hannon BE, Amick TR, Baylin SB, O'Hagan HM. Mismatch repair proteins recruit DNA methyltransferase 1 to sites of oxidative DNA damage. *J Mol Cell Biol* 2016;**8**:244–254.

260. Jung M, Pfeifer GP. Aging and DNA methylation. *BMC Biol* 2015;**13**:1–8.

261. Khokhlova E, Fesenko ZS, Sopova J V, Leonova EI. Features of DNA Repair in the Early Stages of Mammalian Embryonic Development. *Genes (Basel)* 2020;**11**:1–11.

262. Vilenchik MM, Knudson AG. Endogenous DNA double-strand breaks: Production, fidelity of repair, and induction of cancer. *Proc Natl Acad Sci U S A* 2003;**100**:12871–12876.

263. Liu Z, Leung D, Thrush K, Zhao W, Ratliff S, Tanaka T *et al.* Underlying features of epigenetic aging clocks in vivo and in vitro. *Aging Cell* 2020;**19**:1–11.

264. Christiansen C, Castillo-Fernandez JE, Domingo-Relloso A, Zhao W, El-Sayed Moustafa JS, Tsai PC *et al.* Novel DNA methylation signatures of tobacco smoking with trans-ethnic effects. *Clin Epigenetics* 2021;**13**:1–13.

265. Cohen AA. Aging across the tree of life: The importance of a comparative perspective for the use of animal models in aging. *Biochim Biophys Acta Mol Basis Dis* 2018;**1864**:2680–2689.

266. Horvath S, Singh K, Raj K, Khairnar S, Sanghavi A, Shrivastava A *et al.* Reversing age: Dual species measurement of epigenetic age with a single clock. *bioRxiv* 2020 doi:10.1101/2020.05.07.082917.

267. Raj K, Szladovits B, Haghani A, Zoller JA, Li CZ, Black P *et al.* Epigenetic clock and methylation studies in cats. *Geroscience* 2021 doi:10.1007/s11357-021-00445-8.

268. Sugrue VJ, Zoller JA, Narayan P, Lu AT, Ortega-Recalde OJ, Grant MJ *et al.* Castration delays epigenetic aging and feminizes dna methylation at androgen-regulated loci. *Elife* 2021;**10**:1–20.

269. Schachtschneider KM, Schook LB, Meudt JJ, Shanmuganayagam D, Zoller JA, Haghani A *et al.* Epigenetic clock and DNA methylation analysis of porcine models of aging and obesity. *Geroscience* 2021 doi:10.1007/s11357-021-00439-6.

270. Horvath S, Zoller JA, Haghani A, Jasinska AJ, Raj K, Breeze CE *et al.* Epigenetic clock and methylation studies in the rhesus macaque. *Geroscience* 2021 doi:10.1007/s11357-021-00429-8.

271. Arneson A, Haghani A, Thompson MJ, Pellegrini M, Kwon S Bin, Vu H *et al.* A mammalian methylation array for profiling methylation levels at conserved sequences. *Nat Commun* 2022;**13**.

272. Lenart P, Kuruczova D, Joshi PK, Bienertová-Vašků J. Male mortality rates mirror mortality rates of older females. *Sci Rep* 2019;**9**:1–9.

273. Horvath S, Gurven M, Levine ME, Trumble BC, Kaplan H, Allayee H *et al.* An epigenetic clock analysis of race/ethnicity, sex, and coronary heart disease. *Genome Biol* 2016;**17**:1–22.

274. Link C, Jaffe AE, Irizarry RA. Accounting for cellular heterogeneity is critical in epigenome-wide association studies. *Genome Biol* 2014;**15**:1–9.

275. Huang Y, Pastor WA, Shen Y, Tahiliani M, Liu DR, Rao A. The behaviour of 5-hydroxymethylcytosine in bisulfite sequencing. *PLoS One* 2010;**5**:1–9.

276. Jin SG, Kadam S, Pfeifer GP. Examination of the specificity of DNA methylation profiling techniques towards 5-methylcytosine and 5-hydroxymethylcytosine. *Nucleic Acids Res* 2010;**38**:1–7.

277. Spiers H, Hannon E, Schalkwyk LC, Bray NJ, Mill J. 5-Hydroxymethylcytosine Is Highly Dynamic Across Human Fetal Brain Development. *BMC Genomics* 2017;**18**:1–14.

278. Willer CJ, Li Y, Abecasis GR. METAL: Fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* 2010;**26**:2190–2191.

279. Gladyshev VN. The Ground Zero of Organismal Life and Aging. *Trends Mol Med* 2021;**27**:11–19.

280. van Iterson M, van Zwet EW, Heijmans BT, 't Hoen PAC, van Meurs J, Jansen R *et al.* Controlling bias and inflation in epigenome- and transcriptome-wide association studies using the empirical null distribution. *Genome Biol* 2017;**18**:1–13.

281. Sillanpää E, Heikkinen A, Kankaanpää A, Paavilainen A, Kujala UM, Tammelin TH *et al.* Blood and skeletal muscle ageing determined by epigenetic clocks and their associations with physical activity and functioning. *Clin Epigenetics* 2021;**13**:110.

282. Morris TJ, Butcher LM, Feber A, Teschendorff AE, Chakravarthy AR, Wojdacz TK *et al.* ChAMP: 450k Chip Analysis Methylation Pipeline. *Bioinformatics* 2014;**30**:428–430.

283. Pidsley R, Zotenko E, Peters TJ, Lawrence MG, Risbridger GP, Molloy P *et al.* Critical evaluation of the Illumina MethylationEPIC BeadChip microarray for whole-genome DNA methylation profiling. *Genome Biol* 2016;**17**:1–17.

284. Leek JT, Johnson WE, Parker HS, Jaffe AE, Storey JD. The SVA package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics* 2012;**28**:882–883.

285. Du P, Zhang X, Huang CC, Jafari N, Kibbe WA, Hou L *et al.* Comparison of Beta-value and M-value methods for quantifying methylation levels by microarray analysis. *BMC Bioinformatics* 2010;**11**.

286. 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**.

287. Li Z, Zhang Z, Ren Y, Wang Y, Fang J, Yue H *et al.* Aging and age-related diseases: from mechanisms to therapeutic strategies. Biogerontology. 2021;**22**:165–187.

288. Kehler DS. Age-related disease burden as a measure of population ageing. Lancet Public Health. 2019;**4**:e123–e124.

289. Viechtbauer W. Conducting meta-analyses in R with the metafor package. Journal of Statistical Software. *J Stat Softw* 2010;**36**:1–48.

290. Roadmap Epigenomics Consortium, Kundaje A, Meuleman W, Ernst J, Bilenky M, Yen A *et al.* Integrative analysis of 111 reference human epigenomes. *Nature* 2015;**518**:317–329.

291. Phipson B, Maksimovic J, Oshlack A. MissMethyl: An R package for analyzing data from Illumina's HumanMethylation450 platform. *Bioinformatics* 2016;**32**:286–288.

292. Zhou W, Laird PW, Shen H. Comprehensive characterization, annotation and innovative use of Infinium DNA methylation BeadChip probes. *Nucleic Acids Res* 2017;**45**:e22.

293. Benjamin DJ, Berger JO, Johannesson M, Nosek BA, Wagenmakers EJ, Berk R *et al.* Redefine statistical significance. *Nat Hum Behav* 2018;**2**:6–10.

294. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society B* 1995;**57**:289–300.

295. 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**.

296. Garagnani P, Bacalini MG, Pirazzini C, Gori D, Giuliani C, Mari D *et al.* Methylation of ELOVL2 gene as a new epigenetic marker of age. *Aging Cell* 2012;**11**:1132–1134.

297. Marttila S, Kananen L, Häyrynen S, Jylhävä J, Nevalainen T, Hervonen A *et al.* Ageing-associated changes in the human DNA methylome: Genomic locations and effects on gene expression. *BMC Genomics* 2015;**16**.

298. Florath I, Butterbach K, Müller H, Bewerunge-hudler M, Brenner H. Cross-sectional and longitudinal changes in DNA methylation with age: An epigenome-wide analysis revealing over 60 novel age-associated CpG sites. *Hum Mol Genet* 2014;**23**:1186–1201.

299. Tarkhov AE, Lindstrom-Vautrin T, Zhang S, Ying K, Moqri M, Zhang B *et al.* Nature of epigenetic aging from a single-cell perspective. doi:10.1101/2022.09.26.509592.

300. Xu Z, Taylor JA. Genome-wide age-related DNA methylation changes in blood and other tissues relate to histone modification, expression and cancer. *Carcinogenesis* 2014;**35**:356–364.

301. Duffield T, Csuka L, Akalan A, Vega Magdaleno G, Palmer D, Pedro De Magalhães J. Epigenetic fidelity in complex biological systems and implications for ageing. doi:10.1101/2023.04.29.538716.

302. Blanco E, González-Ramírez M, Alcaine-Colet A, Aranda S, di Croce L. The Bivalent Genome: Characterization, Structure, and Regulation. Trends in Genetics. 2020;**36**:118–131.

303. Izgi H, Han D, Isildak U, Huang S, Kocabiyik E, Khaitovich P *et al.* Inter-tissue convergence of gene expression during ageing suggests age-related loss of tissue and cellular identity. *Elife* 2022;**11**.

304. Urban LA, Trinh A, Pearlman E, Siryaporn A, Downing TL. The impact of age-related hypomethylated DNA on immune signaling upon cellular demise. Trends Immunol. 2021;**42**:464–468.

305. Ewald CY. The Matrisome during Aging and Longevity: A Systems-Level Approach toward Defining Matreotypes Promoting Healthy Aging. Gerontology. 2020;**66**:266–274.

306. Moldakozhayev A, Gladyshev VN. Metabolism, homeostasis, and aging. *Trends in Endocrinology & Metabolism* 2023 doi:10.1016/j.tem.2023.01.003.

307. Vaidya H, Jeong HS, Keith K, Maegawa S, Calendo G, Madzo J *et al.* DNA methylation entropy as a measure of stem cell replication and aging. *Genome Biol* 2023;**24**:27.

308. Brunet A, Goodell MA, Rando TA. Ageing and rejuvenation of tissue stem cells and their niches. Nat Rev Mol Cell Biol. 2023;**24**:45–62.

309. Sharpless NE, DePinho RA. How stem cells age and why this makes us grow old. Nat Rev Mol Cell Biol. 2007;**8**:703–713.

310. Qi L, Teschendorff AE. Cell-type heterogeneity: Why we should adjust for it in epigenome and biomarker studies. *Clin Epigenetics* 2022;**14**.

311. Ibáñez-Solé O, Ascensión AM, Araúzo-Bravo MJ, Izeta A. Lack of evidence for increased transcriptional noise in aged tissues. *Elife* 2022;**11**.

312. Mogilenko DA, Shchukina I, Artyomov MN. Immune ageing at single-cell resolution. Nat Rev Immunol. 2022;**22**:484–498.

313. Zheng SC, Breeze CE, Beck S, Teschendorff AE. Identification of differentially methylated cell types in epigenome-wide association studies. *Nat Methods* 2018;**15**:1059–1066.