RESEARCH

Open Access

A comprehensive map of the aging blood methylome in humans

Kirsten Seale[1], Andrew Teschendorff[2,3], Alexander P. Reiner[4], Sarah Voisin[1,5†] and Nir Eynon[5*†]

†Sarah Voisin and Nir Eynon are co-senior authors.

*Correspondence:
Nir.Eynon@monash.edu

[1] Institute for Health and Sport (iHeS), Victoria University, Footscray, VIC 3011, Australia
[2] CAS Key Lab of Computational Biology, Shanghai Institute of Nutrition and Health, Chinese Academy of Sciences, Shanghai 200031, China
[3] UCL Cancer Institute, University College London, London, UK
[4] Department of Epidemiology, University of Washington, Seattle, WA, USA
[5] Australian Regenerative Medicine Institute, Monash University, Clayton, VIC 3800, Australia

Abstract

Background: During aging, the human methylome undergoes both differential and variable shifts, accompanied by increased entropy. The distinction between variably methylated positions (VMPs) and differentially methylated positions (DMPs), their contribution to epigenetic age, and the role of cell type heterogeneity remain unclear.

Results: We conduct a comprehensive analysis of > 32,000 human blood methylomes from 56 datasets (age range = 6–101 years). We find a significant proportion of the blood methylome that is differentially methylated with age (48% DMPs; FDR < 0.005) and variably methylated with age (37% VMPs; FDR < 0.005), with considerable overlap between the two groups (59% of DMPs are VMPs). Bivalent and Polycomb regions become increasingly methylated and divergent between individuals, while quiescent regions lose methylation more uniformly. Both chronological and biological clocks, but not pace-of-aging clocks, show a strong enrichment for CpGs undergoing both mean and variance changes during aging. The accumulation of DMPs shifting towards a methylation fraction of 50% drives the increase in entropy, smoothening the epigenetic landscape. However, approximately a quarter of DMPs exhibit anti-entropic effects, opposing this direction of change. While changes in cell type composition minimally affect DMPs, VMPs and entropy measurements are moderately sensitive to such alterations.

Conclusion: This study represents the largest investigation to date of genome-wide DNA methylation changes and aging in a single tissue, providing valuable insights into primary molecular changes relevant to chronological and biological aging.

Background

All humans experience similar aging symptoms with chronological time; however, 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]. Aging is initiated at the basic level of biological organization and is hypothesized to be underpinned by 12 interconnected hallmarks [2], including alterations to the epigenome, a focus of this research [2]. Specifically,

the age-associated changes in DNA methylation (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]. 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 both a DMP and a VMP), both features may contribute to age-associated DNAm changes, but in fundamentally different ways. Insights into epigenetic aging can also be drawn from quantifying the changes at the whole methylome level using single measurements of entropy, a probability formula that estimates the amount of information in a set of CpGs averaged across a population of cells [3]. In blood, an increase in entropy over time implies increasing age-related "methylation disorder," implying the methylome loses information with age [3]. Entropy can summarize all the age-related changes in DNAm using a single value, for a sample at a particular age, making it a highly useful in capturing the entirety of the age-related changes in the methylome. Previous literature has shown that DMPs shifting towards a methylation fraction of 50% drive entropy [4, 5], but there are also subsets of DMPs that trend away from the mean and may oppose or counteract the effect [6].

The vast majority of studies have solely investigated DMPs, including epigenetic clocks, which may favor DMPs for accuracy of prediction. However, focusing only on patterns of DMPs is limiting when trying to understand aspects of biological aging, particularly when making sense of why individuals of the same age display vastly different aging rates. Yet our knowledge of the extent to which the methylome is differentially and variably methylated with age is limited, since much of our understanding to date has been drawn from smaller, isolated datasets, such as Slieker et al. (~3000 samples) and Hannum et al. (~650 samples) [4, 7]. Another complexity in human studies is that cohorts are also highly heterogenous in their characteristics (e.g., sex distribution, disease status, ethnicity, age), making it challenging to generalize findings drawn from individual cohorts. Furthermore, detecting DMPs and especially VMPs requires a large sample size and a broad age range, so it is possible that important epigenetic aging patterns have been missed in previous studies due to insufficient statistical power. Moreover, no studies to date have investigated the contribution of VMPs to changes in entropy.

To address these gaps, we performed a large-scale epigenome-wide association study (EWAS) meta-analysis of age in whole blood. Leveraging the statistical power from 56 whole blood datasets of > 32,000 samples, we quantified the age-associated DNAm changes, distinguishing between shared changes (DMPs) and divergent changes (VMPs) over the life course. In addition, we sought to understand how entropy changes and how DMPs and VMPs contribute to these entropy dynamics, including those linear changes that trend towards and away from the mean. We also explored the effect of cellular heterogeneity on these age-associated signatures, since bulk tissue analyses reflect the aggregate cell type DNAm changes as well as age-related changes in cellular composition. We performed cell type correction in addition to an unadjusted analysis [8].

With unprecedented statistical power and a comprehensive investigation of epigenetic signatures of aging in blood, our study shows the sheer scale of differential and variable methylation in human aging, with important implications for chronological and biological aging. Such insights will serve as a foundation for studying the effects of lifestyle, dietary, or pharmaceutical interventions on aging signatures, an approach that was recently shown by our group to successfully capture the rejuvenating effects of exercise training on age-associated DMPs in muscle [9].

Results

Methodology overview

Briefly, the first step in the methodology was to gather existing DNAm datasets from public open access and controlled access data repositories, to assemble an exhaustive database of DNAm profiles in blood (Fig. 1, Additional file 1: Fig. S1). We collected 56 whole blood datasets with a combined sample size of 32,136 samples (Additional file 1: Fig. S1, Additional file 2: Table S1) [10–63]. This large database of human methylomes spanned a broad age range (6–101 years) and laid a solid foundation to quantify DMPs, VMPs, and entropy in each dataset.

Our analysis began with identifying the linear age-related DNAm changes at individual CpGs (i.e., DMPs and VMPs). To identify DMPs, we completed an independent EWAS by fitting a linear model for each dataset, regressing DNAm against age and other covariates that are known to modulate DNAm levels (i.e., sex, ethnicity, batch, body mass index (BMI)) (Additional file 2: Table S1). We then extracted the summary statistics (i.e., the age-related change in DNAm level) from each EWAS and conducted an inverse variance-based fixed-effects meta-analysis of differential methylation andage. To
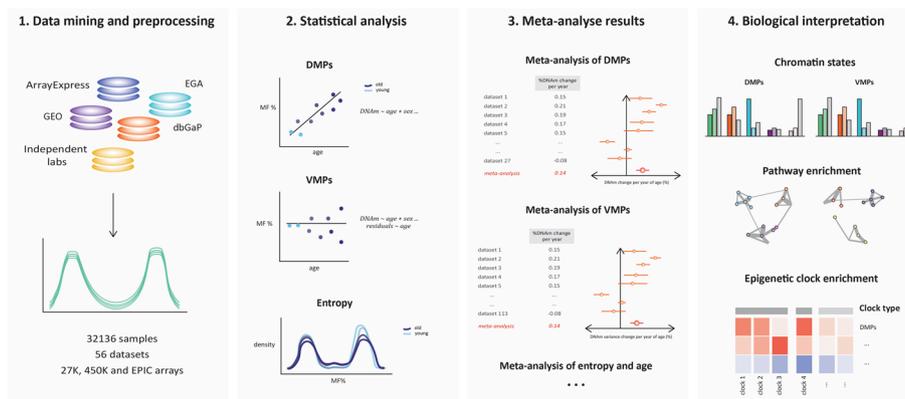

Fig. 1 Overview of methodology. Raw DNAm profiles from the 27 K, 450 K, and EPIC Illumina Array platforms were sourced from open-access and controlled access databases, including ArrayExpress, GEO, EGA, dbGaP, and independent labs. The DNAm profiles of 32,136 samples were collected and pre-processed, for both males and females, from 56 datasets in whole blood (first panel). 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-analyzed across datasets to identify DMPs and VMPs and to determine whether entropy was significantly associated with age. We performed a meta-analysis of age for DMPs, VMPs, and entropy, by pooling the results from the independent EWAS (third panel). We then performed a functional analysis of the DMPs and VMPs to interpret the findings, including pathway analysis, and enrichment in chromatin states (fourth panel)

Seale *et al. Genome Biology*   (2024) 25:240

Page 4 of 24

identify VMPs, we followed a similar approach, using the Breusch-Pagan test for heteroscedasticity in each independent dataset, which models the change in DNAm variance as a function of age. The summary statistics extracted from the independent EWAS (i.e., the age-related change in DNAm variance) were then pooled using a sample size-based fixed-effects meta-analysis of variable methylation and age. We then compared age-related CpGs that were only DMPs (homoscedastic DMPs), those that were only VMPs (constant VMPs), and those that were both DMPs and VMPs (DMPs-VMPs). Finally, we compared DMPs that carried different degrees of information, as defined by Shannon entropy: we compared DMPs that trend towards the mean (entropic DMPs) with those that trend away from the mean towards fully methylated and unmethylated states (anti-entropic DMPs).

Finally, we performed a comprehensive entropy analysis by looking at genome-wide Shannon entropy. 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 Shannon entropy in each independent cohort; then, we pooled these effect sizes across the different cohorts using a fixed-effects meta-analysis to obtain an overall meta-analysis effect size of change in entropy per decade of age (Additional file 2: Table S2). We also calculated entropy on the age-related CpGs (i.e., the complete list of DMPs and VMPs) and the remaining non-age-related CpGs, and meta-analyzed the results to compare the change in entropy with age at age-related CpGs vs non-age-related CpGs, evaluating whether it is these age-related shifts driving changes in entropy. Then, we took a more granular approach and investigated whether different classes of DMPs and VMPs contribute differently to entropy measurements (homo-DMPs, constant VMPs, DMPs-VMPs, or entropic vs anti-entropic DMPs). This allowed us to confirm whether it is the differential or variable shifts in DNAm that increase entropy, or both.

Throughout the pipeline, we compared these different classes of age-related changes in terms of genomic location (chromatin states profiled in PBMCs [64], annotated thank to a comprehensive annotation of the Illumina Methylation arrays [65]), biological pathways (gene ontology (GO) terms, human phenotype ontology (HPO) terms, canonical pathways (CP), expression signatures of genetic and chemical perturbations (CGP), and immunologic signatures (C7)), and overrepresentation in various epigenetic clocks (chronological age: Horvath's pan-tissue clock [66], Hannum's clock [4], the blood clock developed by Zhang et al. [67], the centenarian clock [68], and the mammalian universal clock [69]; biological age: PhenoAge [70]; pace of aging: DunedinPoAm [71], Dunedin-PACE [72]). We also explored the effect of cellular heterogeneity on these age-associated signatures. CpGs that determine cell identity are typically lowly methylated in a given cell type, while being highly methylated in other cell types, and the overall methylation fraction in bulk tissue at those cell-type-specific CpGs would be highly sensitive to changes in the relative proportions of different cell types. Aging is associated with an increase in monocytes, neutrophils, basophils, NK cells, CD4 + and CD8 + T memory cells, with a concomitant decrease in naïve B cells, T-regulatory cells, CD4 + and CD8 + naïve T cells [73]. We deconvoluted the proportions for granulocytes, monocytes, natural killer cells (NK), CD4 + T cells, CD8 + T cells, and B cells for each sample using a reference-based method [74], and repeated all the above-mentioned analyses after adjusting the linear model for blood cell type proportions.

Seale *et al. Genome Biology*   (2024) 25:240

Page 5 of 24

Aging is associated with widespread changes in DNAm levels and increases in DNAm variance in blood

With the unprecedented statistical power granted by > 32,000 samples from 56 datasets, we found that nearly half of all tested sites (333,300 CpGs) were DMPs (48%) at a stringent FDR < 0.005. Two-thirds of DMPs (66%) decrease in DNAm levels ("hypoDMPs"), while the remaining third increase in DNAm levels with age ("hyperDMPs") (Fig. 2A). HyperDMPs increase by an average of 0.027% methylation fraction per year of age, noting a maximum increase of 0.46% per year of age for cg26079664, and hypoDMPs decrease by an average of − 0.034% per year of age, with the maximum decrease of − 0.55% per year of age for cg10501210. Our meta-analysis identified DMPs that were highly consistent across datasets. For example, cg16867657, which is in the promoter of *ELOVL2* and has been associated with aging in a plethora of studies [75–78], was estimated to gain 0.45% DNAm per year of age across the different datasets (Fig. 2B).

There was an inverse correlation between the overall methylation fraction of a CpG and the direction of change during aging: 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 (Additional file 1: Fig. S2). In contrast, DMPs with intermediate DNAm levels trend equally frequently towards
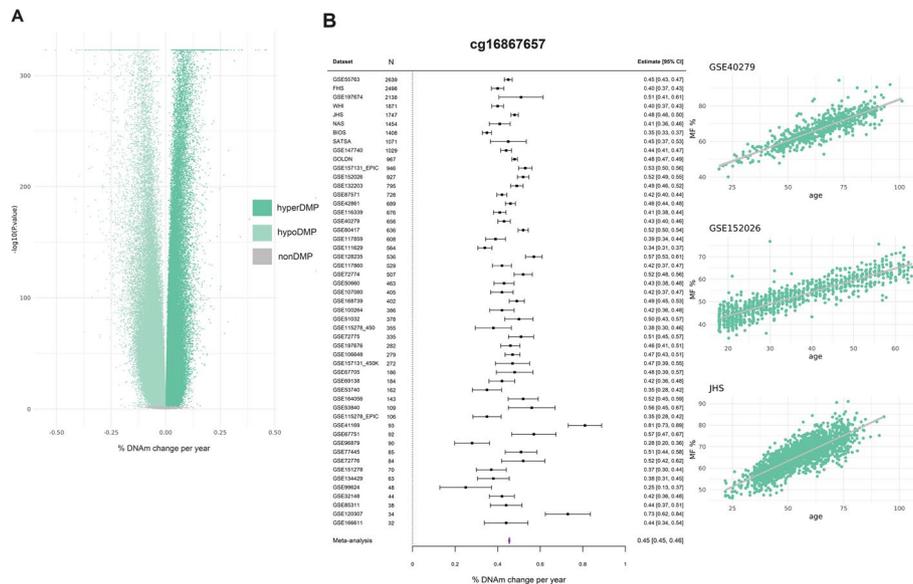

Fig. 2 Meta-analyses of EWAS in blood to identify DMPs. **A** A volcano plot displaying the meta-analysis effect size (x-axis) and significance (y-axis) for the 696,228 tested CpGs in the differential methylation meta-analysis of age in blood. The strongest associations have the smallest *p* values and will be the highest points on the plot. Hypomethylated (hypoDMP), hypermethylated (hyperDMP), and non-DMPs are represented by the colors in the legend. DMPs are classified at a false discovery rate (FDR) < 0.005. The "ceiling" of extremely significant *p* values is an artifact of the software that cannot handle numbers smaller than 2.2×10^{-16}. **B** A forest plot of a highly significant CpG (cg16867657), FDR < 2.2×10^{-16}. 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 (0.45% change per year of age) 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

Seale *et al. Genome Biology*   (2024) 25:240

Page 6 of 24

high and low DNAm levels. For example, in the BIOS dataset ($n = 1408$) [20], 21% of DMPs were considered to have an "intermediate" methylation level, of which 40% gained methylation with age, and 60% lost methylation with age (Additional file 1: Fig. S2).

HypoDMPs were overrepresented in quiescent chromatin regions and those weakly repressed by Polycomb complexes, while hyperDMPs were overrepresented in bivalent promoters and enhancers as well as regions repressed by Polycomb complexes (χ^2 test p value $< 2.2e^{-16}$) (Additional file 1: Fig. S3A). Despite being located in distinct chromatin states, hypoDMPs and hyperDMPs were found in similar genes (Fisher's exact test p value $< 2.2e^{-16}$) (Additional file 1: Fig. S3B), for example, signal transduction and signaling (GO), developmental conditions (HPO), and naïve to memory T-cell (C7, immunologic gene set). With the exception of the universal pan-mammalian clock, all chronological and biological clocks were enriched for hyperDMPs, and all but the Hannum clock for hypoDMPs, with no difference in enrichment for these two classes of DMPs in biological vs chronological clocks (Fisher's exact test FDR < 0.005, Additional file 1: Fig. S3C). Pace of aging clocks did not show any enrichment for DMPs (Fisher's exact test FDR > 0.005, Additional file 1: Fig. S3C).

These results remain largely unchanged when the meta-analysis was adjusted for cellular heterogeneity (Pearson's correlation of meta-analyses effect sizes $= 0.94$, p value $< 2.2e^{-16}$) (Additional file 1: Fig. S4A).

We then meta-analyzed the same 56 whole blood datasets to identify changes in methylation variability (VMPs) during aging. We identified 243,958 VMPs (37% of tested CpGs) at FDR < 0.005, nearly all of which increased in variance (99% of VMPs). 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 (Fig. 3A). There was a large overlap between DMPs and VMPs (i.e., a CpG site whose average DNAm level changed during aging was also more likely to see its variance increase with age; Fischer's exact test p value $< 2.2 \times 10^{-16}$). We identified 196,192 DMPs-VMPs, 137,108 homoscedastic DMPs (i.e., DMPs only), and 47,766 constant VMPs (i.e., VMPs only) (Fig. 3B). Among the DMP-VMPs, 73,357 (37%) increased in both average methylation and variance, and 122,835 (63%) decreased in average methylation but increased in variance.

We compared the distributions of homoscedastic DMPs, constant VMPs, and DMPs-VMPs in different chromatin states profiled in PBMCs [64]. Constant VMPs were enriched in enhancers and Polycomb regions, DMPs-VMPs in bivalent promoters and enhancers as well as Polycomb regions, and homoscedastic DMPs were in quiescent regions (χ^2 test p value $< 2.2e^{-16}$) (Additional file 1: Fig. S5A). The three classes of age-related changes were related to a plethora of pathways, very similar to hypo and hyperDMPs (Additional file 1: Fig. S5B). All chronological and biological clocks were strongly enriched for DMPs-VMPs, while pace of aging clocks was not enriched for any kind of age-related CpGs (Additional file 1: Fig. S5C). Homoscedastic DMPs were overrepresented in the pan-tissue clock but depleted in Zhang et al.'s clock. Constant VMPs were depleted in two chronological clocks but were overrepresented in the PhenoAge.

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
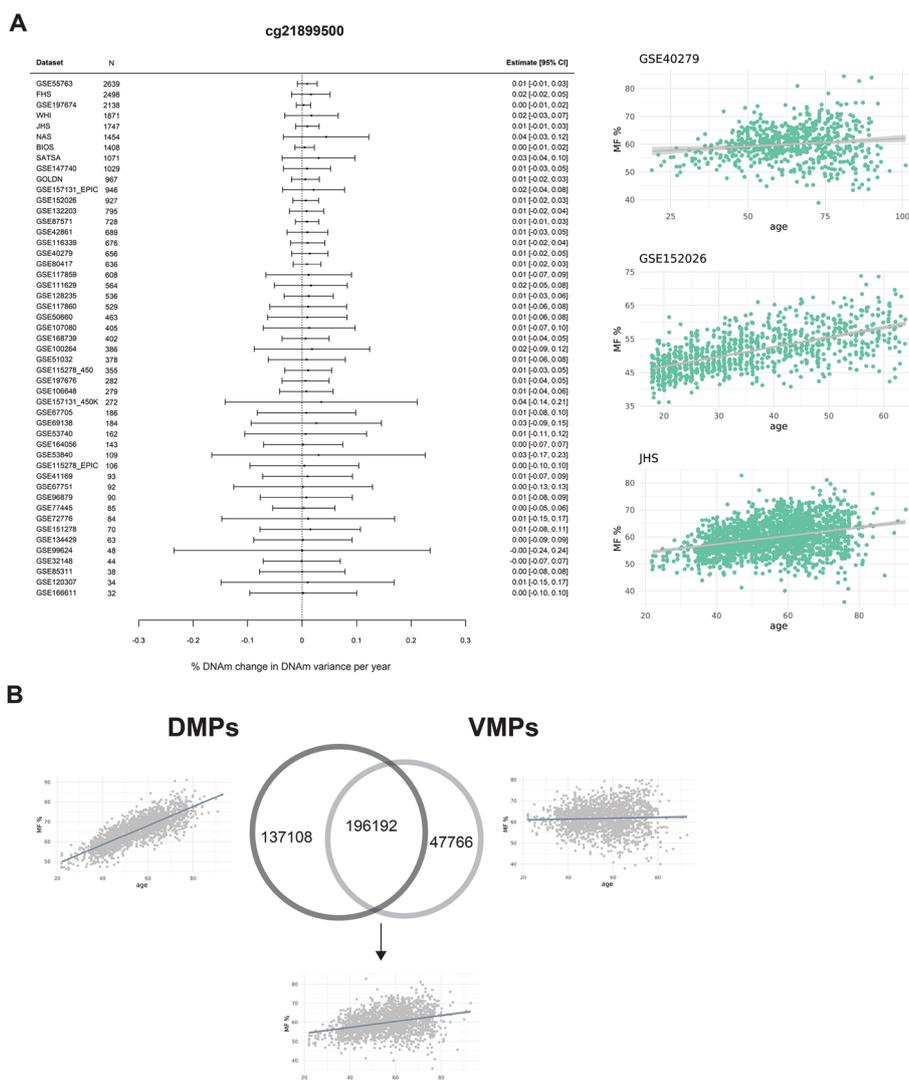
Seale *et al. Genome Biology*   (2024) 25:240

Page 7 of 24

Fig. 3 Meta-analyses of EWAS in blood to identify VMPs. **A** A forest plot of the top VMP (cg21899500), FDR $< 2.2 \times 10^{-16}$, that is hypermethylated with age and increases in variance with age. 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 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. **B** Venn diagram of the overlap between DMPs and VMPs. The numbers in each circle represent the number of CpGs that are DMPs only (left), both DMPs and VMPs (middle), and VMPs only (right). The arrows point to an example of each category of age-related CpG displayed as a methylation plot, with age on the x-axis and methylation fraction (MF) as a percentage on the y-axis. Each point represents a single sample

correlation of meta-analyses Z score $= 0.93$, p value $< 2.2\mathrm{e}^{-16}$) (Additional file 1: Fig. S4B). 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 5913 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.

Entropy increases in the aging blood methylome, driven by the cumulative changes in differential but not variable methylation at entropic CpGs

We determined whether the aging blood methylome increases in entropy ("chaos") with age, and what type of epigenetic changes underpin this phenomenon. Entropy captures the amount of information encoded by the epigenome: if a CpG is highly ($\sim 100\%$) or lowly ($\sim 0\%$) methylated, this implies that said CpGs is highly "predictable" over all cells in a given sample; conversely, if a CpG has a methylation fraction closer to 50%, it is deemed "unpredictable" across cells within a sample. 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% [3]. In these two opposite scenarios, the methylome of a cell is either entirely predictable, or entirely unpredictable.

When taking all CpGs into account (both age- and non-age-related CpGs), we observed a very small but significant increase in entropy of 0.0005 per decade of age (p value < 0.0001), with substantial heterogeneity between cohorts ($I^2 = 88\%$) (Additional file 1: Fig. S6).

As an increase in entropy with age reflects a drift towards a methylation fraction of 50% over multiple CpGs, we hypothesized that the increase in entropy would be driven by age-related CpGs (DMPs and/or VMPs). We re-calculated entropy in each sample from each dataset, but only considering the methylation levels at age-related CpGs (Additional file 2: Table S2). As a "control," we also re-calculated entropy in each sample from each dataset, but only considering the methylation levels of non-age-related CpGs. 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, I^2 statistic 46%) (Fig. 4A). In contrast, age-related CpGs increase in entropy by 0.002 per decade of age (p value < 0.0001, I^2 statistic 85%) (Fig. 4A). 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 (Fig. 4B). This can be explained by the fact that there are more CpGs with intermediate methylation levels among age-related sites ($\sim 20\%$), than non-age-related sites ($\sim 1\%$).

To dissect the respective contributions of differential or variable methylation to changes in entropy, we calculated entropy for two datasets with a large sample size and broad age range (BIOS and FHS) [20, 23] (Additional file 2: Table S1) on homoscedastic DMPs, DMPs-VMPs, and constant VMPs (Additional file 1: Fig. S7A, Additional file 2: Table S3), and regressed entropy against age in each category. DMPs-VMPs display the largest significant increase in entropy during aging. We also observed that entropy significantly increases at homoscedastic DMPs, however, is lower both overall and at baseline than for DMPs-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 ($\sim 28\%$ of VMPs are intermediately methylated vs $\sim 20\%$ of DMPs). While the overall entropy
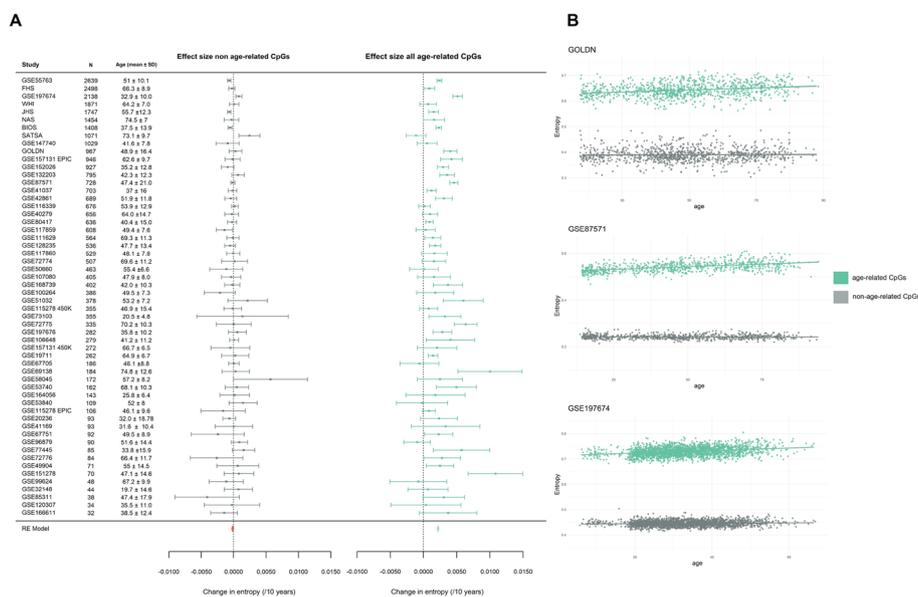
Fig. 4 Meta-analysis of entropy and age. **A** A forest plot of the two meta-analyses comparing the changes in entropy between the non-age-related CpGs (left) and the age-related CpGs (right) identified in blood. The meta-analyses effect sizes are represented by the orange and light green polygons. On the x-axes is the change in entropy per decade of age (left: − 0.0003 change in entropy per 10 years of age (*p* value < 0.0001); right: 0.002 per decade of age (*p* value < 0.0001)), and on the y-axes are the effect sizes and standard error measurements from the independent epigenome-wide association study (EWAS). The dataset name, sample size (*N*), and age ± sd (standard deviation) 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

at CpGs that are only VMPs (i.e., constant VMPs) is high, we observed a decrease in entropy at those sites that was significant in only one of the two examined datasets (Additional file 1: Fig. S7A, Additional file 2: Table S3), 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 [20], which has a large sample size and distribution of samples across a large age range (Additional file 1: Fig. S1, Additional file 2: Table S1). 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 (Fig. 5A). 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 (Fig. 5B). We validated these results in a second dataset, GSE128235 [13], and found highly concordant results in both the proportion of DMPs that converge to (72%) and diverge from (28%) the mean with age, but also the significant increase in entropy at the converging sites of 0.005 per decade of age (*p* value < $2.2e^{-16}$), and significant decrease in entropy of − 0.004 per decade of age at the diverging sites (*p* value < $2.2e^{-16}$).
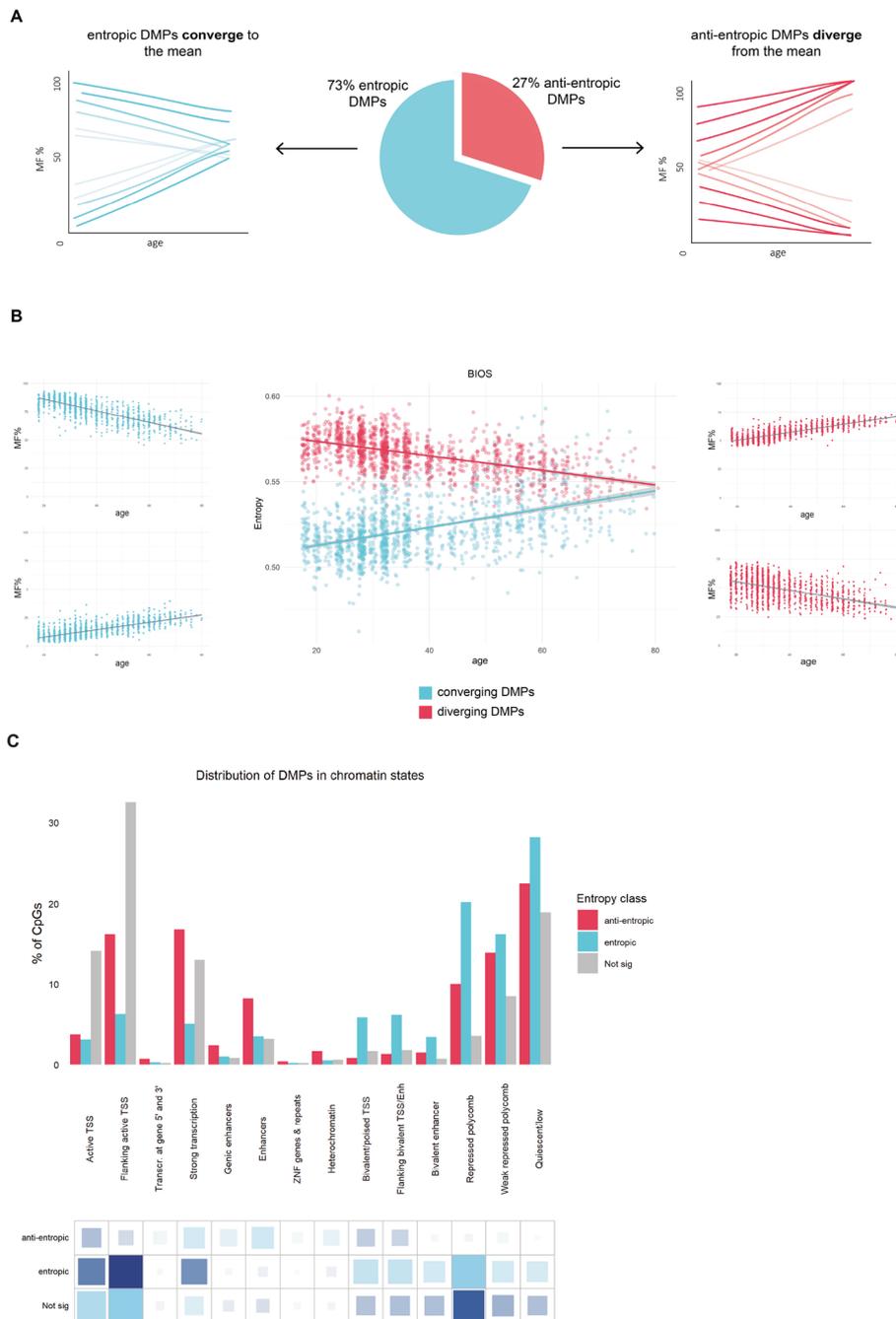
Fig. 5 Contribution of differential methylation to changes in entropy with age. **A** A pie chart of the proportion of DMPs that are entropic and converge to the mean (blue) and the proportion of DMPs that are anti-entropic and diverge from the mean (red) for the BIOS dataset in blood. The arrows point 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. The center plot displays the entropy (y-axis) value plotted against age for the converging (blue) and diverging (red) DMPs in the BIOS dataset. **C** Distribution of entropic (blue) and anti-entropic differentially methylated positions (DMPs) (red) and non-DMPs (gray) in chromatin states from peripheral blood mononuclear cells (PBMCs). The grids under the graph represent the residuals from the χ^2 tests, with the size of the blocks in the grid being proportional to the cell's contribution. Sky blue indicates overrepresentation or enrichment in the chromatin state, and navy represents underrepresentation or depletion in the chromatin state

We then investigated the distribution of the entropic DMPs, anti-entropic DMPs, and non-DMPs in chromatin states of PBMCs (Fig. 5C), noting that entropic DMPs are overrepresented in bivalent promoters and enhancers, regions bound by Polycomb proteins, and quiescent states (χ^2 test p value $< 2.2e^{-16}$). In contrast, anti-entropic DMPs are overwhelmingly found at regions of strong transcription and enhancers (χ^2 test p value $< 2.2e^{-16}$) (Fig. 5C).

We hypothesized that cell type heterogeneity would bias age-related changes in entropy estimates upwards (i.e., the increase in entropy with age would be inflated because of changes in cell type % with age). We first repeated the analyses after adjusting the DNAm profiles for blood cell types in each dataset (see "Methods"). There was a moderate correlation of 0.53 (p value $= 2.9e^{-5}$) between the effect sizes (i.e., the change in entropy per decade of age) before vs after adjustment for cell type proportions (Additional file 1: Fig. S8A), and half of the datasets displaying effect sizes that declined in magnitude after adjusting for cell type proportions (Additional file 1: Fig. S8B). However, the overall meta-analysis effect size remained unchanged after adjustment (0.0005 change in entropy per decade of age) (Additional file 1: Fig. S8C).

We then looked at age-related changes in entropy in datasets containing isolated cell types, speculating that if age-related changes in entropy were solely driven by changes in cell type %, we would fail to see an increase in entropy in these datasets. We looked at monocytes (GSE56046 [79]), CD4+T cells (GSE59065 [80], GSE56581 [81] and GSE137593 [82]), CD8+T cells (GSE59065 [80]), and B cells (GSE137594 [83]) (Additional file 2: Table S4). In nearly all datasets, we failed to detect any change in entropy during aging, but both CD4+T and CD8+T cells in GSE59065 displayed a marked increase in entropy during aging (Additional file 2: Table S4). Finally, we took advantage of the unique design of dataset GSE184269 [84] that contains both mixed and sorted blood cells from the same individuals (naïve B cells, naïve CD4+T cells, naïve CD8+T cells, NK cells, monocytes, and granulocytes in GSE184269 [84]), speculating that if age-related changes in entropy were solely driven by changes in cell type %, PBMCs would show higher entropy levels than sorted cell types. Entropy was markedly higher in PBMCs than NK, naïve CD4+T, naïve CD8+T, and naïve B cells, but the highest entropy levels were in the heterogeneous class of granulocytes that comprise basophils, eosinophils, and neutrophils (Additional file 1: Fig. S9). These results suggest that changes in cell type composition during aging partially account for age-related increases in entropy. Moreover, the measure of entropy that we have used may be extremely sensitive to confounders that induce relatively small shifts in the global DNA methylation distribution.

Discussion

We demonstrated that during aging, large proportions of the blood methylome slowly shift in mean methylation levels, and about half of them also become increasingly divergent between individuals as they get older. Our results are robust and replicable, owing to the unparalleled statistical power generated from >32,000 samples in 56 independent cohorts. Importantly, these age-related changes occur largely independently of changes in the relative proportions of different blood cell types during aging. Considering the widespread distribution of these DNAm changes, aging appears to affect most molecular

pathways, from development, metabolism, and cell-signaling to homeostatic processes. Shannon entropy increases in the blood methylome during aging due to the presence of DMPs and because the majority of DMPs trend towards intermediate methylation levels of 50%. This implies that during aging, it becomes increasingly difficult to predict the methylation state of any given blood cell, in agreement with the smoothening of the epigenetic landscape. However, the overall increase in entropy is small in magnitude, and we uncovered that roughly one third of DMPs, particularly those enriched at enhancers and regions of active transcription, are "anti-entropic."

While we know from previous studies that DMPs are a feature of aging in blood [4, 78, 85–89], our findings are unparalleled, revealing the sheer magnitude and omnipresence of these differential shifts that may have been previously underappreciated. Moreover, our study is 38 times larger than the largest known study of VMPs in blood [7]. Unlike previously reported [7], we noted a significant overlap between DMPs and VMPs (i.e., many CpGs change in both average methylation and variance 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 "anti-entropic" CpGs could represent genes that become increasingly regulated [6], a theory worth investigating further. Since the majority of DMPs are entropic, the net global effect is an increase in entropy with age.

In alignment with previous studies, we identified key processes related to development, differentiation, and cell-signaling are altered during aging [3]. The promoters and bodies of actively transcribed genes were fairly unaffected by epigenetic aging. In contrast, our results support previous evidence of age-related gains in methylation and variance at important regulatory regions, particularly bivalent domains that harbor both active and inactive histone marks and regions repressed by the Polycomb complex [3]. Bivalent regions prepare key developmental genes to be switched on during differentiation in specific cells, such as embryonic stem cells [90]. The epigenetic clock starts ticking upon differentiation [91], and some gains in methylation may lead to loss of hematopoietic stem cell plasticity. However, these age-related epigenetic changes could affect all cell types in a tissue, not just stem cells. Bivalent promoters and poised enhancers are not exclusive to stem cells, and bivalency has been proposed to be a major reversible signature that distinguishes between cell types in adult tissues [90]. This would make sense, as epigenetic clocks work well in a plethora of tissues, even those that do not have many known stem cells [92]. Age-related hypermethylation may therefore hinder all cells from robustly maintaining their identity into old age, a process shared among individuals during chronological aging, but that happens at different rates between people. 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 aging [93]. This phenomenon has also been observed in human tissues based on gene expression data, whereby 40% of the tissues lose their cellular identity during aging [94]. 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 [95]. It is therefore plausible that environmental insults (i.e., sedentary behavior, sun exposure, toxins, inflammation, injury) could accelerate or exacerbate these changes via metabolic stress, damage to the extracellular matrix (ECM), and mitochondrial biogenesis [96], and these would be detected as VMPs. The stochasticity or variability may be a good indicator of cumulative damage of the environment or non-specific damage, whereby the epigenome is adapting to changing environmental cues [97]. It is therefore somewhat puzzling that both chronological and biological clocks were strongly enriched for those CpGs that show changes in average and variance during aging (DMPs-VMPs). Intuitively, chronological clocks, whose objective is to predict time elapsed since birth, should contain an overwhelming number of homoscedastic DMPs as those carry less "noise" between individuals as they get older. Furthermore, the age-related sites we identified were not enriched in pace-of-aging clocks: while those clocks do not measure biological age per se and rather the speed at which one ages across different organ systems, we expected some overlap with age-related sites as the biological processes controlling the rate of aging should be somewhat related to the biological processes controlling biological age.

In contrast to age-related hypermethylation that was accompanied with large divergence between individuals, age-related hypomethylation was more homogeneous across the lifespan and mainly affected quiescent regions, with unclear functional consequences. It may reflect tissue-specific operations [3], and comparing our results across other tissues would be necessary to confirm this. In blood, hypomethylation has been hypothesized to be immunogenic and contribute to inflammation [98]. Our pathway enrichment analysis revealed DMGs and VMGs to be associated with abnormal inflammatory cascades, and multiple organ system abnormalities, consistent with chronic inflammation as a hallmark of aging [2]. This inflammation is systemic and associated with various aging-related phenotypes in multiple tissues.

Cellular heterogeneity in tissues poses a significant challenge in DNAm studies [8]. 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-associated DNAm changes that are either present in all underlying subtypes, or only in a predominant subtype [78], a finding that overlaps with previous studies [7, 99]. While majority of the age-associated changes seem independent of changes in cell composition, about one third of VMPs were significant only in the unadjusted analyses, likely capturing variability related to changes in cell composition with age. It has been suggested that the increase in "transcriptional noise" underlying a loss of cell identity lacks sufficient evidence and may instead be related to the age-related changes in cell composition [98]. As such, these VMPs may be capturing an important phenotype of aging in blood (and potentially other tissues), such as the remodeling of the immune system cell type composition [100]. Importantly, performing cell type correction is recommended in addition to an unadjusted analysis [8], particularly in aging 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 [101], to understand the relative contribution of different cellular compartments in age-related diseases.

The significance of our cohort substructure, comprising extensive individual datasets with samples spanning a wide age range (6–101 years old), was evident when attempting to detect all kinds of age-related changes. Detecting VMPs using the Breusch-Pagan test for heteroscedasticity necessitated not only large sample sizes and a broad age range, but also possibly individuals at the "upper limit" of old age. Interestingly, a recent preprint in mouse blood showed that the variance of DNAm increased only at very old ages [97], implying that previous human studies may have been limited due to small sample sizes, narrow age ranges, and a scarcity of very old individuals, hindering investigations into variance or stochasticity with age [88]. In contrast, the standard linear model used to detect DMPs is less sensitive to the characteristics of the dataset substructure. However, the age range in the datasets was skewed, with varying sex distributions, diseases statuses, and other factors that could influence the effect of age on the methylome. For instance, age-related changes might be more pronounced in different ethnicities or sexes, introducing some variability. Given the substantial sample size, we see this heterogeneity as a strength more than a weakness, because it implies that the DMPs, VMPs, and entropy we detected hold true for a broad range of individuals and could be considered "universal" markers of aging. Moreover, we utilized multiple versions of the Illumina arrays to maximize the inclusion of CpGs across the various datasets. CpGs in the 850 K encompass all those from both the 450 K and the 27 K platforms, ensuring conservation across platforms. Despite not achieving a perfect overlap across datasets, each CpG is individually meta-analyzed, and this approach optimizes the sample size for each meta-analysis. 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.

In conclusion, using an unparalleled cohort of methylomes in humans, we provide a comprehensive picture of the global age-related changes observed in blood. Future work could focus on assessing the effect of longevity interventions, such as exercise, nutrition, and supplementation, on these age-related signatures in vivo humans.

Methods

This study was conducted as a large-scale, multi-tissue EWAS meta-analysis of age. Bioinformatics techniques were applied to analyze 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 [102]. 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).

Data mining

To create the repository of blood DNAm datasets, we have carried out a comprehensive data mining enterprise collecting the methylomes of 32,136 human blood samples from 56 datasets, profiled on the Illumina Methylation array platforms (27 K, 450 K, and EPIC) (Fig. 1, Additional file 2: Table S1). This includes 49 open-access datasets from Gene Expression Omnibus (GEO) repository and 1 from ArrayExpress, 4 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 (Additional file 2: Table S1). 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 [66].

Pre-processing

Datasets with raw DNAm data available were pre-processed, normalized, and filtered using the R statistical software (www.r-project.org), and following the *ChAMP* [65] 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 [74]. 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 sex chromosomes were also filtered out in datasets containing both males and females, as well as probes mapping to single nucleotide polymorphisms (SNPs) [65]. Additional cross-hybridizing probes identified by Pidsley et al. [103] 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 450 K and EPIC Illumina arrays were normalized using the *champ.norm* function [74]. We explored the technical and biological sources of variation in each dataset using a singular value decomposition method provided by the *champ.SVD* function [74]. 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 [104]. 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 [105]. 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 (Additional file 2: Table S1). We used the pre-processed matrices for datasets that we were unable to preprocess ourselves due to missing information (Additional file 2: Table S1).

Statistical framework

DMPs

To identify DMPs, we performed an EWAS in each dataset using linear regression models and moderated Bayesian statistics, as implemented in the *limma* package [106, 107]. 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)) (Additional file 2: Table S1), when this information was available (see below for the inclusion/

exclusion criteria of covariates). 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). Linear models were performed using M values (a logit transformation of β values), which have more favorable statistical properties for differential analysis [108].

To then compare the effect of cell type heterogeneity on age-associated DNAm changes, we repeated all linear models adjusting for cell type proportions in whole blood. We applied the *champ.refbase* method, as implemented in the *ChAMP* package, to estimate the cell type proportions for granulocytes, monocytes, NK cells, CD4 + T cells, CD8 + T cells, and B cells, for each sample [74]. We then repeated the linear model for each blood dataset including the 5 largest cell types to remove the confounding effect of cell type proportion on DMPs.

To assess whether the direction of DNAm with age (hypo/hyper) depends on the baseline DNAm levels, we classified DMPs in "high," "intermediate," and "low" methylation categories using a large dataset (BIOS) [20]. In this dataset and for each DMP, we calculated the average MF and classified a DMP as "high" if MF $\geq$ 75%, "low" if MF $\leq$ 25%, and the remainder "intermediate." This categorization was made separately for young (i.e., < 30 y.o.) and old (> 60 y.o.) individuals, and we assessed whether the DMP trended from "high" in young to "intermediate" or "low" in old, etc.

VMPs

Age-associated 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, 109]. DNAm was first regressed against age and other confounders for each dataset (i.e., the linear model to identify DMPs) to obtain residuals. 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) [4], and subsequently regressed the squared residuals of the remaining markers against age. All analyses were repeated and corrected for cell types to compare the effect of cell type heterogeneity on age-associated VMPs.

Entropy

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%. Shannon entropy was calculated for each sample in each dataset, using a probability formula adapted to handle DNAm data [3, 4, 7].

To calculate the genome-wide Shannon entropy, a linear model is fitted for each dataset on the logit transformed M values, adjusting for dataset covariates where appropriate (e.g., sex, BMI, batch). Age was not included in this model, as we did not want to remove the effect of age on the β values. The mean M values from the original matrix were added to the residuals from the linear model, and then transformed back to obtain adjusted β values, as the Shannon entropy formula has been adapted specifically to handle β values.

Using the adjusted β values, genome-wide Shannon entropy was computed for each sample in each dataset according to the formula [3, 4]:

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

where MF_i is the methylation fraction (e.g., beta value) 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 regressing age against entropy and recorded the summary statistics for each dataset (Additional file 2: Table S2), as follows:

> *Entropy $\sim$ age*

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 aging 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-associated CpGs identified from the meta-analyses below (i.e., a complete list of both DMPs and VMPs), as well as the non-age-associatedCpGs (i.e., the complete list of non-DMPs and non-VMPs). We then repeated the above steps calculating entropy CpGs that were only DMPs, CpGs that were both DMPs andVMPs, CpGs that were only VMPs (Additional file 2: Table S3), entropic DMPs, and anti-entropic DMPs. We also re-calculated entropy on all measurements and corrected for cell type heterogeneity, to compare the effect of cell type composition on age-associated changes in entropy (Additional file 2: Table S2), as follows:

> *Entropy $\sim$ age $+$ CD8T $+$ CD4T $+$ NK $+$ Bcell $+$ Gran . . .*

In addition, we calculated entropy for datasets from sorted cell types on the genome-wide set of CpGs, only on the age-associated CpGs, and on the non-age-associated CpGs (Additional file 2: Table S4).

Adjusting for covariates

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" exists, 92 conditions have been classified as "age-related" based upon an exponential increase in incidence with age using a two-step mathematical modeling technique [110]. To have a direct view of the relationship between age and each disease/condition present in our datasets (Additional file 2: Table S1), 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, anemia, inflammatory bowel disease/Crohn's disease, 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

Seale *et al. Genome Biology*   (2024) 25:240

Page 18 of 24

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 do not depend on age-related functional decline of the body, but on age-related differences in behavior (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 aging (e.g., smoking, BMI, hypertension, exercise training, and bariatric surgery) were also included as covariates.

Meta-analyses of DMPs and age

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* [101]. The results from the independent EWAS were then pooled, using an inverse variance-based fixed-effects meta-analysis implemented in METAL [111]. 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-associated DMPs were then identified using a stringent meta-analysis false discovery rate (FDR) < 0.005. The meta-analysis was repeated with datasets adjusted for cell types.

Meta-analyses of VMPs and age

As for the DMPs, we used METAL to pool results from the independent EWAS, but we followed a sample size-based fixed-effects meta-analysis (instead of an inverse variance method) [111]. This approach was more appropriate to meta-analyze 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. Age-associated VMPs were identified using a stringent meta-analysis FDR < 0.005. The meta-analysis was repeated with datasets adjusted for cell types.

Meta-analyses of entropy and age

To identify the change in entropy with age in each tissue, the summary statistics (i.e., effect size and standard error) extracted from the independent entropy regressions were pooled using a fixed-effects meta-analysis using the R package *metafor* [112] (Additional file 2: Table S2). We meta-analyzed 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.

Chromatin state enrichment, pathway analysis, and epigenetic clock CpGs analysis

To aid in the biological interpretation of the identified age-associated methylation sites, 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 [64].

To gain insights into the cellular and phenotype consequences of aging 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), expression signatures of genetic and chemical perturbations (CGP gene set in MsigDB), and immunological signatures (C7 gene set in MsigDB) were enriched among the VMPs and DMPs using the *gsameth* function from the *missMethyl* package [113]. An improved adaptation of Zhou et al.'s comprehensive annotation was used to assign one or more genes to each VMP and DMP [65]. All GO, HPO, CP, and CGO terms were deemed significant at an FDR < 0.005 [114, 115].

We also tested whether DMPs and VMPs were particularly over- or underrepresented in CpGs that make up the different epigenetic clocks that were developed using elastic net. We tested clocks trained to predict chronological age (Horvath's pan-tissue clock, 353 CpGs [66]; Hannum's clock, 71 CpGs [4]; the blood clock developed by Zhang et al., 514 CpGs [67]; the centenarian clock, 747 CpGs [68]; version 2 of the mammalian universal clock, 816 CpGs [69]), biological age (PhenoAge, 513 CpGs [70]), and the pace of aging (DunedinPoAm, 46 CpGs [71]; DunedinPACE, 173 CpGs [72]). This was done by comparing the distribution of the age-related CpGs with that of non-age-related CpGs, respectively, among the list of CpGs that make up different clocks, with the help of a χ^2 test. We used a different background of probes for each clock, as each clock was developed from a different pool of CpGs (e.g., the pan-tissue clock was developed from a pool of ∼ 21,000 CpGs common to all Illumina HumanMethylation arrays, while DunedinPACE was developed from probes common to the 450 K and EPIC arrays that showed good test–retest reliability). Note that we could not perform this enrichment for clocks whose list of CpGs is not open-access, such as GrimAge [116], and for clocks that use all probes on the Illumina HumanMethylation arrays, such as Altum Age [117], and PC-based clocks [118]. All χ^2 test p values were adjusted for multiple testing and only those tests with FDR < 0.005 were deemed significant.

Figures

R studio ggplot2 from the tidyverse package and Cytoscape.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s13059-024-03381-w.

Additional file 1: The supplementary figures for this publication.

Additional file 2: The supplementary tables for this publication.

Additional file 3: The peer review history.

Seale *et al. Genome Biology*   (2024) 25:240

Page 20 of 24

Disclaimer

The views expressed in this manuscript are those of the authors and do not necessarily represent the views of the National Heart, Lung, and Blood Institute; the National Institutes of Health; or the US Department of Health and Human Services.

Review history

The review history is available as Additional file 3.

Peer review information

Tim Sands was the primary editor of this article and managed its editorial process and peer review in collaboration with the rest of the editorial team.

Authors' contributions

 Conceptualization: K.S., S.V., and N.E.; methodology: K.S., S.V., and A.T.; data collection: K.S. and S.V. (A.P.R. data collection of JHS); data analysis: K.S. and S.V.; writing original draft: K.S.; reviewing and editing: S.V., A.T., N.E.; co-supervision: S.V.; supervision: N.E.

Availability of data and materials

Four datasets are sourced from the dbGaP and include The Genetics of Lipid-Lowering Drugs and Diet Network Lipidomics Study (GOLDN; accession number phs000741.v2.p1) [10], the Women's Health Initiative (WHI; accession number phs001335.v2.p3) [11], the Normative Aging Study (NAS, accession number phs000853.v2.p2) [19], and the Framingham Heart Study (FHS, accession number phs000974.v5.p4) [23]. The BIOS can be requested and downloaded from the European Genome-Phenome Archive (EGA), accession EGAS00001001077 [20]. The Jackson Heart Study (JHS) can be requested from the Jackson Heart Study website [22]. The Swedish Adoption/Twin Study of Aging (SATSA) is available from ArrayExpress under the accession number E-MTAB-7309 [26]. The remaining datasets are sourced from the GEO platform, including GSE55763 [12], GSE128235 [13], GSE99624 [14], GSE115278 [15], GSE87571 [16], GSE53740 [17], GSE58045 [18], GSE77445 [21], GSE49904 [24], GSE80417 [25], GSE42861 [27], GSE51032 [28], GSE67705 [29], GSE32148 [30], GSE106648 [31], GSE69138 [32], GSE50660 [33], GSE40279 [34], GSE41037 [35], GSE41169 [36], GSE53840 [37], GSE67751 [38], GSE72775 [39], GSE111629 [40], GSE72774 [41], GSE72776 [42], GSE166611 [43], GSE164056 [44], GSE85311 [45], GSE151278 [46], GSE96879 [47], GSE134429 [48], GSE120307 [49], GSE20236 [50], GSE19711 [51], GSE157131 [52], GSE117859 [53], GSE117860 [54], GSE147740 [55], GSE152026 [56], GSE132203 [57], GSE100264 [58], GSE107080 [59], GSE116339 [60], GSE168739 [61], GSE197674 [62], GSE197676 [63], GSE56046 [79], GSE59065 [80], GSE56581 [81], GSE137593 [82], GSE137594 [83], and GSE184269 [84]. The source code used to reproduce the methodology and figures is available from https://github.com/kirstenblythe/Map-of-ageing-blood-methylome [119] and at https://zenodo.org/records/12786566 under MIT licenses [120].

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors have declared no competing interests.

References

1. Partridge L, Deelen J, Slagboom PE. Facing up to the global challenges of ageing. Nature. 2018;561:45–56.
2. Lopez-Otin C, Blasco MA, Partridge L, Serrano M, Kroemer G. Hallmarks of aging: an expanding universe. Cell. 2023;186:243–78.

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. Hannum G, Guinney J, Zhao L, Zhang L, Hughes G, Sadda S, et al. Genome-wide methylation profiles reveal quantitative views of human aging rates. Mol Cell. 2013;49:359–67.
5. Martin-Herranz DE, Aref-Eshghi E, Bonder MJ, Stubbs TM, Choufani S, Weksberg R, et al. Screening for genes that accelerate the epigenetic aging clock in humans reveals a role for the H3K36 methyltransferase NSD1. Genome Biol. 2019;20:146.
6. Duffield T, Csuka L, Akalan A, Magdaleno GV, Senez L, Palmer D, Magalhães JPd. Epigenetic fidelity in complex biological systems and implications for ageing. bioRxiv . 2023;2023.2004.2029.538716.
7. Slieker RC, van Iterson M, Luijk R, Beekman M, Zhernakova DV, Moed MH, et al. Age-related accrual of methylomic variability is linked to fundamental ageing mechanisms. Genome Biol. 2016;17:191.
8. Qi L, Teschendorff AE. Cell-type heterogeneity: why we should adjust for it in epigenome and biomarker studies. Clin Epigenetics. 2022;14:31.
9. 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. 2024;23: e13859.
10. Arnett DK. Genetics of Lipid Lowering Drugs and Diet Network (GOLDN) Lipidomics Study. Datasets. dbGaP. 2016. https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs000741.v2.p1.
11. Absher D, Assimes T, Horvath S, Tsao P. Integrative genomics and risk of CHD and related phenotypes in the Women's Health Initiative. Datasets. dbGaP. 2013. https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs001335.v1.p3.
12. Lehne B, Drong A, Loh M, Zhang W, Scott WR, Tan ST, Afzal U, Scott J, Jarvelin MR, Elliott P, McCarthy MI, Kooner JS, Chambers JC. A coherent approach for analysis of the Illumina HumanMethylation450 BeadChip improves data quality and performance in epigenome-wide association studies. Datasets. Gene Expression Omnibus (GEO). 2015. https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE55763.
13. Arloth J, Binder EB. Epigenome analysis of depressed and control subjects. Datasets. Gene Expression Omnibus (GEO). 2019. https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE128235.
14. Fernandez-Rebollo E, Eipel M, Seefried L, Hofmann P, Strathmann K, Jakob F, Wagner W. Primary osteoporosis is not reflected by disease-specific DNA methylation or accelerated epigenetic age in blood. Datasets. Gene Expression Omnibus (GEO). 2018. https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE99624.
15. Arpón A, Milagro FI, Ramos-Lopez O, Mansego ML, Riezu-Boj JI, Martínez JA. Epigenome-wide association study in peripheral white blood cells: Methyl Epigenome Network Association (MENA) project. Datasets. Gene Expression Omnibus (GEO). 2018. https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE115278.
16. Johansson Å. Continuous aging of the human DNA methylome throughout the human lifespan. Datasets. Gene Expression Omnibus (GEO). 2016. https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE87571.
17. Li Y, Chen JA, Sears RL, Gao F, Klein ED, Karydas A, Geschwind MD, Rosen HJ, Boxer AL, Guo W, Pellegrini M, Horvath S, Miller BL, Geschwind DH, Coppola G. An epigenetic signature in peripheral blood associated with neurodegenerative tauopathy and the risk-associated haplotype on 17q21.31. Datasets. Gene Expression Omnibus (GEO). 2014. https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE53740.
18. Tsai P, Bell JT. Epigenome-wide scans identify differentially methylated regions for age and age-related phenotypes in a healthy ageing population. Datasets. Gene Expression Omnibus (GEO). 2014. https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE58045.
19. Baccarelli AA, Schwartz J. Normative Aging Study (NAS). Datasets. dbGaP. 1999. https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs000853.v2.p2.
20. The BIOS Consortium. Biobank-based integrative omics studies. Datasets. European Genome Phenome Archive (EGA). 2018. https://ega-archive.org/dacs/EGAC00001000277.
21. Boks MP, Vinkers CH. Genome-wide DNA methylation levels and altered cortisol stress reactivity following childhood trauma in humans. Datasets. Gene Expression Omnibus (GEO). 2016. https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE77445.
22. Reiner A. Jackson Heart Study. Datasets. 1998. https://www.jacksonheartstudy.org/Research/Study-Data/Data-Access.
23. Ramachandran V, Larson MG, Heard-Costa N. NHLBI TOPMed: genomic activities such as whole genome sequencing and related phenotypes in the Framingham Heart Study. Datasets. dbGaP. 1966. https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs000974.v5.p4.
24. Day K, Waite LL, Thalacker-Mercer A, West A, Bamman MM, Brooks JD, Myers RM, Absher D. Differential DNA methylation with age displays both common and dynamic features across human tissues that are influenced by CpG landscape. Datasets. Gene Expression Omnibus (GEO). 2013. https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE49904.
25. Hannon E, Dempster E, Viana J, Burrage J, Smith AR, Macdonald R, St. Clair D, Mustard C, Breen G, Therman S, Kaprio J, Toulopoulou T, Hulshoff Pol HE, Bohlken MM, Kahn RS, Nenadic I, Hultman CM, Murray RM, Collier DA, Bass N, Gurling H, McQuillin A, Schalkwyk L, Mill J. An integrated genetic-epigenetic analysis of schizophrenia: evidence for co-localization of genetic associations and differential methylation. Datasets. Gene Expression Omnibus (GEO). 2016. https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE80417.
26. Wang Y. DNA methylation of longitudinal samples from The Swedish Adoption/Twin Study of Aging (SATSA). Datasets. ArrayExpress. 2018. https://www.ebi.ac.uk/biostudies/arrayexpress/studies/E-MTAB-7309?query=SATSA.
27. Liu Y, Feinberg AP. Differential DNA methylation in rheumatoid arthritis. Datasets. Gene Expression Omnibus (GEO). 2013. https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=gse42861.
28. Polidoro S, Campanella G, Krogh V, Palli D, Panico S, Tumino R, Vineis P. EPIC-Italy at HuGeF. Datasets. Gene Expression Omnibus (GEO). 2013. https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE51032.
29. Gross AM, Jaeger PA, Kreisberg JF, Licon K, Jepsen KL, Khosroheidari M, Morsey BM, Shen H, Flagg K, Chen D, Zhang K, Fox HS, Ideker T. Methylome-wide analysis of chronic HIV infected patients and healthy controls. Datasets. Gene Expression Omnibus (GEO). 2016. https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE67705.

Seale *et al. Genome Biology*   (2024) 25:240

Page 22 of 24

30. Kellermayer R, Harris RA. DNA methylation in peripheral blood from individuals with Crohns' disease or ulcerative colitis and normal controls. Datasets. Gene Expression Omnibus (GEO). 2012. https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE32148.

31. Liu Y. Differential DNA methylation in multiple sclerosis. Datasets. Gene Expression Omnibus (GEO). 2018. https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE106648.

32. Roquer J, Soriano-Tarraga C, Jimenez-Conde J. Epigenome analysis of ischemic stroke patients. Datasets. Gene Expression Omnibus (GEO). 2015. https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE69138.

33. Tsaprouni LG, Yang TP, Bell JT, Dick KJ, Kanoni S, Nisbet J, Viñuela A, Grundberg E, Nelson CP, Meduri E, Buil A, Cambien F, Hengstenberg C, Erdmann J, Schunkert H, Goodall AH, Ouwehand WH, Dermitzakis ET, Spector TD, Samani NJ, Deloukas P. Cigarette smoking reduces DNA methylation levels at multiple genomic loci but the effect is partially reversible upon cessation. Datasets. Gene Expression Omnibus (GEO). 2014. https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE50660.

34. Zhang K, Ideker T. Genome-wide methylation profiles reveal quantitative views of human aging rates. Datasets. Gene Expression Omnibus (GEO). 2012. https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE40279.

35. Ophoff R, Horvath S. Aging effects on DNA methylation modules in blood tissue. Datasets. Gene Expression Omnibus (GEO). 2012. https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE41037.

36. Ophoff R, Horvath S. Blood DNA methylation profiles in a Dutch population. Datasets. Gene Expression Omnibus (GEO). 2012. https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE41169.

37. Horvath S, Levine AJ. Whole blood DNA methylation samples from HIV positive men. Datasets. Gene Expression Omnibus (GEO). 2014. https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE53840.

38. Horvath S, Levine AJ. Blood Illumina Inf 450 DNA methylation samples from HIV+ and HIV- human subjects. Datasets. Gene Expression Omnibus (GEO). 2015. https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE67751.

39. Horvath S, Ritz BR. DNA methylation profiles of human blood samples from Hispanics and Caucasians. Datasets. Gene Expression Omnibus (GEO). 2016. https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE72775.

40. Ritz B, Horvath S. Genome wide DNA methylation study of Parkinson's disease in whole blood samples. Datasets. Gene Expression Omnibus (GEO). 2018. https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE111629.

41. Horvath S, Ritz BR. DNA methylation profiles of human blood samples from Caucasian subjects with Parkinson's disease. Datasets. Gene Expression Omnibus (GEO). 2018. https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE72774.

42. Horvath S, Ritz BR. DNA methylation profiles of human blood samples from Hispanic subjects with Parkinson's disease. Datasets. Gene Expression Omnibus (GEO). 2018. https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE72776.

43. Nonino CB, Noronha NY, Nicoletti CF, Pinhel MA. Trait related and differential DNA methylation in obese and normal weight Brazilian women. Datasets. Gene Expression Omnibus (GEO). 2021. https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE166611.

44. Wiegand A, Kreifelts B, Munk MJ, Geiselhart N, Ramadori KE, MacIsaac JL, Fallgatter AJ, Kobor MS, Nieratschker V. DNA methylation differences associated with social anxiety disorder and early life adversity. Datasets. Gene Expression Omnibus (GEO). 2021. https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE164056.

45. Martens CR, Lubieniecki KL, McNamara MN, Bohr AD, McQueen MB, Seals DR. Epigenetic patterns with aging and exercise are associated with indicators of healthspan in humans. Datasets. Gene Expression Omnibus (GEO). 2020. https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE85311.

46. Ovejero-Benito MC, Sanz-García A, Llamas-Velasco M, Reolid A, Daudén E, Abad-Santos F. Genome-wide DNA methylation analysis of peripheral blood samples of moderate-to-severe psoriasis patients treated with anti-TNF drugs. Datasets. Gene Expression Omnibus (GEO). 2020. https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE151278.

47. Joseph S, Green-Knox B, George NI, Lyn-Cook B. Epigenome-wide methylation profile in sustemic lupus erythematosus: impact of ethnicity and SLEDAI score. Datasets. Gene Expression Omnibus (GEO). 2020. https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE96879.

48. Rodríguez-Ubreva J, de la Calle-Fabregat C, Li T, Ciudad L, Ballestar ML, Català-Moll F, Morante-Palacios O, García-Gómez A, Celis R, Humby F, Nerviani A, Martín J, Pitzalis C, Cañete JD, Ballestar E. Inflammatory cytokines shape a changing DNA methylome in monocytes mirroring disease activity in rheumatoid arthritis (patients). Datasets. Gene Expression Omnibus (GEO). 2019. https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE134429.

49. Palma-Gudiel H, Córdova-Palomera A, Fañanás L. Genome-wide DNA methylation analysis in peripheral blood of monozygotic twins informative for psychopathology. Datasets. Gene Expression Omnibus (GEO). 2018. https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE120307.

50. Rakyan VK, Down TA, Maslau S, Andrew T, Yang TP, Beyan H, Whittaker P, McCann OT, Finer S, Valdes AM, Leslie RD, Deloukas P, Spector TD. Human aging-associated DNA hypermethylation occurs preferentially at bivalent chromatin domains. Gene Datasets. Expression Omnibus (GEO). 2010. https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE20236.

51. Teschendorff A, Widschwendter M. Genome wide DNA methylation profiling of United Kingdom Ovarian Cancer Population Study (UKOPS). Datasets. Gene Expression Omnibus (GEO). 2010. https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE19711.

52. Kardia SL, Smith JA. Methylation data from stored peripheral blood leukocytes from African American participants in the GENOA study. Datasets. Gene Expression Omnibus (GEO). 2020. https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE157131.

53. Xu K, Zhang X, Justice A. Smoking-associated DNA methylation features link to HIV outcomes [HumanMethylation450 BeadChip]. Datasets. Gene Expression Omnibus (GEO). 2019. https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE117859.

54. Xu K, Zhang X, Justice A. Smoking-associated DNA methylation features link to HIV outcomes [Infinium MethylationEPIC]. Datasets. Gene Expression Omnibus (GEO). 2019. https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE117860.

55. Robinson O, Polidoro S, Fiorito G, Vineis P, Elliott P. DNA methylation analysis of human peripheral blood mononuclear cell collected in the AIRWAVE study. Datasets. Gene Expression Omnibus (GEO). 2020. https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE147740.

56. Hannon E, Mill J. Blood DNA methylation profiles from first episode psychosis patients and controls I. Datasets. Gene Expression Omnibus (GEO). 2021. https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE152026.

57. Kilaru V, Katrinli S, Smith AK, Ressler K. DNA methylation (EPIC) from the Grady Trauma Project. Datasets. Gene Expression Omnibus (GEO). 2019. https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE132203.

58. Xu K, Zhang X, Justice A. DNA methylation signatures of injection illicit drug use (IDU) and hepatitis C (HCV) predict HIV pathophysiologic frailty I. Datasets. Gene Expression Omnibus (GEO). 2017. https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE100264.

59. Xu K, Zhang X, Justice A. DNA methylation signatures of illicit drug injection and hepatitis C are associated with HIV frailty II. Datasets. Gene Expression Omnibus (GEO). 2017. https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE107080.

60. Curtis SW, Kilaru V, Cobb DO, Marcus M, Conneely KN, Smith AK. Exposure to polybrominated biphenyl (PBB) associates with DNA methylation differences across the genome. Datasets. Gene Expression Omnibus (GEO). 2018. https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE116339.

61. de Moura MC, Dávalos V, Esteller M. Epigenome-wide association study of COVID-19 severity with respiratory failure. Datasets. Gene Expression Omnibus (GEO). 2021. https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE168739.

62. Dong Q, Song N, Wang Z, Qin N, Chen C, Li Z, Sun X, Easton J, Mulder H, Plyler E, Neale G, Walker E, Li Q, Ma X, Chen X, Huang I, Yasui Y, Ness KK, Zhang J, Hudson MM, Robison LL. Genome-wide association studies identify novel genetic loci for epigenetic age acceleration among survivors of childhood cancer. Datasets. Gene Expression Omnibus (GEO). 2022. https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE197674.

63. Dong Q, Song N, Wang Z, Qin N, Chen C, Li Z, Sun X, Easton J, Mulder H, Plyler E, Neale G, Walker E, Li Q, Ma X, Chen X, Huang I, Yasui Y, Ness KK, Zhang J, Hudson MM, Robison LL. Genome-wide association studies identify novel genetic loci for epigenetic age acceleration among survivors of childhood cancer. Datasets. Gene Expression Omnibus (GEO). 2022. https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE197676.

64. Roadmap Epigenomics C, Kundaje A, Meuleman W, Ernst J, Bilenky M, Yen A, Heravi-Moussavi A, Kheradpour P, Zhang Z, Wang J, et al. Integrative analysis of 111 reference human epigenomes. Nature. 2015;518:317–30.

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

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

67. Zhang Q, Vallerga CL, Walker RM, Lin T, Henders AK, Montgomery GW, et al. Improved precision of epigenetic clock estimates across tissues and its implication for biological ageing. Genome Med. 2019;11:54.

68. Dec E, Clement J, Cheng K, Church GM, Fossel MB, Rehkopf DH, et al. Centenarian clocks: epigenetic clocks for validating claims of exceptional longevity. Geroscience. 2023;45:1817–35.

69. Lu AT, Fei Z, Haghani A, Robeck TR, Zoller JA, Li CZ, et al. Universal DNA methylation age across mammalian tissues. Nat Aging. 2023;3:1144–66.

70. Levine ME, Lu AT, Quach A, Chen BH, Assimes TL, Bandinelli S, et al. An epigenetic biomarker of aging for lifespan and healthspan. Aging (Albany NY). 2018;10:573–91.

71. Belsky DW, Caspi A, Arseneault L, Baccarelli A, Corcoran DL, Gao X, et al. Quantification of the pace of biological aging in humans through a blood test, the DunedinPoAm DNA methylation algorithm. Elife. 2020;9:9.

72. Belsky DW, Caspi A, Corcoran DL, Sugden K, Poulton R, Arseneault L, et al. DunedinPACE, a DNA methylation biomarker of the pace of aging. Elife. 2022;11:11.

73. Luo Q, Dwaraka VB, Chen Q, Tong H, Zhu T, Seale K, et al. A meta-analysis of immune-cell fractions at high resolution reveals novel associations with common phenotypes and health outcomes. Genome Med. 2023;15:59.

74. Morris TJ, Butcher LM, Feber A, Teschendorff AE, Chakravarthy AR, Wojdacz TK, et al. ChAMP: 450k chip analysis methylation pipeline. Bioinformatics. 2014;30:428–30.

75. Slieker RC, Relton CL, Gaunt TR, Slagboom PE, Heijmans BT. Age-related DNA methylation changes are tissue-specific with ELOVL2 promoter methylation as exception. Epigenetics Chromatin. 2018;11:25.

76. 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;11:1132–4.

77. Marttila S, Kananen L, Hayrynen S, Jylhava 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;16:179.

78. Florath I, Butterbach K, Muller 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;23:1186–201.

79. Liu Y. Transcriptomics and methylomics of human monocytes. Datasets. Gene Expression Omnibus (GEO). https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE56046. 2014.

80. Milani L, Peterson P. Age-related profiling of DNA methylation in CD8+ T cells reveals changes in immune response and transcriptional regulator genes. Datasets. Gene Expression Omnibus (GEO). https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE59065. 2015.

81. Liu Y. Transcriptomics and methylomics of human T cells. Datasets. Gene Expression Omnibus (GEO). https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE56581. 2014.

82. Clark AD, Nair N, Anderson AE, Thalayasingam N, Naamane N, Skelton AJ, Diboll J, Barton A, Eyre S, Isaacs JD, Pratt AG, Reynard LN. Early arthritis CD4+ T cell DNA methylation profiling. Datasets. Gene Expression Omnibus (GEO). https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE137593. 2019.

83. Clark AD, Nair N, Anderson AE, Thalayasingam N, Naamane N, Skelton AJ, Diboll J, Barton A, Eyre S, Isaacs JD, Pratt AG, Reynard LN. Early arthritis B cell DNA methylation profiling. Datasets. Gene Expression Omnibus (GEO). https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE137594. 2019.

84. Kaileh M, Roy R, Ramamoorthy S, Boller S, Grosschedl R, De S, Ferrucci L, Sen R. Specification of human immune cell epigenetic identity by combinations of transcription factors. Datasets. Gene Expression Omnibus (GEO). https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE184269. 2021.

Seale *et al. Genome Biology*   (2024) 25:240

Page 24 of 24

85. 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;35:356–64.

86. 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:R102.

87. 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;20:434–9.

88. 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;109:10522–7.

89. 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;20:440–6.

90. Blanco E, Gonzalez-Ramirez M, Alcaine-Colet A, Aranda S, Di Croce L. The bivalent genome: characterization, structure, and regulation. Trends Genet. 2020;36:118–31.

91. Kabacik S, Lowe D, Fransen L, Leonard M, Ang S-L, Whiteman C, et al. The relationship between epigenetic age and the hallmarks of aging in human cells. Nat Aging. 2022;2:484–93.

92. Ogrodnik M, Gladyshev VN. The meaning of adaptation in aging: insights from cellular senescence, epigenetic clocks and stem cell alterations. Nat Aging. 2023;3:766–75.

93. 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;11:11.

94. Dos Santos GA, Chatsirisupachai K, Avelar RA, de Magalhães JP. Transcriptomic analysis reveals a tissue-specific loss of identity during ageing and cancer. BMC Genomics. 2023;24:644.

95. 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.e327.

96. Moldakozhayev A, Gladyshev VN. Metabolism, homeostasis, and aging. Trends Endocrinol Metab. 2023;34:158–69.

97. Tarkhov AE, Lindstrom-Vautrin T, Zhang S, Ying K, Moqri M, Zhang B, et al. Nature of epigenetic aging from a single-cell perspective. Nat Aging. 2024;4(6):854–70.

98. 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;42:464–8.

99. 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 (Albany NY). 2018;10:3541–57.

100. Mogilenko DA, Shchukina I, Artyomov MN. Immune ageing at single-cell resolution. Nat Rev Immunol. 2022;22:484–98.

101. Zheng SC, Breeze CE, Beck S, Teschendorff AE. Identification of differentially methylated cell types in epigenome-wide association studies. Nat Methods. 2018;15:1059–66.

102. van Iterson M, van Zwet EW, Consortium B, Heijmans BT. Controlling bias and inflation in epigenome- and transcriptome-wide association studies using the empirical null distribution. Genome Biol. 2017;18:19.

103. 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;17:208.

104. 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;28:882–3.

105. 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;30:1363–9.

106. Smyth GK. limma: linear models for microarray data. In Bioinformatics and computational biology solutions using R and Bioconductor. Edited by Gentleman R, Carey VJ, Huber W, Irizarry RA, Dudoit S. New York, NY: Springer New York; 2005: 397–420.

107. 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.

108. 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: 587.

109. Breusch TS, Pagan AR. A simple test for heteroscedasticity and random coefficient variation. Econometrica. 1979;47:1287–94.

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

111. Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. Bioinformatics. 2010;26:2190–1.

112. Viechtbauer W. Conducting meta-analyses in R with the metafor package. J Stat Softw. 2010;36:1–48.

113. Phipson B, Maksimovic J, Oshlack A. missMethyl: an R package for analyzing data from Illumina's HumanMethylation450 platform. Bioinformatics. 2016;32:286–8.

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

115. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat. 1995;57:289–300.

116. Lu AT, Quach A, Wilson JG, Reiner AP, Aviv A, Raj K, et al. DNA methylation GrimAge strongly predicts lifespan and healthspan. Aging (Albany NY). 2019;11:303–27.

117. de Lima Camillo LP, Lapierre LR, Singh R. A pan-tissue DNA-methylation epigenetic clock based on deep learning. npj. Aging. 2022;8:4.

118. Higgins-Chen AT, Thrush KL, Wang Y, Minteer CJ, Kuo PL, Wang M, et al. A computational solution for bolstering reliability of epigenetic clocks: implications for clinical trials and longitudinal tracking. Nat Aging. 2022;2:644–61.

119. Seale, KB. Map of ageing blood methylome. Github. 2024. https://github.com/kirstenblythe/Map-of-ageing-blood-methylome.

120. Seale, KB. A comprehensive map of ageing blood methylome. Zenodo. 2024. https://zenodo.org/records/12786566.

Publisher's Note