

In-depth characterization of the DNA methylation aging landscape links epigenetic noise and genotype-specific epigenetic aging to inflammaging

Zhaozhen Du

Chinese Academy of Sciences

Qi Luo

Chinese Academy of Sciences

Xiaolong Guo

Chinese Academy of Sciences

Qianqian Peng

Chinese Academy of Sciences

Zhipeng Li

Eastern Institute of Technology, Ningbo

Deshuang Huang

Eastern Institute of Technology, Ningbo

Ritambhara Singh

Brown University

Sijia Wang

Chinese Academy of Sciences

Riccardo Marioni

University of Edinburgh

Andrew Erich Teschendorff

andrew@sinh.ac.cn

Chinese Academy of Sciences

Article

Keywords: aging, DNA methylation, deep-learning, mQTL, inflammaging

Posted Date: August 29th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10584417/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.
[Read Full License](#)

Additional Declarations: No competing interests reported.

In-depth characterization of the DNA methylation aging landscape links epigenetic noise and genotype-specific epigenetic aging to inflammaging

Zhaozhen Du ^{1,§}, Qi Luo ^{1,§}, Xiaolong Guo ¹, Qianqian Peng ¹, Zhipeng Li ^{2,3}, Deshuang Huang ², Ritambhara Singh ^{4,5}, Sijia Wang ^{1,*}, Riccardo E. Marioni ^{6,*} and Andrew E. Teschendorff ^{1,*}

1. Shanghai Institute of Nutrition and Health, Shanghai Institute for Biological Sciences, University of Chinese Academy of Sciences, Chinese Academy of Sciences, 320 Yue Yang Road, Shanghai 200031, China.
2. Ningbo Institute of Digital Twin, Eastern Institute of Technology, Ningbo, 315201, Zhejiang, China.
3. School of life sciences, University of Science and Technology of China, Hefei, 230026, Anhui, China.
4. Department of Computer Science, Brown University, Providence, RI, USA.
5. Department of Molecular Biology, Cell Biology and Biochemistry, Brown University, Providence, RI, USA.
6. Centre for Genomic and Experimental Medicine, Institute of Genetics and Cancer, University of Edinburgh, Edinburgh EH4 2XU, UK.

[§] Equal Contribution

* Corresponding authors: Andrew E. Teschendorff: andrew@sinh.ac.cn ; Riccardo Marioni: riccardo.marioni@ed.ac.uk ; Sijia Wang: wangsijia@sinh.ac.cn

Abstract

According to recent estimates, over 50% of the human DNA methylome (DNAm) changes with age. These age-associated changes display a significant diversity of patterns that have not yet undergone any comprehensive characterization. Based on a priori-knowledge we here guide a deep-learning based convolutional neural network to classify over 800,000 CpGs in blood into 10 distinct age-related DNAm patterns, which we extensively characterize in relation to chromatin state, immune-cell composition, inflammaging, epigenetic clocks, genotype and clonal hematopoiesis. One pattern defined by methylation quantitative trait loci (mQTLs) where the age-association is genotype-specific, uncovers a number of potential DNAm mediators of inflammatory diseases, which are distinct to those of existing causal clocks. Another pattern, representing age-dependent noise, is associated with inflammaging and all-cause mortality, and can be dissected into a component driven by variations in the memory CD8⁺ T-cell fraction, and another which is cell-type independent. In summary, by exposing and fully characterizing the diversity of age-related DNAm patterns we here improve our mechanistic understanding of these patterns.

Keywords: aging; DNA methylation; deep-learning; mQTL; inflammaging

Age-associated DNA methylation (DNAm) changes have been implicated in a wide range of complex diseases [1, 2] and are promising aging biomarkers [3-9]. Over the years, studies have quantified the proportion of the DNA methylome that is altered with age [10-13], with the most recent meta-analysis estimating that over 50% of the human DNA methylome is altered [14]. An important emerging insight, still widely underappreciated, is that age-associated DNAm changes come in a large variety of different forms or shapes [15]. For instance, DNAm at many CpGs changes linearly with age, without a corresponding change in variance, whilst other CpGs may additionally display an age-related increase in DNAm variance [15, 16]. Other CpGs may only display an age-related change in DNAm variance without a corresponding change in mean DNAm-levels [17]. Superimposed on this complexity are other CpGs that display clear bi or tri-modal DNAm profiles, which may or may not be age-dependent. CpGs displaying such bi or tri-modal DNAm profiles are often driven by sex and genotype, respectively [13, 18, 19], but other underlying mechanisms are yet to be elucidated. Studying these bi and tri-modal profiles in the context of aging could be important as it could reveal sex and genotype specific age-related changes. Another class of CpGs display remarkable stability except for a few outliers, whose frequency may increase with age and which may be indicative of disease risk [20-25]. Despite our

a-priori knowledge that these patterns exist, their biological and clinical significance remains unclear.

To help elucidate the biological significance of these different age-associated DNAm patterns and how they relate to existing biomarkers such as epigenetic clocks [26], we here aimed to build a CpG classification scheme, which we then extensively characterize in terms of chromatin states, associations with 12 immune cell type fractions, epigenetic clocks, inflammaging and genotype. With this novel map of age-related DNAm patterns, we significantly improve our understanding of the aging DNAm landscape, leading to novel insights into the mechanistic components of various epigenetic clocks. For instance, our map identifies (i) CpGs representing age-dependent DNAm noise, which are underrepresented in current clocks, but that also correlate with inflammaging markers, inflammatory diseases and all-cause-mortality, and (ii) genotype specific age methylation quantitative trait loci (age-mQTLs), that according to Mendelian Randomization, constitute novel causal DNAm mediators of inflammaging and inflammatory diseases.

Results

A deep convolutional neural network to refine a human CpG taxonomy of aging

Previous studies indicate that CpG DNAm profiles come in many different shapes and forms, the biological underpinnings of each still being unclear, with a few exceptions like bi-modal sex-associated differentially methylated cytosines (sex-DMCs) [13] or tri-modal methylation quantitative trait loci (mQTLs) [27, 28]. Given how deep-learning (DL) algorithms have revolutionized scientific and technological challenges in domains where data display non-trivial spatial or temporal patterns [29] [30-34], we reasoned that DL may also be helpful to classify CpG DNAm profiles. To frame this classification task in the context of aging in a way suitable for DL, we transformed CpG DNAm profiles into two-dimensional images, with the DNAm beta-value (a numerical value bounded between 0 and 1) representing one dimension (y-axis), and with age (a numerical value bounded typically between 0 and 100 years) defining the other (x-axis). Under this transformation, the DNAm vs age scatterplots are smoothed out, defining images with pixel intensity/color quantifying the local sample density (samples here represent individuals/donors) (**Fig.1a**). Thus, just like deep convolutional neural networks (CNNs) have proved very successful for image classification tasks, a deep CNN could be trained to successfully classify CpG DNAm profiles into the distinct forms or shapes they take in relation to a phenotype such as age. To accomplish this, we devised an iterative approach, where using a large EPIC DNAm dataset of 3501 whole blood samples (the NSPT cohort), we iteratively train a CNN model called

ResNet18 [35] to classify a total of 811870 CpGs into 10 distinct categories (or taxa) (**Methods, Fig.1a-b**). We focused on 10 categories, motivated by prior-knowledge, which we note is already a substantial improvement over previous studies that have only studied three [14]. We chose *ResNet18* over other deep CNN architectures like VGG16 [36] because of its ability to train stably via residual skip connections and because its depth is sufficient to model hierarchical features in our data without being excessively parameterized [35]. In order to train the CNN-model, we defined the 10 initial categories by using a complex combination of distinct statistical tests with stringent significance thresholds (linear and Breusch-Pagan tests for heteroscedasticity) [21], statistical distributional properties (e.g. kurtosis) [37], and prior biological information (e.g. sex and mQTLs as models for bi and tri-modal patterns respectively), these categories then being refined after multiple iterations of the deep-learning algorithm (**Methods, SI tables S1-S3, SI figs.S1-S2**). Most confusion was seen between the bi and tri-modal age vs non-age categories (**SI fig.S2**), yet this was not a limitation of *ResNet18* as it significantly outperformed a more traditional cluster identification algorithm (Mclust) [38, 39] (**Methods, SI fig.S3**). *ResNet18* also outperformed a more complex deep CNN architecture (VGG16) that does not use residual skip connections (**Methods, SI fig.S4**). At the end of this process, we were able to confidently assign over 600,000 CpGs into one of ten categories (**Fig.1c-d, SI tables S1-S3**), with the rest, i.e. 200,584 CpGs, remaining unassigned due to these CpGs displaying ambiguous hybrid profiles. The ten taxa include (i) unimodal CpGs that change linearly with age without a change in DNAm variance (LinAgeHomo), (ii) unimodal CpGs where both the mean and variance in DNAm change with age (LinAgeHet), (iii) unimodal CpGs where only the DNAm variance changes with age (NonAgeHet), (iv) unimodal stable CpGs where the DNAm level displays ultra-low variance with neither the mean or variance changing with age (UniStable), (v) unimodal unstable CpGs where the DNAm level fluctuates, with neither the mean or variance changing with age (UniUnstable), (vi) CpGs displaying infrequent but distinctive outliers (Outlier) but which are otherwise stable, (vii) bi-modal CpGs where at least one group displays an association with age (BiAge), (viii) bi-modal CpGs where none of the two groups displays an association with age (BiNonAge), (ix) tri-modal CpGs where at least one group displays an association with age (TriAge) and (x) tri-modal CpGs where none of the groups displays an association with age (TriNonAge). The overall accuracy of prediction of CpGs in the blind test set was high at 0.87, with the accuracy over the ten categories ranging from 0.99 (Outlier) to 0.64 (TriAge) (**Fig.1e**).

Influence of sample size and validation in independent cohorts

To test the influence of sample size on the deep-learning predictor, we devised a simulation model generating representative CpGs for each of the 10 CpG categories,

subsequently training a ResNet18 predictor (*SimResNet18*) on simulated datasets containing variable sample sizes, and validating them in independently generated datasets of different sample size (**Methods**). We found that the *SimResNet18* model trained on a set containing variable sample sizes markedly improved prediction accuracy in the independent test sets (**SI fig.S5**). Based on this insight, any validation attempt on independent real DNAm datasets would benefit from a training process that includes variable sample sizes. Hence, we retrained our original *ResNet18* model on the NSPT cohort data but now using CpG profiles defined at two different sample sizes (*MixedResNet18*, **Methods**), subsequently validating the *MixedResNet18* model in two independent EPIC DNAm datasets (HPT & Airwave) [40, 41]. Average validation accuracy over all 10 CpG taxa was around 0.73 (HPT) and 0.79 (Airwave) (**SI fig.S6**). These validation accuracies are noteworthy given that the training set (NSPT cohort-Han Chinese) and validation sets (HPT-African Americans, Airwave-White Caucasians) represent different ancestries, i.e. we note that the true class-label of a CpG may depend on ancestry (e.g. an mQTL in White Caucasians may not be an mQTL in Han Chinese and vice-versa) [19].

ResNet18's age-CpG taxonomy is not identifiable using traditional methods

The above age-related taxonomy can't be recapitulated using more traditional linear methods like PCA or Singular Value Decomposition (SVD). Indeed, a SVD over the same DNAm dataset revealed that singular vectors (i.e. principal components) only captured general bi-modal or tri-modal patterns, mainly linked to sex and genotype, respectively, as well as a component changing linearly and homogeneously with age, but failing to capture non-linear patterns such as those marked by an increase in DNAm variance and outliers (**SI fig.S7**). For all practical purposes, the proposed taxonomy is also exhaustive. For instance, whilst the definition of our 10 CpG categories does not include non-monotonic patterns (e.g. quadratic), we verified using statistical criteria (**Methods**), that these not only constituted a small proportion of all CpGs (~1%), but most importantly that for the overwhelming majority of these, the non-monotonicity is of very low-effect size (**SI fig.S8**). We also compared our 10 CpG taxonomy to a previous 3-class categorization proposed by Seale [14], highlighting the limitations of applying differential means and differential variance statistics to bi and tri-modal DNAm profiles (**SI fig.S9**). For instance, we observed how a significant number of CpGs, which according to Seale's classification scheme display increased variance with age but no change in mean DNAm, were bi or tri-modal in our categorization (**SI fig.S9**), highlighting how differential variance statistics, when applied to multimodal data, can drive artefactual associations with age.

Biological significance of *ResNet18*'s CpG taxonomy

To explore biological relevance of the novel CpG taxonomy, we selected the top 1000 CpGs of each category and tested each for enrichment against 15 chromatin states (as defined over 26 tissue-types [42, 43]), and histone H3 marks (as defined in immune cell-types [44]), using in both cases the eFORGE2 enrichment tool [45, 46]. Focusing first on unimodal CpGs where DNAm changes linearly with age, we observed a striking difference in enrichment between those changing homoscedastically with age (i.e. no change in DNAm variance with age, LinAgeHomo) from heteroscedastic ones (displaying increases in DNAm variance, LinAgeHet) (**Fig.2a**): homoscedastic ones displayed enrichment for enhancer marks and regions flanking active promoters, whilst heteroscedastic ones displayed enrichment for bivalent enhancer states and repressive polycomb group (PCG), the latter being the most strongly enriched term. As expected, the LinAgeHomo enrichment for enhancer regions was strongest for immune cell-types compared to other cell and tissue-types (**SI fig.S10**). CpGs not changing linearly with age but displaying increased DNAm variance (NonAgeHet) were exclusively enriched for repressive PCG states. Put together, and given that enhancers are the most cell-type specific regions [47], this suggests that the linear increases in age are partly associated with age-related linear changes in immune cell-type composition, whereas the increased variability in DNAm is targeting H3K27me3-marked repressive PCG regions. To explore if these age-related patterns are associated with changes in immune-cell type composition, we used EpiDISH [48, 49] to estimate fractions for 12 adult immune-cell types [48, 49] (**SI fig.S11**). This confirmed that age-related linear homoscedastic changes were associated with shifts in the naïve CD4+ T-cells and neutrophil fractions, but interestingly that increased DNAm variance with age is driven by increased variability in the memory CD8+ T-cell fraction (**Fig.2b**).

CpGs displaying variability in DNAm, but whose mean or variance did not associate with age (i.e. the Outlier and UniUnstable taxa), displayed specific enrichment for strongly and weakly transcribed regions, respectively, both marked by the elongation mark H3K36me3 (**Fig.2a**). In contrast, unimodal stable CpGs (UniStable) displayed strongest enrichment for active and flanking promoters, marked by the H3K4me1 and H3K4me3 marks. Of note, the great majority of top-ranked CpGs in this category all displayed low DNAm levels (**SI fig.S12a**) and were enriched for active promoters in all cell and tissue-types (**SI fig.S10**), suggesting that these CpGs map to homeostatically vital genes. In this UniStable category we also observed many CpGs that were fully methylated and strongly enriched for H3K36me3-marked actively transcribed regions (**SI fig.S12b-c**), mapping to the transcribed gene-body regions of the same genes (**SI fig.S12d**). We verified, using two of the largest single-cell RNA-Seq datasets, each profiling over 1 million peripheral blood mononuclear cells [50, 51], that the genes containing UniStable CpGs are indeed actively expressed compared to genes not

containing any such CpGs (**SI fig.S13**). This effect was much stronger in the AIDA dataset of East Asians [50] compared to White Caucasians, consistent with our training set being of East Asian ancestry, and suggesting that ethnicity has some influence on which CpGs are UniStable (**SI fig.S13**).

By construction, bi- and tri-modal CpGs are strongly enriched for sex-DMCs and mQTLs, respectively (**Methods**). Thus, the observed enrichment of tri-modal CpGs for enhancer states (**Fig.2b**) is consistent with mQTLs being enriched for enhancers [18, 19]. We asked if CpGs not used in training and which were confidently assigned to the bi or tri-modal categories upon subsequent iterations of ResNet18, were also driven by sex or genotype, respectively. In the case of newly assigned bi-modal CpGs, the great majority of these were indeed associated with sex (albeit at a lower significance threshold) and those that were not, were preferentially enriched for low-effect size mQTLs (**SI fig.S14a**), probably due to ResNet18 not being able to discern the 3 genotype states when effect sizes and minor allele frequencies are small. In the case of newly assigned tri-modal CpGs, these were effectively all mQTLs characterized by smaller effect sizes (**SI fig.S14b**). This indicates that effectively all tri-modal DNAm patterns in real DNAm datasets are invariably driven by genotype.

Differential enrichment patterns of clock CpGs among CpG taxa

We next explored the representation of CpG taxa among different epigenetic clocks [6, 26]. We assembled DNAm clocks ranging from those highly accurate for chronological age prediction (e.g. Zhang clock [52]) to those that are better in predicting biological age (e.g. PhenoAge and DunedinPACE) [8, 53], as well as mitotic clocks based on CpGs that gain DNAm during cell-division (e.g. epiTOC1/2 [54, 55], stemTOC [56]) and mitotic clocks based on CpGs that lose DNAm with cell-division (e.g. HypoClock [57], RepliTali [58]). We also included recent causality related clocks (DamAge and AdaptAge) [59]. Not surprisingly, the strongest enrichment of LinAgeHomo CpGs was found for the Zhang clock, highlighting how homoscedastic features are favored when the aim is to accurately predict chronological age (**Fig.2c**). Interestingly, the strongest enrichment of LinAgeHet CpGs was observed for mitotic clocks. Of note, like LinAgeHet CpGs, mitotic clock CpGs are strongly enriched for repressive PCG target sites [54]. The TriAge category, representing mQTLs where DNAm also correlates with age, was most strongly enriched for the causality clocks DamAge and AdaptAge, consistent with these clock CpGs being derived from mQTLs [59]. Interestingly, the PhenoAge and DunedinPACE clocks showed enrichment of LinAgeHomo and NonAgeHet categories but not so for LinAgeHet (**Fig.2c**). The observed enrichment patterns were robust to the particular iterative version of ResNet18 (**Methods, SI fig.S15a**). Overall, these enrichment patterns are consistent with the clock's method of construction and their biological significance.

Age-dependent epigenetic noise is associated with inflammaging and mortality

NonAgeHet CpGs can be interpreted as measuring age-dependent noise, with a component of this noise being driven by variations in memory CD8+ T-cell and neutrophil fractions (**Fig.2b**). Except for the PhenoAge and DunedinPACE clocks, current epigenetic clocks are clearly devoid of the NonAgeHet category, thus motivating the construction of a new clock (NoiseClock) that is exclusively enriched for these CpGs (**Fig.2c**). To build this NoiseClock, we identified a subset of 399 NonAgeHet CpGs (**SI table S4, Fig.3a**) using a total of 3 large independent DNAm datasets from healthy individuals [60-62], each encompassing around 700 samples and representative of different ancestries (Han Chinese, White Caucasians, Hispanics). These 399 NonAgeHet CpGs were identified using the Breusch-Pagan test for differential DNAm variance, after adjusting the DNAm data in each of these 3 cohorts for variations in 12 immune-cell fractions (**Methods**). This adjustment was done in order to assess the biological significance of NonAgeHet CpGs beyond their variation attributed to changes in immune-cell composition. The 399 NoiseClock CpGs displayed the same enrichment for H3K27me3 marked repressive PCG loci as the full NonAgeHet set (**SI fig.S16**), and were, generally speaking, less correlated than randomly chosen CpGs, although still significantly higher than 0 (**SI fig.S17**). Based on these 399 CpGs, we then estimated the level of noise or DNAm variation in each sample (**Fig.3a-b, Methods**). This NoiseClock estimator was then applied to a collection of 14 independent blood DNAm datasets (**Methods**), revealing a strong consistent association with age (**Fig.3b**), thus justifying the term ‘clock’. To further demonstrate that our NoiseClock correlates with age independently of variations in cell-type fractions, we applied it to 5 DNAm datasets profiling sorted immune cells (**Methods**): in all cases, the NoiseClock correlated with age, although interestingly the effect was stronger in myeloid cells (**SI fig.S18**). To explore its biological significance, we observed that it correlated fairly strongly with a published inflammaging DNAm-signature consisting of 1765 CpGs whose DNAm-levels are associated with high-sensitive serum C-reactive protein (hsCRP) levels [63] (**SI fig.S19a**). However, the original EWAS meta-analysis identifying these 1765 CpGs did not adjust for shifts in the memory to naïve lymphocyte subsets [63], and correspondingly we found strong associations between the DNAm-based proxy of CRP-levels with immune-cell fractions, notably with the memory CD8+ T-cell fraction (**SI fig.19b**). Thus, we re-computed a DNAm-based proxy for serum CRP-levels restricting to a subset of 62 CpGs that are not discriminative of naïve and memory lymphocyte subsets (Intrinsic Inflammatory Score-IntInflScore, **Methods, Fig.3c**). The NoiseClock still displayed a very clear correlation with the IntInflScore, independent of age, sex and variations in 12 IC-type fractions (**Fig.3d**). Of note, the association of the NoiseClock with the

intrinsic inflammatory CRP-score was weaker than that displayed by biological aging clocks, notably GrimAge2 and DunedinPACE (**Fig.3e**), consistent with these latter clocks having been partly trained on clinical aging biomarkers that included CRP. Besides the correlation with hsCRP-levels, the NoiseClock also displayed positive associations with DNAm-based proxies for CXCL9, TNFRSF1B [64] and the inflammatory component of SystemsAge [65], but interestingly not for IL6 [66] (**SI fig.S20**). The NoiseClock also demonstrated significant associations with prevalent inflammatory diseases in the NSPT cohort, including type-2 diabetes, rheumatoid arthritis, and hypertension (**Fig.3f**). These associations were robust to variations in immune-cell fractions (**SI table S5**). Interestingly, the NoiseClock also displayed an association with smoking exposure, although this was weak and only evident after a meta-analysis over multiple cohorts (**SI fig.S21**). Of note, although the 399 NoiseClock CpGs were derived from 3 separate cohorts, we obtained almost identical results had we built the clock using the ResNet18 NonAgeHet CpGs from the NSPT cohort (**SI fig.S22**). Overall, this supports the robustness of our procedure and further indicates that NonAgeHet CpGs include CpGs whose DNAm-levels are associated with inflammaging in a manner that is also independent of variations in the naïve and memory T-cell subsets.

To further explore potential clinical significance, we computed the NoiseClock in the Generation Scotland (GenS) cohort, which is one of the largest biobanks with available genome-wide DNAm data [67, 68] (**Methods**). The NoiseClock estimate was significantly correlated with age ($R \sim 0.4$, $P < 2e-16$), confirming that it captures DNAm-variance that increases with age. Next, we assessed the NoiseClock in relation to all-cause mortality. Despite the relatively strong correlation with age, the NoiseClock remained significantly associated with all-cause-mortality after adjustment for age, sex, as well as other main risk factors (**Methods, Fig.3g**). Whilst it also added prognostic value over DunedinPACE [53], this was driven entirely by its association with age (**Fig.3g**), and it did not add prognostic value over GrimAge2 [69]. Thus, we conclude that whilst the NoiseClock is an independent predictor of all-cause-mortality (HR=1.07, 95%CI: 1.02-1.13, $P=0.003$), that current mortality predictors like GrimAge2 or DunedinPACE are indirectly capturing this DNAm-noise contribution.

Genotype specific age-mQTLs are depleted of DamAge/AdaptAge CpGs

Another interesting category of CpG profile are those displaying tri-modality where at least one of the modes displays an association with age (TriAge). Since the overwhelming majority of tri-modal CpGs constitute mQTLs where the DNAm level is strongly correlated with a SNP, either *in-cis* or *in-trans* [19], we refer to the subset correlating with age, as age-mQTLs. Whilst the biological significance of mQTLs has

been extensively studied [18, 19], much less is known about these age-mQTLs. Starting out from 273,613 mQTL CpGs identified from 3501 samples of the NSPT cohort (Linear model Bonferroni $P < 1e-14$) and validated in independent cohorts [19], a total of 196,939 were associated with age, independently of genotype, sex and variations in seven major IC fractions ($FDR < 0.05$) (**Fig.4a**). Of note, here we initially did not adjust for 12 IC fractions, because SNPs influencing shifts between naïve and memory T-cell subsets could be important biologically and fairly little is known about such SNPs. We further reasoned that age-mQTLs where only the homozygous genotype for the minor allele (i.e. BB), or where both the heterozygous AB and homozygous BB display an association with age, would be of particular interest, since for these, age and genotype may synergize to increase disease risk. Among the full list of age-mQTLs, a subset of 3,139 displayed an association with age that was specific to the BB ($n=1424$) or AB+BB ($n=1715$) genotypes ($FDR < 0.05$) (**Fig.4a**). Of these BB and AB+BB specific age-mQTLs, around 35% and 32% respectively, displayed a positive interaction between genotype and age, i.e. DNAm changed with genotype (AA=1, AB=2, BB=3) in the same direction as with age.

Since age is the major demographic risk factor for many diseases, we reasoned that the change of DNAm with age in specific high disease risk B genotype groups could help identify causal DNAm mediators of disease. To explore this hypothesis, we first studied the overlap of age-mQTLs with those of recently proposed causality-clocks (DamAge and AdaptAge) [59]. Whilst we observed a very strong enrichment for DamAge and AdaptAge CpGs among age-mQTLs generally, this enrichment was substantially weakened for the BB and AB+BB specific age-mQTLs (**Fig.4b**). Given that the odds ratios of enrichment were also significantly lower for the BB and AB+BB specific age-mQTLs, this means that the differential enrichment is not merely driven by the much larger number of CpGs in the general age-mQTL class. This demonstrates that DamAge and AdaptAge CpGs select for age-mQTLs where the association with age is present in all genotype groups, and hence that they are depleted for age-mQTLs where the association with age is genotype-specific.

Genotype specific age-mQTLs are associated with genotype specific inflammaging

To explore the biological relevance of the BB and AB+BB age-mQTLs, we next asked if the age-related DNAm variation in specific genotypes is reflected in corresponding genotype specific correlations with DNAm-based measures of C-reactive protein (CRP) and interleukin-6 (IL6) levels, two well-known markers of inflammaging. Indeed, since chronic low-grade inflammation ('inflammaging') [70] is known to increase with age, it is plausible that the observed change of DNAm with age in a specific genotype is linked to genotype specific inflammaging. Validating our hypothesis, we observed hundreds of BB and AB+BB age-mQTLs where only the DNAm variation in the BB and

ABBB genotypes correlated significantly with the inflammaging scores (**Fig.4c**). Of note, effectively all of these displayed the same directionality of DNAm change with inflammaging as with age (**Fig.4d**). To assess if the linked SNPs are enriched for disease associated variants, we interrogated GWAS catalogs (**Methods**), which revealed strong enrichment for SNPs associated with inflammation-related conditions, including coronary artery disease (CAD), rheumatoid arthritis (RA), Crohn's disease (CD) and ulcerative colitis (UC) (**Fig.4e**). Examples of ABBB age-mQTLs associated with CAD and type-2 diabetes are depicted in **Fig.4f** (see **SI fig.S23** for other examples). BB and ABBB age-mQTLs displaying associations with CRP/IL6 and with linked SNPs associated with inflammatory disease were also found had we adjusted for naïve and memory lymphocyte fractions using a higher resolution 12 IC-type DNAm reference panel (**SI fig.S24**). Overall, these data support the view that genotype specific age-mQTLs are linked to genotype specific age-associated inflammaging, with the linked SNPs often also associated with inflammatory diseases.

Genotype specific age-mQTLs as potential mediators of inflammatory diseases

We reasoned that these genotype specific joint age and inflammaging mQTLs could be casual mediators of inflammatory disease. To formally test this hypothesis, we used 2-sample Mendelian Randomization (MR) [71, 72], treating the DNAm levels of our genotype specific age-mQTL CpGs as exposures, and with linked SNPs (i.e. SNPs defining mQTLs with the same CpG) as genetic instruments (**Methods, Fig.5a**). As outcome, we considered a range of different GWAS inflammatory traits and diseases including CRP-levels, CAD, RA, type-2 diabetes (T2D), systemic lupus erythematosus (SLE) and ulcerative colitis (UC), obtaining summary statistics of SNP-trait associations from an East Asian database (**Methods**), thus matched to our mQTL East Asian cohort data. Using inverse-variance weighted (IVW) and weighted median (WM) regression, we found a total of 21 exposure-outcome pairs, encompassing 16 candidate causal CpGs (**SI table S6-S7**). Of note, using Steiger-tests [73] we found that our genetic instruments (a total of 370 unique SNPs) were valid, as they explained more of the variance of the exposure (i.e. DNAm) than of the outcome (**SI table S8**). Whilst Egger regression tests indicated little evidence of average directional pleiotropy, pleiotropy-adjusted causal estimates were no longer significant (**SI table S7**). Since Egger-regression lacks power and to explore if the assumptions underlying Egger-regression hold, we performed MR-PRESSO [74] outlier analysis for the 18 CpG-trait pairs with sufficient numbers of SNPs. Whilst for all 18 CpG-trait pairs, causal associations remained significant after adjustment for pleiotropic outliers, 12 pairs displayed few or no outliers, with the remaining 6 pairs displaying pervasive pleiotropy (>20% of SNPs were pleiotropic) (**SI table S9**). Leave-One-Out (LOO) analysis further confirmed that results from IVW and WM were robust to outliers except for

cg19901956-T2D and cg22933800-SLE (**SI fig.S25**). Co-localization analysis using a hybrid SUSIE-ABF strategy [75-77] (**Methods**) revealed a potential common causal variant for 7 CpG-outcome pairs, with another 2 displaying evidence of multiple variants (**SI table S10, SI figs.S26-S27**). Overall, from all the above analyses, we identified the following candidate mediator CpG-trait pairs that merit further investigation. For instance, although cg00119778, with the corresponding SNP being associated with CRP-levels (**Fig.5b**), was not strongly associated with a single causal variant, the region it maps to has been associated with late-onset Alzheimer's disease [78-80], providing a potential link between inflammaging and neurodegeneration [81]. Another example is cg16912114 (**Fig.5b**), which maps to the gene body of *GALNT14*, and with the corresponding trans-mQTL SNPs having been associated with CAD [82]. Thus, DNAm at cg16912114 may mediate risk of CAD via age-associated inflammaging. Yet another example is cg22933800, which maps to the body of the MHC2-class HLA-DQA1 gene (**Fig.5b**), a locus that is strongly linked to ulcerative colitis [83]. Thus, our data points towards age-related DNAm change in HLA-DQA1 as a potential modulator of this molecule's immunogenicity, increasing UC risk. Overall, these results indicate that genotype specific age-mQTLs could be potential mediators of age-associated inflammation predisposing to inflammatory diseases.

Discussion

By transforming age-related DNAm patterns into 2D images and subsequently using a deep convolutional neural network to refine a classification of age-associated DNAm patterns, we have here shown that about 70% of CpGs can be confidently assigned into one of 10 categories. We note that such a categorization would have been difficult if not impossible to construct using traditional statistical tests, clustering algorithms or linear methods, since all of these are unable to capture the full complexity and variety of patterns present in the aging DNA methylome. By then characterizing the inferred taxonomy of age-associated DNAm patterns in terms of chromatin-states, genotype, correlations with immune-cell fractions and enrichment for various epigenetic clocks, we have obtained many novel insights that significantly advance our understanding of the aging DNA-methylome and the mechanistic components of epigenetic clocks. Among these insights we highlight and discuss the following. First, CpGs that display increased DNAm variance with age, and which generally map to repressive PCG and bivalently marked enhancer regions, exhibited surprisingly strong correlations with the memory CD8+ T-cell fraction, thus pointing towards age-associated increased variation of this particular IC fraction as a source of age-related DNAm variability. Interestingly, among these CpGs, the subset displaying an increased average DNAm with age (LinAgeHet), was most strongly enriched for mitotic clock

CpGs. This is consistent with how mitotic clocks have been constructed, as these clocks have not been trained to predict chronological age (hence why the underlying CpGs are not homoscedastic). Instead, the increased DNAm variability with age likely marks subclonal expansions that capture surges in mitotic age [84].

The other subset of heteroscedastic CpGs (NonAgeHet) was generally speaking not enriched for any of the previously published epigenetic clocks, justifying a more in-depth exploration of its biological significance, which we did by building a tailored ‘NoiseClock’. These ‘Noise CpGs’ have recently been proposed, although not proven, to be markers of biological age [17]. Here we have unequivocally shown that noise-CpGs correlate with inflammaging and inflammatory diseases including T2D, RA and hypertension. Importantly, these associations were independent of variations in 12 IC-type fractions, age and sex, suggesting that our 399 NoiseClock CpGs are marking inflammaging-associated DNAm changes in individual IC-types. This is an important insight because previous EWAS meta-analyses for inflammaging markers, such as serum-CRP levels, did not adjust for variations between memory and naïve T-cell fractions [63]. Consequently, we found that of the 1765 CRP-associated CpGs reported by [63], only 62 correlate with serum-CRP levels in a way that is independent of variations between memory and naïve lymphocytes. As assessed in the large Generation Scotland cohort, our NoiseClock also correlated with all-cause mortality, adding prognostic value over age, sex and other main risk factors that included smoking, alcohol consumption and obesity, further demonstrating its biological relevance. However, when combined with age and GrimAge/DunedinPACE, it no longer added prognostic value, suggesting that these epigenetic clocks, by virtue of being trained on clinical biomarkers, are indirectly capturing an inflammatory DNAm-noise component.

Another key insight has been the discovery of numerous genotype specific age-mQTLs where the genotype-specific variation with age appears to be associated with genotype-specific variation in inflammaging. Importantly, these genotype specific age-mQTLs were depleted among CpGs that constitute recently proposed causal clocks [59]. As such, genotype-specific age-mQTLs are likely to constitute an orthogonal set of ‘causal’ CpGs whose mediation effect is specific to certain genotypes. Supporting this, we performed MR, which identified a number of age-mQTLs consistent with DNAm mediating the effect of genotype on inflammaging and inflammatory diseases. Besides the examples mentioned earlier, here we highlight a few more: (i) cg11362640 mapping to the gene-body of *IDE* (Insulin Degrading Enzyme) forms mQTLs with SNPs associated with T2D. Indeed, genetic variation in *IDE* has been associated with T2D [85-87] and insulin secretion is impaired in *Ide* KO mice [88]. Notably, IDE also degrades amyloid β -peptide, the key component of Alzheimer’s disease (AD)-associated senile plaques, and genetic variants in *IDE* have been associated with AD

[89]. Another example is cg17849733, mapping to the TSS1500 region of *PSORS1C1*, a gene which has been implicated in RA risk [90], and for which both outlier pleiotropy and co-localisation analyses were supportive of mediation. Whilst DNAm of the CpG itself has only been implicated in osteoarthritis (OA) [91], inflammation is a major contributing risk factor to OA [92], consistent with the observed correlation with inflammaging.

We acknowledge limitations. First and foremost, around 20% of the 850k EPIC probes could not be confidently assigned to one of our 10 CpG categories. These CpGs were hard to classify because they mapped to a continuum between categories. However, this fuzziness is typical of many classification problems in biomedical science and does not negate the value of our classification scheme. Second, our categorization into 10 taxa should be understood as an initial classification that could be expanded upon. For instance, our classification did not consider baseline DNAm levels. Thus, as shown for the UniStable category, low baseline DNAm levels correspond to CpGs that map to the promoters and 1st exon regions, whilst high baseline levels map to the gene body. In both cases however, this category marks the same actively transcribed genes in blood. Another example is the LinAgeHomo and LinAgeHet categories, as we did not dichotomize these categories further into those displaying increased or decreased average DNAm with age. A third limitation is that our map of age-related DNAm patterns was derived from a cohort representative of only one ancestry (Han Chinese) [93, 94]. Although we were able to successfully validate our ResNet18 classification model in two cohorts representing different ancestries (African Americans and White Caucasians), validation accuracy for some CpG categories was relatively low. However, even according to ordinary statistical criteria, specific CpGs could map to different age-categories depending on ancestry, for instance a UniStable CpG in Han Chinese may belong to the outlier CpG category in African Americans. This therefore highlights the complexity of the human DNA methylome and is not a limitation of the deep-learning approach. We also acknowledge that we did not explicitly consider other CpG subcategories. For instance, there is growing interest in CpGs where DNAm may change super-linearly (i.e. exponentially) with age [95, 96]. However, these CpGs would invariably also display an increased variance with age and hence constitute a subset of our LinAgeHet category. Assessing if this super-linear subset is biologically distinct from our LinAgeHet category, say in terms of chromatin state, epigenetic clock and immune cell fraction associations would be an interesting avenue for future investigation. We have also observed a small fraction of non-monotonic CpGs that display an age-related quadratic pattern, yet the non-monotonicity of these was typically of low effect size (1-5%) and often associated with high variability or multi-modality, raising doubts as to their true biological relevance. Other categories such as outliers, which bear a resemblance to UniStable CpGs, except for a small number of

samples displaying big deviations in DNAm, also justify a more in-depth future exploration. In addition, it should be emphasized that our approach to classify age-related DNAm patterns should be viewed as complementary to gerophysical methods that aim to model such patterns mechanistically [96-103]. Future work should integrate or interpret gerophysical and machine-learning approaches together, as this could lead to important novel insights.

In summary, by performing an in-depth characterization of the aging DNAm landscape into 10 distinct age-related DNAm patterns, we have significantly advanced our biological understanding of these patterns, whilst also shedding new insights into the mechanistic components of various epigenetic clocks, and how these relate to key age-related processes including shifts in immune cell composition, genotype and inflammaging. Demonstrating the value of our approach, we have identified a number of genotype specific age-mQTLs that may mediate the risk of inflammatory diseases.

Methods

DNA methylation datasets for training and validation

The NSPT (EPICv1) DNAm dataset [19] was used for training our convolutional neural network model. Briefly, the NSPT cohort encompasses whole blood samples from Han Chinese individuals. DNAm data was generated with the EPICv1 platform and processed+normalized, as described in Peng et al [19]. After normalization, the DNAm dataset comprised 811,870 probes and 3501 samples, with a mean age of 50 years, standard deviation of +/-13, and an age range of 17 to 83 years. A subset of the NSPT cohort, consisting of 710 samples [60], all from the same batch, is referred to as the 'TZH' cohort.

For independent validation of our CpG classification model, we utilized the HPT-EPIC DNAm dataset derived from whole blood samples of African-American participants in the Genetic Epidemiology Network of Arteriopathy (GENOA) study [40]. Raw IDAT files were downloaded from the NCBI GEO websites under the accession number GSE210255 (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE210255>, Infinium HumanMethylationEPIC BeadChip), corresponding to 1,394 samples. Data preprocessing was performed using the *minfi* package in R [104], where CpG probes failing to meet a detection P-value threshold of <0.05 across all samples were systematically excluded. Following quality control, BMIQ normalization was applied to correct for type-2 probe bias [105]. This yielded a normalized DNAm matrix encompassing 826,512 CpGs and 1,394 samples. The other independent EPIC DNAm dataset derives from the Airwave study, which is predominantly

White Caucasians [41]. This DNAm dataset of 1032 whole blood samples was processed and normalized as described by us previously [48].

Generation of smoothed DNA methylation images

The DNAm beta-valued profile of each CpG site (y-axis) in the NSPT DNAm dataset was plotted against chronological age (x-axis), and converted into a corresponding smoothed density image. In more detail, we first linearly scaled the chronological age with min-max scaling so that the x-axis ranges from 0 to 1. The DNAm beta-value is naturally already bounded between 0 and 1. Both y and x-axes were subsequently removed, as we don't want them in the final image. To create the smoothed density image, we utilized a two-dimensional kernel density estimation (via the `bkde2D` function from `KernSmooth` R-package, with a bandwidth of 0.005 and a grid size of 500) to generate the density plot of beta values against scaled age. To enhance visualization, the density values in each image were normalized to the [0,1] interval using min-max scaling, and mapped to a preset color gradient, transitioning from light green to dark green to reflect increasing sample density levels. Finally, the generated images were saved in PNG format with dimensions of 224x224 pixels at a resolution of 300 DPI. The same procedure was applied to all independent DNAm datasets.

Of note, the density values obtained by a smoothed kernel density procedure depend on the bandwidth used in the x and y dimensions. The ideal bandwidth to use is not only dependent on the spread of datapoints in each dimension, but also on the overall number of datapoints n . Silverman's formula [106] can help identify ideal bandwidths to use. In our context, Silverman's formula specifies the bandwidths (B_x, B_y) in terms of an additional scaling factor sf (typically $sf=1.5$) as follows:

$$B_x = sf * \min(sd(age), IQR(age)/1.34) * (n^{-1/6})$$

$$B_y = sf * \min(sd(beta), IQR(beta)/1.34) * (n^{-1/6})$$

where age and $beta$ are the age and DNAm values of the datapoints, sd is the standard deviation and IQR =inter-quartile range. However, using simulation models we found the Silverman formula to not accurately capture sample-size induced variations in the density value estimates. To adjust for sample size effects and ensure that the ResNet18 models are not overfitting to artefacts caused by sample size, we found that training CpGs using DNAm profiles with variable sample sizes is more effective. Hence, this latter procedure was used here.

ResNet18 model architecture

ResNet-18 is a powerful convolutional neural network (CNN) architecture designed for image classification that leverages residual blocks with skip connections (identity maps) to mitigate multiplicative shrinkage (vanishing gradients) [35]. This enables more effective training of deep CNNs compared to more traditional CNNs like VGG16 [36]. Specifically, with residual

blocks, the output function $H(x)$ to be learned is reformulated as

$$H(x) = \mathcal{F}(x) + x$$

where $\mathcal{F}(x)$ is the residual function. The network processes input images of dimensions $3 \times 224 \times 224$ (channels $\times$ height $\times$ width) through an initial 7×7 convolutional layer (stride 2) and a 3×3 max-pooling layer (stride 2) for feature extraction and spatial reduction. Subsequent processing involves four sequential residual blocks, each composed of two basic blocks, where each basic block contains two 3×3 convolutional layers (3×3 kernels) W_1 and W_2 integrated with skip connections to facilitate residual learning. In more detail, the two-layer basic block takes the form

$$\mathcal{F}(x) = W_2 \text{ReLU}(W_1 x)$$

For a residual block with input x_l , the forward pass is then

$$x_{l+1} = \text{ReLU}(x_l + \mathcal{F}(x_l, \{W\}_l))$$

Backpropagation results in

$$\frac{\partial \mathcal{L}}{\partial x_l} = \frac{\partial \mathcal{L}}{\partial x_{l+1}} \left(1 + \frac{\partial \mathcal{F}(x_l)}{\partial x_l} \right)$$

which avoids multiplicative shrinkage. The overall configuration results in 16 convolutional layers across all residual blocks, combined with one initial convolutional layer and a final fully connected layer, yielding the network's namesake 18-layer structure. The architecture concludes with global average pooling and a fully connected layer that outputs 10-dimensional predictions for CpG category classification, maintaining efficient gradient propagation and hierarchical feature learning through its residual connections.

We implemented the training of the ResNet18 model using PyTorch version 2.4.1, directly sourced from torchvision version 0.19.1. The training process employed the Adam optimizer, with an initial learning rate of 0.01. To dynamically adjust the learning rate during training, a ReduceLROnPlateau scheduler was utilized. This scheduler modifies the learning rate α based on the validation loss. If the validation loss does not exhibit improvement over a specified number of consecutive epochs (patience), the learning rate is reduced. Let α_e be the learning rate at epoch e . Let $L_{val}^{(e)}$ be the validation loss at epoch e . The learning rate update rule is

defined as: if $L_{val}^{(e)}$ has not reduced for p consecutive epochs (where patience $p=3$ in this study), then:

$$\alpha_{e+1} = \alpha_e \times factor \quad (1)$$

In this implementation, the variable *factor* was set to 0.1, thereby reducing the learning rate by 10%. If on the other hand the validation loss is reduced, we set $\alpha_{e+1} = \alpha_e$.

For the multi-class classification task, we selected the cross-entropy loss function. The loss was computed by averaging over each batch, with all classes given equal weight. For a given batch of N CpGs, the overall loss L is the average of the individual losses of each CpG. In detail, for the n -th CpG within a given batch, let $y_{n,j}$ be proportional to its probability belonging to the

j 'th class, and let v_n be the true class index for this CpG. The loss L_n for the n -th CpG is:

$$L_n = -\log\left(\frac{\exp(y_{n,v_n})}{\sum_{k=1}^c \exp(y_{n,k})}\right) \quad (2)$$

where $c = 10$ represents the total number of classes. This value effectively penalizes the model more heavily when it assigns a low probability to the correct class. The total loss for the batch, L , is then the average:

$$L = \frac{1}{N} \sum_{n=1}^N L_n = -\frac{1}{N} \sum_{n=1}^N \log\left(\frac{\exp(y_{n,v_n})}{\sum_{k=1}^c \exp(y_{n,k})}\right) \quad (3)$$

This formulation measures the dissimilarity between the predicted probability distribution and the true distribution (one-hot encoded) of the class labels across the batch. With respect to hyperparameters, a batch size of 64 was selected, and a total of 40 training epochs were conducted.

Defining an initial CpG taxonomy for training ResNet18 models

Using the NSPT DNAm dataset encompassing 3501 samples and 811,870 CpGs, and based on extensive prior experience, we defined an initial set of ten categories/taxa of CpGs: LinAgeHomo (CpGs changing linearly and homoscedastically with age), LinAgeHet (CpGs changing linearly and heteroscedastically (typically increased variance) with age), NonAgeHet (CpGs for which average DNAm does not change with age, but for which the variance changes with age (typically this always involved increases in variance)), Outliers (CpGs that are generally very stable except for a relatively small number of outliers), BiNonAge (bi-modal CpGs where none of the two modes correlates with age), BiAge (bi-modal CpGs where at least one of the two modes correlates linearly with age), TriNonAge (tri-modal CpGs where none of the 3 modes correlates with age), TriAge (tri-modal CpGs where at least one of the 3 modes correlates with age), UniStable (unimodal CpGs displaying ultra-low DNAm variance and thus not correlated with age) and UniUnstable (unimodal CpGs but displaying increased variance and for which neither mean or variance changes with age). To define an initial high-confidence set of CpGs for each of the ten above categories, we used a procedure that involves statistical testing (linear models to test for age-related changes of mean DNAm and Breusch-Pagan tests to test for homoscedasticity vs heteroscedasticity [21]), consideration of higher order statistical moments (notably scaled unbiased Kurtosis [37]) and biological criteria (sex and mQTLs as models for bi and tri-modal CpGs, since this worked better than using statistical models for cluster number prediction). Specific details of the thresholds used for defining each category can be found in **Supplementary Methods**. Finally, CpGs that did not pass various thresholds and hence did not conform to these initial categories were assigned to an initial 'Unclassified' category. The CpG counts for each category following the initial classification were as follows:

LinAgeHomo=5,401; LinAgeHet=139,638; NonAgeHet=10,205; Outlier=1,566; BiNonAge=1,167; BiAge=4,152; TriNonAge=4,003; TriAge=5,200; UniStable=24,435; UniUnstable=1,880, Unclassified=614,223.

Benchmarking of ResNet18 against VGG16

VGG16 is a seminal deep CNN architecture proposed by the Visual Geometry Group at Oxford University [36]. The core innovation of VGG16 lies in the stacking of small-scale (3×3) convolutional filters to replace larger kernels (e.g., 5×5 or 7×7), thereby increasing network depth while preserving the effective receptive field. Structurally, VGG16 comprises 16 weight layers, including 13 convolutional layers and 3 fully connected (FC) layers. Input images with dimensions of 3×3×224 are processed through five sequential convolutional blocks, each followed by a 2×2 max-pooling layer to progressively downsample the feature maps to a final resolution of 7×7. The architecture concludes with three FC layers culminating in a 10-dimensional output. To ensure a rigorous and unbiased comparison with ResNet18, VGG16 was trained and evaluated using the identical data splits (training, validation, and test sets) from the first iteration of the NSPT dataset (i.e., the training and evaluation of CpGs categorization for ResNet18-iteration0). While we aimed for maximum hyperparameter consistency, the initial learning rate for VGG16 was adjusted to 0.0001 (compared to 0.01 used for ResNet18). This modification was empirically necessary, as preliminary trials with a higher learning rate (learning rate=0.01) failed to achieve effective convergence. In those instances, VGG16 remained trapped at a baseline accuracy level equivalent to random chance, suggesting that the weights likely succumbed to vanishing or exploding gradients during initialization—a known susceptibility in deep sequential architectures lacking residual connections. Except for the initial learning rate adaptation, all other training protocols remained strictly aligned with the ResNet18-iteration0 configuration. The training process employed the Adam optimizer coupled with a ReduceLROnPlateau scheduler, which reduced the learning rate to 10% of its current value if the validation loss failed to improve for three consecutive epochs. We utilized cross-entropy loss to supervise the multi-class classification task over 40 training epochs with a batch size of 64.

ResNet18 Iterative Training for CpG categorization

To train the ResNet18 model, we randomly selected 1,000 CpGs from each category, resulting in a total of 10,000 CpGs. Subsequently, these data were randomly divided in an 80:20 ratio into a training set (containing 8,000 CpGs) and validation set (containing 2,000 CpGs). The remaining CpGs from each category constituted the test set (n=187,647). We employed the test set to evaluate the trained ResNet18 model, referred to as *ResNet18-iteration0*, and calculated the confidence for each CpG site within the test set. Confidence was defined as the difference between the highest and the second-highest probabilities in the model predictions. The

ResNet18-iteration0 model was subsequently applied to predict the unclassified CpGs. Initially unclassified CpGs which were predicted to be in a particular category with a confidence greater than 0.9 were assigned to that category, resulting in a cumulative total of 540,145 CpGs assigned to any one category, with 271,725 CpGs remaining unclassified. This procedure was then iterated, now randomly selecting 1,600 CpGs from each category, totaling 16,000 CpGs, and randomly dividing them in an 80:20 training-test set ratio (training set 12,800 CpGs, validation set 3,200 CpGs). The remaining CpGs from each category constituted the test set (n=524,145). We then trained the *ResNet18-iteration1* model, and this model was then applied to the remaining unclassified CpGs, with those exhibiting a confidence greater than 0.9 being added to the identified categories, resulting in a total of 611,286 classified CpGs, with 200,584 CpG remaining unclassified. Subsequently, we randomly drew 3,200 CpGs from each category, resulting in a total of 32,000 CpGs, which were then randomly split into a training set (25,600 CpGs) and a validation set (6,400 CpGs), while the remaining CpGs from each category constituted the test set (n=579,286). Finally, the merged data were input into the ResNet18 model to obtain the trained *ResNet18-iteration2* model. This *ResNet18-iteration2* model however did not lead to substantial improved predictions in the test set, nor did it lead to substantial number of additional unclassified CpGs being confidently assigned to any of the 10 categories. Thus, no further iterations of ResNet18 model were considered. In effect, the final 200,584 unclassified CpGs can be viewed as hybrids between various categories.

Simulated DNA methylation datasets and training/evaluation of *SimResNet18*

To assess the impact of sample size, we used a simulation strategy, generating artificial DNAm datasets of varying sample size and then training ResNet18 models on datasets with a fixed sample size, and datasets with 2 or more sample sizes, subsequently validating them in independently simulated DNAm datasets of sample sizes not seen during training. Further details about specific numbers and the detailed simulation process for each of the ten CpG categories is described in **Supplementary Methods**. Three ResNet18 models (*SimResNet18_V1*, *V2*, *V3*) were then trained on distinct configurations of simulated DNAm datasets. Briefly, *SimResNet18_V1* was trained and evaluated on a simulated dataset consisting of 3,500 samples and 20,000 CpGs. *SimResNet18_V2* was trained and evaluated on two combined simulated datasets, one at sample size 3500 and another at sample size 800, with the total number of CpGs used in training, model selection and testing identical to *V1*-model. *SimResNet18_V3* was trained on a composite dataset aggregating five independent simulated subsets, each containing 4,000 CpGs but with varying sample sizes (500, 1,500, 3,000, 5,000, and 8,000). All models followed an iterative training protocol and hyperparameter configuration as described before. That is, we used PyTorch 2.4.1 with an Adam optimizer, 64 batches, a maximum of 40 epochs, and the ReduceLROnPlateau learning rate scheduler with (Factor=0.1, Patience=3). To assess generalization capability, the trained models were subsequently applied to four independent simulated datasets containing 300, 2,000, 4,000, and

10,000 samples, with each of these independent simulated datasets comprising 10,000 CpGs (1000 CpGs per category). Further details can be found in **Supplementary Methods**.

Training the *MixedResNet18* Model

Based on the simulation results, in order to enhance *ResNet18* model's generalizability, we created a real DNAm dataset with CpG profiles defined at two different sample sizes. Initially, CpGs from the NSPT (EPIC) dataset with 3,501 samples were classified into ten categories. The CpG counts for each category following the initial classification were as follows: LinAgeHomo=2,062; LinAgeHet=67,117; NonAgeHet=4,547; Outlier=1,180; BiNonAge=1,175; BiAge=4,581; TriNonAge=1,766; TriAge=3,982; UniStable=38,753; UniUnstable=1,115, Unclassified=685,592. We then generated a reduced DNAm dataset of 1000 samples, by randomly drawing 1,000-samples from the 3501 ones. For the CpGs, their categories were directly inherited from the definitions based on the 3501 samples. For some categories, e.g. outlier, the subsampling could result in the CpG inheriting the wrong category. Hence, we designed an iterative sampling scheme to ensure the distributional characteristics of outlier CpGs in the subset aligned with the parent dataset (details in **Supplementary Methods**). The *MixedResNet18* model, implemented in PyTorch (v2.5.1) and Torchvision (v0.20.1), adapted the ResNet-18 architecture for 10-class prediction. The model was trained using the Stochastic Gradient Descent (SGD) optimizer with an initial learning rate of 0.01, a momentum of 0.9, and an L2 regularization coefficient of 1×10^{-4} . Its learning rate was dynamically adjusted by a ReduceLROnPlateau scheduler, which monitored the validation loss and reduced the rate by a factor of 0.1 if no improvement was observed for 3 consecutive epochs. Furthermore, early stopping was employed to prevent overfitting by monitoring the validation accuracy. Training was halted if the validation accuracy did not improve by at least 0.0001 for 30 consecutive epochs. Training utilized a batch size of 128 over 500 epochs with cross-entropy loss. Further details on the training protocol, which is similar to that of the previous ResNet18 model, are available in the **Supplementary Methods**.

Non-monotonic patterns

To each CpG, we fitted two models: one of DNAm against age, and another of DNAm against age and an additional covariate term quadratic in age. The first we call LM, the second we call QM. We then compared their Akaike Information Criterion values, computing the difference $\Delta AIC = AIC(QM) - AIC(LM)$. We are interested in those with $\Delta AIC < 0$. Since the relative probability of QM being a better model is $p(QM) = B(QM) / (B(QM) + B(LM))$ where

$$B(QM) = e^{(\min(AIC(QM), AIC(LM)) - AIC(QM))/2}$$

$$B(LM) = e^{(\min(AIC(QM), AIC(LM)) - AIC(LM))/2}$$

it follows, under the assumption $AIC(QM) < AIC(LM)$, that

$$p(QM) = 1/(1 + e^{\text{deltaAIC}/2})$$

Requiring $p(QM) > 0.95$ as the level of evidence in favor of a QM over a LM, means that we require $\text{deltaAIC} < 2 * \log(0.05/0.95)$. We also require the effect size of the quadratic fit (i.e. the DNAm difference between the fitted value at the optimum (maximum or minimum) and the fitted values at the extremes of the age distribution) to be larger than 0.01.

Cross-comparison with 3-CpG categorization from Seale et al.

We followed the procedure from Seale et al [14] to categorize CpGs into 3-4 distinct groups: (i) age-associated differentially methylated CpGs (age-DMCs), (ii) age-associated differentially variable CpGs (age-DVCs) and (iii) joint age-DMCs/ DVCs, with all other CpGs being placed in a fourth category labelled "Other". Briefly, DNAm beta values were first converted to M-values. To identify age-DMCs, we used the limma package [107] with M-values adjusting for sex. For age-DVCs, we assessed age-dependent heteroscedasticity using the Breusch-Pagan test. The model was also adjusted for sex. Statistical significance for both analyses was defined using an FDR threshold of < 0.05 . CpGs significant in both analyses were classified as "joint age-DMCs/DVCs," while those significant exclusively in the mean or variance analysis were labeled as "age-DMCs" or "age-DVCs," respectively. All remaining sites were categorized as "Other."

eFORGE2 enrichment analysis

We selected the top 1000 CpGs for each CpG category and uploaded these lists to the eFORGE2 enrichment webserver tool [45]. The choice of 1000 CpGs is driven by this being the maximum number of CpGs that eFORGE will test enrichment for. We tested enrichment against all major histone marks and chromatin states, as described in the eFORGE2 paper. Enrichment was tested using one-tailed Binomial tests, but with null probabilities estimated using a permutation procedure [45].

Construction of Noise-clock

We constructed NoiseClocks using two different strategies which led to almost identical results. In the first strategy, we selected 3 large blood cohorts comprising healthy individuals from different ancestries and with associated DNAm data. These studies were Johansson et al [62] (Illumina 450k, n=732, White Caucasians), TZH (Illumina EPICv1, n=705, Han Chinese) [60] and Hannum et al [61] (Illumina 450k, n=656, Whites and Hispanics). We used the minfi+BMIQ normalized forms of these data as described by us previously [48]. For each cohort, we then performed linear regressions of DNAm against age whilst adjusting for 12 immune cell type fractions (CTFs), as estimated using EpiDISH [48, 49]. In the Hannum dataset we also adjusted for ethnicity. We selected CpGs that did not associate with age (nominal linear

regression P-value > 0.2). For these selected CpGs, we then extracted the residuals from the linear fit, i.e. the residuals from $\text{DNAm} \sim \text{Age} + \text{CTFs}$, and performed another set of regressions of the squared residuals against age (Breusch-Pagan test), selecting CpGs with $\text{FDR} < 0.001$ and a t-statistic > 0 (increased DNAm variance with age). The overall procedure thus selects CpGs for which DNAm does not change with age in an average sense, but for which DNAm variance increases with age. This analysis resulted in 2,531 CpGs in the TZH dataset, 19,594 CpGs in the Hannum dataset, and 15,784 CpGs in the Johansson dataset. Across all 3 cohorts, this resulted in an overlap of 399 CpGs, that we term ‘Noise-Clock CpGs’. To arrive at a NoiseClock estimate for DNAm samples from an independent cohort, we first center each of the 399 CpGs to have zero mean DNAm. Any missing CpGs or CpGs not present in the independent cohort are removed. We then compute the absolute DNAm deviation matrix. For each column (sample), these absolute deviations from the mean are then averaged over CpGs to arrive at the NoiseClock estimate of the sample. In the second strategy, we constructed a NoiseClock using CpGs derived from the ResNet18 model’s predicted NonAgeHet CpGs that were also present on the 450k platform. The estimation of the NoiseClock value then followed the same procedure as for the previous NoiseClock.

Validation of NoiseClock in sorted DNAm data

We applied the NoiseClock to Illumina 450k DNAm data of sorted naïve CD4T-cells ($n=139$), classical CD14+ monocytes ($n=139$) and CD16+ neutrophils ($n=139$) from BLUEPRINT [108]. We used the DNAm datasets normalized as described in our previous publications, e.g. [56]. We also applied the NoiseClock to Illumina 450k DNAm data of sorted CD14+ monocytes ($n=1199$) and CD4+ T-cells ($n=212$) from Reynolds et al (MESA study) [109]. This DNAm dataset was normalized as done in [56].

Correlation of NoiseClock with inflammatory disease in NSPT-cohort

We correlated estimates from the NoiseClock with prevalence of 19 inflammatory diseases, including self-reported type-2 diabetes and rheumatoid arthritis. Because in natural populations, the incidence of self-reported diseases is an underestimate, we supplemented these with additional clinical definitions including metabolic syndrome, chronic kidney disease, central obesity, dyslipidemia, hypertension, hyperglycemia and obesity. The definition of Metabolic Syndrome (MetS) adheres to the criteria of the revised ATP III. An individual is classified as having MetS if they exhibit three or more of the following conditions: (1) Central obesity, defined as a waist circumference (WC) of ≥ 90 cm for men and ≥ 80 cm for women; (2) Elevated triglycerides (TG) level of ≥ 1.7 mmol/L or undergoing treatment for high triglycerides; (3) Reduced High-Density Lipoprotein (HDL) cholesterol, indicated by levels < 40 mg/dL (1.03 mmol/L) in men and < 50 mg/dL (1.3 mmol/L) in women, or receiving treatment for low HDL; (4) High blood pressure, characterized by systolic blood pressure (SBP)

≥ 130 mmHg or diastolic blood pressure (DBP) ≥ 85 mmHg, or treatment for previously diagnosed hypertension; (5) Elevated fasting plasma glucose (GLU) of ≥ 5.6 mmol/L, or treatment for high glucose levels. Obesity was defined as a Body Mass Index (BMI) of ≥ 28 kg/m². Dyslipidemia was identified by one or more of the following: Total Cholesterol (CH) ≥ 6.22 mmol/L, Low-Density Lipoprotein (LDL) cholesterol ≥ 4.14 mmol/L, HDL cholesterol < 1.04 mmol/L, TG ≥ 2.26 mmol/L, or a self-reported history of hyperlipidemia. Hypertension is defined either by a self-reported history of the condition or by having an SBP ≥ 140 mmHg and/or a DBP ≥ 90 mmHg. Hyperglycemia follows the current American Diabetes Association (ADA) criteria, defined as a fasting plasma glucose level of ≥ 7.0 mmol/L or a self-reported history of diabetes. Chronic kidney disease is defined by eGFR (estimated Glomerular Filtration Rate based on serum creatinine levels) < 60 . In all cases, noise estimates were correlated to these diseases (coded as 1 for cases, 0 for controls) using a one-tailed Wilcoxon rank sum test.

Evaluation of NoiseClock in Generation Scotland

GenS is a family-based cohort of ~24,000 individuals from across Scotland [67, 110]. Volunteers were aged 17-99 years at the point of recruitment (2006 – 2011). Blood samples were taken at a baseline clinic visit, from which DNA methylation was successfully profiled for 18,869 individuals via the Illumina EPIC array [68]. Data linkage to electronic health records spanned the period from January 1980 to April 2022. Death data were obtained via linkage to the National Records of Scotland. While consent to access primary care records was given from all participants, these were only accessible for ~40% of the cohort due to consent constraints with the data holders (individual GP surgeries). When evaluating the NoiseClock in GenS this was done in two ways: adjusting the DNAm data for variations in 12 immune-cell fractions as estimated using EpiDISH, and not adjusting the DNAm data. Because the NoiseClock CpGs were originally selected by adjusting for the IC-fraction variations, results were not dependent on which form of noise estimator we used. Here we report the values for the unadjusted analysis. The NoiseClock estimator was scaled to unit variance before running Cox proportional hazard regression models for all-cause-mortality. Covariates included age, sex, body mass index (BMI, calculated as weight in kg divided by squared height in meters), smoking pack years (calculated by self-reported packs per day multiplied by number of years as a smoker), alcohol units consumed over the last week, years of education and an area-based rank of socioeconomic deprivation (Scottish Index of Multiple Deprivation, SIMD), as described in more detail elsewhere [111]. Schoenfeld residuals were used to examine the proportional hazards assumption at the model level. In total, there were 18,859 samples with 842 death events.

Identification of age-mQTLs

From the NSPT Han Chinese cohort, we previously derived a list of mQTLs (CpG-SNP pairs) (with mQTL P-value $< 1e-14$), encompassing 273,613 unique CpGs [19]. For CpGs associated with multiple SNPs, only the most significant SNP (based on mQTL P-value) was initially retained for downstream analysis. Next, we identified age-mQTLs ('broad age-mQTLs') by correlating DNAm values to age in linear regression models whilst adjusting for sex, 7 immune cell type fractions (estimated with R package EpiDISH [49]) and genotype (encoded as 0=AA, 1=AB/BA, 2=BB, with B representing the minor allele), and demanding FDR <0.05 . To identify genotype specific age-mQTLs, we performed genotype-stratified linear regressions: for each genotype group, we tested the association between DNAm values and age, adjusting for sex and 7 immune cell fractions. Based on genotype-specific associations, we defined seven categories of age-mQTLs: 1) unqBB: mQTLs where DNAm was significantly correlated to age only in BB samples (i.e. age-FDR for BB samples is <0.05 , while age-FDR for AA and AB is >0.2); 2) unqAA: similar to unqBB, with age-FDR for AA samples <0.05 , while age-FDR for AB and BB is >0.2 ; 3) unqAB: similar to unqBB, with age-FDR for AB samples <0.05 , while age-FDR for AA and BB is >0.2 ; 4) AAAB: mQTLs where DNAm was significantly correlated to age in both AA and AB samples but not in BB samples (i.e. age-FDR for AA and AB samples <0.05 , while age-FDR for BB samples >0.2); 5) AABB: similar to AAAB, with age-FDR for AA and BB samples is <0.05 while age-FDR for AB samples >0.2 ; 6) BBAB: similar to AAAB, with age-FDR for BB and AB samples <0.05 , while age-FDR for AA samples >0.2 ; 7) AABBAB: mQTLs where DNAm was significantly correlated to age in all three genotype groups (i.e. age-FDR for AA, AB, BB samples <0.05). For any genotype group with fewer than 10 samples, the regression results were discarded. Note that for age-mQTLs in categories 4) to 7), only CpGs with consistent directions of DNAm-age correlation across the relevant genotype groups were retained.

Calculation of Episcopes for 109 proteins

Gadd et al. trained blood DNAm based predictors for 109 proteins, referred to as Episcopes [64]. Each predictor was trained using an Elastic Net regression model, in which protein levels were predicted from genome-wide DNAm profiles, using matched proteomics and DNAm data from cohorts with sample size ranging from 706 to 944 individuals, derived from the German population-based study, and the Scottish Lothian Birth Cohort 1936 study. The code for calculating Episcopes for samples in a new DNAm dataset can be obtained from: <https://zenodo.org/records/7154750>, via running MethylDetectR shiny app.

Calculation of DNAm based CRP scores

Wielsher et al. performed a multi-ethnic meta-analysis over 20,000 whole blood samples from over 30 cohorts, identifying 1765 marker CpGs significantly associated with serum C-Reactive

protein (CRP) levels at a Bonferroni threshold ($P < 1e-7$) [63]. The CpGs were identified after adjustment for 6 main immune-cell fractions (granulocytes, monocytes, CD4+ and CD8+ T-cells, B-cells and NK-cells). This analysis therefore did not adjust for variations between the naïve and memory T-cell subsets. Each of the 1765 CpGs has a signed coefficient (+1 or -1) indicating whether it is positively or negatively correlated with CRP levels. To compute the CRP inflammaging score (InflScore) in a new DNAm dataset, we first z-scored each CpG across the samples so that the mean is 0 and the standard deviation is 1. The InflScore of each sample is then obtained as the Pearson Correlation Coefficient of each sample's z-score normalized DNAm profile with the signed CpG signature, as defined over the 1765 CpGs or a given subset if not all CpGs are present in the new dataset. This InflScore has already been validated by us previously [112]. Because this InflScore is strongly influenced by variations in the naïve to memory T-cell subsets, we decided to build a new CRP-score (called 'IntInflScore') from a reduced set of CpGs that are not specific for naïve and memory T-cell subsets. To identify these CpGs, we analyzed the normalized sorted immune-cell EPIC DNAm data from Salas et al [48, 113], selecting CpGs that can't discriminate (nominal limma P -value >0.2) naïve from memory B-cells, naïve from memory CD4T-cells and naïve from memory CD8T-cells. The number of CpGs that could not discriminate naïve from memory subtypes across all 3 cell subtypes was 141,827. The overlap with the 1765 CRP-associated CpGs was only 62, indicating that most of the 1765 CpGs are discriminating naïve from memory lymphocytes. Hence, these 62 CRP-associated CpGs constitute CpGs that correlate with serum CRP-levels in a manner that is independent of underlying variations between the memory and naïve lymphocyte subsets. To compute the corresponding IntInflScore of a sample, we repeat the above described procedure for InflScore, but now restricting only to these 62 CpGs.

Calculation of DNAm based IL6 and SystemsAge inflammatory score

Stevenson et al. developed a DNAm-based predictor of serum interleukin-6 (IL-6) levels from a cohort of 875 older adults, using Elastic Net regression, where serum IL-6 level was used as the dependent variable and the CpGs from the 450k array were used as the independent variables [66]. This predictor comprises 35 CpGs and their regression coefficients. We applied this predictor to the TZH cohort to obtain the IL-6 DNAm-based score of each sample. To quantify systemic inflammatory burden, we utilized the inflammation component of SystemsAge developed by Sehgal et al [65].

Identification of age-mQTL CpGs correlating with DNAm based protein scores

Focusing on 1,424 unqBB and 1,715 BBAB age-mQTL CpGs, we explored their associations with 111 DNAm-based protein scores (109 Episcores plus IL6 and CRP scores) in the TZH cohort. For each of these CpGs, we retrieved all correlated SNPs with mQTL $P < 1 \times 10^{-14}$, and retained only those SNPs that constituted unqBB or BBAB age-mQTLs, based on genotype-

specific linear regressions of DNAm values on age (as described previously). For unqBB age-mQTLs, we performed linear regression models to predict each DNAm-based protein score using CpG DNAm values, adjusting for sex and 7 immune cell subtype fractions. These regressions were run separately in BB samples and in the combined AA+AB group. We then selected CpG-SNP pairs significantly associated with protein scores only in the BB group (i.e., $P < 0.05$ in BB and $P > 0.2$ in AA+AB). Similarly, for BBAB age-mQTLs, linear regressions were conducted in BB+AB samples and in AA samples separately, with the same covariate adjustments. CpG-SNP pairs significantly associated with protein scores only in the BB+AB group ($P < 0.05$ in BB+AB and $P > 0.2$ in AA) were selected for downstream analyses.

Enrichment analysis for DamAge and AdaptAge clocks

DamAge and AdaptAge clock CpG lists were provided by Ying et al [59]. Enrichment analyses were performed to assess whether unqBB, BBAB, and broadly defined age-mQTL CpGs (i.e. excluding unqBB and BBAB CpGs) were overrepresented among the DamAge and AdaptAge clock CpGs. One-tailed Fisher exact tests were done. For example, in the case of unqBB CpGs, all remaining 450k CpGs that were not classified as unqBB served as the background set. The proportion of DamAge or AdaptAge CpGs among the unqBB set was then compared to that in the background set. This background restriction was applied because DamAge and AdaptAge clocks were originally constructed using Illumina 450k array data.

Enrichment analysis of genotype specific age-mQTLs for inflammation-associated SNPs

Focusing on the CpGs of CRP/IL6-score correlated BB or BBAB age-mQTLs, we first identified all SNPs defining BB or BBAB age-mQTLs for a given CpG, and whether there was also a directionally consistent genotype-specific association with either CRP or IL6-scores. Thus, for each CpG we obtain a list of such SNPs. We then built the union of all SNPs linked to all CpGs. Using the GWAS Catalog (<https://www.ebi.ac.uk/gwas/>) a total $n=1884$ of such SNPs mapped to any trait. We then also calculated the number of SNPs associated with specific inflammation-related traits, including ulcerative colitis, inflammatory bowel disease, psoriasis, systemic lupus erythematosus, chronic obstructive pulmonary disease, severe acute respiratory syndrome, COVID-19, hypertension, coronary artery disease, rheumatoid arthritis, hypothyroidism, drug-induced liver injury, type 1 diabetes, type 2 diabetes, nephrotic syndrome, non-alcoholic fatty liver disease, ankylosing spondylitis, sarcoidosis, multiple sclerosis, Crohn's disease, asthma, ischemic stroke, colorectal cancer, breast carcinoma and prostate carcinoma. Let x denote the number of CRP/IL6-score-correlated age-mQTL SNPs with association to a given trait. A binomial test was then performed to assess the probability of observing x or more such SNPs out of $n=1884$, using as null probability the proportion of all mQTL SNPs ($P < 1e-14$) associated with that particular trait across all related GWAS catalog traits.

Mendelian randomization analysis

We performed two-sample Mendelian Randomization (MR) on age-mQTLs where the CpG correlates with the CRP/IL6-scores and where the SNP correlates with an inflammation-related trait, as determined by the GWAS catalog (<https://www.ebi.ac.uk/gwas/>). We focused on seven representative traits: C-reactive protein (CRP) measurement, rheumatoid arthritis (RA), hypothyroidism, ulcerative colitis (UC), type 2 diabetes (T2D), systemic lupus erythematosus (SLE), and coronary artery disease (CAD). GWAS summary statistics for these traits were obtained from the BioBank Japan database (<https://pheweb.jp/downloads>), specifically statistics for RA, T2D, SLE, hypothyroidism, UC, and CRP were obtained from a large-scale genetic association atlas covering 220 phenotypes (n=179000) [114], whilst summary statistics for CAD were derived from a separate study of over 149000 individuals [115]. For each CpG, all associated mQTL SNPs ($P < 1 \times 10^{-14}$) were initially included as candidate instruments. To ensure independence among SNPs, we applied the *clump_data()* function from the TwoSampleMR R package [71], using linkage disequilibrium (LD) calculated using EAS data from the 1000 Genomes Phase 3. SNPs in high LD ($r^2 > 0.2$ within a 1000 kb window) were clumped, retaining only the variant with the lowest P-value in each LD block. Only CpGs with at least three independent instrumental SNPs were included in the final MR analysis. MR analyses were then performed with TwoSampleMR using multiple methods, including inverse variance weighted (IVW) and weighted median (WM) models, followed by MR-Egger to test for average directional pleiotropy. Validity of genetic instruments was performed using Steiger's test [73]. Robustness was tested using leave-one-out. We also performed MR-PRESSO analysis [74] to test Egger assumption and detect and adjust for pleiotropic outliers. To evaluate whether the 21 identified exposure-outcome pairs (encompassing 16 candidates causal CpGs) shared a common causal genetic basis, we performed colocalization analysis using the coloc R package [75, 76, 116]. While the standard Bayesian approach (coloc.abf) is widely used, its "single causal variant" assumption often leads to false negatives in the presence of allelic heterogeneity, where multiple independent signals in a locus can mask one another. To overcome this limitation, we employed a hybrid SuSiE-ABF strategy as recommended by Wallace [75]. We prioritized coloc.susie(), which leverages the Sum of Single Effects (SuSiE) framework to disentangle multiple independent causal signals within the same genomic region. In instances where coloc.susie() failed to define valid credible sets, we reverted to the coloc.abf() method. The analysis was conducted within a 1-Mb genomic window (± 500 kb from the target CpG), with GWAS summary statistics for inflammatory diseases retrieved from the BioBank Japan database (<https://pheweb.jp/downloads>). To ensure robust inference, we adopted the approach utilized by the GTEx Consortium [77] for interpreting colocalization results. In scenarios with limited statistical power, both PP.H3 (support for distinct causal variants) and PP.H4 (support for a shared causal variant) can be low, which does not definitively preclude a shared causal etiology. To address this, rather than relying solely on the absolute PP.H4 value, we utilized the posterior probability ratio, defined as $PP.H4 / (PP.H3 + PP.H4) > 0.9$, to identify

reliable colocalization events.

Code Availability

The Python code for training and testing the ResNet18 models are freely available from <https://github.com/Duzhaozhen/AgeTaxoDeepL>. On this site we also provide Python code for predicting the age-taxonomy of an independent CpG DNAm profile. R-scripts computing our NoiseClock and Intrinsic Inflammaging Score are available from <https://github.com/Duzhaozhen/AgeTaxoDeepL>. We have also produced an Ocean Code Capsule, available from <https://codeocean.com/capsule/8686352/tree>, accessible via an access-token.

Data Availability

The main datasets used here are freely available from public repositories. The DNAm datasets of HPT [40], Johansson [62] Airwave [41] and Hannum [61] cohorts, are publicly available from the NCBI GEO website under these accession numbers: GSE210255 (HPT, <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE210255>), GSE87571 (Johansson, <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE87571>), GSE147740 (Airwave, <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE147740>), GSE40279 (Hannum, <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE40279>). The Illumina EPIC DNAm data for NSPT cohort [19] can be requested at <https://ngdc.cncb.ac.cn/omix/release/OMIX004363>. The mQTL database is available for download at <https://www.biosino.org/panmqtl/>. The GWAS results are available on GWAS catalog (<https://www.ebi.ac.uk/gwas/>) and Biobank Japan database (<https://pheweb.jp/downloads>). The scRNA-Seq data from Yazar et al [51] is publicly available on <https://cellxgene.cziscience.com/collections/dde06e0f-ab3b-46be-96a2-a8082383c4a1>. The AIDA scRNA-Seq data [50] is publicly available on <https://cellxgene.cziscience.com/collections/ced320a1-29f3-47c1-a735-513c7084d508>. The DNAm data of sorted immune cells are available from GEO (MESA) under accession number GSE56046 [<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE56046>], and EGA (BLUEPRINT) EGAS00001001456 [<https://ega-archive.org/studies/EGAS00001001456>].

Declarations

Ethics approval and consent to participate: Not applicable here since we have only analyzed existing publicly available data.

Consent for publication: Not applicable.

Competing interests: The authors declare no competing interests.

Funding: AET was supported by National Science Foundation of China (NSFC) grants, grant numbers 31970632 and a RFIS grant RFIS grant W2431024. REM was supported by the Wellcome Trust (104036/Z/14/Z and 221890/Z/20/Z). Generation Scotland received core support from the Chief Scientist Office of the Scottish Government Health Directorates (CZD/16/6) and the Scottish Funding Council (HR03006). DNA methylation profiling of the Generation Scotland samples was carried out by the Genetics Core Laboratory at the Edinburgh Clinical Research Facility, Edinburgh, Scotland, and was funded by the Medical Research Council UK and Wellcome (Wellcome Trust Strategic Award STRatifying Resilience and Depression Longitudinally (STRADL; Reference 104036/Z/14/Z). SW was supported by CAS Young Team Program for Stable Support of Basic Research (Grant No: YSBR-077), the National Natural Science Foundation of China (Grant No: 32325013) and Shanghai Science and Technology Innovation Action Project (Grant No: 24JS2810300). QP was supported by the National Natural Science Foundation of China (Grant No: 32471216).

Authors contributions: AET conceived and directed the study. AET wrote the MS with contributions from ZD and QL. REM contributed and directed the analysis in the Generation Scotland cohort. SW contributed the NSPT-cohort data. QP contributed phenome data of NSPT-cohort. Bioinformatic and statistical analyses were performed by ZD, QL, GX, QP, REM and AET. ZL helped with the deep-learning implementation. DH and RS contributed feedback. All authors read and approved final manuscript.

References

1. Lappalainen T, Grealis JM: **Associating cellular epigenetic models with human phenotypes.** *Nat Rev Genet* 2017, **18**:441-451.
2. Teschendorff AE, West J, Beck S: **Age-associated epigenetic drift: implications, and a case of epigenetic thrift?** *Hum Mol Genet* 2013, **22**:R7-R15.
3. Moqri M, Herzog C, Poganik JR, Biomarkers of Aging C, Justice J, Belsky DW, Higgins-Chen A, Moskalev A, Fuellen G, Cohen AA, et al: **Biomarkers of aging for the identification and evaluation of longevity interventions.** *Cell* 2023, **186**:3758-3775.
4. Moqri M, Herzog C, Poganik JR, Ying K, Justice JN, Belsky DW, Higgins-Chen AT, Chen BH, Cohen AA, Fuellen G, et al: **Validation of biomarkers of aging.** *Nat Med* 2024, **30**:360-372.
5. Horvath S: **DNA methylation age of human tissues and cell types.** *Genome Biol* 2013, **14**:R115.
6. Bell CG, Lowe R, Adams PD, Baccarelli AA, Beck S, Bell JT, Christensen BC, Gladyshev VN, Heijmans BT, Horvath S, et al: **DNA methylation aging clocks: challenges and**

1142 **recommendations.** *Genome Biol* 2019, **20**:249.

1143 7. Marioni RE, Shah S, McRae AF, Chen BH, Colicino E, Harris SE, Gibson J, Henders AK, Redmond
1144 P, Cox SR, et al: **DNA methylation age of blood predicts all-cause mortality in later life.**
1145 *Genome Biol* 2015, **16**:25.

1146 8. Levine ME, Lu AT, Quach A, Chen BH, Assimes TL, Bandinelli S, Hou L, Baccarelli AA, Stewart JD,
1147 Li Y, et al: **An epigenetic biomarker of aging for lifespan and healthspan.** *Aging (Albany NY)*
1148 2018, **10**:573-591.

1149 9. Horvath S, Raj K: **DNA methylation-based biomarkers and the epigenetic clock theory of**
1150 **ageing.** *Nat Rev Genet* 2018, **19**:371-384.

1151 10. Yuan T, Jiao Y, de Jong S, Ophoff RA, Beck S, Teschendorff AE: **An integrative multi-scale**
1152 **analysis of the dynamic DNA methylation landscape in aging.** *PLoS Genet* 2015, **11**:e1004996.

1153 11. Fraga MF, Ballestar E, Paz MF, Ropero S, Setien F, Ballestar ML, Heine-Suner D, Cigudosa JC,
1154 Urioste M, Benitez J, et al: **Epigenetic differences arise during the lifetime of monozygotic**
1155 **twins.** *Proc Natl Acad Sci U S A* 2005, **102**:10604-10609.

1156 12. Christensen BC, Houseman EA, Marsit CJ, Zheng S, Wrensch MR, Wiemels JL, Nelson HH,
1157 Karagas MR, Padbury JF, Bueno R, et al: **Aging and environmental exposures alter tissue-**
1158 **specific DNA methylation dependent upon CpG island context.** *PLoS Genet* 2009, **5**:e1000602.

1159 13. Bergstedt J, Azzou SAK, Tsuo K, Jaquaniello A, Urrutia A, Rotival M, Lin DTS, MacIsaac JL, Kobor
1160 MS, Albert ML, et al: **The immune factors driving DNA methylation variation in human blood.**
1161 *Nat Commun* 2022, **13**:5895.

1162 14. Seale K, Teschendorff A, Reiner AP, Voisin S, Eynon N: **A comprehensive map of the aging blood**
1163 **methyome in humans.** *Genome Biol* 2024, **25**:240.

1164 15. Seale K, Horvath S, Teschendorff A, Eynon N, Voisin S: **Making sense of the ageing methylome.**
1165 *Nat Rev Genet* 2022, **23**:585-605.

1166 16. Sliker RC, van Iterson M, Luijk R, Beekman M, Zhernakova DV, Moed MH, Mei H, van Galen M,
1167 Deelen P, Bonder MJ, et al: **Age-related accrual of methylomic variability is linked to**
1168 **fundamental ageing mechanisms.** *Genome Biol* 2016, **17**:191.

1169 17. Mei X, Blanchard J, Luellen C, Conboy MJ, Conboy IM: **Fail-tests of DNA methylation clocks,**
1170 **and development of a noise barometer for measuring epigenetic pressure of aging and**
1171 **disease.** *Aging (Albany NY)* 2023, **15**:8552-8575.

1172 18. Min JL, Hemani G, Hannon E, Dekkers KF, Castillo-Fernandez J, Luijk R, Carnero-Montoro E,
1173 Lawson DJ, Burrows K, Suderman M, et al: **Genomic and phenotypic insights from an atlas of**
1174 **genetic effects on DNA methylation.** *Nat Genet* 2021, **53**:1311-1321.

1175 19. Peng Q, Liu X, Li W, Jing H, Li J, Gao X, Luo Q, Breeze CE, Pan S, Zheng Q, et al: **Analysis of blood**
1176 **methylation quantitative trait loci in East Asians reveals ancestry-specific impacts on**
1177 **complex traits.** *Nat Genet* 2024, **56**:846-860.

1178 20. Teschendorff AE, Jones A, Fiegl H, Sargent A, Zhuang JJ, Kitchener HC, Widschwendter M:
1179 **Epigenetic variability in cells of normal cytology is associated with the risk of future**
1180 **morphological transformation.** *Genome Med* 2012, **4**:24.

1181 21. Teschendorff AE, Relton CL: **Statistical and integrative system-level analysis of DNA**
1182 **methylation data.** *Nat Rev Genet* 2018, **19**:129-147.

1183 22. Chen GK, Yan Q, Paul KC, Kusters CDJ, Folle AD, Furlong M, Keener A, Bronstein J, Horvath S,
1184 Ritz B: **Stochastic Epigenetic Mutations Influence Parkinson's Disease Risk, Progression, and**
1185 **Mortality.** *J Parkinsons Dis* 2022, **12**:545-556.

- 1186 23. Gagliardi A, Dugue PA, Nost TH, Southey MC, Buchanan DD, Schmidt DF, Makalic E, Hodge AM,
1187 English DR, Doo NW, et al: **Stochastic Epigenetic Mutations Are Associated with Risk of Breast**
1188 **Cancer, Lung Cancer, and Mature B-cell Neoplasms.** *Cancer Epidemiol Biomarkers Prev* 2020,
1189 **29:2026-2037.**
- 1190 24. Gentilini D, Garagnani P, Pisoni S, Bacalini MG, Calzari L, Mari D, Vitale G, Franceschi C, Di Blasio
1191 AM: **Stochastic epigenetic mutations (DNA methylation) increase exponentially in human**
1192 **aging and correlate with X chromosome inactivation skewing in females.** *Aging (Albany NY)*
1193 2015, **7:568-578.**
- 1194 25. Yan Q, Paul KC, Lu AT, Kusters C, Binder AM, Horvath S, Ritz B: **Epigenetic mutation load is**
1195 **weakly correlated with epigenetic age acceleration.** *Aging (Albany NY)* 2020, **12:17863-17894.**
- 1196 26. Teschendorff AE, Horvath S: **Epigenetic ageing clocks: statistical methods and emerging**
1197 **computational challenges.** *Nat Rev Genet* 2025.
- 1198 27. Elliott HR, Shihab HA, Lockett GA, Holloway JW, McRae AF, Smith GD, Ring SM, Gaunt TR, Relton
1199 CL: **Role of DNA Methylation in Type 2 Diabetes Etiology: Using Genotype as a Causal Anchor.**
1200 *Diabetes* 2017, **66:1713-1722.**
- 1201 28. Gaunt TR, Shihab HA, Hemani G, Min JL, Woodward G, Lyttleton O, Zheng J, Duggirala A,
1202 McArdle WL, Ho K, et al: **Systematic identification of genetic influences on methylation across**
1203 **the human life course.** *Genome Biol* 2016, **17:61.**
- 1204 29. LeCun Y, Bengio Y, Hinton G: **Deep learning.** *Nature* 2015, **521:436-444.**
- 1205 30. Webb S: **Deep learning for biology.** *Nature* 2018, **554:555-557.**
- 1206 31. Angermueller C, Parnamaa T, Parts L, Stegle O: **Deep learning for computational biology.** *Mol*
1207 *Syst Biol* 2016, **12:878.**
- 1208 32. Maxmen A: **Deep learning sharpens views of cells and genes.** *Nature* 2018, **553:9-10.**
- 1209 33. Landhuis E: **Deep learning takes on tumours.** *Nature* 2020, **580:551-553.**
- 1210 34. Eraslan G, Avsec Z, Gagneur J, Theis FJ: **Deep learning: new computational modelling**
1211 **techniques for genomics.** *Nat Rev Genet* 2019, **20:389-403.**
- 1212 35. He K, Zhang X, Ren S, Sun J: **Deep Residual Learning for Image Recognition.** In *IEEE Conference*
1213 *on Computer Vision and Pattern Recognition (CVPR); 27-30 June 2016.* 2016: 770-778.
- 1214 36. Simonyan K, Zisserman A: **Very deep convolutional networks for large-scale image**
1215 **recognition.** . *arXiv* 2014, **14091556.**
- 1216 37. Teschendorff AE, Naderi A, Barbosa-Morais NL, Caldas C: **PACK: Profile Analysis using**
1217 **Clustering and Kurtosis to find molecular classifiers in cancer.** *Bioinformatics* 2006, **22:2269-**
1218 **2275.**
- 1219 38. Yeung KY, Fraley C, Murua A, Raftery AE, Ruzzo WL: **Model-based clustering and data**
1220 **transformations for gene expression data.** *Bioinformatics* 2001, **17:977-987.**
- 1221 39. Dempster AP, Laird NM, Rubin DB: **Maximum Likelihood From Incomplete Data Via the EM**
1222 **Algorithm.** *J Royal Stat Soc B* 1977, **39:1--38.**
- 1223 40. Shang L, Zhao W, Wang YZ, Li Z, Choi JJ, Kho M, Mosley TH, Kardia SLR, Smith JA, Zhou X: **meQTL**
1224 **mapping in the GENOA study reveals genetic determinants of DNA methylation in African**
1225 **Americans.** *Nat Commun* 2023, **14:2711.**
- 1226 41. Robinson O, Chadeau Hyam M, Karaman I, Climaco Pinto R, Ala-Korpela M, Handakas E, Fiorito
1227 G, Gao H, Heard A, Jarvelin MR, et al: **Determinants of accelerated metabolomic and**
1228 **epigenetic aging in a UK cohort.** *Aging Cell* 2020, **19:e13149.**
- 1229 42. Ernst J, Kellis M: **Discovery and characterization of chromatin states for systematic annotation**

1230 of the human genome. *Nat Biotechnol* 2010, **28**:817-825.

1231 43. Ernst J, Kellis M: **Large-scale imputation of epigenomic datasets for systematic annotation of**
1232 **diverse human tissues.** *Nat Biotechnol* 2015, **33**:364-376.

1233 44. Adams D, Altucci L, Antonarakis SE, Ballesteros J, Beck S, Bird A, Bock C, Boehm B, Campo E,
1234 Caricasole A, et al: **BLUEPRINT to decode the epigenetic signature written in blood.** *Nat*
1235 *Biotechnol* 2012, **30**:224-226.

1236 45. Breeze CE, Reynolds AP, van Dongen J, Dunham I, Lazar J, Neph S, Vierstra J, Bourque G,
1237 Teschendorff AE, Stamatoyannopoulos JA, Beck S: **eFORGE v2.0: updated analysis of cell type-**
1238 **specific signal in epigenomic data.** *Bioinformatics* 2019, **35**:4767-4769.

1239 46. Breeze CE, Paul DS, van Dongen J, Butcher LM, Ambrose JC, Barrett JE, Lowe R, Rakyan VK,
1240 Iotchkova V, Frontini M, et al: **eFORGE: A Tool for Identifying Cell Type-Specific Signal in**
1241 **Epigenomic Data.** *Cell Rep* 2016, **17**:2137-2150.

1242 47. Loyfer N, Magenheim J, Peretz A, Cann G, Bredno J, Klochendler A, Fox-Fisher I, Shabi-Porat S,
1243 Hecht M, Pelet T, et al: **A DNA methylation atlas of normal human cell types.** *Nature* 2023,
1244 **613**:355-364.

1245 48. Luo Q, Dwaraka VB, Chen Q, Tong H, Zhu T, Seale K, Raffaele JM, Zheng SC, Mendez TL, Chen Y,
1246 et al: **A meta-analysis of immune-cell fractions at high resolution reveals novel associations**
1247 **with common phenotypes and health outcomes.** *Genome Med* 2023, **15**:59.

1248 49. Teschendorff AE, Breeze CE, Zheng SC, Beck S: **A comparison of reference-based algorithms**
1249 **for correcting cell-type heterogeneity in Epigenome-Wide Association Studies.** *BMC*
1250 *Bioinformatics* 2017, **18**:105.

1251 50. Kock KH, Tan LM, Han KY, Ando Y, Jevapatarakul D, Chatterjee A, Lin QXX, Buyamin EV, Sonthalia
1252 R, Rajagopalan D, et al: **Asian diversity in human immune cells.** *Cell* 2025, **188**:2288-2306
1253 e2224.

1254 51. Yazar S, Alquicira-Hernandez J, Wing K, Senabouth A, Gordon MG, Andersen S, Lu Q, Rowson
1255 A, Taylor TRP, Clarke L, et al: **Single-cell eQTL mapping identifies cell type-specific genetic**
1256 **control of autoimmune disease.** *Science* 2022, **376**:eabf3041.

1257 52. Zhang Q, Vallerga CL, Walker RM, Lin T, Henders AK, Montgomery GW, He J, Fan D, Fowdar J,
1258 Kennedy M, et al: **Improved precision of epigenetic clock estimates across tissues and its**
1259 **implication for biological ageing.** *Genome Med* 2019, **11**:54.

1260 53. Belsky DW, Caspi A, Corcoran DL, Sugden K, Poulton R, Arseneault L, Baccarelli A, Chamarti K,
1261 Gao X, Hannon E, et al: **DunedinPACE, a DNA methylation biomarker of the pace of aging.**
1262 *Elife* 2022, **11**.

1263 54. Yang Z, Wong A, Kuh D, Paul DS, Rakyan VK, Leslie RD, Zheng SC, Widschwendter M, Beck S,
1264 Teschendorff AE: **Correlation of an epigenetic mitotic clock with cancer risk.** *Genome Biol*
1265 2016, **17**:205.

1266 55. Teschendorff AE: **A comparison of epigenetic mitotic-like clocks for cancer risk prediction.**
1267 *Genome Med* 2020, **12**:56.

1268 56. Zhu T, Tong H, Du Z, Beck S, Teschendorff AE: **An improved epigenetic counter to track mitotic**
1269 **age in normal and precancerous tissues.** *Nat Commun* 2024, **15**:4211.

1270 57. Zhou W, Dinh HQ, Ramjan Z, Weisenberger DJ, Nicolet CM, Shen H, Laird PW, Berman BP: **DNA**
1271 **methylation loss in late-replicating domains is linked to mitotic cell division.** *Nat Genet* 2018,
1272 **50**:591-602.

1273 58. Endicott JL, Nolte PA, Shen H, Laird PW: **Cell division drives DNA methylation loss in late-**

replicating domains in primary human cells. *Nat Commun* 2022, **13**:6659.

59. Ying K, Liu H, Tarkhov AE, Sadler MC, Lu AT, Moqri M, Horvath S, Kutalik Z, Shen X, Gladyshev VN: **Causality-enriched epigenetic age uncouples damage and adaptation.** *Nat Aging* 2024, **4**:231-246.

60. You C, Wu S, Zheng SC, Zhu T, Jing H, Flagg K, Wang G, Jin L, Wang S, Teschendorff AE: **A cell-type deconvolution meta-analysis of whole blood EWAS reveals lineage-specific smoking-associated DNA methylation changes.** *Nat Commun* 2020, **11**:4779.

61. Hannum G, Guinney J, Zhao L, Zhang L, Hughes G, Sadda S, Klotzle B, Bibikova M, Fan JB, Gao Y, et al: **Genome-wide methylation profiles reveal quantitative views of human aging rates.** *Mol Cell* 2013, **49**:359-367.

62. Johansson A, Enroth S, Gyllenstein U: **Continuous Aging of the Human DNA Methylome Throughout the Human Lifespan.** *PLoS One* 2013, **8**:e67378.

63. Wielscher M, Mandaviya PR, Kuehnelt B, Joeheanes R, Mustafa R, Robinson O, Zhang Y, Bodinier B, Walton E, Mishra PP, et al: **DNA methylation signature of chronic low-grade inflammation and its role in cardio-respiratory diseases.** *Nat Commun* 2022, **13**:2408.

64. Gadd DA, Hillary RF, McCartney DL, Zaghlool SB, Stevenson AJ, Cheng Y, Fawns-Ritchie C, Nangle C, Campbell A, Flaig R, et al: **Epigenetic scores for the circulating proteome as tools for disease prediction.** *Elife* 2022, **11**.

65. Sehgal R, Markov Y, Qin C, Meer M, Hadley C, Shadyab AH, Casanova R, Manson JE, Bhatti P, Moore AZ, et al: **Systems Age: a single blood methylation test to quantify aging heterogeneity across 11 physiological systems.** *Nat Aging* 2025, **5**:1880-1896.

66. Stevenson AJ, Gadd DA, Hillary RF, McCartney DL, Campbell A, Walker RM, Evans KL, Harris SE, Spires-Jones TL, McRae AF, et al: **Creating and Validating a DNA Methylation-Based Proxy for Interleukin-6.** *J Gerontol A Biol Sci Med Sci* 2021, **76**:2284-2292.

67. Milbourn H, McCartney D, Richmond A, Campbell A, Flaig R, Robertson S, Fawns-Ritchie C, Hayward C, Marioni RE, McIntosh AM, et al: **Generation Scotland: an update on Scotland's longitudinal family health study.** *BMJ Open* 2024, **14**:e084719.

68. Walker RM, McCartney DL, Carr K, Barber M, Shen X, Campbell A, Bernabeu E, Aitken E, Fawkes A, Wrobel N, Murphy L: **Data Resource Profile: Whole Blood DNA Methylation Resource in Generation Scotland (MeGS).** *medRxiv* 2024.

69. Lu AT, Binder AM, Zhang J, Yan Q, Reiner AP, Cox SR, Corley J, Harris SE, Kuo PL, Moore AZ, et al: **DNA methylation GrimAge version 2.** *Aging (Albany NY)* 2022, **14**:9484-9549.

70. Franceschi C, Campisi J: **Chronic inflammation (inflammaging) and its potential contribution to age-associated diseases.** *J Gerontol A Biol Sci Med Sci* 2014, **69 Suppl 1**:S4-9.

71. Davey Smith G, Hemani G: **Mendelian randomization: genetic anchors for causal inference in epidemiological studies.** *Hum Mol Genet* 2014, **23**:R89-98.

72. Relton CL, Davey Smith G: **Two-step epigenetic Mendelian randomization: a strategy for establishing the causal role of epigenetic processes in pathways to disease.** *Int J Epidemiol* 2012, **41**:161-176.

73. Hemani G, Tilling K, Davey Smith G: **Orienting the causal relationship between imprecisely measured traits using GWAS summary data.** *PLoS Genet* 2017, **13**:e1007081.

74. Verbanck M, Chen CY, Neale B, Do R: **Detection of widespread horizontal pleiotropy in causal relationships inferred from Mendelian randomization between complex traits and diseases.** *Nat Genet* 2018, **50**:693-698.

1318 75. Wallace C: **A more accurate method for colocalisation analysis allowing for multiple causal**
1319 **variants.** *PLoS Genet* 2021, **17**:e1009440.

1320 76. Giambartolomei C, Vukcevic D, Schadt EE, Franke L, Hingorani AD, Wallace C, Plagnol V:
1321 **Bayesian test for colocalisation between pairs of genetic association studies using summary**
1322 **statistics.** *PLoS Genet* 2014, **10**:e1004383.

1323 77. Consortium GT, Laboratory DA, Coordinating Center -Analysis Working G, Statistical Methods
1324 groups-Analysis Working G, Enhancing Gg, Fund NIHC, Nih/Nci, Nih/Nhgri, Nih/Nimh, Nih/Nida,
1325 et al: **Genetic effects on gene expression across human tissues.** *Nature* 2017, **550**:204-213.

1326 78. Semick SA, Bharadwaj RA, Collado-Torres L, Tao R, Shin JH, Deep-Soboslay A, Weiss JR,
1327 Weinberger DR, Hyde TM, Kleinman JE, Jaffe AE, Mattay VS: **Integrated DNA methylation and**
1328 **gene expression profiling across multiple brain regions implicate novel genes in Alzheimer's**
1329 **disease.** *Acta Neuropathol* 2019, **137**:557-569.

1330 79. Xiong Z, Yang F, Li M, Ma Y, Zhao W, Wang G, Li Z, Zheng X, Zou D, Zong W, et al: **EWAS Open**
1331 **Platform: integrated data, knowledge and toolkit for epigenome-wide association study.**
1332 *Nucleic Acids Res* 2022, **50**:D1004-D1009.

1333 80. Yang F, Xiong Z, Zong W, Kong D, Tang B, Chen X, Wei Y, Yu X, Zhang Y, Zou D, et al: **EWAS Open**
1334 **Platform 2026: a deeply integrated resource for epigenome-wide association studies.** *Nucleic*
1335 *Acids Res* 2025.

1336 81. Gulen MF, Samson N, Keller A, Schwabenland M, Liu C, Gluck S, Thacker VV, Favre L, Mangeat
1337 B, Kroese LJ, et al: **cGAS-STING drives ageing-related inflammation and neurodegeneration.**
1338 *Nature* 2023, **620**:374-380.

1339 82. O'Halloran AM, Patterson CC, Horan P, Maree A, Curtin R, Stanton A, McKeown PP, Shields DC:
1340 **Genetic polymorphisms in platelet-related proteins and coronary artery disease:**
1341 **investigation of candidate genes, including N-acetylgalactosaminyltransferase 4 (GALNT4)**
1342 **and sulphotransferase 1A1/2 (SULT1A1/2).** *J Thromb Thrombolysis* 2009, **27**:175-184.

1343 83. Reppell M, Zheng X, Dreher I, Blaes J, Regan E, Haslberger T, Guay H, Pivorunas V, Smaoui N:
1344 **HLA-DQA1*05 Associates With Anti-Tumor Necrosis Factor Immunogenicity and Low**
1345 **Adalimumab Trough Concentrations in Inflammatory Bowel Disease Patients From the**
1346 **SERENE Ulcerative Colitis and Crohn's Disease Studies.** *J Crohns Colitis* 2025, **19**.

1347 84. Du Z, Ling Y, Tong H, Guo X, Teschendorff AE: **OmniAge: a compendium of aging omic**
1348 **biomarkers links mitotic clocks to clonal hematopoiesis and causality.** zenodo
1349 <https://zenodo.org/records/20199075>; 2026.

1350 85. Rudovich N, Pivovarov O, Fisher E, Fischer-Rosinsky A, Spranger J, Mohlig M, Schulze MB,
1351 Boeing H, Pfeiffer AF: **Polymorphisms within insulin-degrading enzyme (IDE) gene determine**
1352 **insulin metabolism and risk of type 2 diabetes.** *J Mol Med (Berl)* 2009, **87**:1145-1151.

1353 86. Sladek R, Rocheleau G, Rung J, Dina C, Shen L, Serre D, Boutin P, Vincent D, Belisle A, Hadjadj S,
1354 et al: **A genome-wide association study identifies novel risk loci for type 2 diabetes.** *Nature*
1355 2007, **445**:881-885.

1356 87. Zeggini E, Weedon MN, Lindgren CM, Frayling TM, Elliott KS, Lango H, Timpson NJ, Perry JR,
1357 Rayner NW, Freathy RM, et al: **Replication of genome-wide association signals in UK samples**
1358 **reveals risk loci for type 2 diabetes.** *Science* 2007, **316**:1336-1341.

1359 88. Steneberg P, Bernardo L, Edfalk S, Lundberg L, Backlund F, Ostenson CG, Edlund H: **The type 2**
1360 **diabetes-associated gene ide is required for insulin secretion and suppression of alpha-**
1361 **synuclein levels in beta-cells.** *Diabetes* 2013, **62**:2004-2014.

1362 89. Bertram L, McQueen MB, Mullin K, Blacker D, Tanzi RE: **Systematic meta-analyses of Alzheimer**
1363 **disease genetic association studies: the AlzGene database.** *Nat Genet* 2007, **39**:17-23.

1364 90. Sun H, Xia Y, Wang L, Wang Y, Chang X: **PSORS1C1 may be involved in rheumatoid arthritis.**
1365 *Immunol Lett* 2013, **153**:9-14.

1366 91. Steinberg J, Ritchie GRS, Roumeliotis TI, Jayasuriya RL, Clark MJ, Brooks RA, Binch ALA, Shah
1367 KM, Coyle R, Pardo M, et al: **Integrative epigenomics, transcriptomics and proteomics of**
1368 **patient chondrocytes reveal genes and pathways involved in osteoarthritis.** *Sci Rep* 2017,
1369 **7**:8935.

1370 92. Tang S, Zhang C, Oo WM, Fu K, Risberg MA, Bierma-Zeinstra SM, Neogi T, Atukorala I, Malfait
1371 AM, Ding C, Hunter DJ: **Osteoarthritis.** *Nat Rev Dis Primers* 2025, **11**:10.

1372 93. Breeze CE, Beck S, Berndt SI, Franceschini N: **The missing diversity in human epigenomic**
1373 **studies.** *Nat Genet* 2022, **54**:737-739.

1374 94. Breeze CE, Wong JYY, Beck S, Berndt SI, Franceschini N: **Diversity in EWAS: current state,**
1375 **challenges, and solutions.** *Genome Med* 2022, **14**:71.

1376 95. Shen X, Wang C, Zhou X, Zhou W, Hornburg D, Wu S, Snyder MP: **Nonlinear dynamics of multi-**
1377 **omics profiles during human aging.** *Nat Aging* 2024, **4**:1619-1634.

1378 96. Avchaciov K, Antoch MP, Andrianova EL, Tarkhov AE, Menshikov LI, Burmistrova O, Gudkov AV,
1379 Fedichev PO: **Unsupervised learning of aging principles from longitudinal data.** *Nat Commun*
1380 2022, **13**:6529.

1381 97. Pyrkov TV, Avchaciov K, Tarkhov AE, Menshikov LI, Gudkov AV, Fedichev PO: **Longitudinal**
1382 **analysis of blood markers reveals progressive loss of resilience and predicts human lifespan**
1383 **limit.** *Nat Commun* 2021, **12**:2765.

1384 98. Fedichev PO, Gruber J: **A Minimal Model Explains Aging Regimes and Guides Intervention**
1385 **Strategies.** *bioRxiv* 2025, **2025.08.25.671954**.

1386 99. Tarkhov AE, Lindstrom-Vautrin T, Zhang S, Ying K, Moqri M, Zhang B, Tyshkovskiy A, Levy O,
1387 Gladyshev VN: **Nature of epigenetic aging from a single-cell perspective.** *Nat Aging* 2024,
1388 **4**:854-870.

1389 100. Tong H, Dwaraka VB, Chen Q, Luo Q, Lasky-Su JA, Smith R, Teschendorff AE: **Quantifying the**
1390 **stochastic component of epigenetic aging.** *Nat Aging* 2024, **4**:886-901.

1391 101. Ciarchi M, Simons BD, Rulands S: **A temporal hierarchy determines epigenetic aging** *bioRxiv*
1392 2025, **2025.08.13.669476**.

1393 102. Hashtroud A, Bonder M, Stegle O, von Meyen F, Reik W, Rulands S: **Collective dynamics of DNA**
1394 **methylation during ageing.** *bioRxiv* 2024, **2024.12.15.628564**.

1395 103. Meyer DH, Schumacher B: **Aging clocks based on accumulating stochastic variation.** *Nat Aging*
1396 2024, **4**:871-885.

1397 104. Aryee MJ, Jaffe AE, Corrada-Bravo H, Ladd-Acosta C, Feinberg AP, Hansen KD, Irizarry RA: **Minfi:**
1398 **a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA**
1399 **methylation microarrays.** *Bioinformatics* 2014, **30**:1363-1369.

1400 105. Teschendorff AE, Marabita F, Lechner M, Bartlett T, Tegner J, Gomez-Cabrero D, Beck S: **A beta-**
1401 **mixture quantile normalization method for correcting probe design bias in Illumina Infinium**
1402 **450 k DNA methylation data.** *Bioinformatics* 2013, **29**:189-196.

1403 106. Silverman BW: *Density Estimation for Statistics and Data Analysis.* London: Chapman & Hall;
1404 1986.

1405 107. Wettenhall JM, Smyth GK: **limmaGUI: a graphical user interface for linear modeling of**

microarray data. *Bioinformatics* 2004, **20**:3705-3706.

108. Chen L, Ge B, Casale FP, Vasquez L, Kwan T, Garrido-Martin D, Watt S, Yan Y, Kundu K, Ecker S, et al: **Genetic Drivers of Epigenetic and Transcriptional Variation in Human Immune Cells.** *Cell* 2016, **167**:1398-1414 e1324.
109. Reynolds LM, Taylor JR, Ding J, Lohman K, Johnson C, Siscovick D, Burke G, Post W, Shea S, Jacobs DR, Jr., et al: **Age-related variations in the methylome associated with gene expression in human monocytes and T cells.** *Nat Commun* 2014, **5**:5366.
110. Dabrowski JK, Yang EJ, Crofts SJC, Hillary RF, Simpson DJ, McCartney DL, Marioni RE, Kirschner K, Latorre-Crespo E, Chandra T: **Probabilistic inference of epigenetic age acceleration from cellular dynamics.** *Nat Aging* 2024.
111. Guo X, Robertson JA, Aparicio A, Seale K, Chen Q, Richmond A, Du Z, Sulaiman M, Zheng SC, Ballestar E, et al: **Variations in Innate Immune Cell Subtypes Correlate with Epigenetic Clocks, Inflammaging and Health Outcomes.** *Adv Sci (Weinh)* 2025:e05922.
112. Guo X, Teschendorff AE: **Epigenetic clocks and inflammaging: pitfalls caused by ignoring cell-type heterogeneity.** *Geroscience* 2025.
113. Salas LA, Zhang Z, Koestler DC, Butler RA, Hansen HM, Molinaro AM, Wiencke JK, Kelsey KT, Christensen BC: **Enhanced cell deconvolution of peripheral blood using DNA methylation for high-resolution immune profiling.** *Nat Commun* 2022, **13**:761.
114. Sakaue S, Kanai M, Tanigawa Y, Karjalainen J, Kurki M, Koshiba S, Narita A, Konuma T, Yamamoto K, Akiyama M, et al: **A cross-population atlas of genetic associations for 220 human phenotypes.** *Nat Genet* 2021, **53**:1415-1424.
115. Kanai M, Ulirsch JC, Karjalainen J, Kurki M, Karczewski KJ, Fauman E, Wang QS, Jacobs H, Aguet F, Ardlie KG, Kerimov N, al. e: **Insights from complex trait fine-mapping across diverse populations.** *medRxiv* 2021.
116. Wang G, Sarkar A, Carbonetto P, Stephens M: **A simple new approach to variable selection in regression, with application to genetic fine mapping.** *J R Stat Soc Series B Stat Methodol* 2020, **82**:1273-1300.

Figure Legends

Fig.1. A deep-learning derived CpG taxonomy of age-related DNAm patterns. a) Starting out from an EPICv1 DNAm dataset of approximately 3500 whole blood samples and over 811,870 CpGs, we used statistical criteria to define an initial list of CpG categories or taxa, each one characterized by a particular age or non age-dependent DNAm pattern. DNAm scatterplots against age were transformed into 2-dimensional images, and CpGs split into a training and a test set. **b)** The 2D images of the training set were then used as input to a deep convolutional neural network (ResNet18) to learn

features that distinguish the different patterns from each other. The resulting ResNet18 predictor was then validated on the independent test set. Application of this predictor to originally ‘unclassified’ CpGs allowed re-assignment of these into various categories, thus allowing for an iterative procedure which sequentially assigns more CpGs to various CpG categories. **c)** Examples of 2D images (smoothed scatterplots of DNAm vs scaled age) for one CpG in each of 10 identified categories. **d)** Barplots display for each CpG category and iteration, the number of CpGs that are assigned to that category. We also display the total number of assigned (classified) and unassigned (unclassified) CpGs as a function of iteration index. **e)** Evaluation of the ResNet18-iteration1 classification model on the test set of CpGs (i.e. CpGs that were confidently assigned to each category using statistical criteria to serve as ground truth).

Fig.2. Biological characterization of the 10 CpG age-related categories. **a)** Heatmap displaying statistical over-enrichment of each CpG category for 15 chromatin states and 5 histone marks as defined in differentiated immune cells (top panel) and hematopoietic stem-cells (bottom panel), as computed using eFORGE2 for the top-1000 CpGs of each category. Statistical significance is estimated using q-values derived from a one-tailed Binomial test. Heatmap displays the average $-\log_{10}(\text{q-values})$ across all immune cells or hematopoietic stem cells. **b)** Heatmap displaying the average of absolute t-statistics from linear regressions correlating the DNAm profiles of CpGs in each CpG category to estimated immune-cell fractions, as assessed over the 3500 blood samples. The average of absolute t-statistics is taken over all CpGs in a given CpG category. **c)** Balloon plot displaying statistical over-enrichment of each CpG category for various epigenetic clocks. The number of CpGs in each epigenetic clock is given below the panel. Balloon size represents ratio of observed to expected overlap. P-values are derived from a one-tailed Binomial test.

Fig.3. Biological significance of NoiseClock CpGs. **a)** Flowchart depicting the derivation of the 399 NoiseClock CpGs. In each of the 3 cohorts, we selected noise CpGs as those not displaying any change of average DNAm with age, but displaying increased variance. The overlap of noise CpGs (399) between the 3 cohorts was highly statistically significant under the SuperExactTest. Smoothed scatterplot of an example noise CpG is shown on the top right. P(LM) and P(BP) are the P-values obtained from linear regression and Breusch-Pagan tests, respectively. Deviations of DNAm from the mean (y-axis) for the 399 noise-CpGs (x-axis) for a sample displaying little noise and a sample displaying high noise is shown at the bottom right. **b)** Mathematical definition of the NoiseClock estimator. Smoothed scatterplot of the calculated NoiseClock in the TZH cohort (y-axis) vs chronological age (x-axis) with the green dots representing the best exponential fit. Forest plot of effect size associations between the noise-level estimate of the 399 CpG NoiseClock with age across 14 blood cohorts. Bars represent 95% confidence intervals. Fixed and random effect meta-analyses are given together with their P-values. **c)** Flowchart depicting the derivation of 62 serum CRP-associated CpGs that are not discriminative of naïve and memory lymphocyte subsets, and from

the intrinsic inflammatory score (IntInflScore) is then computed. **d)** As **b)** but for the NoiseClock against the IntInflScore, adjusting for age, sex and 12 immune-cell type fractions. **e)** As **d)** but for the extrinsic age-acceleration (EAA) measure of 4 clocks against IntInflScore adjusting for sex and 12 immune-cell fractions. **f)** Violin plots of NoiseClock levels in the NSPT cohort with samples stratified by presence (red) or absence (blue) of 4 inflammatory diseases, including metabolic syndrome, hypertension, type-2-diabetes and rheumatoid arthritis. P-values are from a one-tailed Wilcoxon-test. **g)** Forest plot of Hazard Ratios for NoiseClock estimator against all-cause-mortality (842 death events) in Generation Scotland cohort (n=18859) for an unadjusted model, for a model adjusted for only age and sex, for a model adjusted for all main risk factors, for a model adjusted by DunedinPACE, for a model adjusted by DunedinPACE in addition to sex and age and finally for models adjusted for GrimAge2, and additionally also by age and sex.

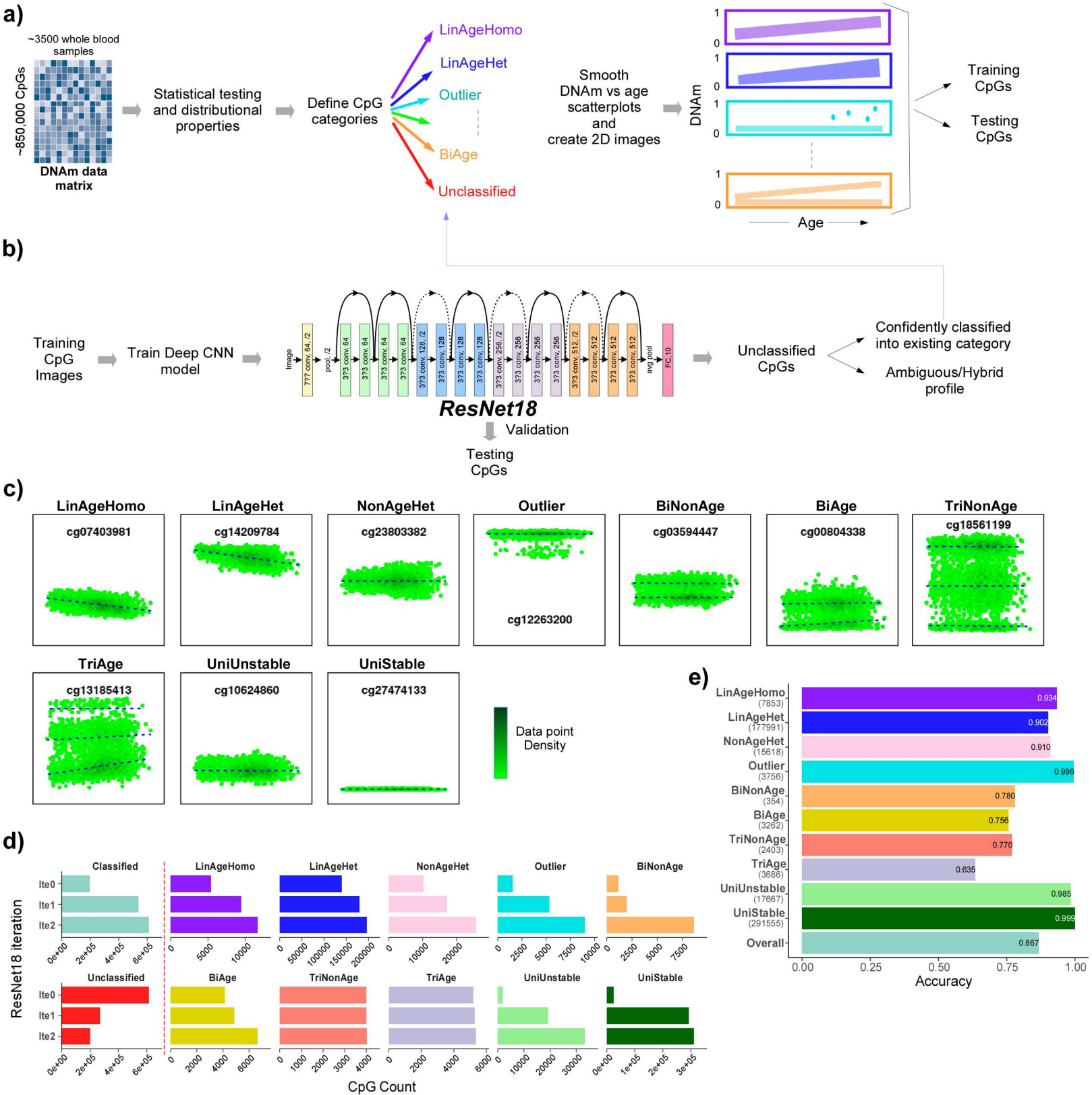
Fig.4. Identification and biological significance of genotype specific age-mQTLs.

a) Identification flowchart of genotype specific age-mQTLs, which are mQTLs displaying an association with age independently of variations in genotype, sex and 7 main immune-cell fractions, and where the association with age is present in only the BB or AB+BB genotypes (B=minor allele). Two examples are given. **b)** Enrichment barplot displaying odds ratio (OR) of DamAge and AdaptAge causal clock CpGs among all age-mQTLs and the genotype specific ones. P-values are from a one-tailed Fisher's exact test. **c)** Flowchart indicates the numbers that are hypermethylated and hypomethylated with age and how many of these also display associations with a DNAm-based proxy of serum CRP levels. **d)** Scatterplot of the t-statistic of association with age with the corresponding t-statistic of association with CRP-proxy for genotype specific age-mQTLs that correlate with CRP. P-value is from a one-tailed Fisher-test. **e)** Barplots of observed to expected enrichment ratios of SNPs linked to genotype specific age-mQTLs displaying correlations with CRP/IL6, among GWAS SNPs associated with inflammatory disease traits, as indicated. P-values are from one-tailed Binomial tests. **f)** Two examples of genotype-specific age-mQTLs correlating with CRP or IL6 and where the linked SNP has been associated with an inflammatory disease, as indicated. P-values are from a linear regression.

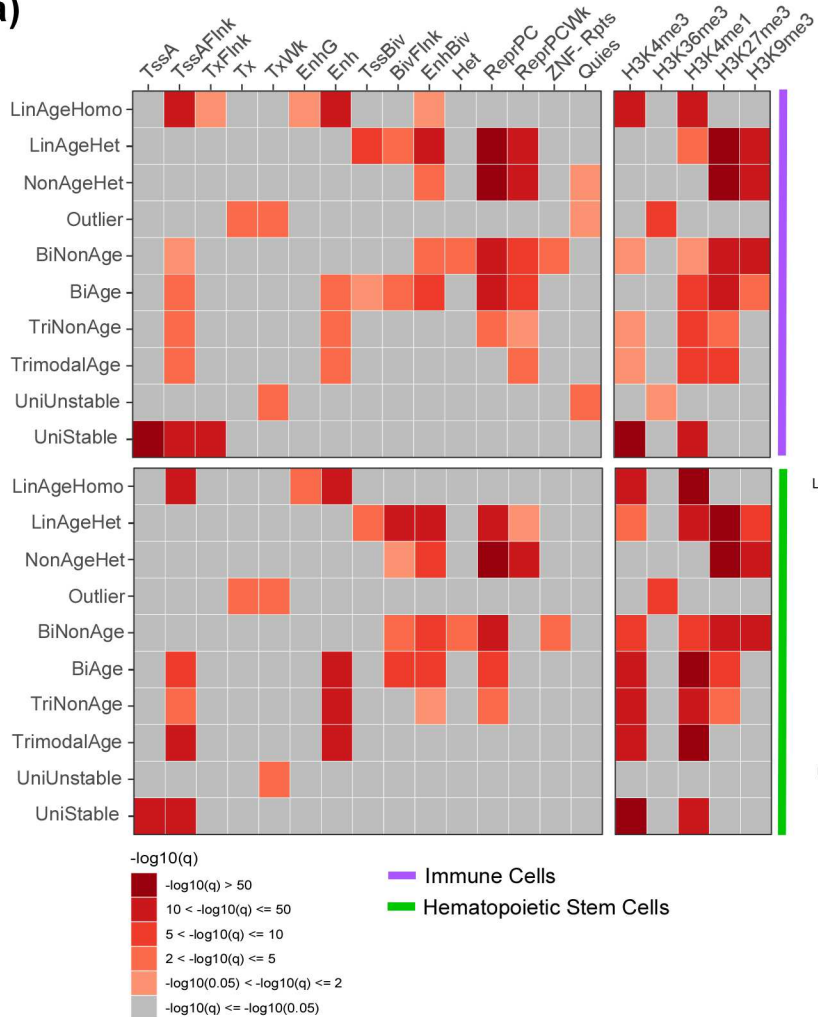
Fig.5. Genotype specific age-mQTLs as mediators of inflammatory diseases. a)

Flowchart describing the selection of CpGs and SNPs for 2-sample Mendelian Randomization analysis. Briefly, our starting point are the BB and BBAB age-mQTLs where the CpG's DNAm level correlates not only with age in the given genotype group(s) but also with CRP or IL6 levels, and where additionally the linked SNP is associated with inflammatory traits. For each CpG, we then find all correlated SNPs, using only independent SNPs as instrumental variables (IVs). For these, TwoSample Mendelian Randomization is then performed. **b)** Scatterplots displaying SNP effect sizes on traits (y-axis) against SNP effect sizes on CpG DNAm (x-axis). Standard errors

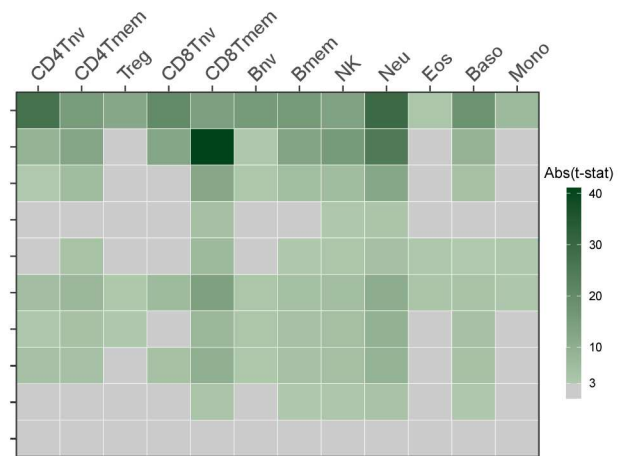
for both SNP–trait and SNP–CpG associations are shown. Red dashed line is from the inverse variance weighted (IVW) model. P-value in red is are from the IVW model. Blue dashed line and P-values are from MR-PRESSO regression after outlier removal. CpG-trait paper is indicated above each panel. CRP=C-reactive protein, CAD=coronary artery disease, UC=ulcerative colitis.



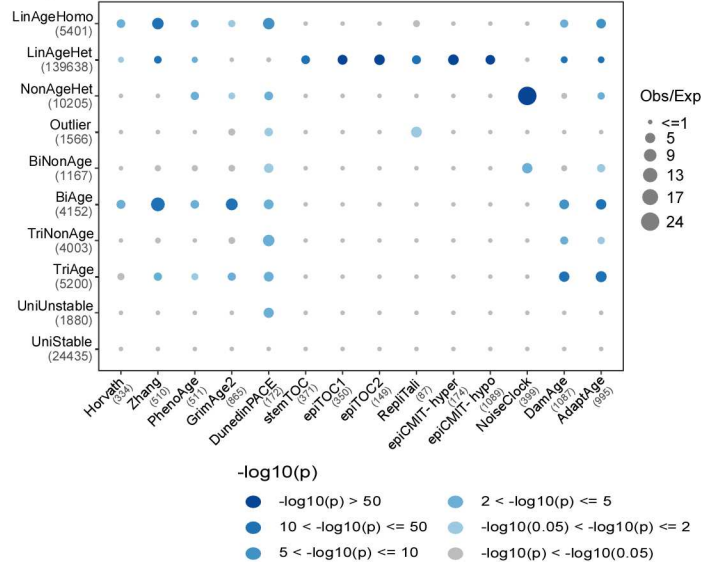
a)

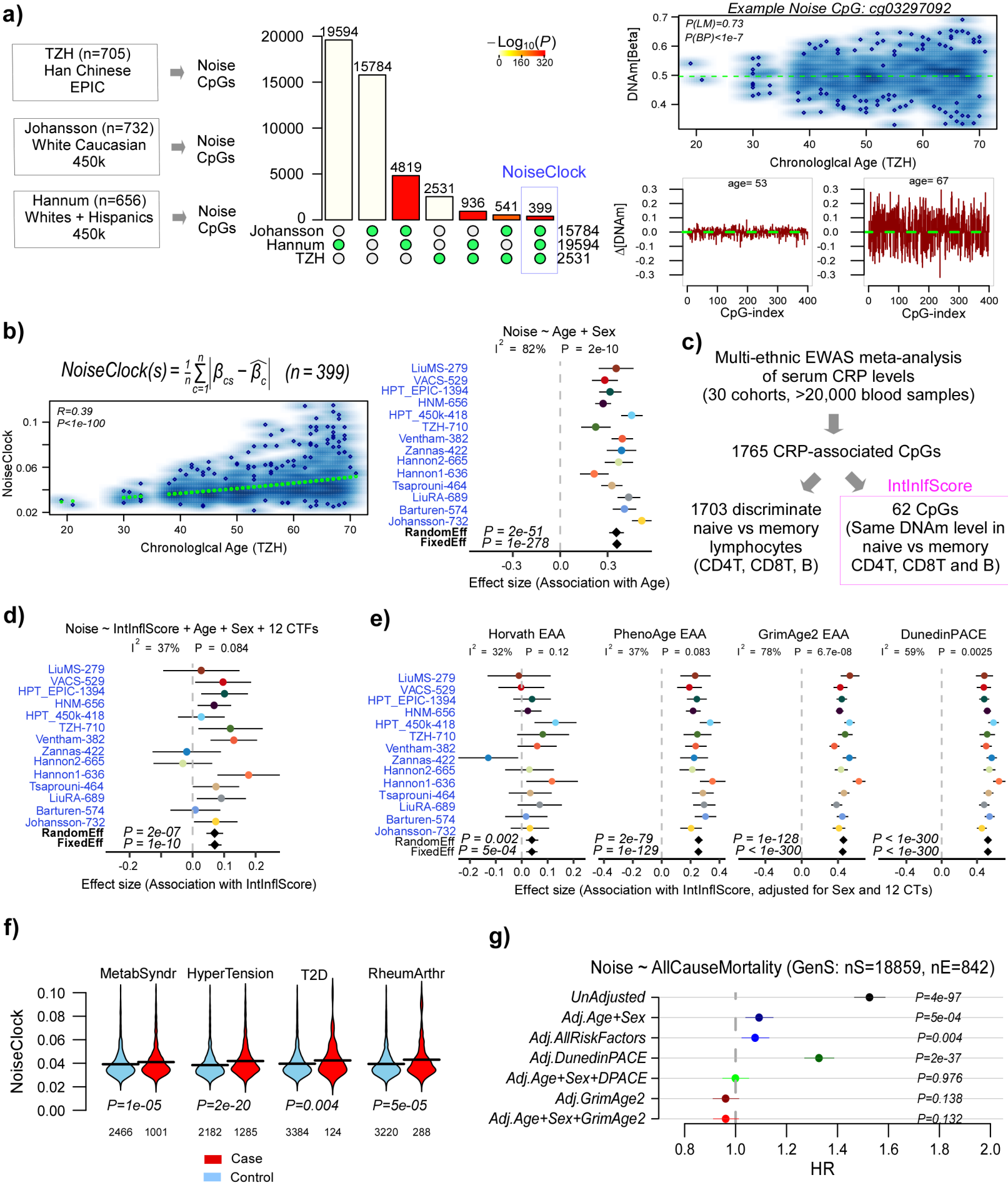


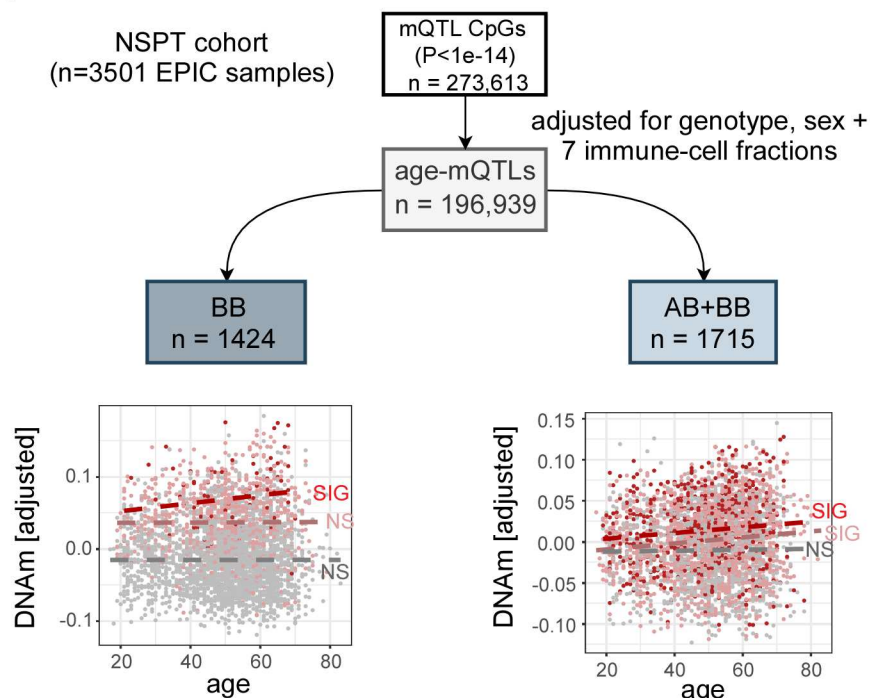
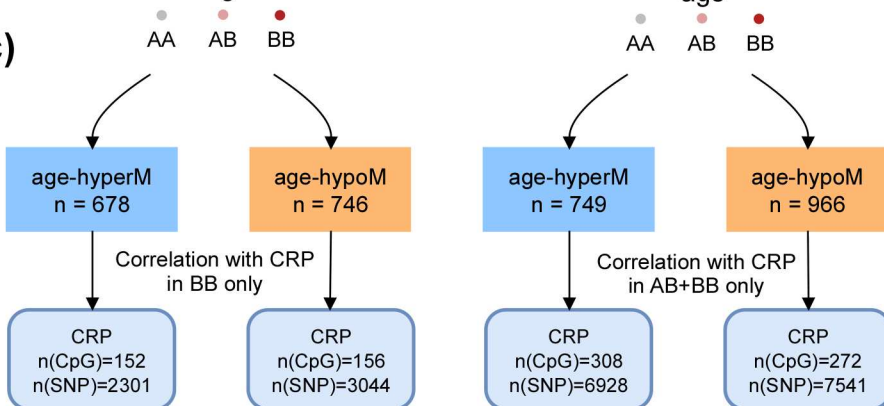
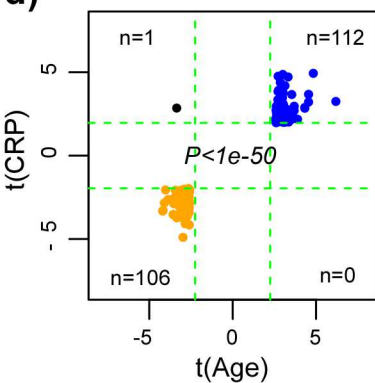
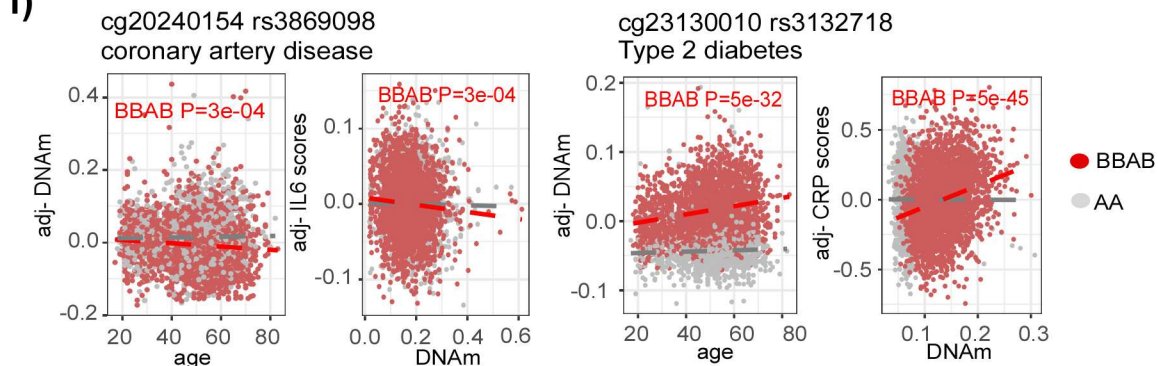
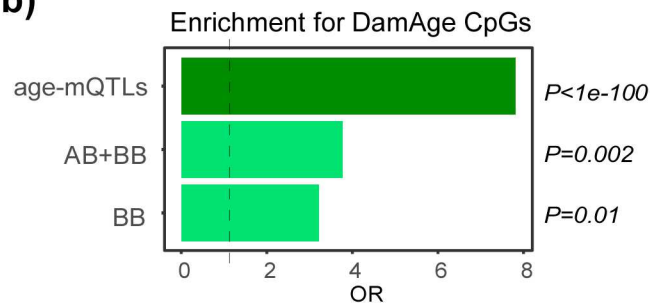
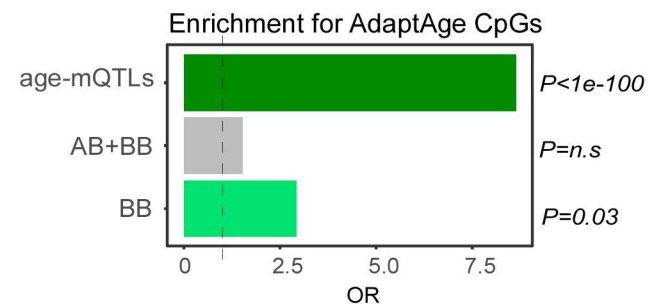
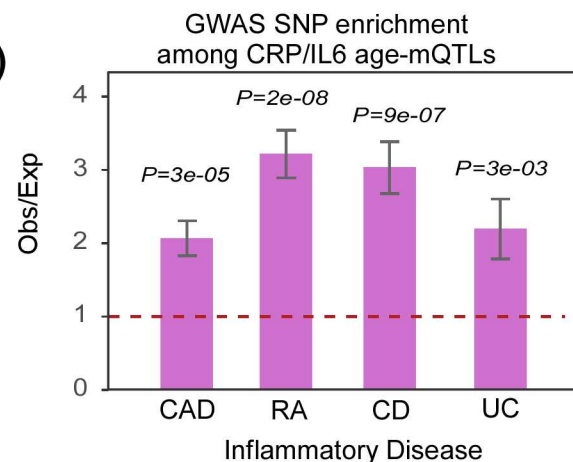
b)

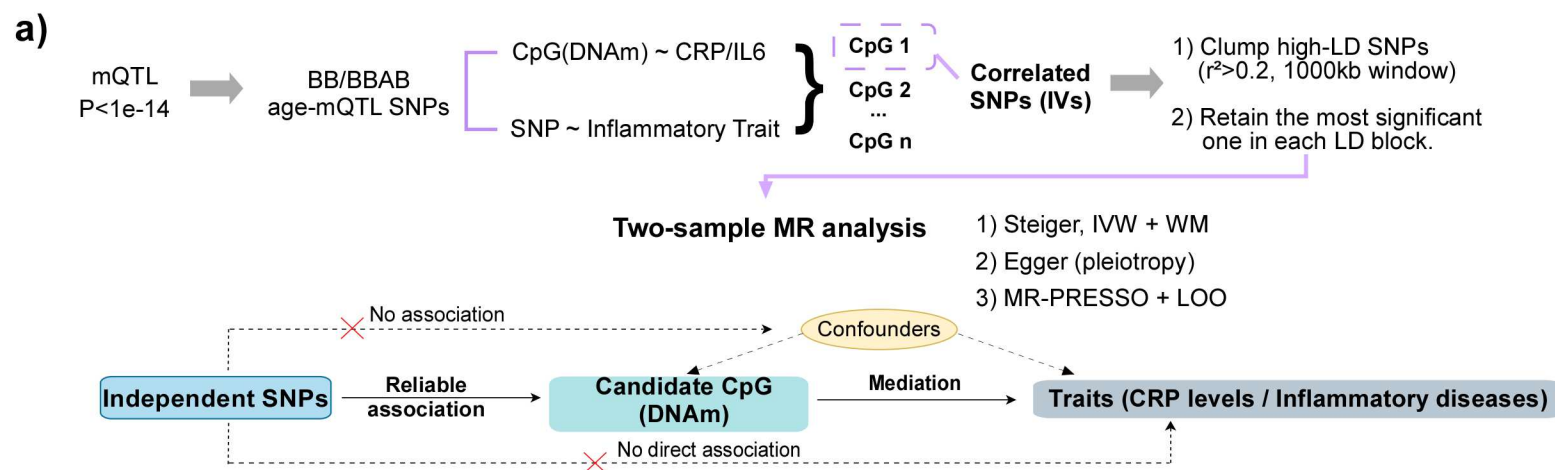
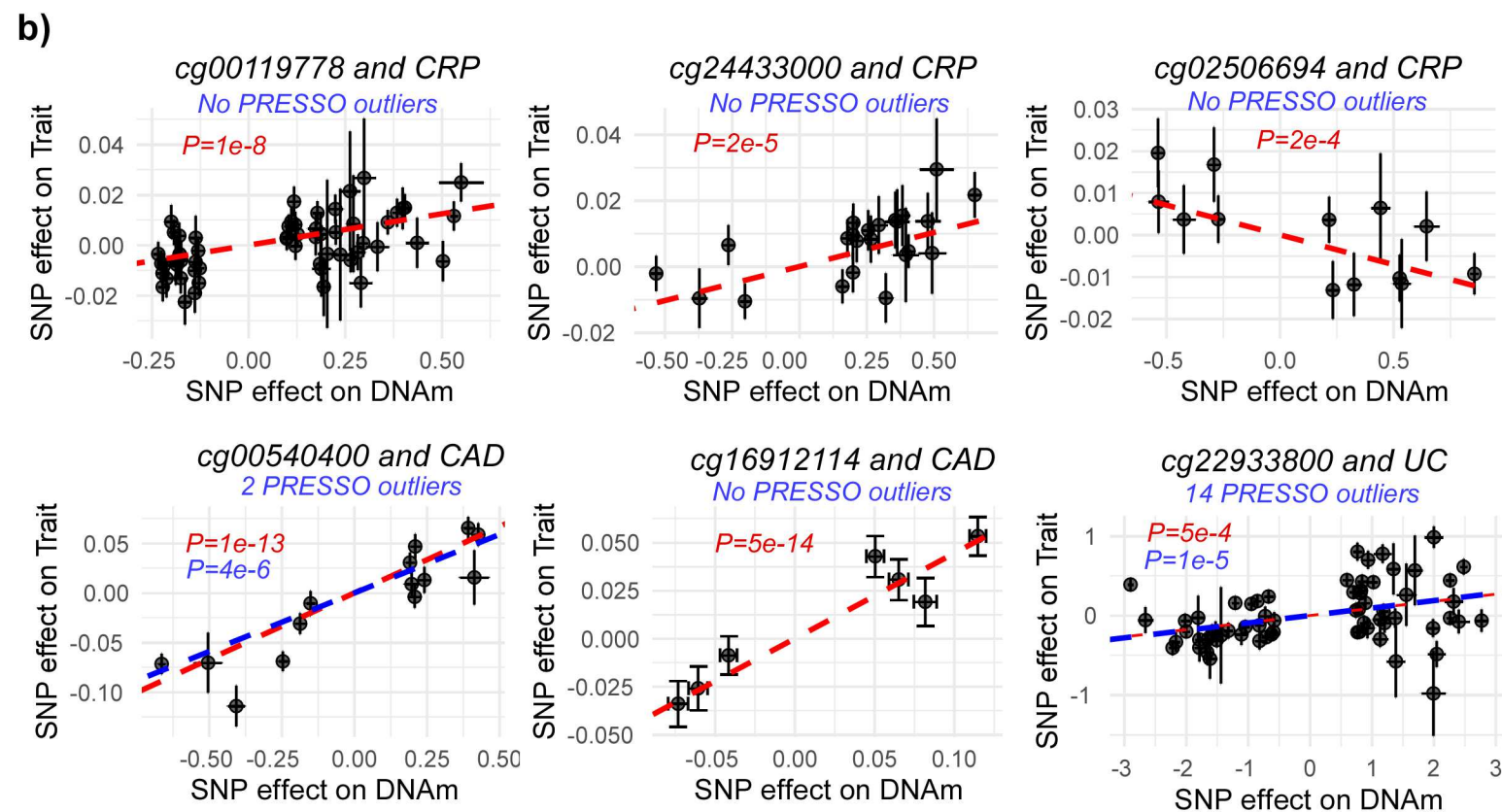


c)





a)**c)****d)****f)****b)****e)**



Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SIAgeTaxoDeepLNpjAging.pdf](#)
- [SuppTables.xlsx](#)
- [SupplementaryMethods.docx](#)