

Cell-type specific epigenetic clocks to quantify biological age at cell-type resolution

Huige Tong^{1,*}, Xiaolong Guo^{1,*}, Magsue Jacques², Qi Luo¹, Nir Eynon², Andrew E. Teschendorff¹

¹CAS Key Laboratory of Computational Biology, CAS-MPG Partner Institute for Computational Biology, Shanghai Institute of Nutrition and Health, Shanghai Institute for Biological Sciences, University of Chinese Academy of Sciences, Chinese Academy of Sciences, Shanghai 200031, China

²Australian Regenerative Medicine Institute (ARMI), Faculty of Medicine, Nursing and Health Sciences, Monash University, Clayton, Victoria 3800, Australia

*Equal contribution

Correspondence to: Andrew E. Teschendorff; **email:** andrew@sinh.ac.cn

Keywords: DNA methylation, epigenetic clocks, cell-type deconvolution, biological aging, Alzheimer's disease, obesity

Received: August 12, 2024

Accepted: December 12, 2024

Published: December 29, 2024

Copyright: © 2024 Tong et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/) (CC BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

ABSTRACT

The ability to accurately quantify biological age could help monitor and control healthy aging. Epigenetic clocks have emerged as promising tools for estimating biological age, yet they have been developed from heterogeneous bulk tissues, and are thus composites of two aging processes, one reflecting the change of cell-type composition with age and another reflecting the aging of individual cell-types. There is thus a need to dissect and quantify these two components of epigenetic clocks, and to develop epigenetic clocks that can yield biological age estimates at cell-type resolution. Here we demonstrate that in blood and brain, approximately 39% and 12% of an epigenetic clock's accuracy is driven by underlying shifts in lymphocyte and neuronal subsets, respectively. Using brain and liver tissue as prototypes, we build and validate neuron and hepatocyte specific DNA methylation clocks, and demonstrate that these cell-type specific clocks yield improved estimates of chronological age in the corresponding cell and tissue-types. We find that neuron and glia specific clocks display biological age acceleration in Alzheimer's Disease with the effect being strongest for glia in the temporal lobe. Moreover, CpGs from these clocks display a small but significant overlap with the causal DamAge-clock, mapping to key genes implicated in neurodegeneration. The hepatocyte clock is found accelerated in liver under various pathological conditions. In contrast, non-cell-type specific clocks do not display biological age-acceleration, or only do so marginally. In summary, this work highlights the importance of dissecting epigenetic clocks and quantifying biological age at cell-type resolution.

INTRODUCTION

Individuals of the same chronological age may display marked differences in health and future disease risk [1]. These differences are attributed to variations in biological age, and whilst a number of biomarkers for biological age have been developed, these are mostly measured in, or restricted to, easily accessible tissues like blood [1–4]. More recently, epigenetic clocks based on DNA methylation (DNAm) marks [5–7] have

emerged as powerful tools to predict chronological age, and to some extent also biological age [3, 8–16]. However, many challenges remain that limit the practical utility and interpretation of these epigenetic clocks [9, 10, 17].

Prominent among these challenges is cell-type heterogeneity (CTH) [18, 19]: most if not all epigenetic clocks have been developed or trained from heterogeneous bulk-tissues that are composed of many

different cell-types. This CTH has two undesirable effects, as recently stressed by others [20–22]. First, it can distort biological age estimates by underlying variations in cell-type fractions. As a concrete example, whole blood, as well as solid tissues with a significant immune-cell infiltration, are all characterized by an age-related shift from naïve to mature T-cell fractions [23–26], which as shown by Jonkman et al. can confound biological age estimates in immune cell subtypes [23]. Whole blood tissue is also subject to an age-related skew towards the myeloid lineage [27]. Another example is brain, with early studies reporting correlations of an accelerated Horvath-clock [28] with neurodegenerative features (e.g. neurotic plaques or amyloid load) [29, 30] in the prefrontal cortex, or with genetic variants associated with neurodegenerative diseases [31–33], but with more recent studies attributing such age-accelerations to changes in underlying cell-type proportions [34–36]. The second undesirable effect of CTH is that, even if we adjust for it *a-posteriori*, as it is often done in blood tissue by adjusting epigenetic clock values for variations in estimated immune cell fractions [28], the ensuing epigenetic clock estimates are still averages over all the distinct cell-types in the tissue. Thus, standard *a-posteriori* adjustment for cell-type fractions does not inform us about the biological aging of individual cell-types. This is a serious limitation for understanding which cell-types age faster and contribute most to age-related diseases, or for identifying the cell-types that respond best to cellular rejuvenation and other healthy-age promoting interventions [37–42].

Thus, current epigenetic clocks are composites, reflecting a mixture of two underlying age-related processes: an extrinsic one reflecting age-related variations of a tissue's cell-type composition, and an intrinsic one, reflecting age-related DNAm changes in the individual cell-types of the tissue, the latter being itself a composite process, as each cell-type may age at a different rate. Importantly, the relative contribution of extrinsic and intrinsic aging processes to an epigenetic clock's accuracy has not yet been quantified, and the only attempts so far at quantifying biological age at cell-type resolution have come from studies that have applied epigenetic clocks derived from bulk-tissue (e.g. liver) to sparse and unreliable single-cell DNAm data (e.g. hepatocytes) [43].

Here we advance the epigenetic clock field in two ways. First, we dissect epigenetic clocks into their extrinsic and intrinsic components, rigorously quantifying these aging components in two of the most frequently profiled tissues (whole blood and brain), demonstrating that extrinsic aging can account for a substantial proportion of an epigenetic clock's predictive accuracy. Second,

we present a computational strategy for building cell-type specific epigenetic clocks that can yield biological age estimates at cell-type resolution. This strategy explicitly adjusts for CTH whilst training the clock, leveraging recent state-of-the-art cell-type deconvolution methods [24, 44] to identify cell-type specific age-associated DNAm changes [45]. As such this is one of the first studies to characterize age-associated DNAm changes at cell-type resolution. We illustrate our strategy in brain and liver tissue, demonstrating how cell-type specific epigenetic clocks from these tissues can improve tissue-specific estimation of chronological and biological age.

RESULTS

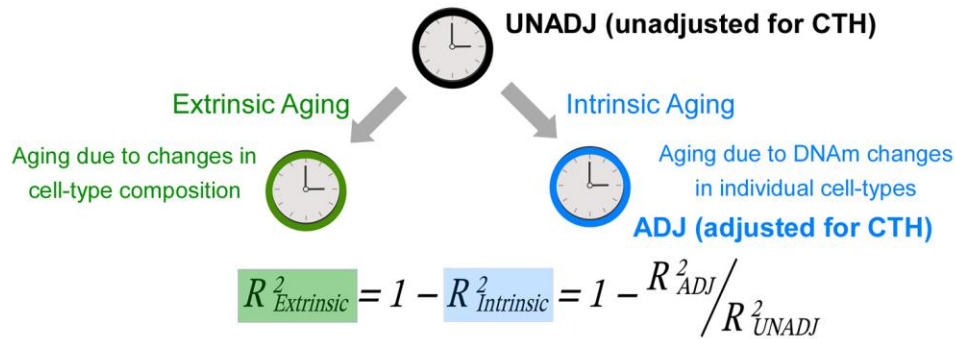
Dissecting and quantifying the two components of epigenetic clocks

Epigenetic clocks are composites of two underlying age-related processes, an extrinsic one that reflects age-associated variations in cell-type composition and an intrinsic one that captures age-associated DNAm changes in individual cell-types (Figure 1A). Thus, the predictive accuracy (R^2 value) of a clock can be composed as the sum of an extrinsic and intrinsic accuracy, so that the extrinsic R^2 value can be estimated from the predictive accuracies of two clocks: one adjusted for CTH, and another which is not (Figure 1A). To quantify the relative rates of extrinsic and intrinsic aging, we assembled a large collection of whole blood Illumina DNAm datasets, using two of the largest cohorts to train Elastic Net age-predictors unadjusted and adjusted for CTH, respectively (Methods). Adjustment for CTH was performed at the resolution of 12 immune cell subtypes using a validated DNAm reference matrix encompassing naïve and memory B-cells, naïve and memory CD4T-cells, naïve and memory CD8T-cells, T-regulatory cells, natural-killer cells, monocytes, neutrophils, eosinophils and basophils [24] (Methods). We also built an elastic net clock adjusting for 9 immune-cell subtypes, by merging together the estimated fractions of the naïve and memory lymphocyte subsets, allowing us to assess the importance of the known age-related shift from naïve to mature T-cell subsets in driving an epigenetic clock's accuracy (Methods) [24]. We then estimated R^2 values of the 3 clocks in 11 independent whole blood DNAm datasets (Methods). This revealed that the unadjusted clock achieved substantially higher R^2 values than the two adjusted clocks, especially when compared to the clock adjusted for all 12 immune cell types (Figure 1B and Supplementary Figures 1, 2). We estimated that the intrinsic aging component accounts for approximately 61% of the predictive accuracy of the unadjusted clock (Figure 1B), which means that

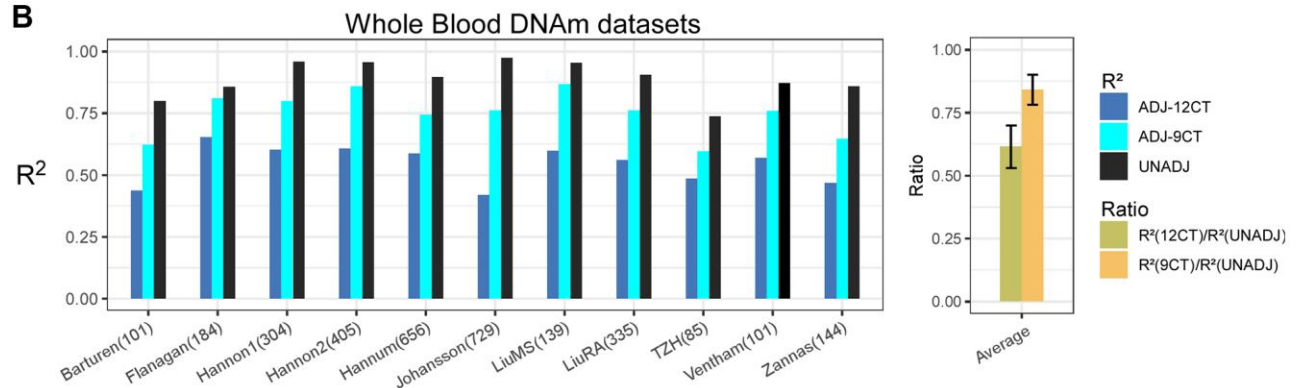
extrinsic aging accounts for a substantial fraction of a clock's accuracy (39%). Of note, not adjusting for relative shifts in naïve to memory subsets, we would have over-and-under estimated the intrinsic and extrinsic aging components to be approximately 84% and 16%, respectively (Figure 1B). By assembling a large collection of brain DNAm datasets (Methods, Supplementary Table 1) and adjusting for 7 brain cell-

type fractions (GABAergic excitatory neurons, glutamatergic inhibitory neurons, astrocytes, microglia, oligodendrocytes, endothelial cells and stromal cells), as estimated with the HiBED algorithm [46], we verified that the intrinsic component of aging in brain is around 88%, higher than the estimate obtained in blood (Figure 1C and Supplementary Figure 3). To see which brain cell-types are driving the extrinsic aging component,

A Quantification of the two components of tissue-specific epigenetic clocks



B



C

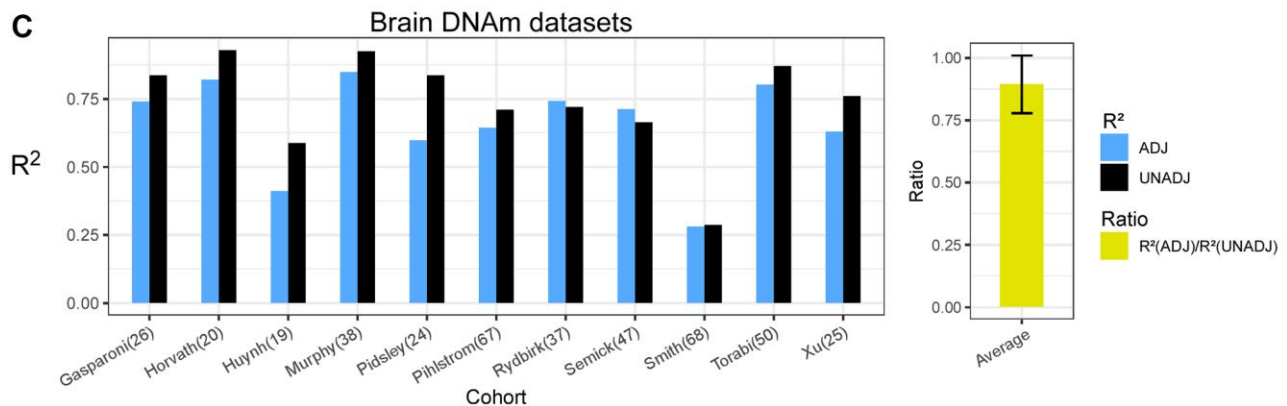


Figure 1. Quantification of intrinsic and extrinsic aging. (A) Graphical depiction of how the predictive accuracy of an epigenetic clock derived from bulk-tissue can be decomposed in terms of an extrinsic and intrinsic aging process, as shown. The accuracy, measured by the R^2 value, of a clock reflecting the extrinsic process can be estimated from the given formula, if one can estimate the accuracies of the full unadjusted clock and the one fully adjusted for cell-type heterogeneity (CTH), the latter reflecting the intrinsic aging of the individual cell-types. (B) Barplots compare the R^2 values of the unadjusted clock, and clocks adjusted for 12 and 9 immune cell types, in each of 11 whole blood cohorts. Right panel displays the average and standard deviation of the two indicated ratios across the 11 cohorts. These ratios measure the fraction of the unadjusted clock's accuracy that can be attributed to intrinsic aging. (C) As (B) but for brain DNAm datasets, where the adjusted clock was adjusted for 7 brain cell-types.

we performed a fixed and random effects meta-analysis over 13 Illumina DNAm brain datasets, encompassing over 2,000 samples (Supplementary Table 1), which revealed an age-associated increase of excitatory neurons, whilst inhibitory neurons and astrocytes decreased (Figure 2). Thus, this age-related shift in neuronal subtypes accounts for the 12% extrinsic aging component. Of note, these findings are broadly consistent with recent single-nucleus DNAm studies reporting reduced inhibitory neuron fractions in the brain of aged donors [47, 48]. Overall, these results obtained in blood and brain tissue highlight the critical importance of adjusting for CTH when building epigenetic clocks.

A computational strategy to build cell-type specific epigenetic clocks

In order to dissect the intrinsic aging component, we next devised a computational strategy to build cell-type specific epigenetic clocks (Figure 3A). After estimation of the cell-type fractions in a given bulk-tissue DNAm dataset, we apply the CellDMC algorithm [45] to a training subset of the data to identify age-associated differentially methylated cytosines in each cell-type of interest (age-DMCTs) (Figure 3A). CellDMC uses the estimated cell-type fractions as weights in a model that

expresses a CpG's DNAm profile across samples as a corresponding weighted mixture of latent phenotype-dependent DNAm profiles. In the present context, the phenotype is age and the described model includes interaction terms between age and the cell-type fractions (Methods). These interaction terms allow cell-type specific age-DMCTs to be found (see Figure 3A for a hypothetical example). Assuming there is a substantial number of age-DMCTs for a given cell-type, we then apply an Elastic Net Regression model [49] to train predictive models for chronological age parameterized by a penalty parameter (Figure 3B). This training is restricted to age-DMCTs of the given cell-type and can be performed either on the residuals of the DNAm data matrix, obtained by regressing out the effect of cell-type composition, or on the unadjusted DNAm data matrix (Methods). In the former case, we define the clocks as being “intrinsic”, otherwise they are “semi-intrinsic”. The best performing models in a blinded model-selection set define our cell-type specific intrinsic and semi-intrinsic epigenetic clocks, one for each cell-type (Figure 3B). These clocks are then validated in independent DNAm datasets.

The reasoning behind defining separate intrinsic and semi-intrinsic clocks can be understood with simulation models. We generated simulated DNAm mixtures of 3

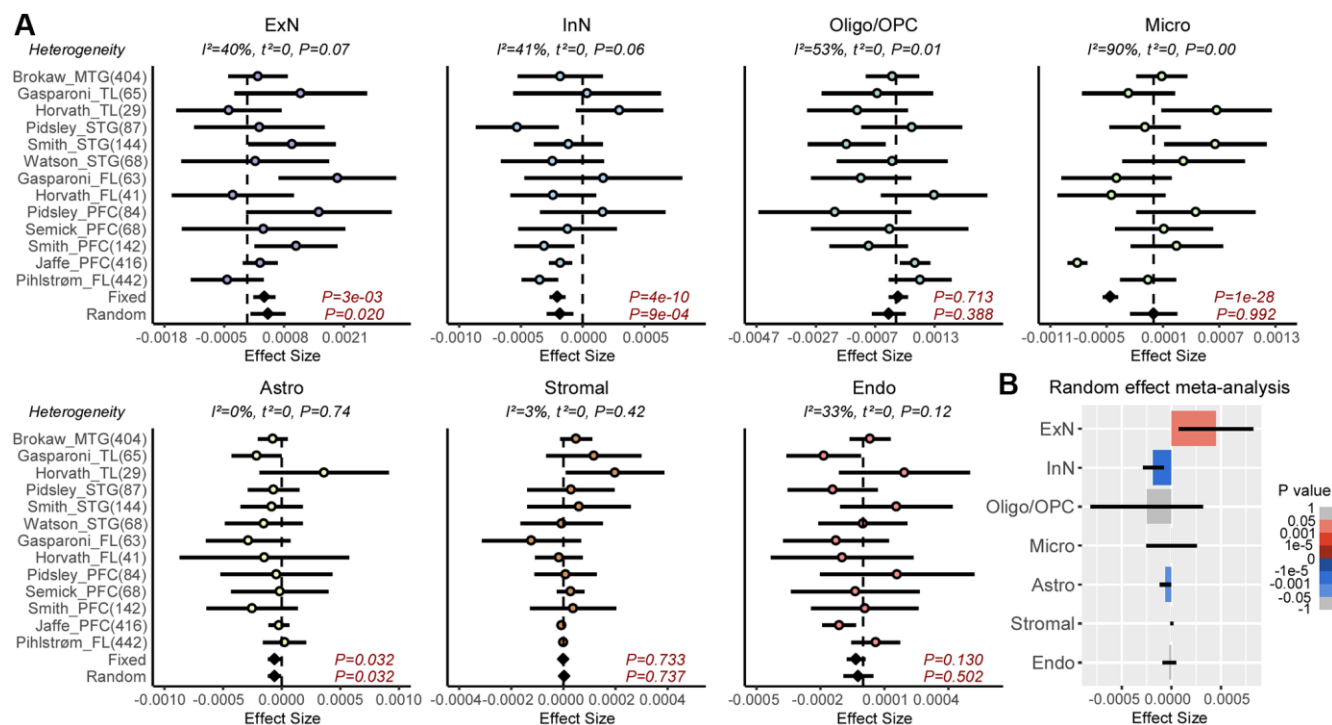


Figure 2. Meta-analysis of brain cell-type fractions with age. (A) Forest plots of associations between brain cell-type fractions and age, for 7 brain cell-types (Abbreviations: ExN: excitatory neurons; InN: inhibitory neurons; Oligo/OPC: oligodendrocytes/oligo precursor cells; Micro: microglia; Astro: astrocytes, Stromal; Endo: endothelial) across 13 independent DNAm datasets. Number of samples in each cohort is given in brackets. *P*-values from a Fixed and Random Effects models are also given. Heterogeneity statistics and *P*-values of heterogeneity are given above each plot. (B) Effect sizes and *P*-values from the Random Effects meta-analysis model.

underlying cell-types (granulocytes, monocytes and lymphocytes), and in which age-DMCTs are introduced into only one cell-type (lymphocytes) under four different scenarios (Methods): (i) age-DMCTs

do not map to cell-type specific CpGs and no cell-type fractions change with age, (ii) as (i) but with the lymphocyte fraction changing with age, (iii) age-DMCTs discriminate lymphocytes from the other two

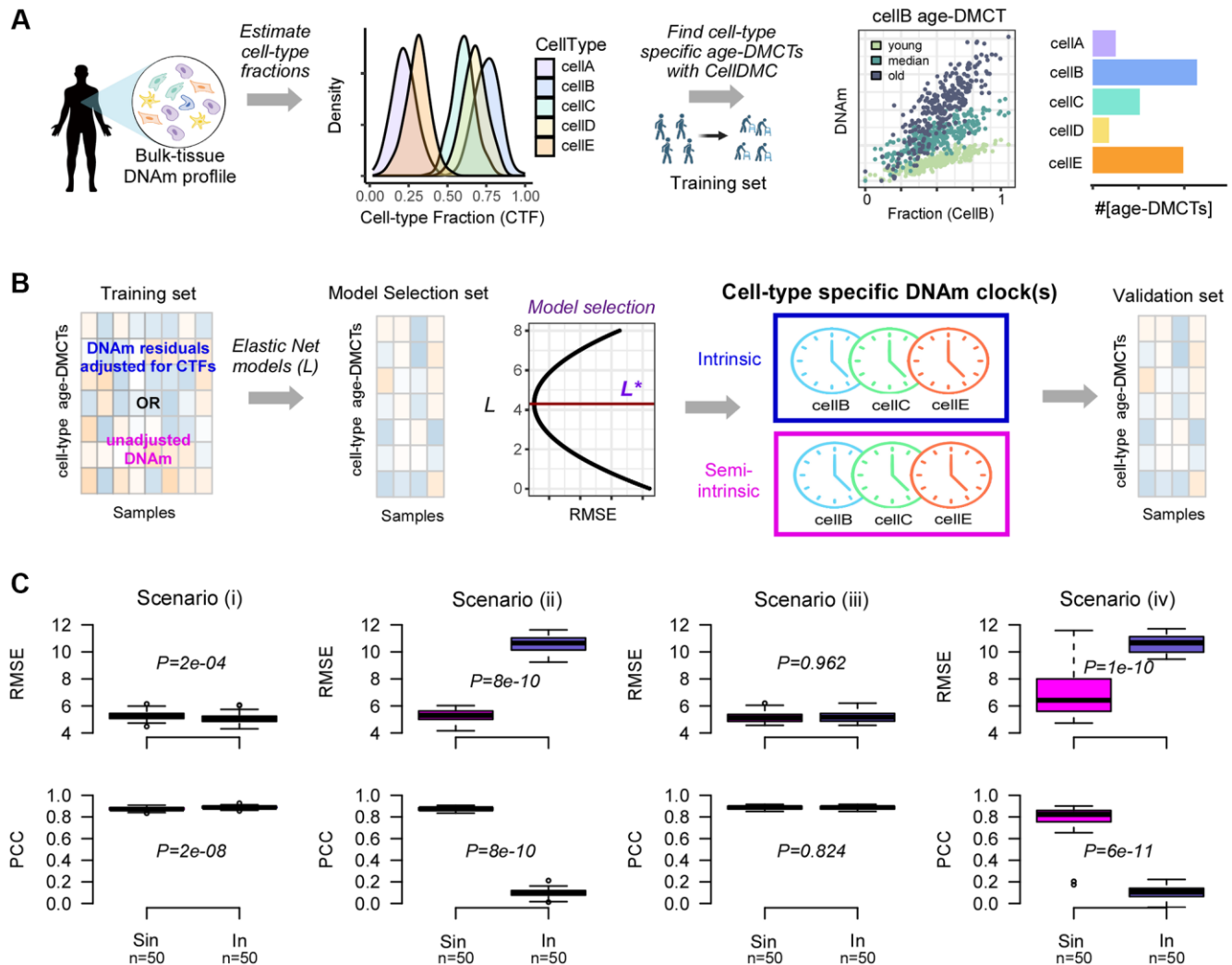


Figure 3. Computational strategy for building cell-type specific epigenetic clocks. (A) Given a tissue-type of interest with DNAm profiles measured over a relatively large number of samples, we first estimate the underlying cell-type proportions in each sample using an existing tissue-specific DNAm reference matrix. Density plots depict the distribution of cell-type fractions across all samples. Next, using a sufficiently large training set of samples encompassing a relatively wide age-range, we apply the CellDMC algorithm to infer age-associated DNAm changes in each of the underlying cell-types (age-DMCTs). Barplot depicts the number of age-DMCTs in each cell-type. (B) The construction of cell-type specific clocks then proceeds by restricting to age-DMCTs of one cell-type: an intrinsic clock is built by adjusting the DNAm training dataset for variations in cell-type fractions (CTFs), defining a matrix of DNAm residuals. Alternatively, the training over the age-DMCTs can be done on the DNAm data matrix without adjustment for CTFs which will result in a “semi-intrinsic” clock. In either case, Elastic Net models are learned for each choice of a penalty parameter L , and the optimal model L^* is selected based on the best generalization performance (smallest root mean square error (RMSE)) obtained in a blinded model selection set. This optimal model then defines the corresponding cell-type specific DNAm-clock. This procedure can be done for each cell-type separately, assuming that sufficient numbers of age-DMCTs in that cell-type can be identified. Once the cell-type specific clocks are built, these are then validated in independent DNAm datasets. (C) Top row: boxplots display the root mean square error (RMSE) between predicted and true ages, for semi-intrinsic (Sin) and intrinsic (In) clocks, as assessed in 50 simulated validation sets of 200 mixtures for four different scenarios. Cell-type specific clocks were constructed from 50 simulated training sets of 200 mixtures (mixing together 3 cell-types that we call granulocytes, monocytes and lymphocytes) with age-DMCTs occurring only in one cell-type (lymphocytes). In scenarios (i) and (iii), no cell-type fractions change with age, in scenarios (ii) and (iv) the lymphocyte fraction changes with age. In scenarios (i) and (ii), age-DMCTs do not discriminate immune-cell types from each other, in scenarios (iii) and (iv) age-DMCTs discriminate lymphocytes from granulocytes and monocytes. P-values from a two-tailed paired Wilcoxon test are given. Bottom row: as top-row but now displaying the PCC of predicted vs. true age.

cell-types and no cell-type fraction changes with age, (iv) as (iii) but with the lymphocyte fraction changing with age. For each scenario, we generated training and validation sets, deriving lymphocyte specific intrinsic and semi-intrinsic clocks, subsequently validating them in the independent test sets. Interestingly, whilst the intrinsic clock displayed marginally better performance in the scenarios where cell-type fractions did not change with age, the semi-intrinsic clock was substantially better in cases where the lymphocyte fraction did change with age (Figure 3C). Thus, the use of residuals in scenarios where a given cell-type fraction changes with age can dilute out the biological signal in that cell-type, and under these circumstances, the cell-type specific semi-intrinsic clock would be preferable.

Construction and validation of neuron-specific epigenetic clocks

Next, we applied this strategy to the case of brain tissue. For training the clocks, we used one of the largest collections of prefrontal cortex (PFC) samples profiled with Illumina 450k technology [50], encompassing 416 samples with a wide age range (18–97 years) (Supplementary Table 1 and Figure 4A). Although the HiBED algorithm [46] inferred fractions for 7 brain cell-types (Supplementary Figure 4A), for subsequent CellDMC analysis, power considerations [45, 51] required us to collapse the inferred fractions to 3 broad cell-types: neurons, glia and endothelial/stromal (Figure 4A). Applying CellDMC with the estimated fractions, disease status and sex as covariates, resulted in 16,814 neuron-specific and 11,157 glia-specific age-DMCTs (Figure 4A, Methods). Visualization of representative age-DMCTs confirmed the cell-type specific nature of their age-associated DNAm changes (Figure 4B). Interestingly, neuron and glia age-DMCTs displayed a weak marginal co-localization with the cell-type specific CpGs in the brain DNAm reference matrix (Supplementary Figure 5). Focusing on the neuron-DMCTs and adjusting the DNAm data for variations in cell-type fractions, we next applied the Elastic Net regression framework to learn different models specified by a penalty parameter. For model (i.e. optimal parameter) selection, we used a completely independent Illumina DNAm dataset from Philstrom et al. [52] which included 67 PFC control samples with an age range 49–100 years (Figure 4C and Supplementary Table 1). We declared the model achieving best generalization performance in the Philstrom dataset, as defining our neuron-specific intrinsic (“Neu-In”) clock (Figure 4C, Supplementary Figure 4B and Supplementary Table 2). This neuron-specific clock was further validated in 5 independent datasets of sorted neuron samples, encompassing a total of 143 samples (Figure 4D, Supplementary Figure 4C and Supplementary Table 1).

An analogous neuron “semi-intrinsic” clock (“Neu-Sin”) was obtained by applying the same training procedure to the same age-DMCTs in the Jaffe et al. DNAm data but now without adjusting the DNAm data for variations in cell-type fractions (Methods, Figure 4D, Supplementary Figure 4D and Supplementary Table 3). To assess specificity of the clocks to neurons and brain tissue, we assembled a large collection of Illumina DNAm datasets encompassing other brain cell-types (glia), other sorted non-brain cell-types and bulk tissue-types (Supplementary Table 1, Methods, and Supplementary Methods). As a further control, we also benchmarked our Neu-In/Sin clocks to two other clocks: a clock obtained by running the Elastic Net on all CpGs but adjusting the DNAm data for cell-type fractions (Methods, Figure 4D, Supplementary Figure 4E and Supplementary Table 4), thus defining a brain, but not neuron-specific, clock (“BrainClock”), and the multi-tissue Horvath clock. We restricted the assessment of all clocks to control samples only, to avoid potential confounding by disease. First, we observed that the neuron intrinsic and semi-intrinsic clocks also displayed good correlations with chronological age in sorted glia datasets, although the largest of these glia datasets displayed a significantly weaker association, suggesting that our neuron-clocks are indeed neuron-specific (Figure 4D). In line with this, a multivariate linear regression of correlation strength against cell-type (neuron vs. glia), weighted by cohort size, revealed a significantly higher correlation in neurons compared to glia for the neuron-intrinsic clock but not for the other 3 clocks (Supplementary Figure 6A). When assessed between brain and non-brain tissue/cell-types, we observed that the two neuron-specific clocks as well as the brain-specific clock achieved better predictions of chronological age in the brain-tissue datasets compared to non-brain ones, whilst, as expected for a multi-tissue clock, Horvath’s clock performed similarly between brain and non-brain tissues (Figure 4E, 4F). Importantly, and despite the relatively small number of sorted neuron datasets, the neuron-specific clocks also displayed stronger correlations with age in the sorted neuron datasets compared to the brain-specific clock (Supplementary Figure 6B). Thus, overall, our data support the view that the neuron-specific clocks are indeed optimized to predict chronological age in brain cell subtypes and that they are less confounded by changes in cell-type composition compared to e.g. the brain-specific clock.

Neuron and glia-specific clocks are accelerated in Alzheimer’s Disease

Next, we asked if the clocks predict age-acceleration in neurodegeneration. To this end, we collated eleven DNAm-studies of Alzheimer’s Disease encompassing

two brain regions, frontal lobe (FL) and temporal lobe (TL) (Supplementary Table 5), amounting to a total of 213 AD cases and 185 controls in FL, and 456 cases and 341 controls in TL. As before, we first performed a meta-analysis of cell-type fractions to see if any of these fractions change in AD (Methods). This revealed a significant decrease of excitatory neurons

and corresponding increases of oligodendrocyte/OPC and microglia fractions in AD (Supplementary Figure 7). Whilst the decreased excitatory neuron fraction was observed in both TL and FL, the microglia and oligo/OPC increases were more specific to the frontal and temporal lobes, respectively (Supplementary Figure 8). Thus, only the excitatory neuron fraction is observed

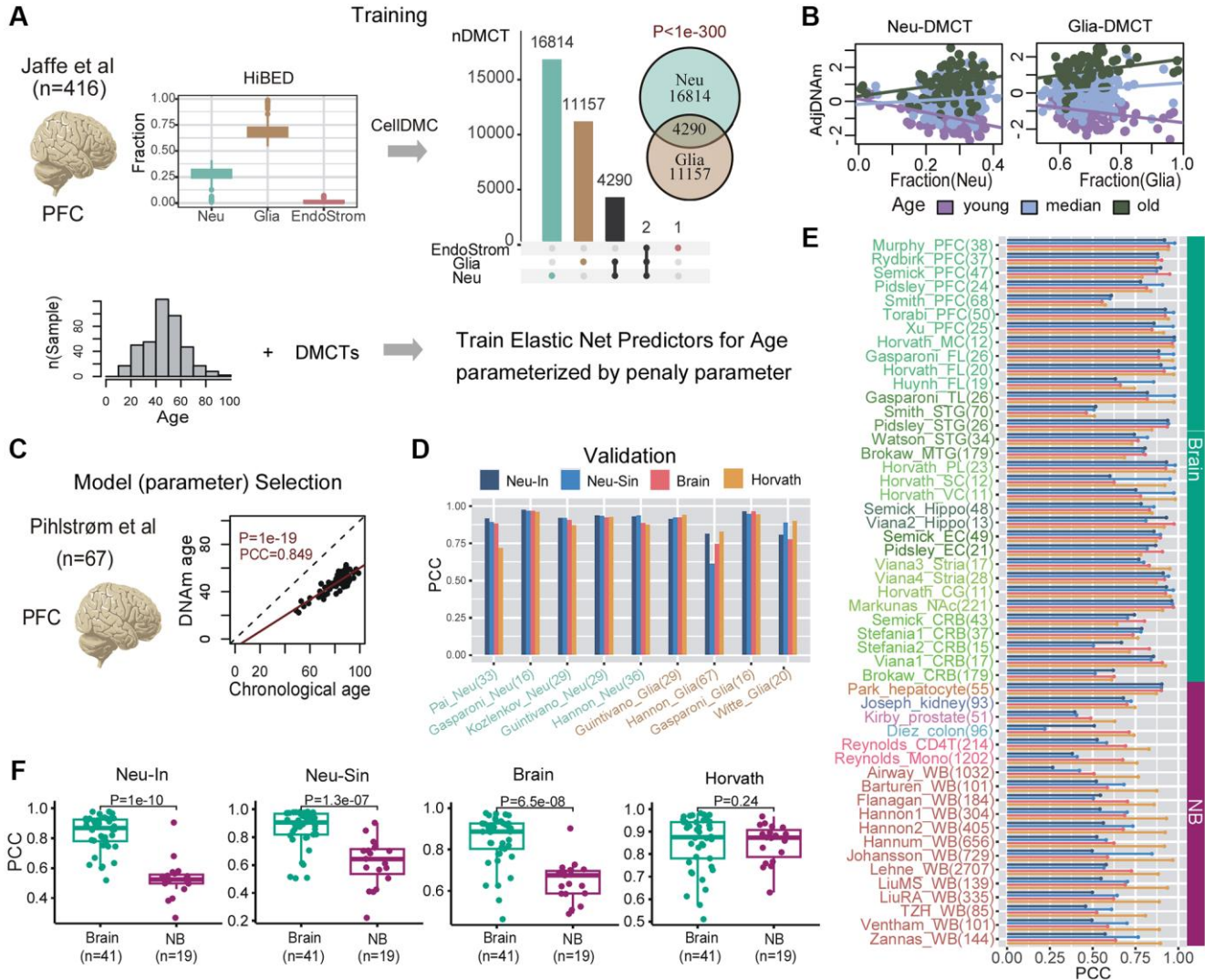


Figure 4. Construction and validation of neuron-clocks. (A) The training is done on the Jaffe et al. dataset encompassing 416 prefrontal cortex (PFC) samples. Fractions for 3 broad cell-types (neurons, glia and endothelial/stromal) are inferred using the HiBED algorithm. Subsequently, CellDMC is applied to infer DMCTs in each cell-type. The Fisher-test one-tailed P -value of overlap between Neu-DMCTs and Glia-DMCTs is given. Finally, elastic net predictors of chronological age are trained from the Neu-DMCTs parameterized by a penalty parameter. (B) Scatterplots of adjusted DNAm values (adjusted for variations in cell-type fractions) of one Neu-DMCT and one Glia-DMCT against the corresponding Neuron and Glia fraction, respectively. Samples have been colored according to age-group and regression lines for each age-group have been added. (C) Optimal model (parameter) selection is performed using a completely independent PFC dataset from Pihlström et al. Scatterplot displays the predicted vs. chronological age for the optimal model. (D) Barplot of Pearson Correlation Coefficients (PCC) for 4 clocks (Neu-In, Neu-Sin, Brain and Horvath) across sorted neuron and glia datasets. Number of samples in each dataset is given in brackets. (E) Barplot of Pearson Correlation Coefficients (PCC) for the same 4 clocks (Neu-In, Neu-Sin, Brain and Horvath) across brain and non-brain DNAm datasets. Each cohort is labeled with the number of samples in brackets. Abbreviations: WB: whole blood; PFC: prefrontal cortex; TL: temporal lobe; FL: frontal lobe; MTG/STG: middle/superior temporal gyrus; PL: parietal lobe; SC: sensory cortex; VC: visual cortex; MC: motor cortex; Hippo: hippocampus; EC: entorhinal cortex; Stria: striatum; CG: cingulate gyrus; NAC: nucleus accumbens; CRB: cerebellum; Mono: monocyte. (F) For each of the 4 clocks, boxplots of PCCs comparing the values obtained in brain vs. non-brain datasets. The number of datasets in each group is given below x-axis. P -value is from a one-tailed Wilcoxon rank sum test.

to change with both age and AD (Figure 2 and Supplementary Figure 7), although interestingly the increased fraction seen with age is reversed in AD cases. For each of the 4 clocks in each brain region we then performed a fixed and random effects meta-analysis for AD (Methods), revealing a weak but statistically significant age-acceleration in AD cases for the neuron-intrinsic clock in the TL, but not for brain-specific and Horvath's clocks (Figure 5A, 5B). In the case of Horvath's clock, the association remained non-significant upon adjustment for brain cell-type fractions (Supplementary Figure 9). The association of the neuron-intrinsic clock in the PFC/FL was marginally significant. Given that the I^2 test for heterogeneity showed great consistency in the effect sizes between cohorts and regions (Figure 5A, 5B), we extended the meta-analysis to include all datasets from both regions, now revealing significant associations for both

intrinsic and semi-intrinsic neuron-clocks (Figure 5C). Because all of the AD datasets considered in this analysis are from bulk-tissue, we wondered if a similar finding would apply to the other major cell-type in brain, i.e. glia. To this end, we used the previous Jaffe and Philstrom DNAm datasets to build and validate analogous glia intrinsic (Glia-In) and semi-intrinsic (Glia-Sin) clocks (Methods, Supplementary Figure 10, Supplementary Tables 6, 7), subsequently computing their DNAm-Ages in the AD datasets. Interestingly, this also revealed age-acceleration effects in the glia compartment of both FL and TL regions, with the effect being strongest for the Glia-Sin clock (Figure 5A–5C). Comparing the predicted age-acceleration of the Neu-In with the Glia-In clock, and that of the Neu-Sin with the Glia-Sin clocks, revealed moderate but significant correlations between the Neu and Glia clock estimates (Supplementary Figure 11). Thus, age-acceleration

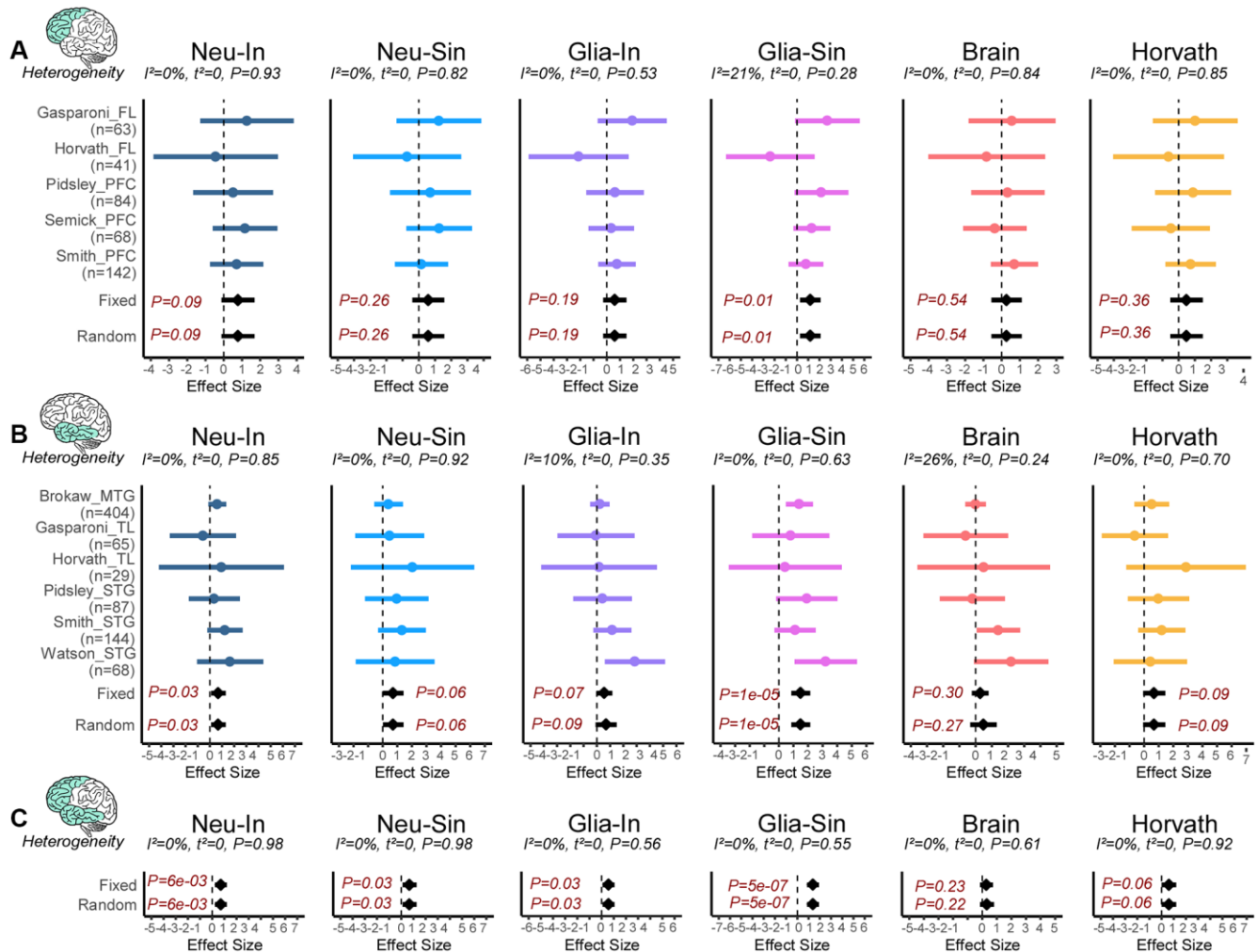


Figure 5. Neuron and glia specific clocks predict age-acceleration in AD. (A) Forest plot of associations between DNAmAge and Alzheimer's Disease (AD) for 6 different epigenetic clocks across 5 independent DNAm datasets from frontal lobe (FL) or prefrontal cortex (PFC). The effect sizes and P -values of a fixed and random effects models are also given. Statistics and P -values of heterogeneity are shown above each panel. Number of samples in each dataset is given to the left. (B) As (A) but for DNAm datasets profiling regions in the temporal lobe (TL), which includes middle and superior temporal gyrus (MTG and STG). (C) As (A), but combining FL and TL regions.

patterns in neurons and glia are similar, despite the two clocks being made up of entirely distinct sets of CpGs. In summary, these data indicate that both neuron and glia specific clocks display age-acceleration in AD, and that this acceleration is more prominent in the TL glia fraction.

Construction and validation of a hepatocyte-specific epigenetic clock

As a second example, we considered the case of liver tissue, aiming to build a hepatocyte-specific epigenetic clock. We used our EpiSCORE algorithm [53] and liver DNAm reference matrix defined over 5 cell-types (hepatocytes, cholangiocytes, Kupffer cells, endothelial cells and lymphocytes) [44] to estimate corresponding cell-type fractions in a series of 210 normal liver specimens (age range = 18–75, mean age = 47, sd = 12) with available Illumina EPIC DNAm profiles [54] (Methods). Liver cell-type fractions changed with chronological age, although this was mainly restricted to cholangiocytes and endothelial cells, a pattern that was replicated in independent cohorts (Supplementary Figure 12). Next, we assessed if we can reliably identify age-DMCTs. To this end we applied CellDMC [45] to all 210 samples, revealing that most age-DMCTs were in either hepatocytes or cholangiocytes, with hepatocyte and cholangiocyte age-DMCTs displaying strong overlap but with opposite directionality of change (Supplementary Figure 13A–13D). As with the age-DMCTs in neurons and glia, here too we observed a preferential co-localization to marker CpGs in our liver DNAm reference matrix (Supplementary Figure 5). To validate the hepatocyte age-DMCTs, we computed the average DNAm over hepatocyte age-hypermethylated DMCTs in an independent set of 55 primary hepatocyte cultures from donors of different ages (Methods), which revealed the expected positive correlation with age (Supplementary Figure 13E). Correspondingly, hepatocyte age-hypomethylated DMCTs displayed the expected anti-correlation, with the pattern exactly inverted for cholangiocyte age-DMCTs (Supplementary Figure 13E). To further validate the hepatocyte and cholangiocyte age-DMCTs, we collected three additional liver DNAm datasets (Methods) and repeated the CellDMC analysis in each one. This confirmed the strong overlap between hepatocyte and cholangiocyte age-DMCTs in each validation cohort (Supplementary Figures 13F and 14). This strong validation supports the view that we can confidently identify age-associated DMCTs in hepatocytes and that these display opposite directionality of change in cholangiocytes.

Having demonstrated that we can successfully identify hepatocyte age-DMCTs, we next applied a 10-fold cross-validation procedure to build a hepatocyte-

specific semi-intrinsic clock (Methods). Of note, this procedure involved re-applying the CellDMC algorithm on the full set of CpGs but now using only on 9 folds, with the leave-one-out bag being used for model selection (Methods). This procedure was iterated 10 times, each time using a different bag as the leave-one-out test set, a procedure that avoids overfitting [55], which resulted in a final age-predictor (“HepClock”) (Supplementary Table 8 and Figure 6A). We also trained an analogous clock but now using age-DMCTs derived from a linear model that only adjusts for cell-type fractions, the resulting clock thus being liver-specific but not hepatocyte-specific (“LiverClock”, Methods, Supplementary Table 9). To validate and benchmark these clocks, we applied them to the DNAm dataset of primary hepatocyte cultures (Methods). Although R-values (Pearson Correlation Coefficients) were similar for all clocks, HepClock attained a significantly lower median absolute error of 4.9 years compared to values of 20 and 10 for the liver and Horvath clocks, respectively (Figure 6B). This suggests that improved predictions of chronological age in sorted hepatocytes are indeed possible with a corresponding hepatocyte specific clock. Next, we applied HepClock and LiverClock to a collection of DNAm datasets from bulk-liver tissue, sorted immune-cell types, as well as other tissue-types including blood, skin, prostate, colon and kidney (Methods). This confirmed the clear specificity of HepClock and LiverClock to liver tissue (Figure 6C). We note that Horvath’s clock was excluded from this analysis, because some of the liver DNAm datasets were used for the original training of the Horvath clock itself [28].

Hepatocyte specific clock predicts biological age-acceleration in liver pathologies

We next applied HepClock, LiverClock and Horvath’s clock to 3 separate DNAm studies representing a variety of liver pathologies, in order to assess the ability of these clocks to predict biological age-acceleration. One study encompassed 115 liver samples from patients with non-alcoholic fatty liver disease (NAFLD) [54], another set comprised 67 liver tissue samples from obese individuals (BMI >35) [56], with the 3rd cohort comprising 94 liver tissue samples from patients with Alpha-1 Antitrypsin Deficiency (AATD) [57], a genetically inherited disorder that increases the risk of cirrhosis (Methods). Whilst on the NAFLD samples all three clocks were accelerated, the observed acceleration was strongest for the Liver and HepClocks (Figure 6D). On the clinically obese samples only HepClock displayed a very significant age-acceleration effect, with Horvath’s clock being marginally associated (Figure 6D). In the AATD-cohort, only the Liver and HepClocks displayed age-acceleration (Figure 6D).

In another independent liver-tissue DNAm dataset encompassing healthy non-obese and obese individuals [58, 59], HepClock and Horvath clock age-acceleration measures displayed a correlation with BMI, but not so for the LiverClock (Figure 6E), mirroring the results in the previous cohort of 67 obese individuals (Figure 6D). The age-acceleration effect in the hepatocytes of NAFLD and obese livers is noteworthy, because the hepatocyte fraction itself decreases substantially in NAFLD and obese individuals (Supplementary Figure 15). Overall, as assessed over the 4 studies, the hepatocyte-specific epigenetic clock improves the sensitivity to detect biological age acceleration in the

hepatocytes of liver tissue in variety of different liver pathologies including an inherited one where Horvath's clock did not predict age-acceleration.

Minimal but significant overlaps with chronological age clocks

Having demonstrated how cell-type specific clocks predict age-acceleration in relevant conditions and cell-types, it is important to study the nature of the specific CpGs making up these clocks. We first assessed overlaps of our clock CpGs with previous clocks including Horvath [28], Zhang [11] and PhenoAge [3].

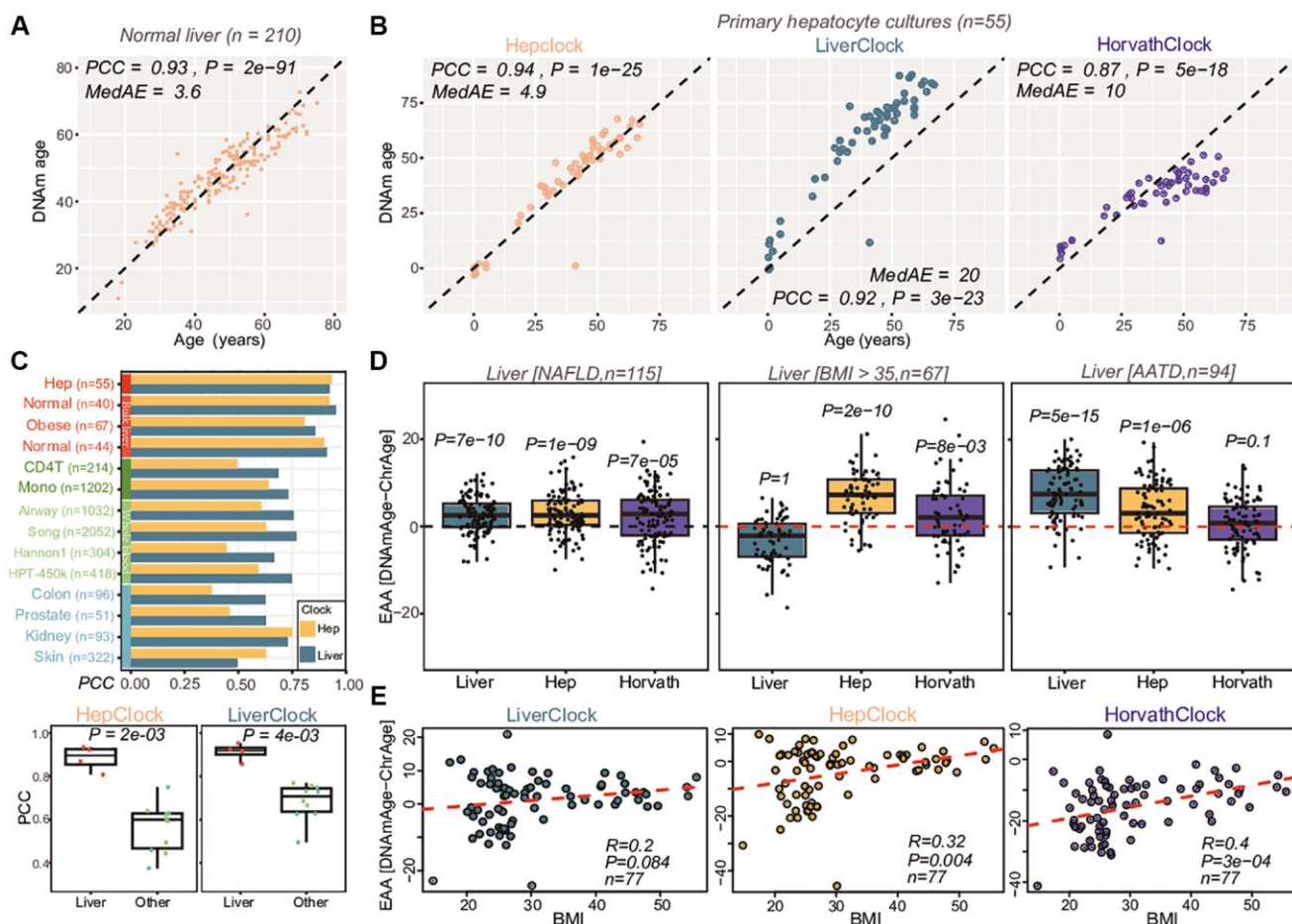


Figure 6. Construction and validation of hepatocyte clock. (A) Overall strategy used to train a hepatocyte specific clock from bulk liver tissue samples. Plots show how the root mean square error (RMSE) changes as a function of the penalty parameter in the elastic net regression model, and scatterplot at the bottom compares the predicted and chronological ages for the optimal model defining the hepatocyte-clock. Pearson Correlation Coefficient (PCC), Median Absolute Error (MedAE) and regression P -value are given. (B) Validation of the hepatocyte and liver clocks (latter trained on age-DMCs, not hepatocyte-DMCTs) in a primary hepatocyte culture. Also shown is the result for Horvath's clock. PCC, MedAE and P -values from a linear regression are given. (C) Barplot comparing the PCC-values of the hepatocyte and liver clocks across several liver and non-liver tissue datasets. Boxplots compare the PCC values of the hepatocyte and liver clocks in the liver/hepatocyte dataset compared to all others. P -value is from a one-tailed Wilcoxon rank sum test. (D) Boxplots compare the extrinsic age-acceleration (EAA) of the liver, hepatocyte and Horvath clocks in two independent liver DNAm datasets, one profiling non-alcoholic fatty liver disease (NAFLD) cases and the other profiling liver-tissue from obese (BMI >35) individuals. In each case, the P -value is from a one-tailed Wilcoxon rank sum test comparing the distribution to a mean value of 0 (red dashed line). (E) Boxplots of EAA from the three clocks in an independent DNAm liver dataset, with samples stratified according to BMI levels, as shown. For each panel, we give the PCC and linear regression test P -value.

In general, although overlaps were small in absolute terms, some were statistically significant (Supplementary Figure 16). For instance, most of our cell-type specific clock CpGs displayed a significant overlap with CpGs from Zhang and Horvath's clocks, two clocks trained to predict chronological age. The smaller non-significant overlap with PhenoAge clock CpGs may indicate that our cell-type specific clocks in brain and liver are capturing different facets of biological aging compared to PhenoAge.

Minimal but significant overlap with DamAge causality clock

The overlap with the recently proposed causal DamAge and AdaptAge clocks [60] was also small in absolute terms, but interestingly the neuron and glia-clock CpGs displayed a significant overlap with the DamAge clock (Binomial $P < 0.001$) and not with AdaptAge (Binomial $P > 0.05$) (Supplementary Figure 16). The overlapping CpGs from Neu-In and Neu-Sin with DamAge mapped to 4 genes (*SESN2*, *EPS8L2*, *SGMS1*, *ZNF642*), whilst the corresponding overlapping DamAge-CpGs from Glia-In and Glia-Sin mapped to 3 genes (*ST3GAL4*, *CBX7*, *TOMM40L*) (Supplementary Table 10). Supporting the statistical significance of these overlaps with the neuron and glia clocks, these overlapping genes have previously been implicated in AD. For instance, *SESN2* plays an important role in AD progression [61, 62] and is regarded as a diagnostic AD biomarker [63]. The expression level of *EPS8L2* is significantly reduced in an AD mouse model [64], whilst *SGMS1* is functionally implicated in AD pathogenesis [65], displaying elevated expression levels in the brain of AD cases [66]. Interestingly, *ZNF642* is upregulated in Parkinson's disease [67], and Zinc finger proteins are well-known to be associated with neurological diseases [68]. One of the shared DamAge glia-clock CpGs mapped to the promoter of *ST3GAL4*, a gene associated with brain atrophy in AD patients [69, 70] and that correlates with Braak staging [71]. Expression of PRC1 component *CBX7* has been shown to regulate the AD related INK4b-ARF-INK4a locus [72] and is associated with suppression of axon growth and regeneration [73]. Finally, *TOMM40L* is in linkage disequilibrium with APOE and contributes synergistically to the risk of Alzheimer's Disease [74–77]. Specifically, the *APOEε4-TOMM40' 523 L* haplotype is associated with a higher risk and shorter times of conversion from mild cognitive impairment (MCI) to AD [78], possibly driven by mitochondrial dysfunction [75]. Overall, these data demonstrate that our brain cell-type specific clocks do map to CpGs and genes that have been causally and negatively implicated in neurodegeneration and AD, which is consistent with our neuron and glia clocks displaying age-acceleration in AD.

The Neu-In clock CpGs are enriched for AD EWAS hits and mQTLs

Given the associations of our clocks with AD and BMI, we next asked if our clock CpGs overlap with AD or BMI associated differentially methylated cytosines (DMCs), as determined by corresponding EWAS [79, 80] (Methods). Whilst for the brain-clocks we did observe a significant overlap with AD-DMCs, this was not evident for the Hep/LiverClocks and BMI-DMCs (Supplementary Figure 17).

Next, we asked if our clock CpGs are associated with mQTLs. We performed an enrichment against three broad classes of mQTLs: adult blood derived mQTLs as given by the ARIES database [81], mQTLs derived from the eGTEx consortium [82] which are available for a number of different tissue-types including colon, prostate, ovary, lung, breast, kidney, skeletal muscle, testis and whole blood, and mQTLs derived for fetal and adult brain [83, 84] (Methods). Using the ARIES database, this revealed a significant enrichment for mQTLs among brain clock CpGs, but only marginal associations for the Hep/LiverClocks (Supplementary Figure 18A), probably because of the smaller number of Hep/LiverClock CpGs, itself reflecting the smaller training set in the case of liver tissue. Interestingly, all brain clock CpGs displayed significant enrichment for mQTLs as defined in adult brain, but not fetal brain, although this negative finding could be due to the smaller number of mQTLs in fetal brain (Supplementary Figure 18B). For the mQTLs as defined by eGTEx we also observed enrichment, but this was restricted mainly to the neuron-clocks and tissues with the largest numbers of mQTLs (e.g. colon, lung, prostate and ovary) (Supplementary Figure 18C), suggesting that other associations may become significant if we had more CpGs in either the clock or mQTL list. Supporting this, we note that we did not observe any enrichment for mQTLs as derived from whole blood in eGTEx, whilst we did observe enrichment against the larger mQTL ARIES database. In the case of the Neu-In clock and blood mQTLs from ARIES, we observed 2 clock-CpGs defining cis-mQTLs where the linked SNP has been associated with AD [85] (Supplementary Figure 18D and Supplementary Table 11). Interestingly, these cis-mQTLs map to the gene *CASS4*, a CAS scaffolding protein family member involved in the formation of neuritic plaques and neurofibrillary tangles, as well as in the disruption of synaptic connections in AD [86, 87]. Of note, Neu-In clock CpGs defining cis-mQTLs in adult brain also included SNPs that mapped to *CASS4*. In addition, another Neu-In/Neu-Sin clock CpG defines a cis-mQTL in brain mapping to the *TSPAN14* gene (Supplementary Figure 18D), which is associated with increased AD risk [85, 88, 89]. Thus, a minor

component of the neuron-specific clocks involves CpGs whose DNAm-levels are also influenced by genetic variants associated with AD.

DISCUSSION

Here we have quantified, within the context of epigenetic clocks, the relative contributions of extrinsic and intrinsic aging, demonstrating that the extrinsic component in blood can account for up to around 40% of an epigenetic clock's accuracy, whilst in brain this fraction is much lower at around 12%. Thus, whilst most of an epigenetic clock's accuracy is intrinsic, driven by age-associated DNAm changes in individual cell-types, the age-related shifts in tissue composition can contribute a substantial amount. For instance, based on our analysis, epigenetic clocks like Horvath or Zhang clocks [11, 28], which typically attain an R^2 value of 0.8 or higher, would likely only display R^2 values of approximately 0.5 if one were to rederive the clocks adjusting for 12 immune-cell type fractions. Whilst in the case of blood, the extrinsic component is driven by an age-related increase of the mature to naïve T-cell fractions, in brain, this is driven by an age-related shift in the excitatory to inhibitory neuron fractions.

Our second key contribution is the dissection of the intrinsic component of aging through construction of cell-type specific epigenetic clocks. As shown here in the case of brain and liver tissue, the resulting cell-type specific clocks demonstrate clear cell-type and tissue-specificity, achieving improved predictions of chronological age in the corresponding sorted cells and tissue-types. Most importantly, our cell-type specific clocks yield improved estimates of biological age in the relevant cell-types and tissues (here brain and liver), as compared to Horvath's multi-tissue clock and the non-cell-type specific (but tissue-specific) brain and liver clocks, the latter result highlighting the importance of distinguishing cell-type specific from tissue-specific clocks.

The case of Alzheimer's Disease is important to discuss further, since there has been substantial controversy and debate as to whether epigenetic clocks display age-acceleration in AD or not. Here we have significantly advanced the field by constructing neuron and glia-specific clocks, and demonstrating an age-acceleration effect for both neurons and glia, with the effect being strongest for the Glia-Sin clock in the temporal lobe, suggesting that epigenetic age-acceleration associated with AD may be particularly prominent in the glia TL compartment. As far as the age-acceleration of the Neu-In and Neu-Sin clocks is concerned, it is worth noting that the excitatory neuron fraction was observed to increase with age, whilst it was decreased in AD.

This pattern would suggest that the observed age-acceleration of the neuron-compartment in AD is not due to shifts in underlying neuronal subtype proportions, but instead caused by DNAm changes in these specific neuronal subtypes. It is also noteworthy that one of the CpGs in the Neu-In clock defines a cis-mQTL linked with a SNP that has been associated with AD and which maps to *CASS4*, a gene which has previously been implicated in neurotic plaque formation and loss of synaptic connectivity in AD [86, 87]. Another CpG in the Neu-In clock mapped to a cis-mQTL implicating *TSPAN14*, which is associated with increased AD risk [85, 88, 89]. These findings are important, because it is conceivable that clock-CpGs whose DNAm levels track age-acceleration in AD due to environmental risk factors (e.g. social isolation, sleep deprivation) may also be under the influence of genetic variants that determine AD-risk, which would thus point towards a shared causal pathway to AD [17].

The neuron and glia specific clock CpGs also displayed a statistically significant overlap with CpGs of the causal DamAge clocks. Remarkably, the few overlapping CpGs mapped to genes that have been strongly implicated in AD. Alongside *CBX7* and *TOMM40L* discussed earlier, it is worth highlighting here also *SESN2* and *SGMS1*. Inactivation of Sestrin-2 (*SESN2*) results in diverse metabolic pathologies, including oxidative damage, fat accumulation, mitochondrial dysfunction, and muscle degeneration, that resemble accelerated tissue aging [90, 91]. *SESN2* is also induced by amyloid- β , playing a protective role against amyloid- β neurotoxicity [61]. Knockdown of sphingomyelin synthase-1 (*SGMS1*) attenuates AD-like pathology through promoting lysosomal degradation of BACE1 [65]. Moreover, *SGMS1* is significantly elevated in the hippocampus of AD brain and inhibition significantly reduces the level of amyloid- β [66]. Of note, for *SGMS1*, the implicated CpG maps to the 5'UTR, where an age-related decrease in DNAm (as observed here in our brain clocks, Supplementary Table 10) would normally be associated with increased expression, hence consistent with the damaging effect as implicated in the DamAge clock. However, it should also be noted that the relation between DNAm and gene-expression is complex.

The ability to quantify biological age at cell-type resolution can also help address whether age-acceleration in various cell-types of a tissue correlate with each other or not. Here we explored this important question in the context of the neuron and glia clocks, revealing that the age-acceleration of these two clocks are indeed significantly correlated. Despite this, the age-acceleration measures were also sufficiently distinct,

with the glia clocks displaying stronger acceleration in AD compared to the neuron-clocks. In the case of liver, it should be noted that most of the identified age-associated DNAm changes occurred in hepatocytes and cholangiocytes, with the underlying DNAm changes in these two cell-types displaying a very strong anti-correlation. Thus, a putative cholangiocyte clock would be composed of overwhelmingly the exact same CpGs making up the hepatocyte clock, and thus be effectively analogous to it, which is why we did not consider it here.

Whilst the computational framework presented here could be extended to construct cell-type specific epigenetic clocks for other tissue and cell-types, it is worth highlighting some of the key limitations. First, it must be stressed that our strategy hinges critically on the ability to detect age-associated DNAm changes in individual cell-types and that this requires appropriately powered datasets [26, 51]. Indeed, the need for large sample sizes (typically at least many hundreds) has recently been emphasized by us [51]. The required sample size to ensure adequate power is a complex function of many parameters, including the number of cell-types (i.e. cellular resolution) at which the inference is to be made and the distribution (notably means and variances) of their proportions across samples. At higher cellular resolution, there is a corresponding need that the cohort be big enough so that all possible combinations of cell-type fractions are realized, which in practice may be difficult because of intrinsic correlations between cell-types. Thus, the need for fairly large DNAm datasets currently limits our strategy to a few tissue-types (e.g. brain/liver) and a low cellular resolution (e.g. neuron, glia and stroma). In order to be able to quantify biological age in additional subsets, say glia subtypes in brain, or for the different immune cell subtypes in liver, one would need on the order of thousands of samples, which in future could be realized by merging DNAm datasets together. For instance, for an age-related disease like NAFLD that is associated with molecular alterations in hepatocytes and which leads to a marked reduction in the hepatocyte fraction with age, other cell-types (e.g. Kupffer macrophages) may also undergo biological aging, marking the persistent inflammation associated with NAFLD [92], but here we were underpowered to detect age-DMCTs in macrophages. Ultimately, increased cellular resolution may be critical as for instance, adjustment for a broad cell-type that encompasses two cell subtypes cannot account for age-related relative variations in these subtypes. A related important question is the relevant cellular resolution or definition of cell-types [93] to use in such analyses, as this is often dependent on the specific biological questions to be addressed.

Another key unresolved issue is the relative quantification of cell-type specific versus shared/common age-associated DNAm changes. For instance, as shown by us, most age-associated DNAm changes in blood seem to be shared between distinct immune cell-types [94], implying that the number of truly cell-type specific age-associated DNAm changes may be small in a tissue like blood. Ideally, cell-type specific clocks should be built from those CpGs that only change in one cell-type of the tissue and not others, yet confidently identifying these is challenging and would require the aforementioned very large DNAm datasets. Certainly, an in-depth study in a tissue like blood, where large numbers of DNAm datasets are available would merit further investigation, although given that the relative fractions of naïve and mature T-cell subtypes change with age [24], this may require on the order of tens of thousands of samples. In the case of brain too, we have observed a substantial overlap between neuron and glia age-associated DMCs, the resulting lineage-specificity of the corresponding clocks also being less evident than in the case of liver and the hepatocyte specific clock. Indeed, in the case of liver it is worth stressing that we were able to confidently validate the hepatocyte specific age-DMCs in sorted hepatocytes, as well as the hepatocyte and cholangiocyte age-DMCs in 3 independent bulk liver DNAm datasets.

In summary, we have quantified the relative contributions of extrinsic and intrinsic aging to an epigenetic clock's predictive accuracy, whilst also presenting a computational strategy to dissect the intrinsic component, building cell-type specific epigenetic clocks that clearly demonstrate cell-type and tissue specificity in relation to prediction of chronological and biological age. The construction of such cell-type specific clocks will be critical to advance our understanding of biological aging at cell-type resolution and for evaluating in which cell-types cellular rejuvenation interventions work best.

METHODS

Brief description of main datasets analyzed

Human liver DNA methylation datasets

Johnson et al. [54]: This is an Illumina EPIC DNAm dataset of liver-tissue, comprising 341 samples, with 16 duplicates among them. The dataset includes samples from four stages representing the progression of non-alcoholic fatty liver disease (NAFLD): stage 0, stage 3, stage 3–4, and stage 4. Stage 0 is normal liver. Data is available at GEO (<https://www.ncbi.nlm.nih.gov/geo/>; accession number GSE180474). We downloaded a preprocessed data matrix following QC, including

P-values of detection for each probe. To select high-quality probes for clock construction, we utilized the `dropLociWithSnps()` function of `minfi` v.1.44.0 to remove SNP probes [95]. Sex chromosome CpGs were also excluded to minimize gender disparities. Any beta value with a *P*-value greater than 0.01 was considered as a missing value, and any probe with a coverage-fraction less than 0.99 was discarded. Missing values were then imputed using `impute.knn` ($k = 5$) of `impute` v.1.72.3 [96]. Type-2 bias was corrected using BMIQ algorithm [97], and finally duplicates were averaged, resulting in a final data matrix of 325 samples encompassing 812,168 CpGs on autosomes. Of these 325 samples, 210 were from normal healthy liver, whereas the remaining 115 represent NAFLD of grade-3 or higher. In this work, the 210 normal liver samples were used to build and validate the Liver and HepClocks in a 10-fold CV-strategy, whereas the 115 NAFLD samples were used to test for biological age-acceleration.

Park et al. [98]: Illumina HM850 (EPIC) dataset encompassing 56 primary hepatocyte cultures. Raw IDAT files were downloaded from GSE123995. Data was processed as before. Any probe with missing values was excluded, resulting in 809,289 probes. A sample was removed due to very low predicted hepatocyte fraction (as estimated using EpiSCORE).

Bonder et al. [56]: Illumina HM450k dataset profiling 67 obese liver samples. All individuals have a BMI greater than 35. Raw IDAT files were downloaded from GSE61446. `minfi::preprocessRaw()` was employed to process the raw IDAT data. The remaining steps in the preprocessing pipeline align with the aforementioned description, with a probe coverage threshold set at 0.90. A total of 411,233 probes remained.

Bacalini et al. [99]: Illumina HM450 dataset profiling 40 normal liver samples. Raw IDAT files were downloaded from GSE107038 and processed as the other datasets above. Any probe with missing values was excluded, resulting in 446,651 probes. A sample was removed due to very low predicted hepatocyte fraction (as estimated using EpiSCORE).

Horvath and Ahrens et al. [58, 59]: Illumina HM450 datasets comprising 77 samples from healthy non-obese ($n = 44$) and healthy obese ($n = 33$) individuals. Processed data was downloaded from GEO under accession numbers GSE61258 and GSE48325. For both datasets, we only used probes with no missing values (i.e., significant detection *P*-value < 0.01) across all samples. Datasets were then normalized with BMIQ. The merged dataset encompassed 77 samples and 417,213 probes. A subset of samples was obtained following bariatric surgery, but since surgery had no

effect on clock estimates [58, 59] we kept these samples in. Of note, the subset of 44 healthy non-obese samples was used to assess the predictive accuracy of the clocks with chronological age.

Wang et al. AATD dataset [57]: Illumina HM850 (EPIC) dataset encompassing 108 liver samples from 94 Alpha-1 Antitrypsin Deficiency (AATD) patients and 14 AATD patients with cirrhosis. Only the normal liver samples from the 94 AATD patients without cirrhosis were used. Processed normalized data was downloaded from GEO under accession number GSE119100. This normalized data is already adjusted for type-2 probe bias (adjusted with Subset-quantile Within Array Normalization (SWAN)), and probes that failed to achieve significant detection ($p < 0.01$) in all samples alongside cross-reactive probes were filtered out, leaving a total of 866,836 probes.

Human brain DNA methylation datasets used for training and model selection

Jaffe: The 450k dataset from Jaffe et al. [50], which profiled DNAm in human dorsolateral prefrontal cortex from controls and patients with schizophrenia (SZ), was obtained from the NCBI GEO website under the accession number GSE74193 (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE74193>). The file “GSE169156_RAW.tar” which contains the IDAT files was downloaded and processed with `minfi` package [95]. Probes with $>25\%$ NAs (defined by $P > 0.01$) were discarded. Samples younger than 18 years old were also discarded. The filtered data was subsequently normalized with BMIQ [97], resulting in a normalized data matrix for 473,536 probes and 416 samples, of which 226 are controls and 190 are SZ cases, age range 18–97 years.

Pihlstrøm: The EPIC dataset from Pihlstrøm et al. [52], which profiled DNAm in human frontal cortex, was obtained from the NCBI GEO website under the accession number GSE203332 (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE203332>). The file “GSE203332_RAW.tar” which contains the IDAT files was downloaded and processed with `minfi` package. Probes with NAs (defined by $P > 0.01$) were discarded. The filtered data was subsequently normalized with BMIQ, resulting in a normalized data matrix for 832,899 probes and 442 samples, of which 67 are controls with an age-range of 49 to 99 years.

Human brain DNA methylation datasets used for validation

Due to the large number of brain DNAm datasets analyzed, these are all summarized in Supplementary Methods.

Human DNA methylation datasets from other tissues

All the whole blood datasets used here have been preprocessed and analyzed previously by us [24]. In addition to whole blood, we also analyzed the following DNAm datasets from other tissue-types:

Colon [100]: Illumina HM450 dataset comprising 96 normal colon mucosa. Raw data is available from GSE131013. IDAT files were downloaded and processed with minfi package. Probes with NAs (defined by $P > 0.01$) were discarded. The filtered data was subsequently normalized with BMIQ, resulting in a normalized data matrix for 448,286 probes and 240 samples. 144 tumor samples were discarded.

Prostate [101]: Illumina HM450 dataset comprising 51 normal prostate samples. Raw data is available from GSE76938. IDAT files were downloaded and processed with minfi package. Probes with NAs (defined by $P > 0.01$) were discarded. The filtered data was subsequently normalized with BMIQ, resulting in a normalized data matrix for 471,112 probes and 123 samples and 72 prostate cancer samples were discarded.

Kidney [102]: Illumina HM450 dataset profiling 31 normal kidney tissue samples. Raw data is available from GSE79100. There are three technical replicates per tissue. IDAT files were downloaded and processed with minfi package. Probes with NAs (defined by $P > 0.01$) were discarded. The filtered data was subsequently normalized with BMIQ, resulting in a normalized data matrix for 454,671 probes and 93 samples.

Skin [103]: Illumina HM450 dataset profiling 322 normal skin samples from a peri-umbilical punch biopsy. From GSE90124, we downloaded the processed type-2 probe bias adjusted DNAm data with 450,531 probes and 322 samples.

Estimating cell-type fractions in whole blood, liver and brain

For a given bulk-tissue sample with a genome-wide DNAm profile, it is possible to estimate the cell-type composition of the tissue with a corresponding tissue-specific DNAm reference matrix [104]. Such a DNAm reference matrix consists of cell-type specific CpGs and cell-types, and is typically built from genome-wide DNAm data of sorted cell-types [105]. The cell-type fractions themselves are estimated by regressing the DNAm profile of the bulk-tissue samples against the reference profiles of the DNAm reference matrix, a procedure that only uses the highly discriminatory cell-type specific CpGs as defined in the DNAm reference matrix [106]. For whole blood, we estimated fractions for 12 immune cell subtypes (naïve and memory B-

cells, naïve and memory CD4+ T-cells, naïve and memory CD8+ T-cells, T-regulatory cells, natural killer cells, monocytes, neutrophils, eosinophils, basophils) using EpiDISH and our validated DNAm reference matrix for these 12 cell-types [24]. In the case of liver tissue, we used our validated EpiSCORE algorithm and the associated liver DNAm reference matrix defined for hepatocytes, cholangiocytes, endothelial cells, Kupffer macrophages and lymphocytes [53]. In the case of brain-tissue, we applied the HiBED algorithm [46], since the EpiSCORE brain DNAm reference matrix led to the microglia fraction always being zero. We applied HiBED to estimate proportions for 7 brain cell-types including GABAergic excitatory neurons, glutamatergic inhibitory neurons, astrocytes, microglia, oligodendrocytes, endothelial cells and stromal cells. When applying CellDMC we collapsed these estimates to 3 broad cell-types: neurons, glia and endothelial/stromal.

Relative quantification of extrinsic and intrinsic aging in whole blood and brain

Whole blood: As training data we used two of the largest whole blood DNAm datasets, the Airway [107] and Lehne [108] cohorts, which we have previously normalized and analyzed [24]. Briefly, the Airway EPIC DNAm dataset (GSE147740) included 1032 samples with age information. The Lehne dataset is a 450k DNAm dataset (GSE55763) encompassing 2707 samples in total. The two datasets were merged, resulting in 314,436 CpGs and 3739 samples. In the Airway dataset, age is cofounded by gender (Pearson correlation test, P -value = 0.003896). To train the unadjusted and adjusted clocks, we applied the Elastic Net regression to the training dataset with $\alpha = 0.5$ and a 10-fold cross-validation to infer the optimal penalty parameter λ . In the unadjusted case, we applied the Elastic Net to the standardized DNAm residuals after adjusting for sex and cohort only. In the adjusted case, we applied the Elastic Net to the standardized DNAm residuals after adjusting for sex, cohort and the 12 estimated immune-cell type fractions. We also trained another clock adjusted for 9 immune cell-type fractions by merging the naïve and memory lymphocyte subsets together, in other words, by adding the fractions of naïve and memory B-cells to yield a fraction for B-cells, and similarly for CD4T and CD8T-cells. The 3 optimal clocks were then applied to 11 independent whole blood datasets (using only control/healthy samples, see Supplementary Table 1), all previously normalized as in Luo et al. [24]. In the case of the adjusted clocks, the clocks were applied to the standardized DNAm residuals after adjustment for the 12 or 9 immune-cell fractions. Finally, the predicted ages were linearly regressed to the true chronological ages in each cohort, resulting in R^2 values for the

unadjusted and the two adjusted clocks. The ratio of the adjusted clock's R^2 value to that of the unadjusted clock, R_{ADJ}^2 / R_{UNADJ}^2 quantifies the fraction of the unadjusted clock's accuracy that can be attributed to intrinsic aging. For the extrinsic component, its fractional contribution is given by $1 - R_{ADJ}^2 / R_{UNADJ}^2$.

Brain: For training, we used the Jaffe et al. dataset [50] encompassing 491 samples, including 300 controls and 191 samples from schizophrenia patients, with an age range 0.003 ~ 96.98 years. We used HiBED to estimate the fractions for 7-cell-types (ExN, InN, Oligo/OPC, Micro, Astro, Endo, Strom), as described earlier. To train the Elastic Net clocks with $\alpha = 0.5$, we used a 10-fold cross-validation to infer the optimal penalty parameter λ . In the unadjusted case, we applied the Elastic Net to the standardized DNAm residuals obtained after adjusting for sex and disease-status, whilst in the adjusted case, it was applied on the standardized DNAm residuals obtained after adjusting for sex, disease-status and 7 brain cell-type fractions. The two clocks were then applied to 11 independent brain DNAm datasets (using control/healthy samples only, see Supplementary Table 1), obtaining in each cohort corresponding R_{ADJ}^2 & R_{UNADJ}^2 values. The fraction of the unadjusted clock's accuracy attributed to intrinsic and extrinsic aging was then estimated as described above for whole blood.

Construction of the neuron/glia/brain-specific epigenetic clocks

We used HiBED [46] to estimate fractions for 7 brain cell-types in the Jaffe and Philstrom datasets, in each case summing up fractions to yield estimates for 3 broad cell-types: neurons (Neu), glia (Glia) and endothelial/stromal. We then applied CellDMC [45], as implemented in the EpiDISH package [109], to the Jaffe dataset in order to infer age-associated DMCTs in each of these 3 broad cell-types. Briefly, the CellDMC model postulates the following relationship for the DNAm level β_{cs} of a CpG c in sample s :

$$\beta_{cs} = \sum_{t=1}^3 \hat{f}_{ts} \beta_{cts} = \sum_{t=1}^3 \hat{f}_{ts} (\mu_{ct} + Age_s \gamma_{ct}) + \varepsilon_{cs}$$

where \hat{f}_{ts} are the estimated cell-type fractions for cell-type t in sample s . This can be re-expressed as a multivariate regression model with interaction terms between age and cell-type fractions:

$$\beta_{cs} = \sum_{t=1}^3 \mu_{ct} \hat{f}_{ts} + \sum_{t=1}^3 \gamma_{ct} \hat{f}_{ts} Age_s + \varepsilon_{cs}$$

This estimates the regression coefficients μ_{ct} and γ_{ct} so as to minimize the error in a least squares sense. When applying CellDMC to the Jaffe dataset, we also adjusted for schizophrenia (SZ) status and gender, i.e., we ran the model:

$$\beta_{cs} = \sum_{t=1}^3 \mu_{ct} \hat{f}_{ts} + \sum_{t=1}^3 \gamma_{ct} \hat{f}_{ts} Age_s + \theta_c Sex_s + \varphi_c SZ_s + \varepsilon_{cs}$$

We included both SZ cases and controls, adjusting for case/control status, so as to increase power. Having identified the significant neuron-DMCTs (those CpGs c with a significant γ_{ct} with $t = neuron$, t -test FDR < 0.05), we next trained Elastic Net regression (glmnet R package) [49] models ($\alpha = 0.5$) for age, each parameterized by a different penalty parameter (λ) value. This training was done on the residuals obtained by regressing out the cell-type fractions from the DNAm beta matrix. To select the optimal model, we then applied the age-predictors for each choice of λ to the Philstrom DNAm dataset (using the residuals after adjusting this DNAm data matrix for cell-type fractions), selecting the λ that minimized the root-mean-square error (RMSE). This defines a neuron-specific epigenetic clock that is applicable to residuals after adjusting for cell-type fraction, a clock we call "Neu-In" (neuron-intrinsic).

We also constructed a similar clock called "Neu-Sin" (neuron semi-intrinsic), following the same procedure as before, but now not adjusting the DNAm data matrix for cell-type fractions. In other words, whilst the Neu-In and Neu-Sin clocks are both built from CpGs that change with age in neurons, the Neu-In clock uses the residuals for construction, whilst the Neu-Sin clock does not. A third non-cell-type specific clock (Brain) was also built using the same procedure as before, but not using CellDMC to identify CpGs changing with age, instead identifying age-DMCs by running a linear model of DNAm versus age adjusting for cell-type fractions, i.e. in this case the regression model is:

$$\beta_{cs} = \sum_{t=1}^3 \mu_{ct} \hat{f}_{ts} + \gamma_c Age_s + \theta_c Sex_s + \varphi_c SZ_s + \varepsilon_{cs}$$

and one finds age-DMCs, as those CpGs c with a significant γ_c (t -test, FDR < 0.05). The age-predictor itself was then also trained on residuals obtained after regressing out the effect of cell-type fractions. Of note, the model selection step to define the Neu-Sin and Brain clocks was also done using the independent Philstrom DNAm dataset.

Finally, an analogous procedure describing the construction of the Neu-In and Neu-Sin clocks was followed to build corresponding glia-specific intrinsic and semi-intrinsic clocks (Glia-In and Glia-Sin).

Validation of the neuron/glia/brain specific clocks

To validate and demonstrate specificity of the Neu-In/Sin, Glia-In/Sin and Brain clocks, we applied them to sorted neuron, sorted glia, brain and non-brain DNAm cohorts. When applying the Neu-In, Glia-In and Brain clocks, we adjusted the DNAm datasets for underlying cell-type fractions, which were estimated using HiBED in the case of brain tissue, EpiDISH in whole blood cohorts and EpiSCORE in other non-brain tissues.

Construction of HepClock and LiverClock

We first describe the construction of the hepatocyte-specific clock (HepClock). We selected 210 normal liver samples from Johnson et al.'s dataset for training and model selection. EpiSCORE was used to estimate fractions for 5 liver cell-types including hepatocytes, as described above. The 210 samples were then randomly divided into 10 bags (each bag with 21 samples), with stratified sampling within each age group to ensure a similar age distribution in each bag. To identify CpGs changing with age in hepatocytes, we then applied the CellDMC algorithm [45] to a subset composed of 9 bags, resulting in a set of hepatocyte age-DMCTs. The specific CellDMC model run was:

$$\beta_{cs} = \sum_{t=1}^5 \mu_{ct} \hat{f}_{ts} + \sum_{t=1}^5 \gamma_{ct} \hat{f}_{ts} Age_s + \varepsilon_{cs}$$

Having identified the significant hepatocyte-DMCTs (those CpGs c with a significant γ_{ct} with $t = hepatocyte$, t -test FDR < 0.05), and using the same 9 bags, we then trained a range of lasso regression models parameterized by a penalty parameter lambda. Each of these models (parameterized by lambda) were subsequently applied to the left-out-bag to yield age-predictions for the samples in this left-out-bag. This procedure was repeated 10-times, each time using a different subset of 9 bags for training and one left-out-bag for model selection. The age-predictions over all 10 folds were then combined and compared to the true ages computing the RMSE for each choice of lambda. The model minimizing the RMSE was then selected as the optimal model. Of note, the above procedure applies CellDMC a total of 10 times, once for each fold. The number of hepatocyte age-DMCTs is variable between folds, so we identified the minimum number (k) obtained across all folds, and subsequently only used the top- k hepatocyte age-DMCTs from each fold before

running the lasso regression model. Lasso was implemented using the glmnet function from the glmnet R-package (version 4.7.1) [49], with alpha set to 1. Having identified the optimal model (lambda parameter), we finally reran the lasso model with this optimal parameter on the full set of 210 samples using the top- k hepatocyte age-DMCTs as inferred using CellDMC on all samples.

The LiverClock was built in an analogous fashion with only one key difference: instead of applying CellDMC in each fold, we ran a linear regression of DNAm against age with all the liver cell-type fractions as covariates. That is, the model in this case is:

$$\beta_{cs} = \sum_{t=1}^5 \mu_{ct} \hat{f}_{ts} + \gamma_{ct} Age_s + \varepsilon_{cs}$$

and one finds age-DMCTs, as those CpGs c with a significant γ_c (t -test, FDR < 0.05). After optimization of lambda, the final age lasso predictor was built from such age-DMCTs derived using all 210 samples.

Simulation analysis on mixtures of sorted cell-types

We used a simulation framework to assess and compare the intrinsic and semi-intrinsic paradigms for cell-type specific clock construction. We first generated basis DNAm data matrices for 3 hypothetical cell-types "A", "B" and "C" each one defined over 10,000 CpGs and 200 samples, drawing DNAm beta-values from Beta (10,40), i.e. from a Beta-distribution with a mean value of 0.25. We then declared the first 1000 CpGs to be cell-type specific for a cell-type "A", by drawing the DNAm-values for two other cell-types ("B" and "C") from Beta (40,10), i.e. from a Beta-distribution with mean 0.75. Thus, the first 1000 CpGs are unmethylated in cell-type "A" compared to "B" and "C". Likewise, another 1000 non-overlapping CpGs were defined to be unmethylated in cell-type "B" and another 1000 non-overlapping CpGs for cell-type "C". The remaining 7000 CpGs are unmethylated in all cell-types. To generate realistic mixtures of these 3 cell-types, we randomly sampled cell-type fractions from true estimated fractions in whole blood by applying our EpiDISH algorithm [109] to estimate 7 immune cell-type fractions in a large dataset of 656 whole blood samples [110]. We declared 3 cell-types to be granulocytes, monocytes and lymphocytes, by adding together the fractions of B-cells, CD4T-cells, CD8 T-cells and NK-cells to represent "lymphocytes" and adding together the fractions of neutrophils and eosinophils to represent "granulocytes". To define the ages of the 200 samples we bootstrapped 200 values from the ages of 139 donors from the BLUEPRINT consortium [111] in

order to mimic an age distribution of a real human cohort. Before mixing together the artificially generated DNAm data matrix for the 3 hypothetical cell-types, we introduced age-DMCTs in the cell-type defining lymphocytes. This was done by drawing DNAm-values for these age-DMCTs to increase with the age of the sample. In more detail, because the minimum age of the BLUERPINT samples was 30, we first assigned M-values according to the formula:

$$Mval = Mval0 + (age - 30) \times effS + \sigma$$

where $effS = 0.1$ and $Mval0 = \log_2(0.1/(1-0.1))$ and σ representing a Gaussian deviate with a standard deviation value equal to $effS$. These values were chosen to mimic age-related DNAm changes as seen in real data, although with a smaller noise component in order to better understand and interpret downstream results.

In total we considered 4 different scenarios depending on the choice of age-DMCTs in lymphocytes and whether the fraction of “lymphocytes” increases with age or not. In scenarios (i) and (ii), 1000 age-DMCTs were chosen to not overlap with any of the 3000 cell-type specific marker CpGs. In scenarios (iii) and (iv), 900 age-DMCTs in lymphocytes were chosen to be lymphocyte-specific markers. Reason we did not pick all 1000 lymphocyte markers, is to leave 100 markers that remain cell-type specific to allow for reference construction and estimation of cell-type fractions using EpiDISH. In scenarios (i) and (iii), cell-type fractions were chosen so that no fraction changed significantly with age, whilst in scenarios (ii) and (iv) the lymphocyte-fraction was chosen to increase with age.

For each of the 4 scenarios, we generated training and validation mixture sets. DNAm reference matrices were built from different realizations of the basis DNAm data matrices but using the same parameter values. These reference matrices were then applied to the training dataset to infer cell-type fractions. With the estimated cell-type fractions we then applied CellDMC [45] to the training set to infer age-DMCTs. Since age-DMCTs are only introduced in lymphocytes, we focus on the predicted age-DMCTs in this cell-type, applying an Elastic Net ($\alpha = 0.5$) penalized regression framework in a 10-fold cross-validation framework to learn a predictor of chronological age using either an intrinsic or semi-intrinsic paradigm: in the semi-intrinsic paradigm the Elastic Net predictor was trained on the original DNAm-values, whilst in the intrinsic case it was applied to residuals obtained by regressing out the effect of the cell-type fractions. The optimal penalty parameter was selected via the 10-fold CV procedure and the optimal predictor then applied to the validation dataset. Of note, in the semi-intrinsic

setting the validation set is defined on the original DNAm-values, whereas the intrinsic model is applied to the residuals of the validation set. For each scenario we ran 50 different simulations, recording the accuracy (root mean square error-RMSE and Pearson Correlation Coefficient-PCC) of the semi-intrinsic and intrinsic lymphocyte-specific clock predictors in the corresponding validation set.

Meta-analyses

In the case of brain tissue datasets, we performed two separate meta-analyses. One of these explored if brain cell-type fractions change with age, whilst the other was done to explore if our cell type specific clocks (extrinsic age-acceleration-EAA) display correlations with Alzheimer’s Disease (AD) status. The former analysis was done for 7 brain cell types over 13 independent DNAm datasets encompassing frontal and temporal lobe regions, in each dataset adjusting for disease status and sex. The latter meta-analysis was done for frontal lobe ($n = 5$) and temporal lobe ($n = 6$) separately, and then again treating both regions together ($n = 11$). In each cohort, associations between EAA and AD status was assessed using a multivariate linear regression with sex and age as covariates, AD status encoded as ordinal with 0 = control, 1 = AD case. We extracted the corresponding effect size, standard error, Student’s t -test, and P -value for each regression. Meta-analysis was then performed using both fixed effect and random effect inverse variance method using the metagen function implemented in the meta R-package [112, 113].

Co-localization of age-DMCTs with DNAm reference matrix CpGs

The liver DNAm reference matrix used to estimate cell type proportions before training the clocks contains 171 marker gene promoters. We calculated the number of hepatocyte and cholangiocyte age-DMCTs that mapped to the promoter regions of these 171 marker genes (where the region is defined as within 1kb upstream or downstream of the transcription start site). To generate a null distribution, we did the following: from the DNAm dataset that we used to infer the age-DMCTs, we randomly selected a matched number of age-DMCTs and checked how many mapped to the promoter regions of these 171 marker genes. This process was repeated 1,000 times to obtain the null distribution, yielding an empirical P -value. In the case of the HiBED DNAm reference matrix, we focused on the CpGs in the first layer of deconvolution, and computed overlaps with neuron and glia age-DMCTs using a 1kb window. A null distribution was estimated in the same way as for liver.

Enrichment analyses

We assessed enrichment of clock-CpGs against EWAS-DMCs for AD and BMI, as determined by the EWAS catalog [79] and EWAS-atlas [80]. Enrichment analysis was done using a one-tailed Fisher exact test. We also assessed enrichment of our clock-CpGs for (i) blood mQTLs as determined by the ARIES mQTL database ($n = 123010$ is the number of distinct CpGs in mQTLs present in the training cohorts of the clocks) [81], focusing on the mQTLs identified in the adult population, (ii) mQTLs ($n(\text{breast}) = 13256$, $n(\text{colon}) = 156419$, $n(\text{kidney}) = 18578$, $n(\text{lung}) = 166242$, $n(\text{skeleton muscle}) = 14437$, $n(\text{ovary}) = 134215$, $n(\text{prostate}) = 68724$, $n(\text{testis}) = 14894$, $n(\text{blood}) = 21222$) identified by the eGTEx consortium [82] and (iii) mQTLs identified from adult [84] ($n = 71965$) and fetal [83] ($n = 30707$) brain DNAm datasets. Numbers of mQTLs mapping to the datasets used to train the clocks were computed. Since the GTEx project identified mQTLs using the Illumina EPIC array, while our clocks were trained using 450k probes, we only used mQTLs that map to 450k probes. Significance was evaluated using a one-tailed Binomial test. Overlap with GWAS AD SNPs from Bellenguez et al. [85] and GWAS BMI SNPs from Pulit et al. [114] was done by restricting to SNPs that were part of mQTLs that included clock-CpGs.

Code availability

All the cell-type specific clocks presented here are freely available as part of the R-package CTSclocks, downloadable from <https://github.com/HGT-UwU/CTSclocks>.

Data availability

The Illumina DNA methylation datasets analyzed here are all freely available from GEO (<https://www.ncbi.nlm.nih.gov/geo>). Details: Liver-NAFLD ($n = 325$, GEO: GSE180474); Hepatocyte cultures ($n = 56$, GEO: GSE123995); Liver-obese ($n = 67$, GEO: GSE61446); Liver-Horvath/Ahrens ($n = 77$, GEO: GSE61258 and GSE48325); Liver tissue (40 normal, GEO: GSE107038); Liver-AATD ($n = 94$, GEO: GSE119100); MESA (214 purified CD4+ T-cells and 1202 Monocyte samples, GEO: GSE56046 and GSE56581); Colon (96 normal and 144 adenocarcinoma, GEO: GSE131013); Prostate (123 normal and 63 benign, GEO: GSE76938); Kidney (93 normal, GEO: GSE79100); Skin (322 normal, GEO: GSE90124); Brain sorted-Pai (33 normal and 67 BD, SCZ, GEO: GSE112179); Brain sorted-Gasparoni (32 normal and 30 AD, GEO: GSE66351); Brain sorted-Kozlenkov (29 normal and 59 Heroin or suicide, GEO:

GSE98203); Brain sorted-Guintivano (58 normal, GEO: GSE41826); Brain sorted-Hannon (103 normal, GEO: GSE234520); Brain sorted-Witte (20 normal and 36 BD, SCZ, MDD, GEO: GSE191200); Brain-Jaffe (226 normal and 190 SCZ, GEO: GSE74193); Brain-Philstrom (67 normal and 375 Various mental disease and neurodegenerative disease, GEO: GSE203332); Brain-Semick (187 normal and 82 AD, GEO: GSE125895); Brain-Pidsley (71 normal and 179 AD, GEO: GSE43414); Brain-Smith (138 normal and 148 AD, GEO: GSE80970); Brain-Horvath (95 normal and 205 AD, HD, GEO: GSE72778); Brain-Gasparoni (52 normal and 76 AD, GEO: GSE66351); Brain-Watson (34 normal and 34 AD, GEO: GSE76105); Brain-Brokaw (358 normal and 450 AD, GEO: GSE134379); Brain-Murphy (38 normal and 37 MDD, GEO: GSE88890); Brain-Rydbirk (37 normal and 41 MSA, GEO: GSE143157); Brain-Torabi (50 normal and 75 SCZ, GEO: GSE128601); Brain-Xu (25 normal and 23 AUD, GEO: GSE49393); Brain-Huynh (19 normal and 27 Multiple sclerosis, GEO: GSE40360); Brain-Viana2 (13 normal and 14 SCZ, GEO: GSE89703); Brain-Viana3 (17 normal and 16 SCZ, GEO: GSE89705); Brain-Viana4 (28 normal and 21 SCZ, GEO: GSE89706); Brain-Markunas (221 normal, GEO: GSE147040); Brain-Viana1 (17 normal and 16 SCZ, GEO: GSE89702); Brain-Stefania1 (37 normal and 21 Suicide, GEO: GSE137222); Brain-Stefania2 (15 normal and 18 Suicide, GEO: GSE137223); Whole blood-Airway (1032 normal, GEO: GSE147740); Whole blood-Barturen (101 normal and 473 SARS, GEO: GSE179325); Whole blood-Flanagan (184 normal, GEO: GSE61151); Whole blood-Hannon1 (304 normal and 332 SCZ, GEO: GSE80417); Whole blood-Hannon2 (405 normal and 260 SCZ, GEO: GSE84727); Whole blood-Hannum (656 normal, GEO: GSE40279); Whole blood-Johansson (729 normal, GEO: GSE87571); Whole blood-Lehne (2707 normal, GEO: GSE55763); Whole blood-LiuMS (139 normal and 140 Multiple Sclerosis, GEO: GSE106648); Whole blood-LiuRA (335 normal and 354 Rheumatoid arthritis, GEO: GSE42861); Whole blood-TZH (85 normal and 620 Different health conditions, GEO: Requested Access); Whole blood-Ventham (101 normal and 279 Inflammatory bowel disease, GEO: GSE87648); Whole blood-Zannas (144 normal and 278 Trauma, GEO: GSE72680). The DNAm-atlas encompassing DNAm reference matrices for 13 tissue-types encompassing over 40 cell-types is freely available from the EpiSCORE R-package <https://github.com/aet21/EpiSCORE>.

AUTHOR CONTRIBUTIONS

TH, GX and AET performed the statistical and bioinformatic analyses. AET conceived the study, and wrote the manuscript with contributions from TH and

GX. QL contributed to the data analyses. MJ and NE contributed pointers to data and helped revise the manuscript.

ACKNOWLEDGMENTS

We thank everyone who supports open-access data.

CONFLICTS OF INTEREST

The authors declare no conflicts of interest related to this study.

FUNDING

This work was supported by NSFC (National Science Foundation of China) grants, Grant Numbers: 31970632, 32170652 and 32370699.

REFERENCES

1. Baker GT 3rd, Sprott RL. Biomarkers of aging. *Exp Gerontol.* 1988; 23:223–39.
[https://doi.org/10.1016/0531-5565\(88\)90025-3](https://doi.org/10.1016/0531-5565(88)90025-3)
PMID:[3058488](https://pubmed.ncbi.nlm.nih.gov/3058488/)
2. Hadley EC, Lakatta EG, Morrison-Bogorad M, Warner HR, Hodes RJ. The future of aging therapies. *Cell.* 2005; 120:557–67.
<https://doi.org/10.1016/j.cell.2005.01.030>
PMID:[15734687](https://pubmed.ncbi.nlm.nih.gov/15734687/)
3. Levine ME, Lu AT, Quach A, Chen BH, Assimes TL, Bandinelli S, Hou L, Baccarelli AA, Stewart JD, Li Y, Whitsel EA, Wilson JG, Reiner AP, et al. An epigenetic biomarker of aging for lifespan and healthspan. *Aging (Albany NY).* 2018; 10:573–91.
<https://doi.org/10.18632/aging.101414>
PMID:[29676998](https://pubmed.ncbi.nlm.nih.gov/29676998/)
4. Moqri M, Herzog C, Poganik JR, Justice J, Belsky DW, Higgins-Chen A, Moskalev A, Fuellen G, Cohen AA, Bautmans I, Widschwendter M, Ding J, Fleming A, et al, and Biomarkers of Aging Consortium. Biomarkers of aging for the identification and evaluation of longevity interventions. *Cell.* 2023; 186:3758–75.
<https://doi.org/10.1016/j.cell.2023.08.003>
PMID:[37657418](https://pubmed.ncbi.nlm.nih.gov/37657418/)
5. Beck S. Taking the measure of the methylome. *Nat Biotechnol.* 2010; 28:1026–8.
<https://doi.org/10.1038/nbt1010-1026>
PMID:[20944589](https://pubmed.ncbi.nlm.nih.gov/20944589/)
6. Bell CG. Epigenomic insights into common human disease pathology. *Cell Mol Life Sci.* 2024; 81:178.
<https://doi.org/10.1007/s00018-024-05206-2>
PMID:[38602535](https://pubmed.ncbi.nlm.nih.gov/38602535/)
7. Seale K, Horvath S, Teschendorff A, Eynon N, Voisin S. Making sense of the ageing methylome. *Nat Rev Genet.* 2022; 23:585–605.
<https://doi.org/10.1038/s41576-022-00477-6>
PMID:[35501397](https://pubmed.ncbi.nlm.nih.gov/35501397/)
8. Horvath S, Raj K. DNA methylation-based biomarkers and the epigenetic clock theory of ageing. *Nat Rev Genet.* 2018; 19:371–84.
<https://doi.org/10.1038/s41576-018-0004-3>
PMID:[29643443](https://pubmed.ncbi.nlm.nih.gov/29643443/)
9. Field AE, Robertson NA, Wang T, Havas A, Ideker T, Adams PD. DNA Methylation Clocks in Aging: Categories, Causes, and Consequences. *Mol Cell.* 2018; 71:882–95.
<https://doi.org/10.1016/j.molcel.2018.08.008>
PMID:[30241605](https://pubmed.ncbi.nlm.nih.gov/30241605/)
10. Bell CG, Lowe R, Adams PD, Baccarelli AA, Beck S, Bell JT, Christensen BC, Gladyshev VN, Heijmans BT, Horvath S, Ideker T, Issa JJ, Kelsey KT, et al. DNA methylation aging clocks: challenges and recommendations. *Genome Biol.* 2019; 20:249.
<https://doi.org/10.1186/s13059-019-1824-y>
PMID:[31767039](https://pubmed.ncbi.nlm.nih.gov/31767039/)
11. Zhang Q, Vallerga CL, Walker RM, Lin T, Henders AK, Montgomery GW, He J, Fan D, Fowdar J, Kennedy M, Pitcher T, Pearson J, Halliday G, et al. Improved precision of epigenetic clock estimates across tissues and its implication for biological ageing. *Genome Med.* 2019; 11:54.
<https://doi.org/10.1186/s13073-019-0667-1>
PMID:[31443728](https://pubmed.ncbi.nlm.nih.gov/31443728/)
12. Lu AT, Quach A, Wilson JG, Reiner AP, Aviv A, Raj K, Hou L, Baccarelli AA, Li Y, Stewart JD, Whitsel EA, Assimes TL, Ferrucci L, Horvath S. DNA methylation GrimAge strongly predicts lifespan and healthspan. *Aging (Albany NY).* 2019; 11:303–27.
<https://doi.org/10.18632/aging.101684>
PMID:[30669119](https://pubmed.ncbi.nlm.nih.gov/30669119/)
13. Levine ME, Hosgood HD, Chen B, Absher D, Assimes T, Horvath S. DNA methylation age of blood predicts future onset of lung cancer in the women's health initiative. *Aging (Albany NY).* 2015; 7:690–700.
<https://doi.org/10.18632/aging.100809>
PMID:[26411804](https://pubmed.ncbi.nlm.nih.gov/26411804/)
14. Levine ME, Lu AT, Chen BH, Hernandez DG, Singleton AB, Ferrucci L, Bandinelli S, Salfati E, Manson JE, Quach A, Kusters CD, Kuh D, Wong A, et al. Menopause accelerates biological aging. *Proc Natl Acad Sci U S A.* 2016; 113:9327–32.
<https://doi.org/10.1073/pnas.1604558113>
PMID:[27457926](https://pubmed.ncbi.nlm.nih.gov/27457926/)

15. Marioni RE, Shah S, McRae AF, Chen BH, Colicino E, Harris SE, Gibson J, Henders AK, Redmond P, Cox SR, Pattie A, Corley J, Murphy L, et al. DNA methylation age of blood predicts all-cause mortality in later life. *Genome Biol.* 2015; 16:25.
<https://doi.org/10.1186/s13059-015-0584-6>
PMID:[25633388](https://pubmed.ncbi.nlm.nih.gov/25633388/)
16. Marioni RE, Shah S, McRae AF, Ritchie SJ, Muniz-Terrera G, Harris SE, Gibson J, Redmond P, Cox SR, Pattie A, Corley J, Taylor A, Murphy L, et al. The epigenetic clock is correlated with physical and cognitive fitness in the Lothian Birth Cohort 1936. *Int J Epidemiol.* 2015; 44:1388–96.
<https://doi.org/10.1093/ije/dyu277>
PMID:[25617346](https://pubmed.ncbi.nlm.nih.gov/25617346/)
17. de Magalhães JP. Distinguishing between driver and passenger mechanisms of aging. *Nat Genet.* 2024; 56:204–11.
<https://doi.org/10.1038/s41588-023-01627-0>
PMID:[38242993](https://pubmed.ncbi.nlm.nih.gov/38242993/)
18. Teschendorff AE, Relton CL. Statistical and integrative system-level analysis of DNA methylation data. *Nat Rev Genet.* 2018; 19:129–47.
<https://doi.org/10.1038/nrg.2017.86>
PMID:[29129922](https://pubmed.ncbi.nlm.nih.gov/29129922/)
19. Jaffe AE, Irizarry RA. Accounting for cellular heterogeneity is critical in epigenome-wide association studies. *Genome Biol.* 2014; 15:R31.
<https://doi.org/10.1186/gb-2014-15-2-r31>
PMID:[24495553](https://pubmed.ncbi.nlm.nih.gov/24495553/)
20. Zhang Z, Reynolds SR, Stolrow HG, Chen JQ, Christensen BC, Salas LA. Deciphering the role of immune cell composition in epigenetic age acceleration: Insights from cell-type deconvolution applied to human blood epigenetic clocks. *Aging Cell.* 2024; 23:e14071.
<https://doi.org/10.1111/accel.14071>
PMID:[38146185](https://pubmed.ncbi.nlm.nih.gov/38146185/)
21. Marttila S, Rajić S, Ciantar J, Mak JKL, Junttila IS, Kummola L, Hägg S, Raitoharju E, Kananen L. Biological aging of different blood cell types. *Geroscience.* 2024. [Epub ahead of print].
<https://doi.org/10.1007/s11357-024-01287-w>
PMID:[39060678](https://pubmed.ncbi.nlm.nih.gov/39060678/)
22. Tomusiak A, Floro A, Tiwari R, Riley R, Matsui H, Andrews N, Kasler HG, Verdin E. Development of an epigenetic clock resistant to changes in immune cell composition. *Commun Biol.* 2024; 7:934.
<https://doi.org/10.1038/s42003-024-06609-4>
PMID:[39095531](https://pubmed.ncbi.nlm.nih.gov/39095531/)
23. Jonkman TH, Dekkers KF, Slieker RC, Grant CD, Ikram MA, van Greevenbroek MMJ, Franke L, Veldink JH, Boomsma DI, Slagboom PE, Heijmans BT, and BIOS Consortium. Functional genomics analysis identifies T and NK cell activation as a driver of epigenetic clock progression. *Genome Biol.* 2022; 23:24.
<https://doi.org/10.1186/s13059-021-02585-8>
PMID:[35031073](https://pubmed.ncbi.nlm.nih.gov/35031073/)
24. Luo Q, Dwaraka VB, Chen Q, Tong H, Zhu T, Seale K, Raffaele JM, Zheng SC, Mendez TL, Chen Y, Carreras N, Begum S, Mendez K, et al. A meta-analysis of immune-cell fractions at high resolution reveals novel associations with common phenotypes and health outcomes. *Genome Med.* 2023; 15:59.
<https://doi.org/10.1186/s13073-023-01211-5>
PMID:[37525279](https://pubmed.ncbi.nlm.nih.gov/37525279/)
25. Maity AK, Teschendorff AE. Cell-attribute aware community detection improves differential abundance testing from single-cell RNA-Seq data. *Nat Commun.* 2023; 14:3244.
<https://doi.org/10.1038/s41467-023-39017-z>
PMID:[37277399](https://pubmed.ncbi.nlm.nih.gov/37277399/)
26. Zheng SC, Webster AP, Dong D, Feber A, Graham DG, Sullivan R, Jevons S, Lovat LB, Beck S, Widschwendter M, Teschendorff AE. A novel cell-type deconvolution algorithm reveals substantial contamination by immune cells in saliva, buccal and cervix. *Epigenomics.* 2018; 10:925–40.
<https://doi.org/10.2217/epi-2018-0037>
PMID:[29693419](https://pubmed.ncbi.nlm.nih.gov/29693419/)
27. Mejia-Ramirez E, Florian MC. Understanding intrinsic hematopoietic stem cell aging. *Haematologica.* 2020; 105:22–37.
<https://doi.org/10.3324/haematol.2018.211342>
PMID:[31806687](https://pubmed.ncbi.nlm.nih.gov/31806687/)
28. Horvath S. DNA methylation age of human tissues and cell types. *Genome Biol.* 2013; 14:R115.
<https://doi.org/10.1186/gb-2013-14-10-r115>
PMID:[24138928](https://pubmed.ncbi.nlm.nih.gov/24138928/)
29. Levine ME, Lu AT, Bennett DA, Horvath S. Epigenetic age of the pre-frontal cortex is associated with neuritic plaques, amyloid load, and Alzheimer's disease related cognitive functioning. *Aging (Albany NY).* 2015; 7:1198–211.
<https://doi.org/10.18632/aging.100864>
PMID:[26684672](https://pubmed.ncbi.nlm.nih.gov/26684672/)
30. Levine AJ, Quach A, Moore DJ, Achim CL, Soontornniyomkij V, Masliah E, Singer EJ, Gelman B, Nemanic N, Horvath S. Accelerated epigenetic aging in brain is associated with pre-mortem HIV-associated neurocognitive disorders. *J Neurovirol.* 2016; 22:366–75.
<https://doi.org/10.1007/s13365-015-0406-3>
PMID:[26689571](https://pubmed.ncbi.nlm.nih.gov/26689571/)

31. Levine AJ, Panos SE, Horvath S. Genetic, transcriptomic, and epigenetic studies of HIV-associated neurocognitive disorder. *J Acquir Immune Defic Syndr*. 2014; 65:481–503.
<https://doi.org/10.1097/QAI.000000000000069>
PMID:[24583618](https://pubmed.ncbi.nlm.nih.gov/24583618/)
32. Lu AT, Hannon E, Levine ME, Hao K, Crimmins EM, Lunnon K, Kozlenkov A, Mill J, Dracheva S, Horvath S. Genetic variants near MLST8 and DHX57 affect the epigenetic age of the cerebellum. *Nat Commun*. 2016; 7:10561.
<https://doi.org/10.1038/ncomms10561>
PMID:[26830004](https://pubmed.ncbi.nlm.nih.gov/26830004/)
33. Lu AT, Hannon E, Levine ME, Crimmins EM, Lunnon K, Mill J, Geschwind DH, Horvath S. Genetic architecture of epigenetic and neuronal ageing rates in human brain regions. *Nat Commun*. 2017; 8:15353.
<https://doi.org/10.1038/ncomms15353>
PMID:[28516910](https://pubmed.ncbi.nlm.nih.gov/28516910/)
34. Murthy M, Shireby G, Miki Y, Viré E, Lashley T, Warner TT, Mill J, Bettencourt C. Epigenetic age acceleration is associated with oligodendrocyte proportions in MSA and control brain tissue. *Neuropathol Appl Neurobiol*. 2023; 49:e12872.
<https://doi.org/10.1111/nan.12872>
PMID:[36542090](https://pubmed.ncbi.nlm.nih.gov/36542090/)
35. Shireby G, Dempster EL, Policicchio S, Smith RG, Pishva E, Chioza B, Davies JP, Burrage J, Lunnon K, Seiler Vellame D, Love S, Thomas A, Brookes K, et al. DNA methylation signatures of Alzheimer's disease neuropathology in the cortex are primarily driven by variation in non-neuronal cell-types. *Nat Commun*. 2022; 13:5620.
<https://doi.org/10.1038/s41467-022-33394-7>
PMID:[36153390](https://pubmed.ncbi.nlm.nih.gov/36153390/)
36. Shireby GL, Davies JP, Francis PT, Burrage J, Walker EM, Neilson GWA, Dahir A, Thomas AJ, Love S, Smith RG, Lunnon K, Kumari M, Schalkwyk LC, et al. Recalibrating the epigenetic clock: implications for assessing biological age in the human cortex. *Brain*. 2020; 143:3763–75.
<https://doi.org/10.1093/brain/awaa334>
PMID:[33300551](https://pubmed.ncbi.nlm.nih.gov/33300551/)
37. Brunet A, Goodell MA, Rando TA. Ageing and rejuvenation of tissue stem cells and their niches. *Nat Rev Mol Cell Biol*. 2023; 24:45–62.
<https://doi.org/10.1038/s41580-022-00510-w>
PMID:[35859206](https://pubmed.ncbi.nlm.nih.gov/35859206/)
38. Chondronasiou D, Gill D, Mosteiro L, Urduinguio RG, Berenguer-Llargo A, Aguilera M, Durand S, Aprahamian F, Nirmalathasan N, Abad M, Martin-Herranz DE, Stephan-Otto Attolini C, Prats N, et al. Multi-omic rejuvenation of naturally aged tissues by a single cycle of transient reprogramming. *Aging Cell*. 2022; 21:e13578.
<https://doi.org/10.1111/ace1.13578>
PMID:[35235716](https://pubmed.ncbi.nlm.nih.gov/35235716/)
39. de Magalhães JP, Ocampo A. Cellular reprogramming and the rise of rejuvenation biotech. *Trends Biotechnol*. 2022; 40:639–42.
<https://doi.org/10.1016/j.tibtech.2022.01.011>
PMID:[35190201](https://pubmed.ncbi.nlm.nih.gov/35190201/)
40. Gill D, Parry A, Santos F, Okkenhaug H, Todd CD, Hernando-Herraez I, Stubbs TM, Milagre I, Reik W. Multi-omic rejuvenation of human cells by maturation phase transient reprogramming. *Elife*. 2022; 11:e71624.
<https://doi.org/10.7554/elife.71624>
PMID:[35390271](https://pubmed.ncbi.nlm.nih.gov/35390271/)
41. Cole JJ, Robertson NA, Rather MI, Thomson JP, McBryan T, Sproul D, Wang T, Brock C, Clark W, Ideker T, Meehan RR, Miller RA, Brown-Borg HM, Adams PD. Diverse interventions that extend mouse lifespan suppress shared age-associated epigenetic changes at critical gene regulatory regions. *Genome Biol*. 2017; 18:58.
<https://doi.org/10.1186/s13059-017-1185-3>
PMID:[28351383](https://pubmed.ncbi.nlm.nih.gov/28351383/)
42. Wang T, Tsui B, Kreisberg JF, Robertson NA, Gross AM, Yu MK, Carter H, Brown-Borg HM, Adams PD, Ideker T. Epigenetic aging signatures in mice livers are slowed by dwarfism, calorie restriction and rapamycin treatment. *Genome Biol*. 2017; 18:57.
<https://doi.org/10.1186/s13059-017-1186-2>
PMID:[28351423](https://pubmed.ncbi.nlm.nih.gov/28351423/)
43. Trapp A, Kerepesi C, Gladyshev VN. Profiling epigenetic age in single cells. *Nat Aging*. 2021; 1:1189–201.
<https://doi.org/10.1038/s43587-021-00134-3>
PMID:[36211119](https://pubmed.ncbi.nlm.nih.gov/36211119/)
44. Zhu T, Liu J, Beck S, Pan S, Capper D, Lechner M, Thirlwell C, Breeze CE, Teschendorff AE. A pan-tissue DNA methylation atlas enables in silico decomposition of human tissue methylomes at cell-type resolution. *Nat Methods*. 2022; 19:296–306.
<https://doi.org/10.1038/s41592-022-01412-7>
PMID:[35277705](https://pubmed.ncbi.nlm.nih.gov/35277705/)
45. Zheng SC, Breeze CE, Beck S, Teschendorff AE. Identification of differentially methylated cell types in epigenome-wide association studies. *Nat Methods*. 2018; 15:1059–66.
<https://doi.org/10.1038/s41592-018-0213-x>
PMID:[30504870](https://pubmed.ncbi.nlm.nih.gov/30504870/)
46. Zhang Z, Wiencke JK, Kelsey KT, Koestler DC, Molinaro AM, Pike SC, Karra P, Christensen BC, Salas LA. Hierarchical deconvolution for extensive cell type

- resolution in the human brain using DNA methylation. *Front Neurosci.* 2023; 17:1198243.
<https://doi.org/10.3389/fnins.2023.1198243>
PMID:37404460
47. Chien JF, Liu H, Wang BA, Luo C, Bartlett A, Castanon R, Johnson ND, Nery JR, Osteen J, Li J, Altshul J, Kenworthy M, Valadon C, et al. Cell-type-specific effects of age and sex on human cortical neurons. *Neuron.* 2024; 112:2524–39.e5.
<https://doi.org/10.1016/j.neuron.2024.05.013>
PMID:38838671
48. Cai M, Zhou J, McKennan C, Wang J. scMD facilitates cell type deconvolution using single-cell DNA methylation references. *Commun Biol.* 2024; 7:1.
<https://doi.org/10.1038/s42003-023-05690-5>
PMID:38168620
49. Friedman J, Hastie T, Tibshirani R. Regularization Paths for Generalized Linear Models via Coordinate Descent. *J Stat Softw.* 2010; 33:1–22.
PMID:20808728
50. Jaffe AE, Gao Y, Deep-Soboslay A, Tao R, Hyde TM, Weinberger DR, Kleinman JE. Mapping DNA methylation across development, genotype and schizophrenia in the human frontal cortex. *Nat Neurosci.* 2016; 19:40–7.
<https://doi.org/10.1038/nn.4181>
PMID:26619358
51. 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.
<https://doi.org/10.1038/s41467-020-18618-y>
PMID:32963246
52. Pihlstrøm L, Shireby G, Geut H, Henriksen SP, Rozemuller AJM, Tunold JA, Hannon E, Francis P, Thomas AJ, Love S, Mill J, van de Berg WDJ, Toft M. Epigenome-wide association study of human frontal cortex identifies differential methylation in Lewy body pathology. *Nat Commun.* 2022; 13:4932.
<https://doi.org/10.1038/s41467-022-32619-z>
PMID:35995800
53. Teschendorff AE, Zhu T, Breeze CE, Beck S. EPISCOPE: cell type deconvolution of bulk tissue DNA methylomes from single-cell RNA-Seq data. *Genome Biol.* 2020; 21:221.
<https://doi.org/10.1186/s13059-020-02126-9>
PMID:32883324
54. Johnson ND, Wu X, Still CD, Chu X, Petrick AT, Gerhard GS, Conneely KN, DiStefano JK. Differential DNA methylation and changing cell-type proportions as fibrotic stage progresses in NAFLD. *Clin Epigenetics.* 2021; 13:152.
<https://doi.org/10.1186/s13148-021-01129-y>
PMID:34353365
55. Teschendorff AE. Avoiding common pitfalls in machine learning omic data science. *Nat Mater.* 2019; 18:422–7.
<https://doi.org/10.1038/s41563-018-0241-z>
PMID:30478452
56. Bonder MJ, Kasela S, Kals M, Tamm R, Lökk K, Barragan I, Buurman WA, Deelen P, Greve JW, Ivanov M, Rensen SS, van Vliet-Ostaptchouk JV, Wolfs MG, et al. Genetic and epigenetic regulation of gene expression in fetal and adult human livers. *BMC Genomics.* 2014; 15:860.
<https://doi.org/10.1186/1471-2164-15-860>
PMID:25282492
57. Wang L, Marek GW 3rd, Hlady RA, Wagner RT, Zhao X, Clark VC, Fan AX, Liu C, Brantly M, Robertson KD. Alpha-1 Antitrypsin Deficiency Liver Disease, Mutational Homogeneity Modulated by Epigenetic Heterogeneity With Links to Obesity. *Hepatology.* 2019; 70:51–66.
<https://doi.org/10.1002/hep.30526>
PMID:30681738
58. Horvath S, Erhart W, Brosch M, Ammerpohl O, von Schönfels W, Ahrens M, Heits N, Bell JT, Tsai PC, Spector TD, Deloukas P, Siebert R, Sipos B, et al. Obesity accelerates epigenetic aging of human liver. *Proc Natl Acad Sci U S A.* 2014; 111:15538–43.
<https://doi.org/10.1073/pnas.1412759111>
PMID:25313081
59. Ahrens M, Ammerpohl O, von Schönfels W, Kolarova J, Bens S, Itzel T, Teufel A, Herrmann A, Brosch M, Hinrichsen H, Erhart W, Egberts J, Sipos B, et al. DNA methylation analysis in nonalcoholic fatty liver disease suggests distinct disease-specific and remodeling signatures after bariatric surgery. *Cell Metab.* 2013; 18:296–302.
<https://doi.org/10.1016/j.cmet.2013.07.004>
PMID:23931760
60. 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; 4:231–46.
<https://doi.org/10.1038/s43587-023-00557-0>
PMID:38243142
61. Chen YS, Chen SD, Wu CL, Huang SS, Yang DI. Induction of sestrin2 as an endogenous protective mechanism against amyloid beta-peptide neurotoxicity in primary cortical culture. *Exp Neurol.* 2014; 253:63–71.
<https://doi.org/10.1016/j.expneurol.2013.12.009>
PMID:24368194

62. Rai N, Kumar R, Desai GR, Venugopalan G, Shekhar S, Chatterjee P, Tripathi M, Upadhyay AD, Dwivedi S, Dey AB, Dey S. Relative Alterations in Blood-Based Levels of Sestrin in Alzheimer's Disease and Mild Cognitive Impairment Patients. *J Alzheimers Dis.* 2016; 54:1147–55.
<https://doi.org/10.3233/JAD-160479>
PMID:[27567861](https://pubmed.ncbi.nlm.nih.gov/27567861/)
63. Wang B, Fu C, Wei Y, Xu B, Yang R, Li C, Qiu M, Yin Y, Qin D. Ferroptosis-related biomarkers for Alzheimer's disease: Identification by bioinformatic analysis in hippocampus. *Front Cell Neurosci.* 2022; 16:1023947.
<https://doi.org/10.3389/fncel.2022.1023947>
PMID:[36467613](https://pubmed.ncbi.nlm.nih.gov/36467613/)
64. Yu M, Nie Y, Yang J, Yang S, Li R, Rao V, Hu X, Fang C, Li S, Song D, Guo F, Snyder MP, Chang HY, et al. Integrative multi-omic profiling of adult mouse brain endothelial cells and potential implications in Alzheimer's disease. *Cell Rep.* 2023; 42:113392.
<https://doi.org/10.1016/j.celrep.2023.113392>
PMID:[37925638](https://pubmed.ncbi.nlm.nih.gov/37925638/)
65. Lu MH, Ji WL, Xu DE, Yao PP, Zhao XY, Wang ZT, Fang LP, Huang R, Lan LJ, Chen JB, Wang TH, Cheng LH, Xu RX, et al. Inhibition of sphingomyelin synthase 1 ameliorates alzheimer-like pathology in APP/PS1 transgenic mice through promoting lysosomal degradation of BACE1. *Exp Neurol.* 2019; 311:67–79.
<https://doi.org/10.1016/j.expneurol.2018.09.012>
PMID:[30243987](https://pubmed.ncbi.nlm.nih.gov/30243987/)
66. Hsiao JH, Fu Y, Hill AF, Halliday GM, Kim WS. Elevation in sphingomyelin synthase activity is associated with increases in amyloid-beta peptide generation. *PLoS One.* 2013; 8:e74016.
<https://doi.org/10.1371/journal.pone.0074016>
PMID:[23977395](https://pubmed.ncbi.nlm.nih.gov/23977395/)
67. Garcia-Esparcia P, Koneti A, Rodríguez-Oroz MC, Gago B, Del Rio JA, Ferrer I. Mitochondrial activity in the frontal cortex area 8 and angular gyrus in Parkinson's disease and Parkinson's disease with dementia. *Brain Pathol.* 2018; 28:43–57.
<https://doi.org/10.1111/bpa.12474>
PMID:[27984680](https://pubmed.ncbi.nlm.nih.gov/27984680/)
68. Bu S, Lv Y, Liu Y, Qiao S, Wang H. Zinc Finger Proteins in Neuro-Related Diseases Progression. *Front Neurosci.* 2021; 15:760567.
<https://doi.org/10.3389/fnins.2021.760567>
PMID:[34867169](https://pubmed.ncbi.nlm.nih.gov/34867169/)
69. Deng Y, He T, Fang R, Li S, Cao H, Cui Y. Genome-Wide Gene-Based Multi-Trait Analysis. *Front Genet.* 2020; 11:437.
<https://doi.org/10.3389/fgene.2020.00437>
PMID:[32508874](https://pubmed.ncbi.nlm.nih.gov/32508874/)
70. Yuan S, Li H, Xie J, Sun X. Quantitative Trait Module-Based Genetic Analysis of Alzheimer's Disease. *Int J Mol Sci.* 2019; 20:5912.
<https://doi.org/10.3390/ijms20235912>
PMID:[31775305](https://pubmed.ncbi.nlm.nih.gov/31775305/)
71. Arefin AS, Mathieson L, Johnstone D, Berretta R, Moscato P. Unveiling clusters of RNA transcript pairs associated with markers of Alzheimer's disease progression. *PLoS One.* 2012; 7:e45535.
<https://doi.org/10.1371/journal.pone.0045535>
PMID:[23029078](https://pubmed.ncbi.nlm.nih.gov/23029078/)
72. Popov N, Gil J. Epigenetic regulation of the INK4b-ARF-INK4a locus: in sickness and in health. *Epigenetics.* 2010; 5:685–90.
<https://doi.org/10.4161/epi.5.8.12996>
PMID:[20716961](https://pubmed.ncbi.nlm.nih.gov/20716961/)
73. Duan RS, Tang GB, Du HZ, Hu YW, Liu PP, Xu YJ, Zeng YQ, Zhang SF, Wang RY, Teng ZQ, Liu CM. Polycomb protein family member CBX7 regulates intrinsic axon growth and regeneration. *Cell Death Differ.* 2018; 25:1598–611.
<https://doi.org/10.1038/s41418-018-0064-0>
PMID:[29459770](https://pubmed.ncbi.nlm.nih.gov/29459770/)
74. Caselli RJ, Dueck AC, Huentelman MJ, Lutz MW, Saunders AM, Reiman EM, Roses AD. Longitudinal modeling of cognitive aging and the TOMM40 effect. *Alzheimers Dement.* 2012; 8:490–5.
<https://doi.org/10.1016/j.jalz.2011.11.006>
PMID:[23102119](https://pubmed.ncbi.nlm.nih.gov/23102119/)
75. Lee EG, Chen S, Leong L, Tulloch J, Yu CE. *TOMM40* RNA Transcription in Alzheimer's Disease Brain and Its Implication in Mitochondrial Dysfunction. *Genes (Basel).* 2021; 12:871.
<https://doi.org/10.3390/genes12060871>
PMID:[34204109](https://pubmed.ncbi.nlm.nih.gov/34204109/)
76. Tirozzi A, Quicciono MS, Cerletti C, Donati MB, de Gaetano G, Iacoviello L, Gialluisi A. A Multi-Trait Association Analysis of Brain Disorders and Platelet Traits Identifies Novel Susceptibility Loci for Major Depression, Alzheimer's and Parkinson's Disease. *Cells.* 2023; 12:245.
<https://doi.org/10.3390/cells12020245>
PMID:[36672180](https://pubmed.ncbi.nlm.nih.gov/36672180/)
77. Zhu Z, Yang Y, Xiao Z, Zhao Q, Wu W, Liang X, Luo J, Cao Y, Shao M, Guo Q, Ding D. TOMM40 and APOE variants synergistically increase the risk of Alzheimer's disease in a Chinese population. *Aging Clin Exp Res.* 2021; 33:1667–75.
<https://doi.org/10.1007/s40520-020-01661-6>
PMID:[32725468](https://pubmed.ncbi.nlm.nih.gov/32725468/)
78. Cardoso R, Lemos C, Oliveiros B, Almeida MR, Baldeiras I, Pereira CF, Santos A, Duro D, Vieira D,

- Santana I, Oliveira CR. APOE ϵ 4-TOMM40L Haplotype Increases the Risk of Mild Cognitive Impairment Conversion to Alzheimer's Disease. *J Alzheimers Dis.* 2020; 78:587–601.
<https://doi.org/10.3233/JAD-200556>
PMID:33016906
79. Battram T, Yousefi P, Crawford G, Prince C, Sheikhali Babaei M, Sharp G, Hatcher C, Vega-Salas MJ, Khodabakhsh S, Whitehurst O, Langdon R, Mahoney L, Elliott HR, et al. The EWAS Catalog: a database of epigenome-wide association studies. *Wellcome Open Res.* 2022; 7:41.
<https://doi.org/10.12688/wellcomeopenres.17598.2>
PMID:35592546
80. Xiong Z, Yang F, Li M, Ma Y, Zhao W, Wang G, Li Z, Zheng X, Zou D, Zong W, Kang H, Jia Y, Li R, et al. EWAS Open Platform: integrated data, knowledge and toolkit for epigenome-wide association study. *Nucleic Acids Res.* 2022; 50:D1004–9.
<https://doi.org/10.1093/nar/gkab972>
PMID:34718752
81. Gaunt TR, Shihab HA, Hemani G, Min JL, Woodward G, Lyttleton O, Zheng J, Duggirala A, McArdle WL, Ho K, Ring SM, Evans DM, Davey Smith G, Relton CL. Systematic identification of genetic influences on methylation across the human life course. *Genome Biol.* 2016; 17:61.
<https://doi.org/10.1186/s13059-016-0926-z>
PMID:27036880
82. Oliva M, Demanelis K, Lu Y, Chernoff M, Jasmine F, Ahsan H, Kibriya MG, Chen LS, Pierce BL. DNA methylation QTL mapping across diverse human tissues provides molecular links between genetic variation and complex traits. *Nat Genet.* 2023; 55:112–22.
<https://doi.org/10.1038/s41588-022-01248-z>
PMID:36510025
83. Hannon E, Spiers H, Viana J, Pidsley R, Burrage J, Murphy TM, Troakes C, Turecki G, O'Donovan MC, Schalkwyk LC, Bray NJ, Mill J. Methylation QTLs in the developing brain and their enrichment in schizophrenia risk loci. *Nat Neurosci.* 2016; 19:48–54.
<https://doi.org/10.1038/nn.4182>
PMID:26619357
84. Ng B, White CC, Klein HU, Sieberts SK, McCabe C, Patrick E, Xu J, Yu L, Gaiteri C, Bennett DA, Mostafavi S, De Jager PL. An xQTL map integrates the genetic architecture of the human brain's transcriptome and epigenome. *Nat Neurosci.* 2017; 20:1418–26.
<https://doi.org/10.1038/nn.4632>
PMID:28869584
85. Bellenguez C, Küçükali F, Jansen IE, Kleindam L, Moreno-Grau S, Amin N, Naj AC, Campos-Martin R, Grenier-Boley B, Andrade V, Holmans PA, Boland A, Damotte V, et al, and EADB, and GR@ACE, and DEGESCO, and EADI, and GERAD, and Demgene, and FinnGen, and ADGC, and CHARGE. New insights into the genetic etiology of Alzheimer's disease and related dementias. *Nat Genet.* 2022; 54:412–36.
<https://doi.org/10.1038/s41588-022-01024-z>
PMID:35379992
86. Andrade-Guerrero J, Santiago-Balmaseda A, Jeronimo-Aguilar P, Vargas-Rodríguez I, Cadena-Suárez AR, Sánchez-Garibay C, Pozo-Molina G, Méndez-Catalá CF, Cardenas-Aguayo MD, Diaz-Cintra S, Pacheco-Herrero M, Luna-Muñoz J, Soto-Rojas LO. Alzheimer's Disease: An Updated Overview of Its Genetics. *Int J Mol Sci.* 2023; 24:3754.
<https://doi.org/10.3390/ijms24043754>
PMID:36835161
87. Hassan M, Shahzadi S, Alashwal H, Zaki N, Seo SY, Moustafa AA. Exploring the mechanistic insights of Cas scaffolding protein family member 4 with protein tyrosine kinase 2 in Alzheimer's disease by evaluating protein interactions through molecular docking and dynamic simulations. *Neurol Sci.* 2018; 39:1361–74.
<https://doi.org/10.1007/s10072-018-3430-2>
PMID:29789968
88. Sáez ME, García-Sánchez A, de Rojas I, Alarcón-Martín E, Martínez J, Cano A, García-González P, Puerta R, Olivé C, Capdevila M, García-Gutiérrez F, Castilla-Martí M, Castilla-Martí L, et al. Genome-wide association study and polygenic risk scores of retinal thickness across the cognitive continuum: data from the NORFACE cohort. *Alzheimers Res Ther.* 2024; 16:38.
<https://doi.org/10.1186/s13195-024-01398-8>
PMID:38365752
89. Yang X, Wen J, Yang H, Jones IR, Zhu X, Liu W, Li B, Clelland CD, Luo W, Wong MY, Ren X, Cui X, Song M, et al. Functional characterization of Alzheimer's disease genetic variants in microglia. *Nat Genet.* 2023; 55:1735–44.
<https://doi.org/10.1038/s41588-023-01506-8>
PMID:37735198
90. Lee JH, Budanov AV, Karin M. Sestrins orchestrate cellular metabolism to attenuate aging. *Cell Metab.* 2013; 18:792–801.
<https://doi.org/10.1016/j.cmet.2013.08.018>
PMID:24055102
91. Lee JH, Budanov AV, Park EJ, Birse R, Kim TE, Perkins GA, Ocorr K, Ellisman MH, Bodmer R, Bier E, Karin M. Sestrin as a feedback inhibitor of TOR that prevents age-related pathologies. *Science.* 2010; 327:1223–8.
<https://doi.org/10.1126/science.1182228>
PMID:20203043

92. Barreby E, Chen P, Aouadi M. Macrophage functional diversity in NAFLD - more than inflammation. *Nat Rev Endocrinol*. 2022; 18:461–72.
<https://doi.org/10.1038/s41574-022-00675-6>
PMID:35534573
93. Arendt D, Musser JM, Baker CVH, Bergman A, Cepko C, Erwin DH, Pavlicev M, Schlosser G, Widder S, Laubichler MD, Wagner GP. The origin and evolution of cell types. *Nat Rev Genet*. 2016; 17:744–57.
<https://doi.org/10.1038/nrg.2016.127>
PMID:27818507
94. Zhu T, Zheng SC, Paul DS, Horvath S, Teschendorff AE. Cell and tissue type independent age-associated DNA methylation changes are not rare but common. *Aging (Albany NY)*. 2018; 10:3541–57.
<https://doi.org/10.18632/aging.101666>
PMID:30482885
95. Aryee MJ, Jaffe AE, Corrada-Bravo H, Ladd-Acosta C, Feinberg AP, Hansen KD, Irizarry RA. Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. *Bioinformatics*. 2014; 30:1363–9.
<https://doi.org/10.1093/bioinformatics/btu049>
PMID:24478339
96. Troyanskaya O, Cantor M, Sherlock G, Brown P, Hastie T, Tibshirani R, Botstein D, Altman RB. Missing value estimation methods for DNA microarrays. *Bioinformatics*. 2001; 17:520–5.
<https://doi.org/10.1093/bioinformatics/17.6.520>
PMID:11395428
97. Teschendorff AE, Marabita F, Lechner M, Bartlett T, Tegner J, Gomez-Cabrero D, Beck S. A beta-mixture quantile normalization method for correcting probe design bias in Illumina Infinium 450 k DNA methylation data. *Bioinformatics*. 2013; 29:189–96.
<https://doi.org/10.1093/bioinformatics/bts680>
PMID:23175756
98. Park CS, De T, Xu Y, Zhong Y, Smithberger E, Alarcon C, Gamazon ER, Perera MA. Hepatocyte gene expression and DNA methylation as ancestry-dependent mechanisms in African Americans. *NPJ Genom Med*. 2019; 4:29.
<https://doi.org/10.1038/s41525-019-0102-y>
PMID:31798965
99. Bacalini MG, Franceschi C, Gentilini D, Ravaioli F, Zhou X, Remondini D, Pirazzini C, Giuliani C, Marasco E, Gensous N, Di Blasio AM, Ellis E, Gramignoli R, et al. Molecular Aging of Human Liver: An Epigenetic/Transcriptomic Signature. *J Gerontol A Biol Sci Med Sci*. 2019; 74:1–8.
<https://doi.org/10.1093/gerona/gly048>
PMID:29554203
100. Díez-Villanueva A, Sanz-Pamplona R, Carreras-Torres R, Moratalla-Navarro F, Alonso MH, Paré-Brunet L, Aussó S, Guinó E, Solé X, Cordero D, Salazar R, Berdasco M, Peinado MA, Moreno V. DNA methylation events in transcription factors and gene expression changes in colon cancer. *Epigenomics*. 2020; 12:1593–610.
<https://doi.org/10.2217/epi-2020-0029>
PMID:32957849
101. Kirby MK, Ramaker RC, Roberts BS, Lasseigne BN, Gunther DS, Burwell TC, Davis NS, Gulzar ZG, Absher DM, Cooper SJ, Brooks JD, Myers RM. Genome-wide DNA methylation measurements in prostate tissues uncovers novel prostate cancer diagnostic biomarkers and transcription factor binding patterns. *BMC Cancer*. 2017; 17:273.
<https://doi.org/10.1186/s12885-017-3252-2>
PMID:28412973
102. Huan T, Mendelson M, Joehanes R, Yao C, Liu C, Song C, Bhattacharya A, Rong J, Tanriverdi K, Keefe J, Murabito JM, Courchesne P, Larson MG, et al. Epigenome-wide association study of DNA methylation and microRNA expression highlights novel pathways for human complex traits. *Epigenetics*. 2020; 15:183–98.
<https://doi.org/10.1080/15592294.2019.1640547>
PMID:31282290
103. Roos L, Sandling JK, Bell CG, Glass D, Mangino M, Spector TD, Deloukas P, Bataille V, Bell JT. Higher Nevus Count Exhibits a Distinct DNA Methylation Signature in Healthy Human Skin: Implications for Melanoma. *J Invest Dermatol*. 2017; 137:910–20.
<https://doi.org/10.1016/j.jid.2016.11.029>
PMID:27993549
104. Houseman EA, Accomando WP, Koestler DC, Christensen BC, Marsit CJ, Nelson HH, Wiencke JK, Kelsey KT. DNA methylation arrays as surrogate measures of cell mixture distribution. *BMC Bioinformatics*. 2012; 13:86.
<https://doi.org/10.1186/1471-2105-13-86>
PMID:22568884
105. 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.
<https://doi.org/10.1038/s41467-021-27864-7>
PMID:35140201
106. Teschendorff AE, Zheng SC. Cell-type deconvolution in epigenome-wide association studies: a review and recommendations. *Epigenomics*. 2017; 9:757–68.
<https://doi.org/10.2217/epi-2016-0153>
PMID:28517979

107. Robinson O, Chadeau Hyam M, Karaman I, Climaco Pinto R, Ala-Korpela M, Handakas E, Fiorito G, Gao H, Heard A, Jarvelin MR, Lewis M, Pazoki R, Polidoro S, et al. Determinants of accelerated metabolomic and epigenetic aging in a UK cohort. *Aging Cell*. 2020; 19:e13149.
<https://doi.org/10.1111/accel.13149>
PMID:[32363781](https://pubmed.ncbi.nlm.nih.gov/32363781/)
108. Lehne B, Drong AW, Loh M, Zhang W, Scott WR, Tan ST, Afzal U, Scott J, Jarvelin MR, Elliott P, McCarthy MI, Kooner JS, Chambers JC. A coherent approach for analysis of the Illumina HumanMethylation450 BeadChip improves data quality and performance in epigenome-wide association studies. *Genome Biol*. 2015; 16:37.
<https://doi.org/10.1186/s13059-015-0600-x>
PMID:[25853392](https://pubmed.ncbi.nlm.nih.gov/25853392/)
109. Teschendorff AE, Breeze CE, Zheng SC, Beck S. A comparison of reference-based algorithms for correcting cell-type heterogeneity in Epigenome-Wide Association Studies. *BMC Bioinformatics*. 2017; 18:105.
<https://doi.org/10.1186/s12859-017-1511-5>
PMID:[28193155](https://pubmed.ncbi.nlm.nih.gov/28193155/)
110. Hannum G, Guinney J, Zhao L, Zhang L, Hughes G, Sada S, Klotzle B, Bibikova M, Fan JB, Gao Y, Deconde R, Chen M, Rajapakse I, et al. Genome-wide methylation profiles reveal quantitative views of human aging rates. *Mol Cell*. 2013; 49:359–67.
<https://doi.org/10.1016/j.molcel.2012.10.016>
PMID:[23177740](https://pubmed.ncbi.nlm.nih.gov/23177740/)
111. Chen L, Ge B, Casale FP, Vasquez L, Kwan T, Garrido-Martín D, Watt S, Yan Y, Kundu K, Ecker S, Datta A, Richardson D, Burden F, et al. Genetic Drivers of Epigenetic and Transcriptional Variation in Human Immune Cells. *Cell*. 2016; 167:1398–414.e24.
<https://doi.org/10.1016/j.cell.2016.10.026>
PMID:[27863251](https://pubmed.ncbi.nlm.nih.gov/27863251/)
112. Borenstein M, Hedges LV, Higgins JP, Rothstein HR. A basic introduction to fixed-effect and random-effects models for meta-analysis. *Res Synth Methods*. 2010; 1:97–111.
<https://doi.org/10.1002/jrsm.12>
PMID:[26061376](https://pubmed.ncbi.nlm.nih.gov/26061376/)
113. Balduzzi S, Rücker G, Schwarzer G. How to perform a meta-analysis with R: a practical tutorial. *Evid Based Ment Health*. 2019; 22:153–60.
<https://doi.org/10.1136/ebmental-2019-300117>
PMID:[31563865](https://pubmed.ncbi.nlm.nih.gov/31563865/)
114. Pulit SL, Stoneman C, Morris AP, Wood AR, Glastonbury CA, Tyrrell J, Yengo L, Ferreira T, Marouli E, Ji Y, Yang J, Jones S, Beaumont R, et al, and GIANT Consortium. Meta-analysis of genome-wide association studies for body fat distribution in 694 649 individuals of European ancestry. *Hum Mol Genet*. 2019; 28:166–74.
<https://doi.org/10.1093/hmg/ddy327>
PMID:[30239722](https://pubmed.ncbi.nlm.nih.gov/30239722/)

SUPPLEMENTARY MATERIALS

Supplementary Methods

DNA methylation datasets used in validation of the neuron/glia and brain-specific clocks

To validate all clocks and to demonstrate their specificity, we assembled a large collection of Illumina DNAm datasets, encompassing sorted brain cell (6 independent datasets) [1–6], bulk brain tissue (19 independent datasets) [2, 7–20] and other sorted non-brain cell-types and tissue-types. We restricted the assessment of the clocks to the control samples, to avoid potential confounding by disease. Below we provide details of the processing of the brain-related DNAm datasets used in the validation.

Sorted neuron and glia datasets

Pai

The EPIC dataset from Pai et al. [1] profiled DNAm in human frontal cortex sorted neuron from controls was obtained from the NCBI GEO website under the accession number GSE112179 (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE112179>). The file “GSE112179_RAW.tar” which contains the IDAT files was downloaded and processed with minfi package. Probes with NAs (defined by $P > 0.01$) were discarded. The filtered data was subsequently normalized with BMIQ, resulting in a normalized data matrix for 855735 probes and 100 samples.

Gasparoni

The 450k dataset from Gasparoni et al. [2] profiled DNAm in human occipital cortex sorted neuron and gliocyte from controls was obtained from the NCBI GEO website under the accession number GSE66351 (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE66351>). The file “GSE66351_RAW.tar” which contains the IDAT files was downloaded and processed with minfi package. Probes with NAs (defined by $P > 0.01$) were discarded. The filtered data was subsequently normalized with BMIQ, resulting in a normalized data matrix for 482176 probes and 62 samples.

Kozlenkov

The 450k dataset from Kozlenkov et al. [3] profiled DNAm in human frontal cortex sorted neuron from controls was obtained from the NCBI GEO website under the accession number GSE98203 (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE98203>). The file “GSE98203_RAW.tar” which contains the IDAT files was downloaded and processed

with minfi package. Probes with NAs (defined by $P > 0.01$) were discarded. The filtered data was subsequently normalized with BMIQ, resulting in a normalized data matrix for 469902 probes and 88 samples.

Guintivano

The 450k dataset from Guintivano et al. [4] profiled DNAm in human frontal cortex sorted neuron and gliocyte from controls was obtained from the NCBI GEO website under the accession number GSE41826 (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE41826>). The file “GSE41826_RAW.tar” which contains the IDAT files was downloaded and processed with minfi package. Probes with NAs (defined by $P > 0.01$) were discarded. The filtered data was subsequently normalized with BMIQ, resulting in a normalized data matrix for 482469 probes and 58 samples.

Hannon

The EPIC dataset from Hannon et al. [5] profiled DNAm in human frontal cortex sorted neuron and gliocyte from controls was obtained from the NCBI GEO website under the accession number GSE234520 (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE234520>). The file “GSE234520_RAW.tar” which contains the IDAT files was downloaded and processed with minfi package. Probes with NAs (defined by $P > 0.01$) were discarded. The filtered data was subsequently normalized with BMIQ, resulting in a normalized data matrix for 812778 probes and 103 samples.

Witte

The EPIC dataset from Witte et al. [6] profiled DNAm in human four different brain regions, medial frontal gyrus (MFG), superior temporal gyrus (STG), subventricular zone (SVZ) and thalamus (THA) sorted gliocyte from controls was obtained from the NCBI GEO website under the accession number GSE191200 (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE191200>). The file “GSE191200_RAW.tar” which contains the IDAT files was downloaded and processed with minfi package. Probes with NAs (defined by $P > 0.01$) were discarded. The filtered data was subsequently normalized with BMIQ, resulting in a normalized data matrix for 838988 probes and 56 samples.

Brain bulk datasets

Murphy

The 450k dataset from Murphy et al. [7] profiled DNAm in human PFC from controls was obtained from the NCBI GEO website under the accession

number GSE88890 (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE88890>). The files “GSE88890_methylatedIntensities.csv.gz”, “GSE88890_unmethylatedIntensities.csv.gz”, “GSE88890_detectionP.csv.gz” were downloaded and processed with minfi package. Probes with > 3% NAs (defined by $P > 0.01$) were discarded. The filtered data was subsequently normalized with BMIQ, resulting in a normalized data matrix for 416054 probes and 75 samples.

Rydbirk

The EPIC dataset from Rydbirk et al. [8] profiled DNAm in human PFC from controls was obtained from the NCBI GEO website under the accession number GSE143157 (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE143157>). The file “GSE143157_RAW.tar” which contains the IDAT files was downloaded and processed with minfi package. Probes with NAs (defined by $P > 0.01$) were discarded. The filtered data was subsequently normalized with BMIQ, resulting in a normalized data matrix for 831166 probes and 78 samples.

Torabi

The 450K dataset from Torabi et al. [9] profiled DNAm in human PFC from controls was obtained from the NCBI GEO website under the accession number GSE128601 (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE128601>). The file “GSE128601_Matrix_signal_intensities_scz.txt.gz” was downloaded and processed with minfi package. Probes with NAs (defined by $P > 0.01$) were discarded. The filtered data was subsequently normalized with BMIQ, resulting in a normalized data matrix for 478722 probes and 125 samples.

Xu

The 450K dataset from Xu et al. [10] profiled DNAm in human PFC from controls was obtained from the NCBI GEO website under the accession number GSE49393 (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE49393>). The file “GSE49393_signal_intensities.txt.gz” was downloaded and processed with minfi package. Probes with NAs (defined by $P > 0.01$) were discarded. The filtered data was subsequently normalized with BMIQ, resulting in a normalized data matrix for 482931 probes and 48 samples.

Huynh

The 450K dataset from Huynh et al. [11] profiled DNAm in human frontal lobe from controls was obtained from the NCBI GEO website under the accession number GSE40360 (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE40360>). The file “GSE40360_RAW_nonorm_nobg.txt.gz” was downloaded and processed with minfi package. Probes with

NAs (defined by $P > 0.01$) were discarded. The filtered data was subsequently normalized with BMIQ, resulting in a normalized data matrix for 474136 probes and 46 samples.

Viana2

The 450K dataset from Viana et al. [12] profiled DNAm in human hippocampus from controls was obtained from the NCBI GEO website under the accession number GSE89703 (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE89703>). The files “GSE89703_methylatedintensities_HC_LNDBB.csv.gz”, “GSE89703_unmethylatedintensities_HC_LNDBB.csv.gz”, “GSE89703_detectionP_HC_LNDBB.csv.gz” were downloaded and processed with minfi package. Probes with NAs (defined by $P > 0.01$) were discarded. The filtered data was subsequently normalized with BMIQ, resulting in a normalized data matrix for 413242 probes and 27 samples.

Viana3

The 450K dataset from Viana et al. [12] profiled DNAm in human striatum, putamen from controls was obtained from the NCBI GEO website under the accession number GSE89705 (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE89705>). The files “GSE89705_methylatedintensities_STR_DBCBB.csv.gz”, “GSE89705_unmethylatedintensities_STR_DBCBB.csv.gz”, “GSE89705_detectionP_STR_DBCBB.csv.gz” were downloaded and processed with minfi package. Probes with NAs (defined by $P > 0.01$) were discarded. The filtered data was subsequently normalized with BMIQ, resulting in a normalized data matrix for 413468 probes and 33 samples.

Viana4

The 450K dataset from Viana et al. [12] profiled DNAm in human striatum, putamen from controls was obtained from the NCBI GEO website under the accession number GSE89706 (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE89706>). The files “GSE89706_methylatedintensities_STR_LNDBB.csv.gz”, “GSE89706_unmethylatedintensities_STR_LNDBB.csv.gz”, “GSE89706_detectionP_STR_LNDBB.csv.gz” were downloaded and processed with minfi package. Probes with NAs (defined by $P > 0.01$) were discarded. The filtered data was subsequently normalized with BMIQ, resulting in a normalized data matrix for 418738 probes and 49 samples.

Markunas

The EPIC dataset from Markunas et al. [13] profiled DNAm in human nucleus accumbens from controls was obtained from the NCBI GEO website under the accession number GSE147040 (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE147040>). The file

“GSE147040_matrix_intensities_for_geo.txt.gz”, was downloaded and processed with minfi package. Probes with NAs (defined by $P > 0.01$) were discarded. The filtered data was subsequently normalized with BMIQ, resulting in a normalized data matrix for 850934 probes and 221 samples.

Viana1

The 450K dataset from Viana et al. [12] profiled DNAm in human cerebellum from controls was obtained from the NCBI GEO website under the accession number GSE89702 (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE89702>). The files “GSE89702_methylatedintensities_CER_DBCBB.csv.gz”, “GSE89702_unmethylatedintensities_CER_DBCBB.csv.gz”, “GSE89702_detectionP_CER_DBCBB.csv.gz” were downloaded and processed with minfi package. Probes with NAs (defined by $P > 0.01$) were discarded. The filtered data was subsequently normalized with BMIQ, resulting in a normalized data matrix for 413811 probes and 33 samples.

Stefania1

The EPIC dataset from Stefania et al. [14] profiled DNAm in human cerebellum from controls was obtained from the NCBI GEO website under the accession number GSE137222 (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE137222>). The files “GSE137222_methylatedIntensities_AI.csv.gz”, “GSE137222_unmethylatedIntensities_AI.csv.gz”, “GSE137222_detectionP_AI.csv.gz” were downloaded and processed with minfi package. Probes with NAs (defined by $P > 0.01$) were discarded. The filtered data was subsequently normalized with BMIQ, resulting in a normalized data matrix for 794359 probes and 58 samples.

Stefania2

The 450k dataset from Stefania et al. [14] profiled DNAm in human cerebellum from controls was obtained from the NCBI GEO website under the accession number GSE137223 (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE137223>). The files “GSE137223_methylatedIntensities_ED.csv.gz”, “GSE137223_unmethylatedIntensities_ED.csv.gz”, “GSE137223_detectionP_ED.csv.gz” were downloaded and processed with minfi package. Probes with NAs (defined by $P > 0.01$) were discarded. The filtered data was subsequently normalized with BMIQ, resulting in a normalized data matrix for 427624 probes and 33 samples.

Semick

The 450k dataset from Semick et al. [15] profiled DNAm in human dorsolateral prefrontal cortex (DLPFC), hippocampus (HIPPO), entorhinal cortex

(ERC), and cerebellum (CRB) from controls and AD cases was obtained from the NCBI GEO website under the accession number GSE125895 (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE125895>). The file “GSE125895_RAW.tar” which contains the IDAT files was downloaded and processed with minfi package. Probes with NAs (defined by $P > 0.01$) were discarded. The filtered data was subsequently normalized with BMIQ, resulting in a normalized data matrix for 437426 probes and 269 samples.

Pidsley

The 450k dataset from Pidsley et al. [16] profiled DNAm in human prefrontal cortex (PFC), entorhinal cortex (EC), superior temporal gyrus (STG) from controls and AD cases was obtained from the NCBI GEO website under the accession number GSE43414 (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE43414>). The file “GSE43414_signal_intensities.csv.gz” was downloaded and processed with minfi package. Probes with NAs (defined by $P > 0.01$) were discarded. The filtered data was subsequently normalized with BMIQ, resulting in a normalized data matrix for 474819 probes and 250 samples.

Smith

The 450k dataset from Smith et al. [17] profiled DNAm in human prefrontal cortex (PFC), superior temporal gyrus (STG) from controls and AD cases was obtained from the NCBI GEO website under the accession number GSE80970 (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE80970>). The files “GSE80970_MSACF_raw_methylated.txt.gz”, “GSE80970_MSACF_raw_unmethylated.txt.gz”, “GSE80970_MSACF_raw_pvals.txt.gz” were downloaded and processed with minfi package. Probes with NAs (defined by $P > 0.01$) were discarded. The filtered data was subsequently normalized with BMIQ, resulting in a normalized data matrix for 400811 probes and 286 samples.

Horvath

The 450k dataset from Horvath et al. [18] profiled DNAm in human frontal lobe (FL), temporal lobe (TL), parietal lobe (PL), motor cortex (MC), sensory cortex (SC), visual cortex (VC), cingulate gyrus (CG) from controls and AD cases was obtained from the NCBI GEO website under the accession number GSE72778 (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE72778>). The file “GSE72778_datSignal.csv.gz” was downloaded and processed with minfi package. Probes with NAs (defined by $P > 0.01$) were discarded. The filtered data was subsequently normalized with BMIQ, resulting in a normalized data matrix for 420402 probes and 300 samples.

Gasparoni

The 450k dataset from Gasparoni et al. [2] profiled DNAm in human frontal lobe (FL), temporal lobe (TL) from controls and AD cases was obtained from the NCBI GEO website under the accession number GSE66351 (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE66351>). The file “GSE66351_RAW.tar” which contains the IDAT files was downloaded and processed with minfi package. Probes with NAs (defined by $P > 0.01$) were discarded. The filtered data was subsequently normalized with BMIQ, resulting in a normalized data matrix for 474413 probes and 128 samples.

Watson

The 450k dataset from Watson et al. [19] profiled DNAm in superior temporal gyrus (STG) from controls and AD cases was obtained from the NCBI GEO website under the accession number GSE76105 (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE76105>). The file “GSE76105_Raw_Data.txt.gz” was downloaded and processed with minfi package. Probes with NAs (defined by $P > 0.01$) were discarded. The filtered data was subsequently normalized with BMIQ, resulting in a normalized data matrix for 479076 probes and 68 samples.

Brokaw

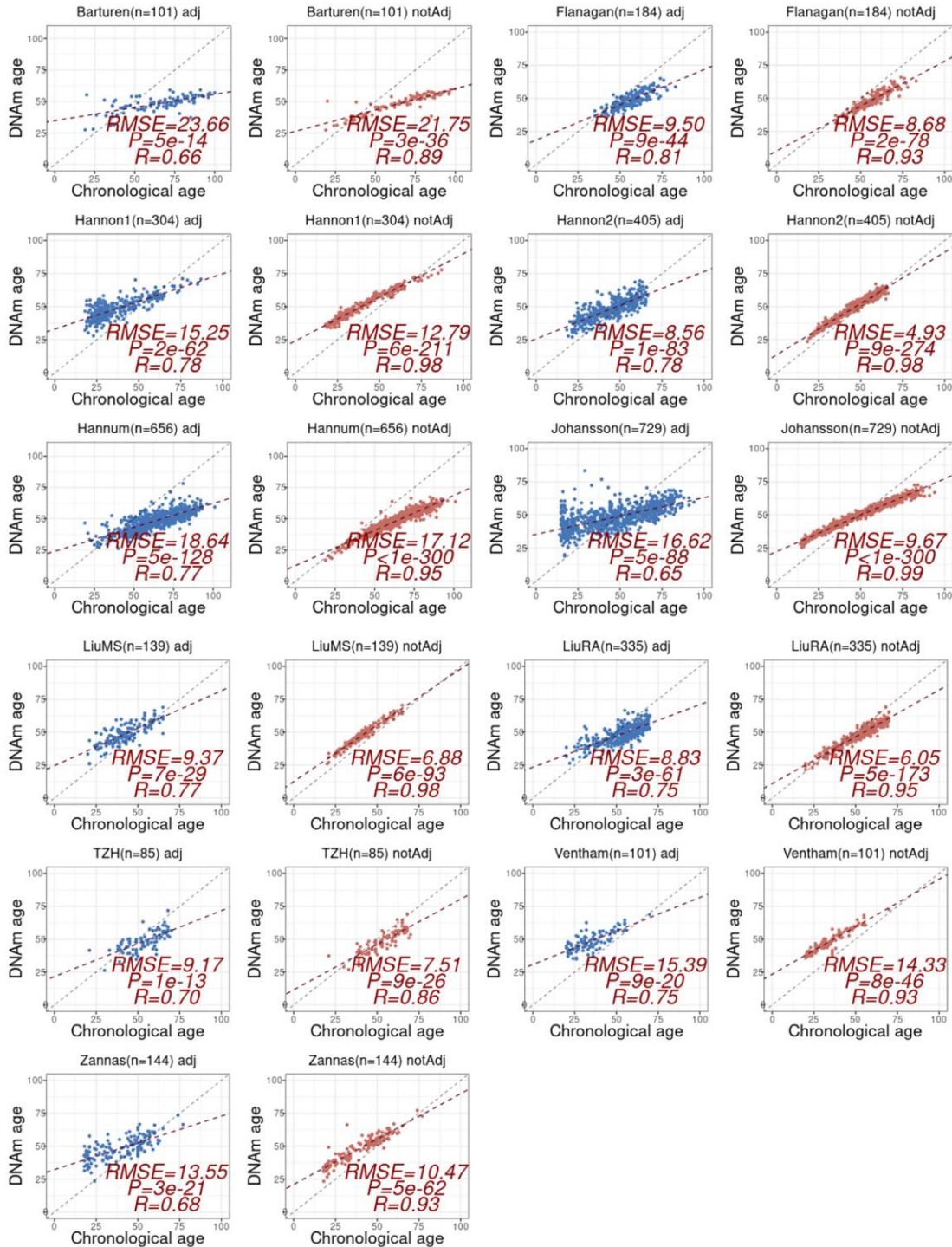
The 450k dataset from Brokaw et al. [20] profiled DNAm in middle temporal gyrus (MTG) and the cerebellum (CRB) from controls and AD cases was obtained from the NCBI GEO website under the accession number GSE134379 (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE134379>). The file “GSE134379_RAW.tar” which contains the IDAT files was downloaded and processed with minfi package. Probes with NAs (defined by $P > 0.01$) were discarded. The filtered data was subsequently normalized with BMIQ, resulting in a normalized data matrix for 370843 probes and 808 samples.

Supplementary References

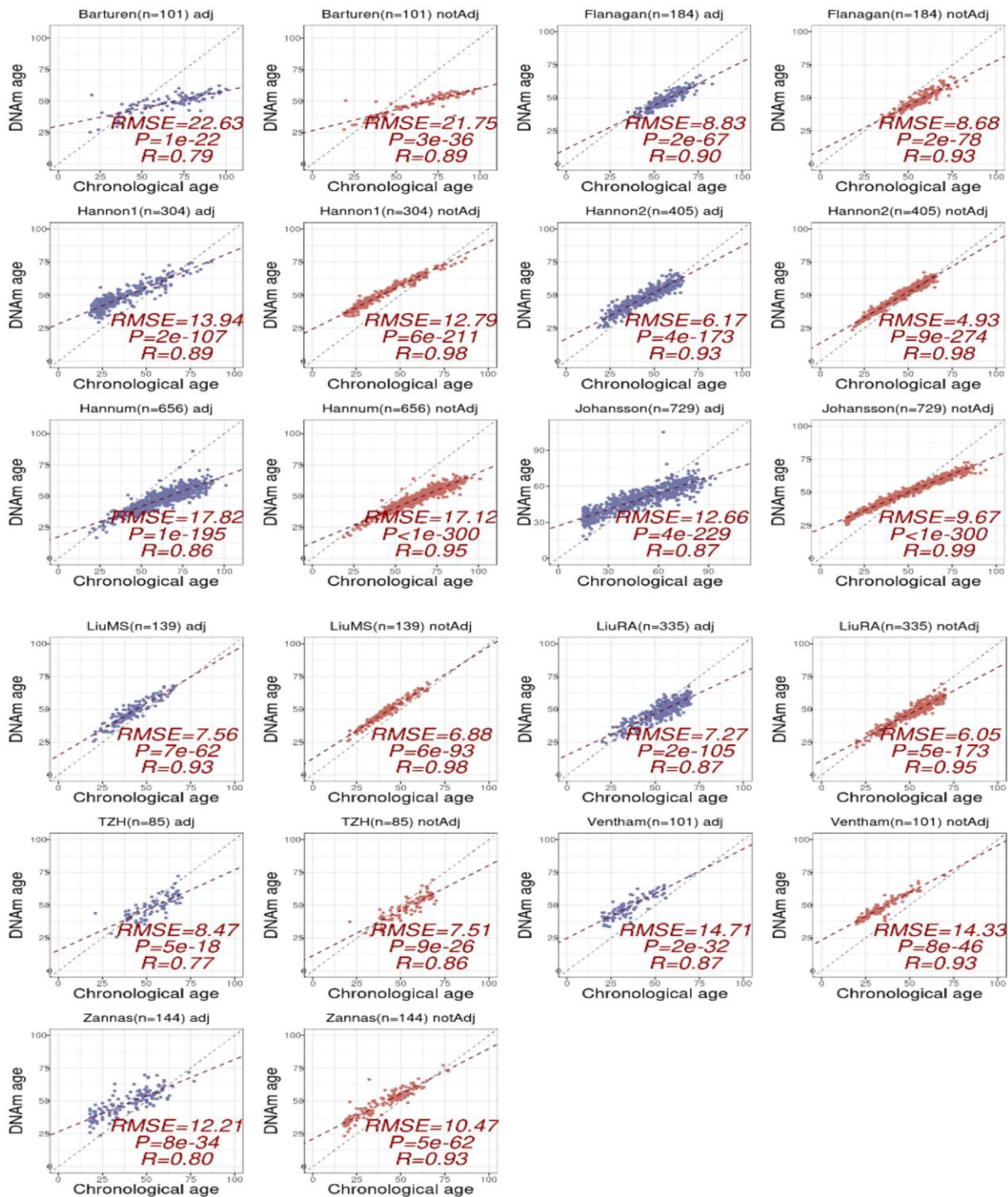
1. Pai S, Li P, Killinger B, Marshall L, Jia P, Liao J, Petronis A, Szabó PE, Labrie V. Differential methylation of enhancer at IGF2 is associated with abnormal dopamine synthesis in major psychosis. *Nat Commun.* 2019; 10:2046. <https://doi.org/10.1038/s41467-019-09786-7> PMID:[31053723](https://pubmed.ncbi.nlm.nih.gov/31053723/)
2. Gasparoni G, Bultmann S, Lutsik P, Kraus TFJ, Sordon S, Vlcek J, Dietinger V, Steinmaurer M, Haider M, Mulholland CB, Arzberger T, Roeber S, Riemenschneider M, et al. DNA methylation analysis on purified neurons and glia dissects age and Alzheimer's disease-specific changes in the human cortex. *Epigenetics Chromatin.* 2018; 11:41. <https://doi.org/10.1186/s13072-018-0211-3> PMID:[30045751](https://pubmed.ncbi.nlm.nih.gov/30045751/)
3. Kozlenkov A, Jaffe AE, Timashpolsky A, Apontes P, Rudchenko S, Barbu M, Byne W, Hurd YL, Horvath S, Dracheva S. DNA Methylation Profiling of Human Prefrontal Cortex Neurons in Heroin Users Shows Significant Difference between Genomic Contexts of Hyper- and Hypomethylation and a Younger Epigenetic Age. *Genes (Basel).* 2017; 8:152. <https://doi.org/10.3390/genes8060152> PMID:[28556790](https://pubmed.ncbi.nlm.nih.gov/28556790/)
4. Guintivano J, Aryee MJ, Kaminsky ZA. A cell epigenotype specific model for the correction of brain cellular heterogeneity bias and its application to age, brain region and major depression. *Epigenetics.* 2013; 8:290–302. <https://doi.org/10.4161/epi.23924> PMID:[23426267](https://pubmed.ncbi.nlm.nih.gov/23426267/)
5. Hannon E, Dempster EL, Davies JP, Chioza B, Blake GET, Burrage J, Policicchio S, Franklin A, Walker EM, Bamford RA, Schalkwyk LC, Mill J. Quantifying the proportion of different cell types in the human cortex using DNA methylation profiles. *BMC Biol.* 2024; 22:17. <https://doi.org/10.1186/s12915-024-01827-y> PMID:[38273288](https://pubmed.ncbi.nlm.nih.gov/38273288/)
6. de Witte LD, Wang Z, Snijders GLJL, Mendeleev N, Liu Q, Sneebouer MAM, Boks MPM, Ge Y, Haghighi F. Contribution of Age, Brain Region, Mood Disorder Pathology, and Interindividual Factors on the Methyloome of Human Microglia. *Biol Psychiatry.* 2022; 91:572–81. <https://doi.org/10.1016/j.biopsych.2021.10.020> PMID:[35027166](https://pubmed.ncbi.nlm.nih.gov/35027166/)
7. Murphy TM, Crawford B, Dempster EL, Hannon E, Burrage J, Turecki G, Kaminsky Z, Mill J. Methyloomic profiling of cortex samples from completed suicide cases implicates a role for PSORS1C3 in major depression and suicide. *Transl Psychiatry.* 2017; 7:e989. <https://doi.org/10.1038/tp.2016.249> PMID:[28045465](https://pubmed.ncbi.nlm.nih.gov/28045465/)
8. Rydbirk R, Folke J, Busato F, Roché E, Chauhan AS, Løkkegaard A, Hejl AM, Bode M, Blaabjerg M, Møller M, Danielsen EH, Brudek T, Pakkenberg B, et al. Epigenetic modulation of AREL1 and increased HLA expression in brains of multiple system atrophy patients. *Acta Neuropathol Commun.* 2020; 8:29. <https://doi.org/10.1186/s40478-020-00908-7> PMID:[32151281](https://pubmed.ncbi.nlm.nih.gov/32151281/)

9. Torabi Moghadam B, Etemadikhah M, Rajkowska G, Stockmeier C, Grabherr M, Komorowski J, Feuk L, Carlström EL. Analyzing DNA methylation patterns in subjects diagnosed with schizophrenia using machine learning methods. *J Psychiatr Res*. 2019; 114:41–7.
<https://doi.org/10.1016/j.jpsychires.2019.04.001>
PMID:[31022588](https://pubmed.ncbi.nlm.nih.gov/31022588/)
10. Xu H, Wang F, Liu Y, Yu Y, Gelernter J, Zhang H. Sex-biased methylome and transcriptome in human prefrontal cortex. *Hum Mol Genet*. 2014; 23:1260–70.
<https://doi.org/10.1093/hmg/ddt516>
PMID:[24163133](https://pubmed.ncbi.nlm.nih.gov/24163133/)
11. Huynh JL, Garg P, Thin TH, Yoo S, Dutta R, Trapp BD, Haroutunian V, Zhu J, Donovan MJ, Sharp AJ, Casaccia P. Epigenome-wide differences in pathology-free regions of multiple sclerosis-affected brains. *Nat Neurosci*. 2014; 17:121–30.
<https://doi.org/10.1038/nn.3588>
PMID:[24270187](https://pubmed.ncbi.nlm.nih.gov/24270187/)
12. Viana J, Hannon E, Dempster E, Pidsley R, Macdonald R, Knox O, Spiers H, Troakes C, Al-Saraj S, Turecki G, Schalkwyk LC, Mill J. Schizophrenia-associated methylomic variation: molecular signatures of disease and polygenic risk burden across multiple brain regions. *Hum Mol Genet*. 2017; 26:210–25.
<https://doi.org/10.1093/hmg/ddw373>
PMID:[28011714](https://pubmed.ncbi.nlm.nih.gov/28011714/)
13. Markunas CA, Semick SA, Quach BC, Tao R, Deep-Soboslay A, Carnes MU, Bierut LJ, Hyde TM, Kleinman JE, Johnson EO, Jaffe AE, Hancock DB. Genome-wide DNA methylation differences in nucleus accumbens of smokers vs. nonsmokers. *Neuropsychopharmacology*. 2021; 46:554–60.
<https://doi.org/10.1038/s41386-020-0782-0>
PMID:[32731254](https://pubmed.ncbi.nlm.nih.gov/32731254/)
14. Policicchio S, Washer S, Viana J, Iatrou A, Burrage J, Hannon E, Turecki G, Kaminsky Z, Mill J, Dempster EL, Murphy TM. Genome-wide DNA methylation meta-analysis in the brains of suicide completers. *Transl Psychiatry*. 2020; 10:69.
<https://doi.org/10.1038/s41398-020-0752-7>
PMID:[32075955](https://pubmed.ncbi.nlm.nih.gov/32075955/)
15. Semick SA, Bharadwaj RA, Collado-Torres L, Tao R, Shin JH, Deep-Soboslay A, Weiss JR, Weinberger DR, Hyde TM, Kleinman JE, Jaffe AE, Mattay VS. Integrated DNA methylation and gene expression profiling across multiple brain regions implicate novel genes in Alzheimer's disease. *Acta Neuropathol*. 2019; 137:557–69.
<https://doi.org/10.1007/s00401-019-01966-5>
PMID:[30712078](https://pubmed.ncbi.nlm.nih.gov/30712078/)
16. Pidsley R, Y Wong CC, Volta M, Lunnon K, Mill J, Schalkwyk LC. A data-driven approach to preprocessing Illumina 450K methylation array data. *BMC Genomics*. 2013; 14:293.
<https://doi.org/10.1186/1471-2164-14-293>
PMID:[23631413](https://pubmed.ncbi.nlm.nih.gov/23631413/)
17. Smith RG, Hannon E, De Jager PL, Chibnik L, Lott SJ, Condliffe D, Smith AR, Haroutunian V, Troakes C, Al-Sarraj S, Bennett DA, Powell J, Lovestone S, et al. Elevated DNA methylation across a 48-kb region spanning the HOXA gene cluster is associated with Alzheimer's disease neuropathology. *Alzheimers Dement*. 2018; 14:1580–8.
<https://doi.org/10.1016/j.jalz.2018.01.017>
PMID:[29550519](https://pubmed.ncbi.nlm.nih.gov/29550519/)
18. Horvath S, Langfelder P, Kwak S, Aaronson J, Rosinski J, Vogt TF, Eszes M, Faull RL, Curtis MA, Waldvogel HJ, Choi OW, Tung S, Vinters HV, et al. Huntington's disease accelerates epigenetic aging of human brain and disrupts DNA methylation levels. *Aging (Albany NY)*. 2016; 8:1485–512.
<https://doi.org/10.18632/aging.101005>
PMID:[27479945](https://pubmed.ncbi.nlm.nih.gov/27479945/)
19. Watson CT, Roussos P, Garg P, Ho DJ, Azam N, Katsel PL, Haroutunian V, Sharp AJ. Genome-wide DNA methylation profiling in the superior temporal gyrus reveals epigenetic signatures associated with Alzheimer's disease. *Genome Med*. 2016; 8:5.
<https://doi.org/10.1186/s13073-015-0258-8>
PMID:[26803900](https://pubmed.ncbi.nlm.nih.gov/26803900/)
20. Brokaw DL, Piras IS, Mastroeni D, Weisenberger DJ, Nolz J, Delvaux E, Serrano GE, Beach TG, Huentelman MJ, Coleman PD. Cell death and survival pathways in Alzheimer's disease: an integrative hypothesis testing approach utilizing -omic data sets. *Neurobiol Aging*. 2020; 95:15–25.
<https://doi.org/10.1016/j.neurobiolaging.2020.06.022>
PMID:[32745806](https://pubmed.ncbi.nlm.nih.gov/32745806/)

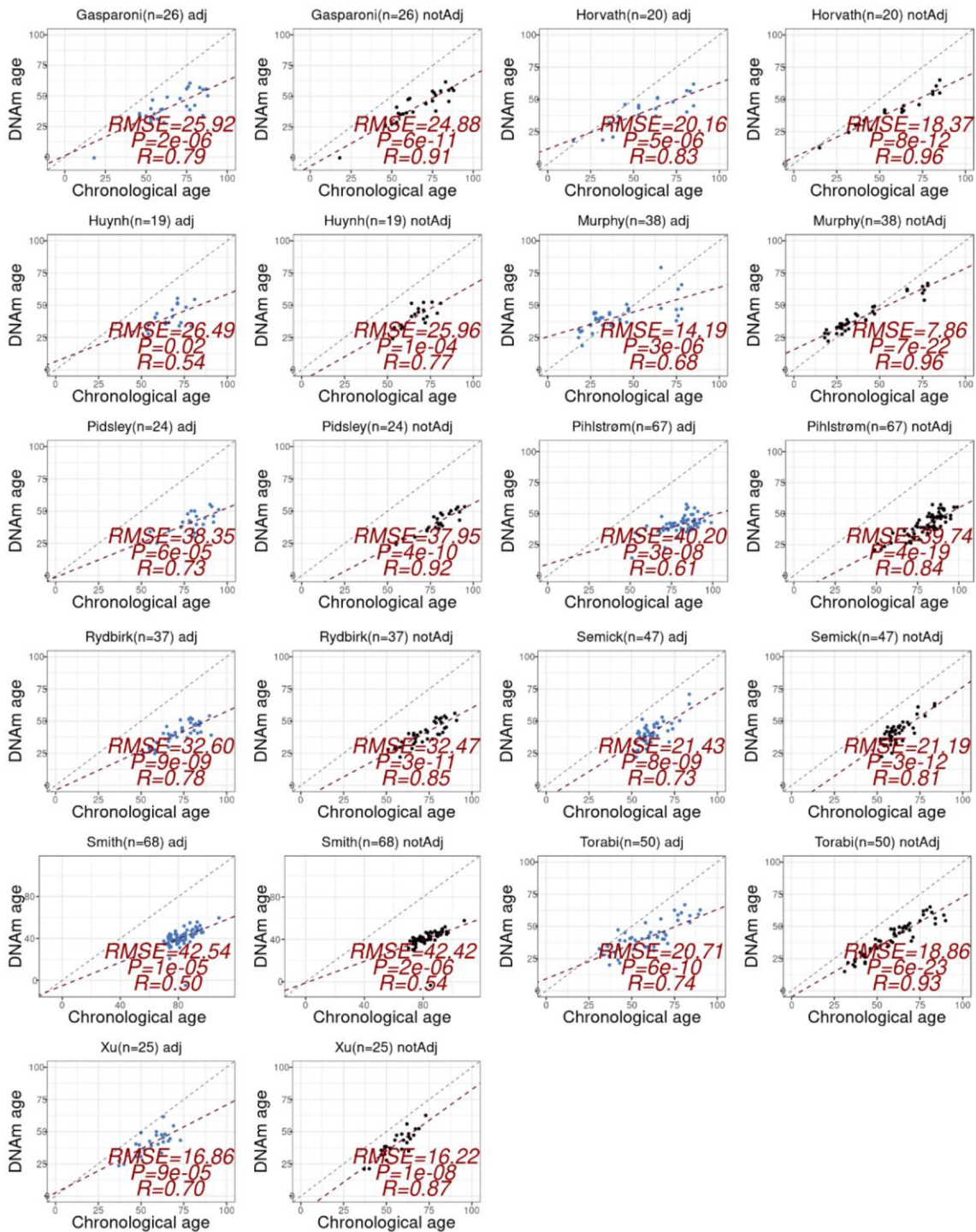
Supplementary Figures



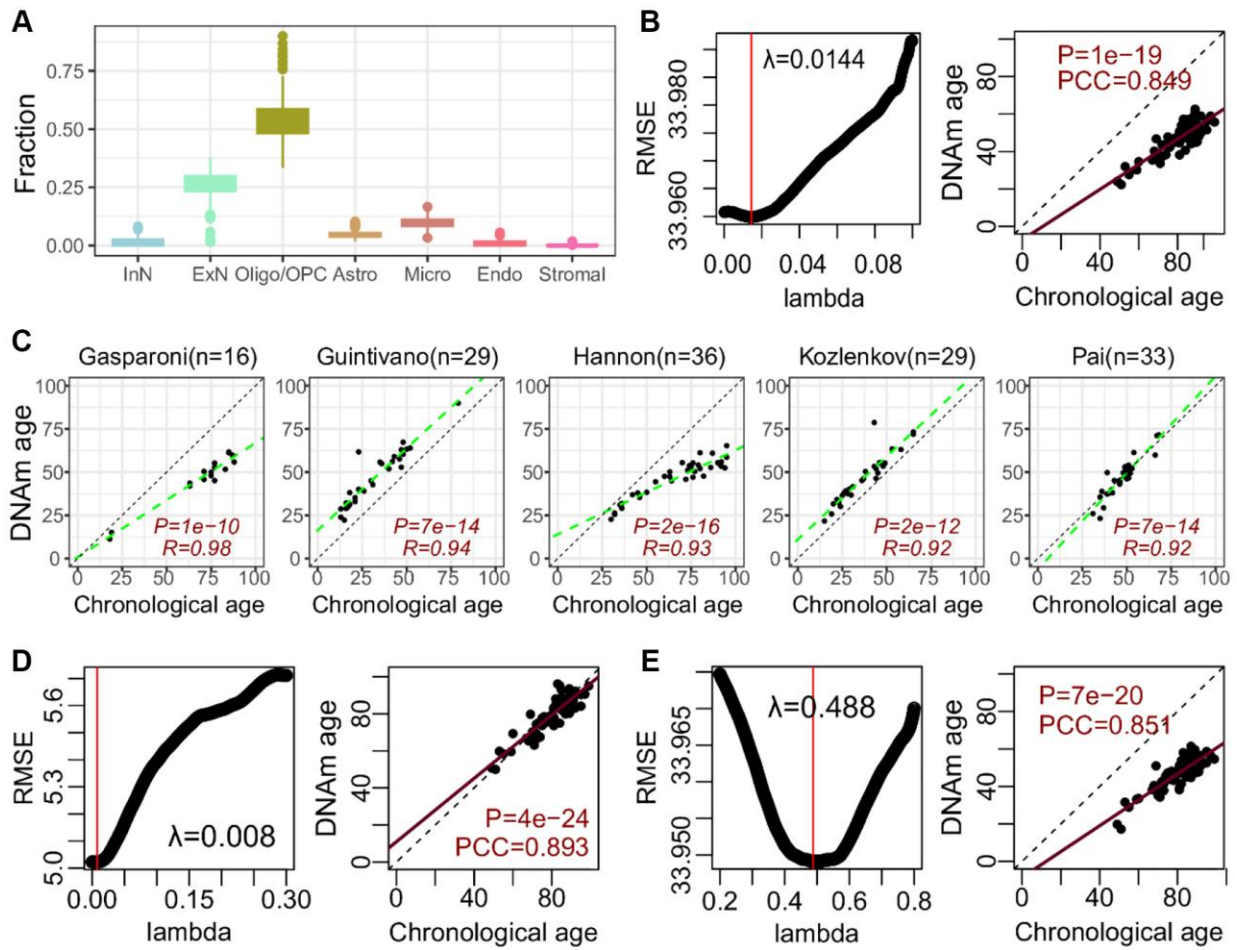
Supplementary Figure 1. Predicted DNAmAge vs. chronological age for CTH adjusted (12 immune-cell types) and unadjusted clocks in 11 whole blood cohorts. Each scatterplot is labeled by cohort, cohort-size and whether it is adjusted or unadjusted clock. Root Mean Square Error (RMSE), two-tailed P -value of a linear regression and R (Pearson Correlation Coefficient) value is given. Adjusted clock was adjusted for variations in naïve + memory B-cells, naïve + memory CD4T, naïve + memory CD8T-cells, T-regulatory cells, NK-cells, monocytes, neutrophils, eosinophils and basophils.



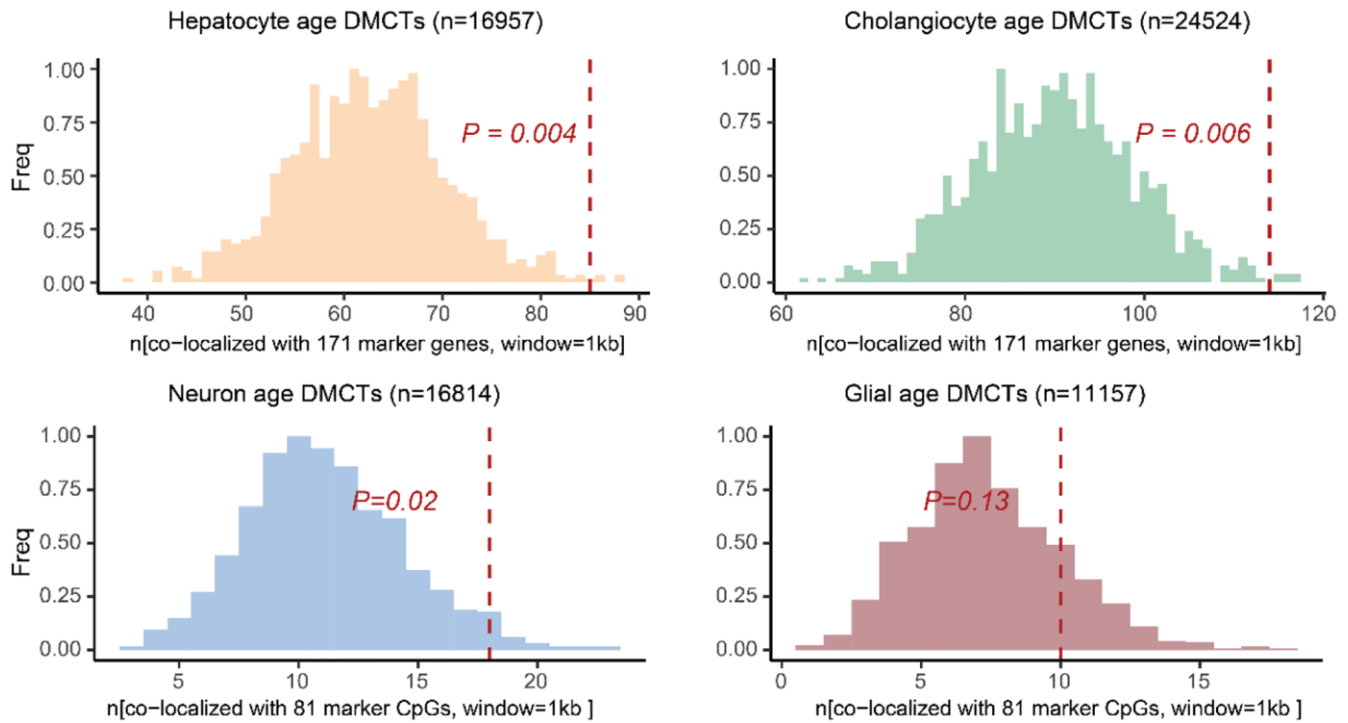
Supplementary Figure 2. Predicted DNAmAge vs. chronological age for CTH adjusted (9 immune-cell types) and unadjusted clocks in 11 whole blood cohorts. Each scatterplot is labeled by cohort, cohort-size and whether it is adjusted or unadjusted clock. Root Mean Square Error (RMSE), two-tailed *P*-value of a linear regression and *R* (Pearson Correlation Coefficient) value is given. Adjusted clock was adjusted for variations in B-cells, CD4T and CD8T-cells, T-regulatory cells, NK-cells, monocytes, neutrophils, eosinophils and basophils.



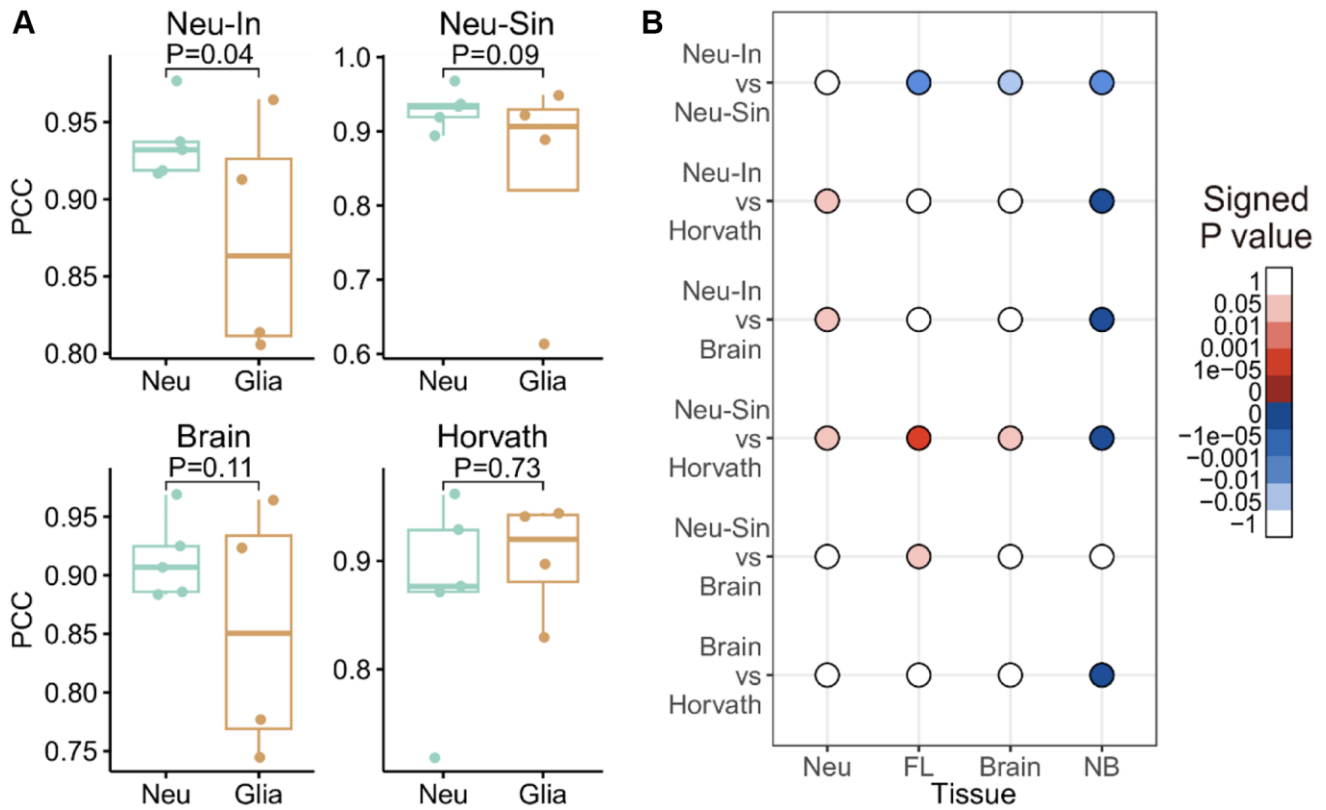
Supplementary Figure 3. Predicted DNAmAge vs. chronological age for CTH adjusted (7 brain cell types) and unadjusted clocks in 11 brain-tissue cohorts. Each scatterplot is labeled by cohort, cohort-size and whether it is adjusted or unadjusted clock. Root Mean Square Error (RMSE), two-tailed *P*-value of a linear regression and R (Pearson Correlation Coefficient) value is given. Adjusted clock was adjusted for fractions of excitatory and inhibitory neurons, astrocytes, endothelial cells, microglia, oligodendrocytes+oligodendrocyte progenitor cells and other stromal cells.



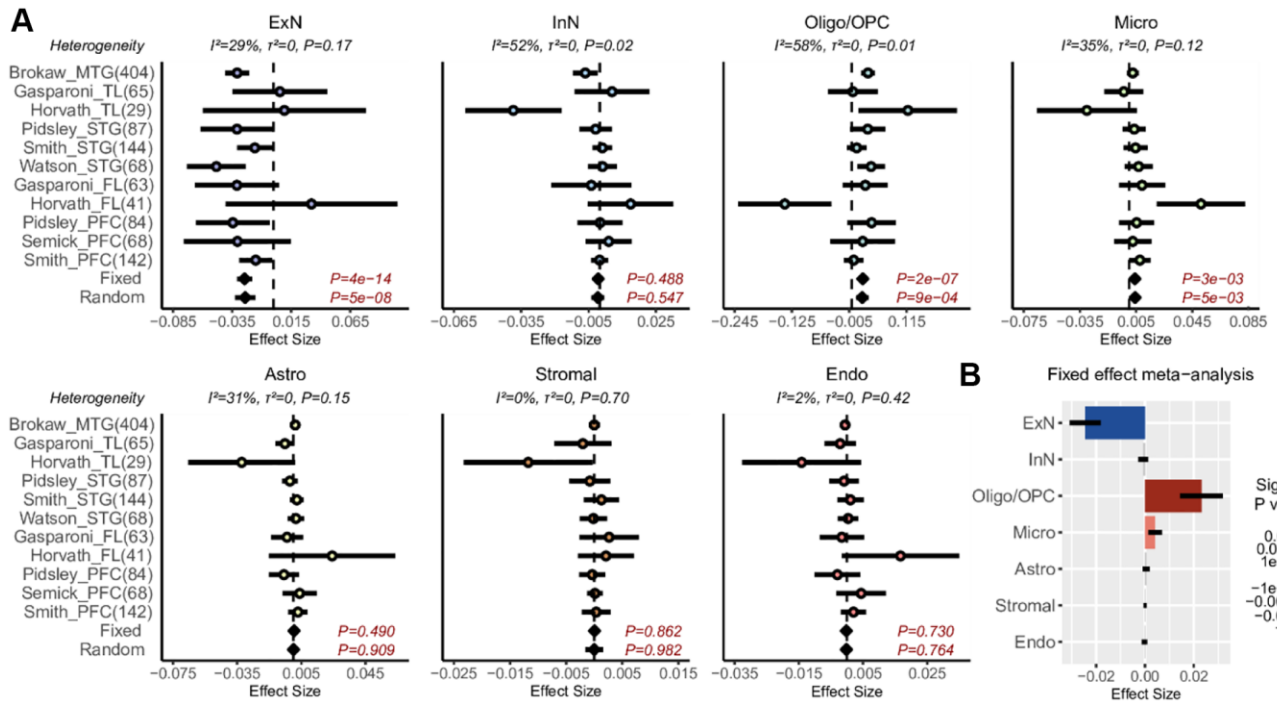
Supplementary Figure 4. Construction and validation of Neu-In, Neu-Sin and brain-clocks. (A) Boxplot shows the estimated fractions of 7 brain cell types in the Jaffe cohort, as estimated using HiBED. InN: inhibitory neuron; ExN: excitatory neuron; Oligo/OPC: oligodendrocyte and oligodendrocyte precursor cell; Micro: microglia; Endo: endothelial cell; Stromal: stromal cell. (B) Optimization and validation of Neu-In clock. Panel on the left shows how RMSE changes with the penalty parameter lambda. Red vertical line represents the optimal lambda value. Panel on the right shows the validation of the optimal clock in Philstrom cohort. PCC-value and associated two-tailed P-value are given. (C) Scatterplots of the chronological age (x-axis) vs. predicted DNAm-age (y-axis) from the Neu-In clock in 5 sorted neuron cohorts. Cohort sizes are indicated above each plot. R-value and associated two-tailed P-value from a linear regression test are given. (D) Same as (B), but for the Neu-Sin clock. (E) As (B) but for the BrainClock.



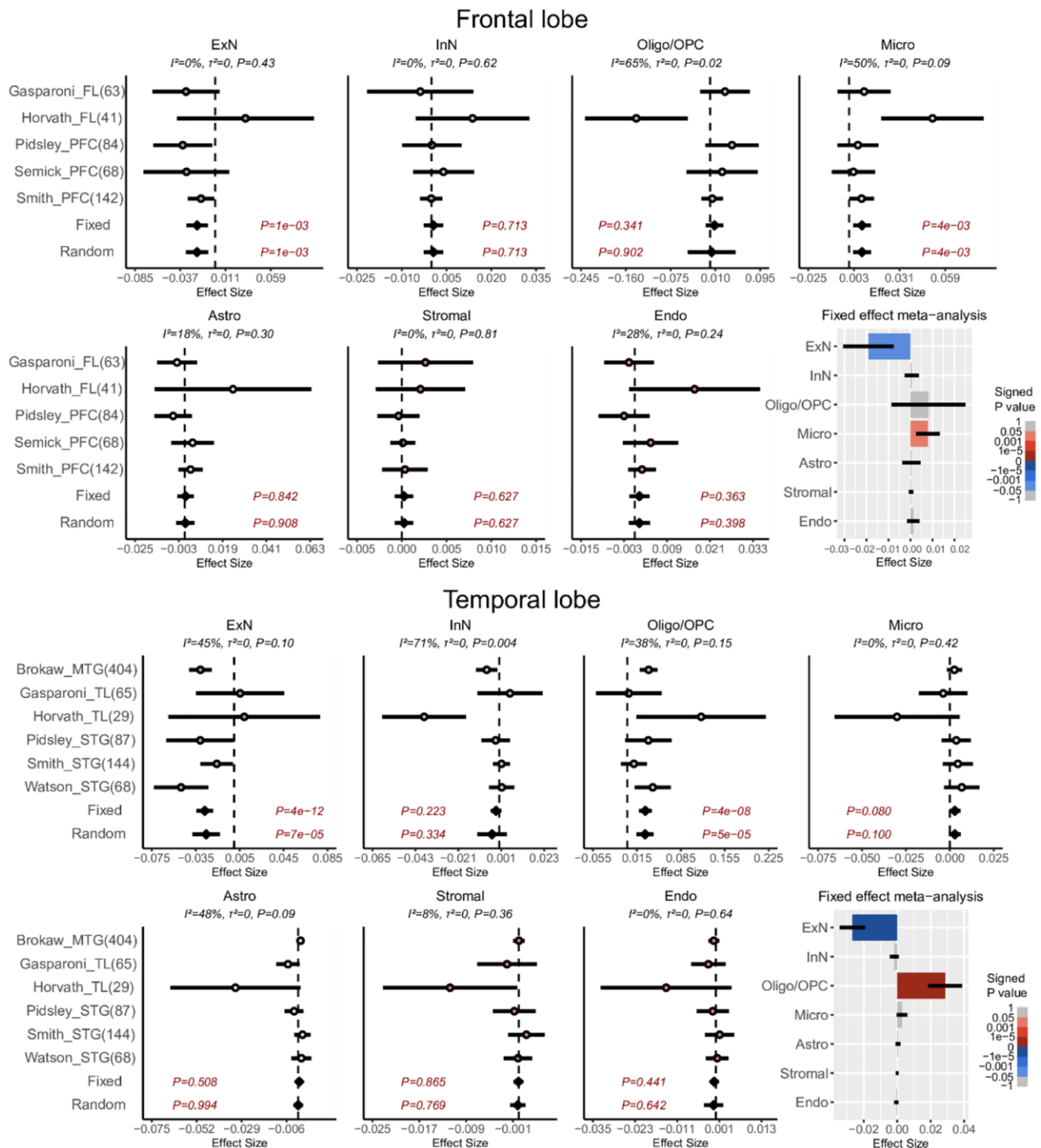
Supplementary Figure 5. Co-localization of age-DMCTs with DNAm reference matrix CpGs. For four classes of age-DMCTs (hepatocyte, cholangiocyte for liver and neuron, glia for brain) we display their overlap with the CpGs in the liver and brain DNAm reference matrices, respectively. We used a 1kb window, so that if an age-DMCT is within 1kb of a corresponding marker CpG in the DNAm reference matrix, this was counted as a “hit”. We then compared the observed overlap (red vertical dashed line) with the overlap expected had we chosen an exact same number of “age-DMCTs” randomly from the array. The latter randomization was done 1000 times, yielding a null distribution that we can compare to, to hence derive an empirical *P*-value.



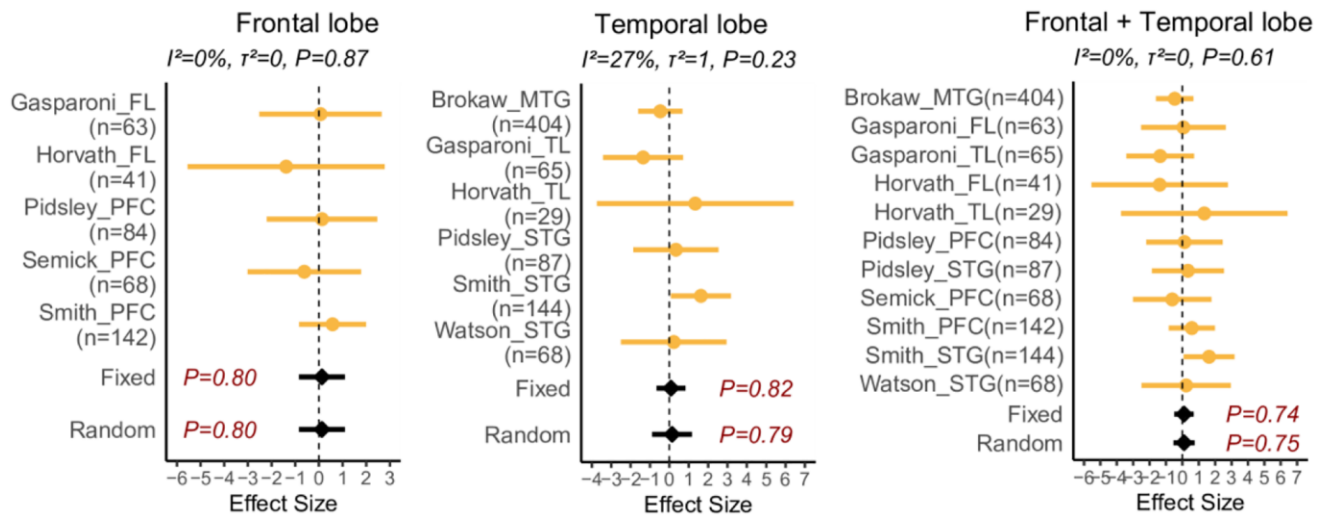
Supplementary Figure 6. Specificity of neuron-clocks. (A) Neuron specific clocks show specificity to neurons according to a weighted linear regression of PCC values against cell type. Boxplots compare PCC values between predicted age and chronological age (y-axis) in cohorts from 2 cell types (x-axis) for the Neu-In clock, Neu-Sin clock, Brain clock and Horvath clock. There are 5 sorted neuron cohorts and 4 glia cell cohorts in total. P -values are from a linear regression of PCC against cell-type (Neu vs. Glia) weighted by cohort size. (B) Comparison of Neu-In, Neu-Sin, Brain and Horvath clock in sorted neuron cohorts (Neu), bulk frontal lobe cohorts (FL), all brain tissue cohorts (brain) and non-brain-tissue cohorts (NB). Balloon plot shows signed P -values from a one-tailed paired Wilcoxon test comparing PCC values of one clock to another. PCC-values represent the correlation between predicted DNAm-Age and chronological age, as obtained with a given clock. Because there are 4 clocks (Neu-In clock, Neu-Sin clock, Brain clock and Horvath clock), there are 6 pairwise comparisons (y-axis). x-axis labels the datasets over which the P -values were estimated. Positive signed P -values indicate higher PCC values for the first clock in the pairwise comparison (e.g. for “Neu-In vs. Neu-Sin”, a positive signed P -value means PCC values of Neu-In are bigger than Neu-Sin).



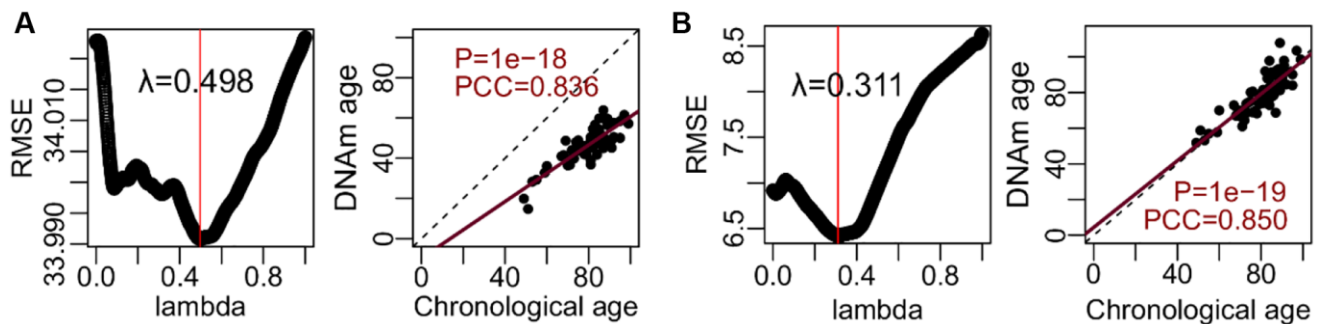
Supplementary Figure 7. Meta-analysis of brain cell-type fraction associations with Alzheimer’s Disease (AD). (A) Forest plots of associations between cell-type fractions and AD, for 7 brain cell-types (Abbreviations: ExN: excitatory neurons; InN: inhibitory neurons; Oligo/OPC: oligodendrocytes/oligo precursor cells; Micro: microglia; Astro: astrocytes, Stromal; Endo: endothelial) across 11 independent DNAm datasets. Cohorts are labeled by name and brain region. Brain regions: Middle and Superior temporal gyrus (MTG and STG), temporal lobe (TL), frontal lobe (FL) and prefrontal cortex (PFC). Number of samples in each cohort is given in brackets. P -values from a Fixed and Random Effects models are also given. Heterogeneity statistics and P -values of heterogeneity are given above each plot. (B) Effect sizes and P -values from the Fixed Effects meta-analysis model.



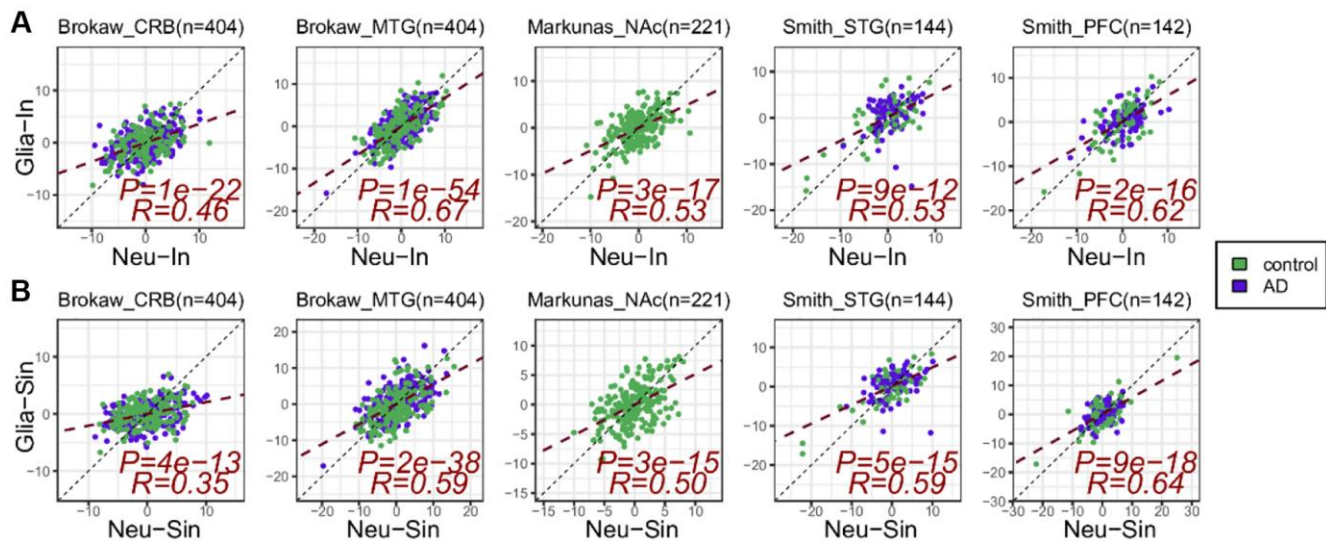
Supplementary Figure 8. Meta-analysis of brain cell-type fraction associations with Alzheimer’s Disease (AD) stratified by brain region. Top panel: Forest plots of associations between cell-type fractions and AD, for 7 brain cell-types (Abbreviations: ExN: excitatory neurons; InN: inhibitory neurons; Oligo/OPC: oligodendrocytes/oligo precursor cells; Micro: microglia; Astro: astrocytes, Stromal; Endo: endothelial) across 6 independent frontal lobe DNAm datasets. Cohorts are labeled by name and brain region/subregion. Brain regions: Middle and Superior temporal gyrus (MTG and STG), temporal lobe (TL), frontal lobe (FL) and prefrontal cortex (PFC). Number of samples in each cohort is given in brackets. P -values from a Fixed and Random Effects models are also given. Heterogeneity statistics and P -values of heterogeneity are given above each plot. Effect sizes and P -values from the Fixed Effects meta-analysis model are given separately. Bottom panel: as top, but for a meta-analysis over 5 independent temporal lobe DNAm datasets.



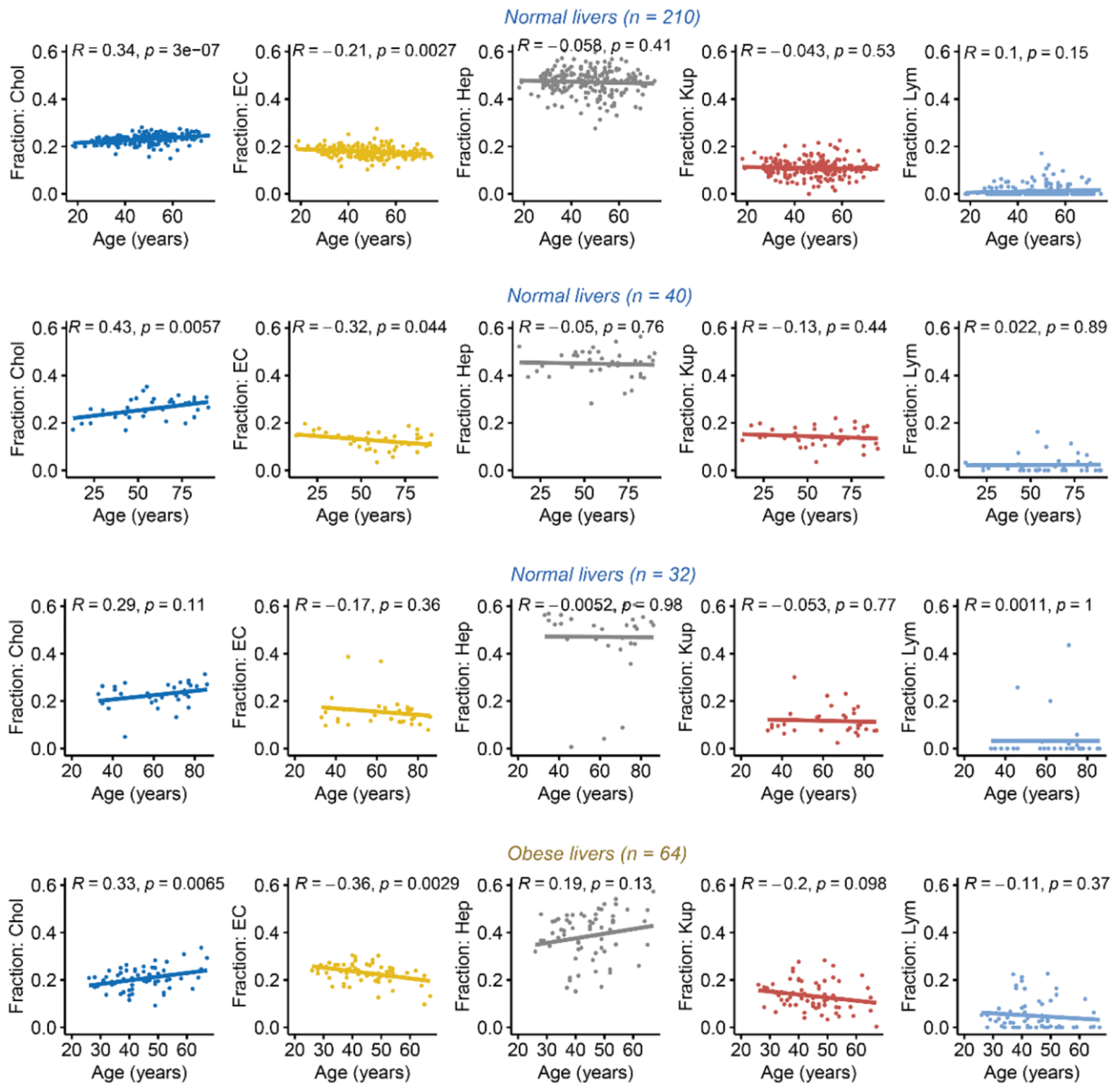
Supplementary Figure 9. Meta-analysis of Horvath clock associations with Alzheimer's Disease (AD) stratified by brain region. Top panel: Forest plots of associations of Horvath's clock age-acceleration adjusted for age, sex and brain cell-type fractions with AD. Number of samples in each study is given in brackets. P -value of heterogeneity is given above each panel. P -values of a fixed and random effect models are given.



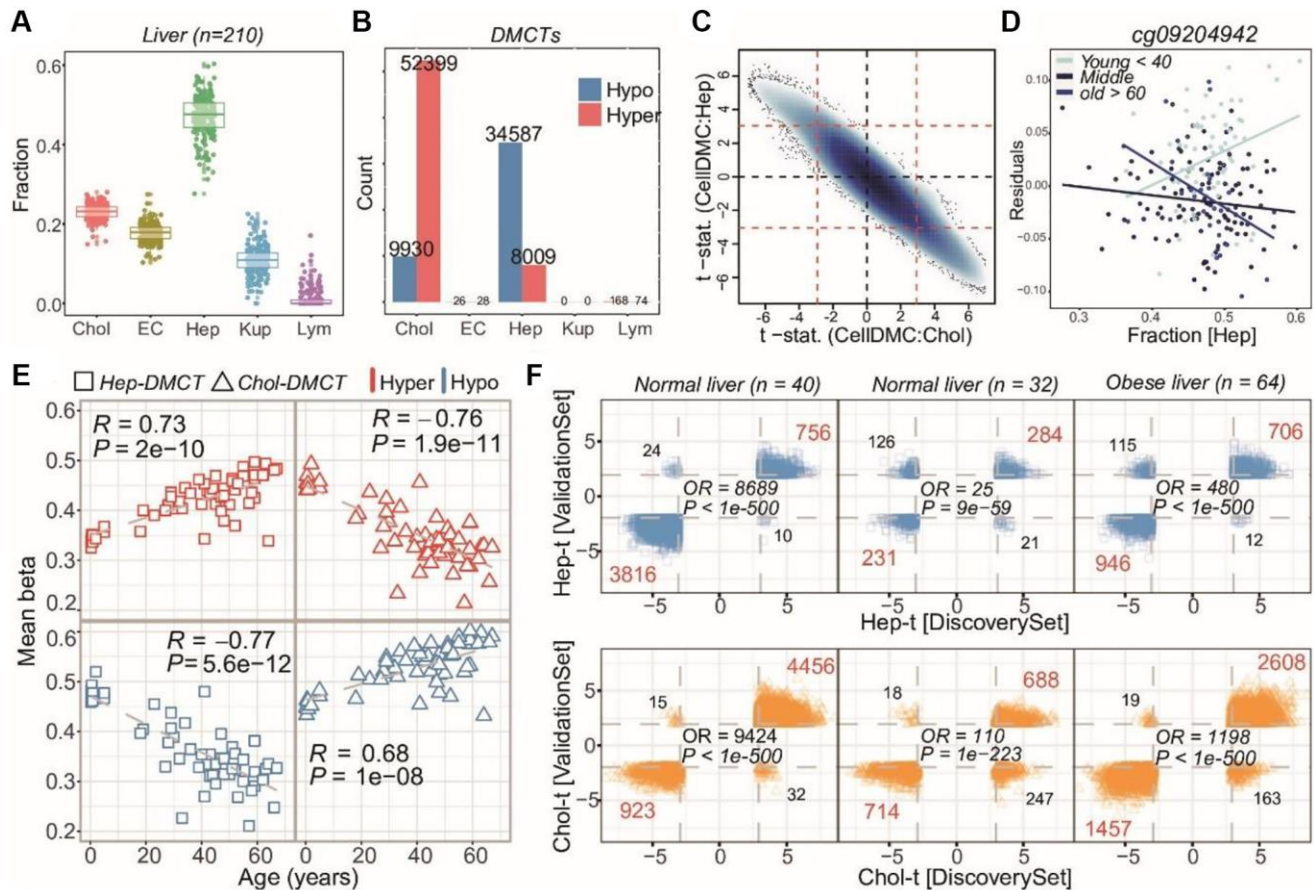
Supplementary Figure 10. Optimization and validation of Glia-In and Glia-Sin clocks. (A) Panel on the left shows how RMSE changes with the penalty parameter λ for the Glia-In clock. Red vertical line represents the optimal λ value. Panel on the right shows the validation of the optimal clock in Philstrom cohort. PCC-value and associated linear regression two-tailed P -value are given. (B) As (A) but for the glia semi-intrinsic (Glia-Sin) clock.



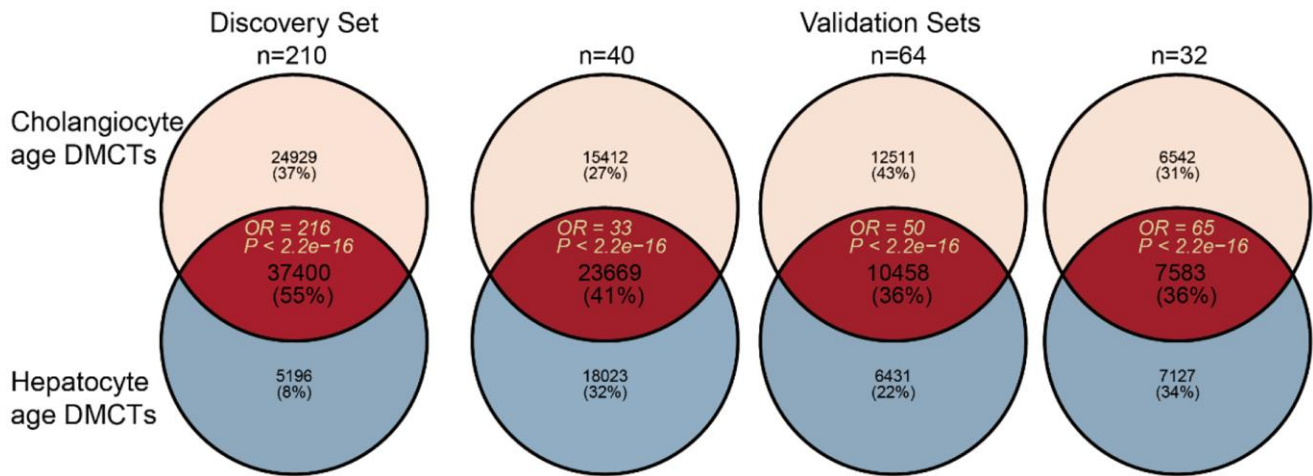
Supplementary Figure 11. Correlation of neuron with glia clocks in AD cohorts. (A) For the 5 largest brain-tissue cohorts with AD cases and controls, we display a scatterplot of the age-acceleration (extrinsic age acceleration-EAA) of the Glia-In clock (y-axis) vs. the age-acceleration (EAA) of the Neu-In clock (x-axis). The R-values and two-tailed P -values from a linear regression are given. The number of samples in cohort is indicated above the plot. (B) As (A) but for the Glia-Sin and Neu-Sin clocks. In the two Smith datasets and for both (A, B) an outlier control sample with very negative age-acceleration has been capped at a value given by the minimum across all other samples.



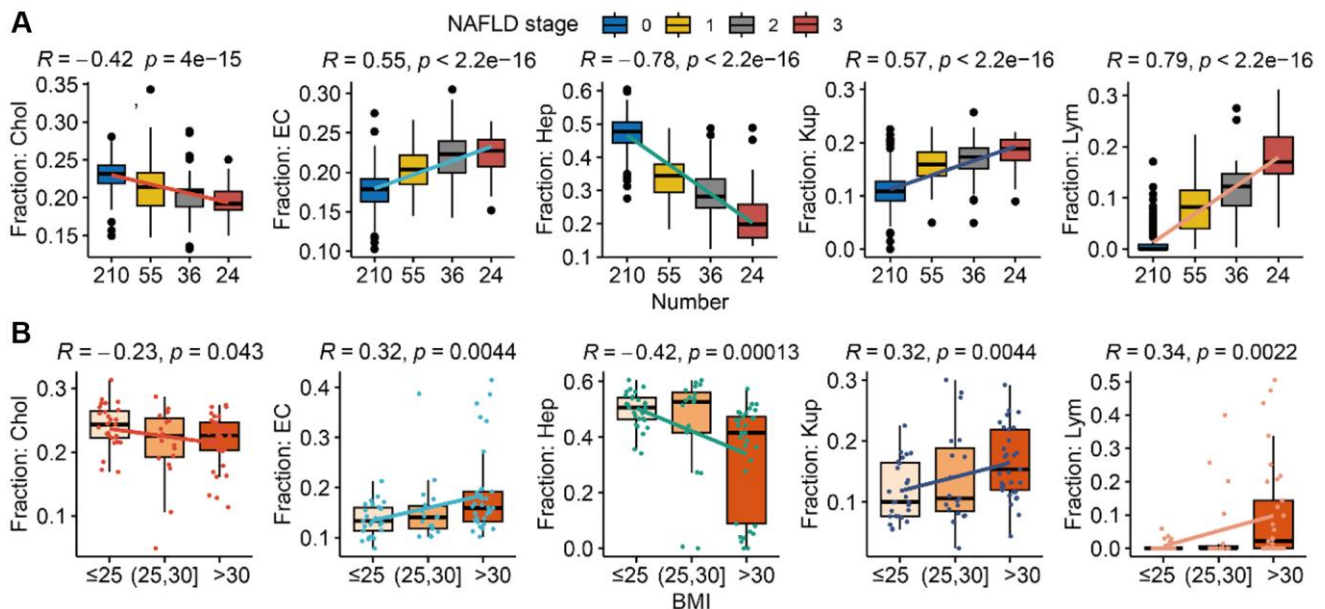
Supplementary Figure 12. Liver cell-type fraction variations with age. For four independent DNAm liver datasets, we plot the fractions of five liver cell-types (as estimated using EpiSCORE) as a function of age. R-values and P-values (two-tailed linear regression) of statistical significance are given. From left to right, cell-types are Cholangiocytes (Chol), endothelial cells (EC), Hepatocytes (Hep), Kupffer macrophages (Kup) and Lymphocytes (Lym).



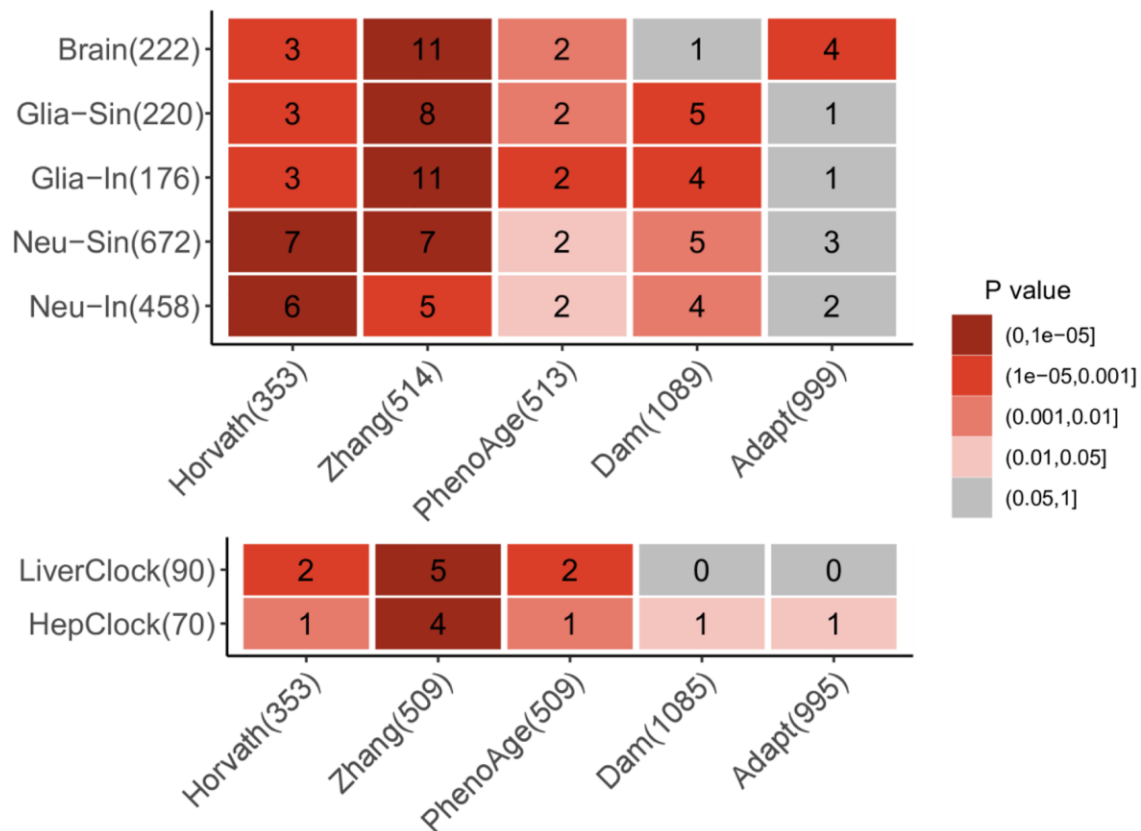
Supplementary Figure 13. Identification and validation of hepatocyte-specific age-DMCs. (A) Boxplots of estimated cell-type fractions (y-axis) using EpiSCORE's DNAm liver reference matrix of 5 liver cell-types (Abbreviations: Chol: cholangiocytes; EC: endothelial cells; Hep: hepatocytes; Kup: Kupffer macrophages; Lym: lymphocytes, x-axis) in a series of 210 normal liver samples. (B) Barplots displaying the number of hypo and hypermethylated age-DMCTs for each cell-type, as inferred with CellDMC using the estimated fractions in (A). (C) Smoothed scatterplot of CellDMC t-statistics for age-DMCTs in hepatocytes (y-axis) vs. cholangiocytes (x-axis). Red dashed lines indicate the FDR = 0.05 significance level. (D) Example of a hepatocyte-specific age-DMCT with y-axis labeling the residual after adjusting for cell-type fractions and x-axis labeling the hepatocyte fraction of the sample, with samples colored by age-group. (E) Scatterplot of average DNAm levels (Mean Beta, y-axis) of four sets of age-associated DMCTs against chronological age (x-axis) in an independent DNAm dataset comprising 55 hepatocyte cultures. The four sets of CpGs are age-associated hypermethylated and hypomethylated DMCTs in hepatocytes and cholangiocytes, respectively. In each panel, we give the Pearson Correlation Coefficient (PCC) and associated correlation-test P -value. (F) Top row: Scatterplots of CellDMC t-statistics for hepatocyte-specific age-DMCTs in the discovery set (x-axis) vs. validation set (y-axis) for 3 separate validation sets encompassing bulk liver tissue samples, as shown. Red dashed lines indicate FDR = 0.05 (discovery set) and $P = 0.05$ (validation set). In each panel, we give the number of age-DMCTs that fall in each significant quadrant, and report the Odds Ratio (OR) and associated P -value from a one-tailed Fisher-test. Bottom row: as top row but for cholangiocyte-specific age-DMCTs.



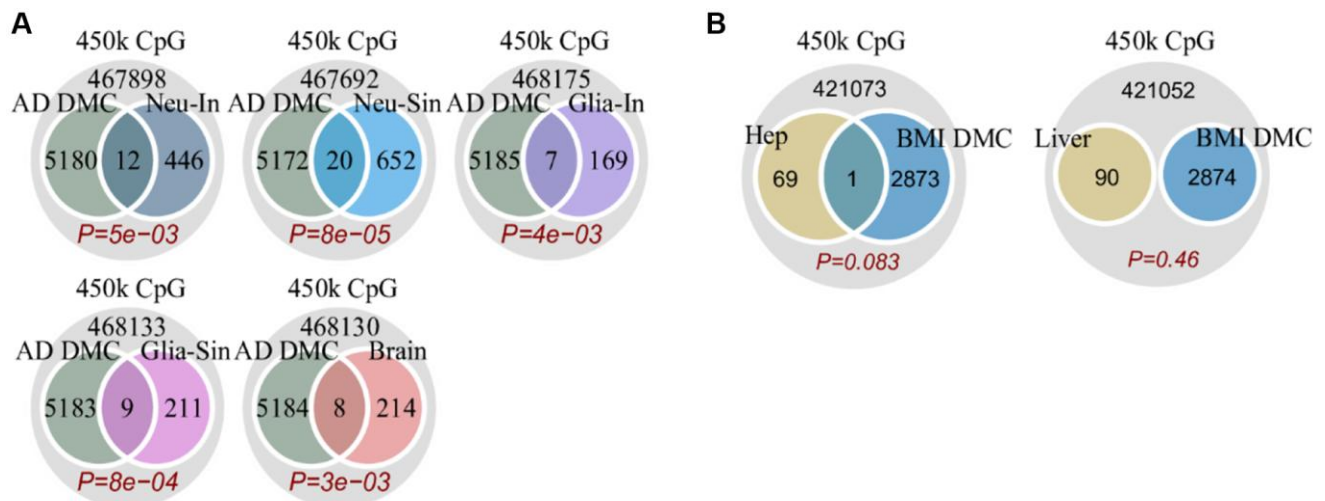
Supplementary Figure 14. Venn diagrams of overlap between hepatocyte and cholangiocyte age-DMCTs. For the discovery liver DNAm dataset, as well as 3 independent validation liver sets, we display Venn diagrams showing the overlap of hepatocyte and cholangiocyte age-DMCTs, as inferred using CellDMC. Odds Ratios (OR) and P-values of overlap were obtained from one-tailed Fisher exact tests.



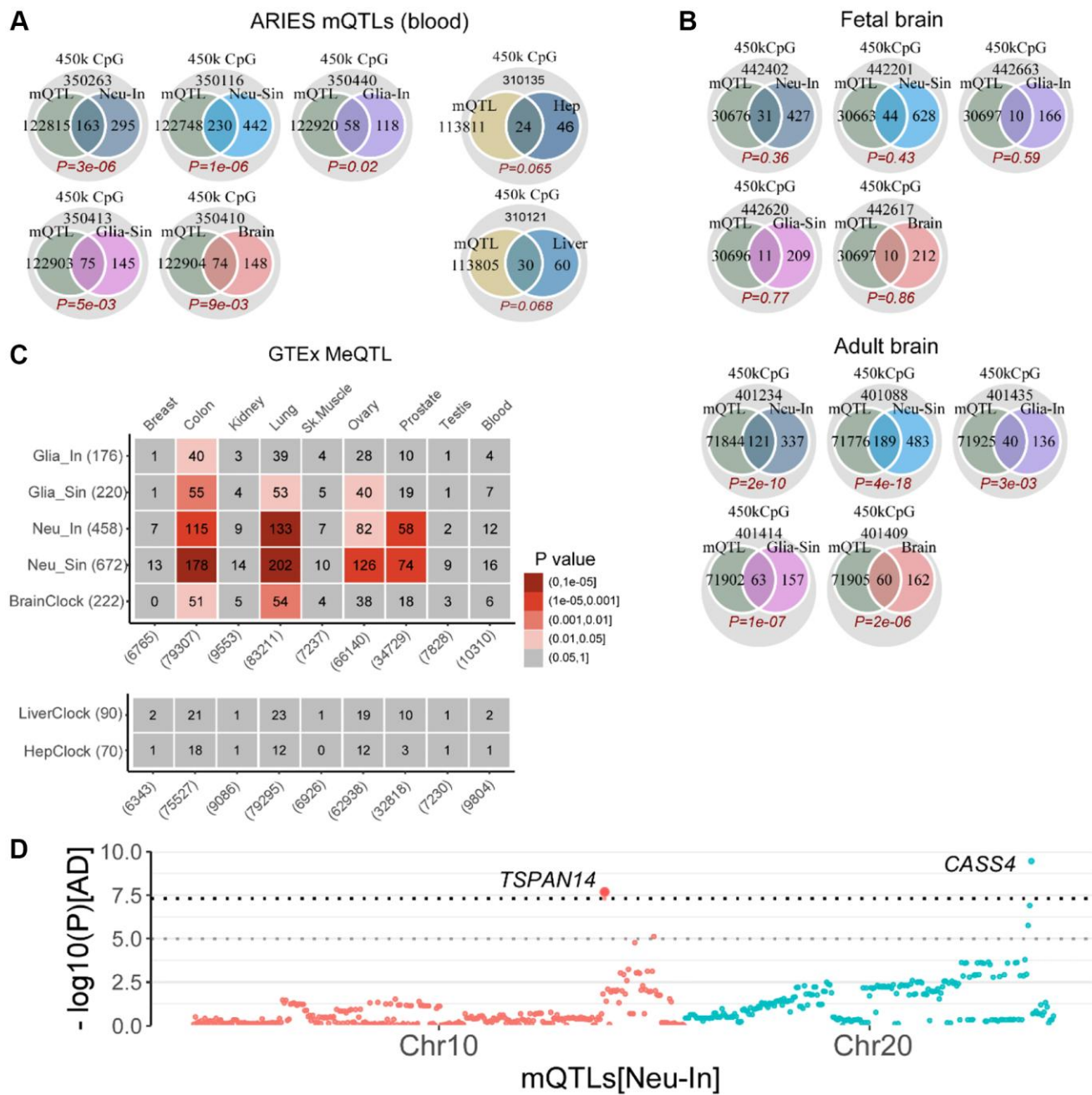
Supplementary Figure 15. Variations of liver cell-type fractions with NAFLD and obesity. (A) Boxplots display the estimated fraction of a given cell-type as a function of NAFLD stage (0 = normal) for cholangiocytes (Chol), endothelial cells (EC), hepatocytes (Hep), Kupffer macrophages (Kup) and lymphocytes (Lym). R-values and two-tailed P-value of a linear regression are given. Number of samples in each stage is given below x-axis. (B) As (A) but with the boxplots displaying the fractions as a function of body mass index (BMI) with BMI stratified into 3 categories, as shown. Of note, (A, B) are for independent DNAm datasets, as described in Methods.



Supplementary Figure 16. CpG overlap diagram of cell-type specific clocks with other clocks. For 5 cell-type specific clocks as well as 2 tissue-specific ones (Brain and Liver), depicted on the y-axis, we display the overlapping CpG number as well as their statistical significance, with the CpGs making up well-known epigenetic clocks (x-axis). The *P*-values derive from a one-tailed binomial test. The number in brackets is the number of clock CpGs present in the DNAm dataset.



Supplementary Figure 17. Enrichment analysis of clock-CpGs for EWAS DMCs. (A) Venn diagrams displaying the overlaps of 5 brain-related clocks with Alzheimer's Disease (AD) associated DMCs, with the one-tailed *P*-values derived from a Binomial-test. (B) As (A) but for the Hep and Liver clocks.



Supplementary Figure 18. Enrichment analysis of clock-CpGs mQTLs. (A) Venn diagrams displaying the overlaps of 5 brain-related and 2 liver-related clocks with blood mQTLs as defined by the ARIES database. *P*-values computed using a one-tailed Binomial test (B) As (A) but only for brain-related clocks and mQTLs as defined from fetal and adult brain. (C) As (A) but for mQTLs as defined in various tissue-types from eGTEx. The numbers in brackets indicate the number of clock-CpGs or the number of distinct mQTLs (CpGs) of each tissue present in the corresponding DNAm dataset. (D) Manhattan like plot of two Neu-In clock CpGs that map to cis-mQTLs in blood and adult brain and where the associated gene has been previously implicated in Alzheimer's Disease (AD). Y-axis labels the significance levels of the SNP associations with AD.

Supplementary Tables

Please browse Full Text version to see the data of Supplementary Tables 1–7.

Supplementary Table 1. DNAm datasets used in neuron and brain clock construction and validation. Columns label accession number, cohort name, cell and/or tissue-type, diseases included, number of sample used in this work, number of control (healthy) samples, number of cases (disease), for what purpose the cohort was used, data-type, technology, and whether the cohort was used in the various meta-analysis (N = No, Y = Yes), including the association of cell-type fraction (CTF) with age, with sex, with Alzheimers Disease (AD), and the association of age-acceleration (AA) with AD.

Supplementary Table 2. The optimal Elastic Net predictor representing the Neu-In clock. Columns label the probeID, the estimated regression coefficient, gene symbol and NCBI Identifier.

Supplementary Table 3. The optimal Elastic Net predictor representing the Neu-Sin clock. Columns label the probeID, the estimated regression coefficient, gene symbol and NCBI Identifier.

Supplementary Table 4. The optimal Elastic Net predictor representing the "BrainClock". Columns label the probeID, the estimated regression coefficient, gene symbol and NCBI Identifier.

Supplementary Table 5. This table lists all cohorts used in the meta-analysis for assessing age-acceleration in Alzheimers Disease (AD).

Supplementary Table 6. The optimal Elastic Net predictor representing the Glia-In clock. Columns label the probeID, the estimated regression coefficient, gene symbol and NCBI Identifier.

Supplementary Table 7. The optimal Elastic Net predictor representing the Glia-Sin clock. Columns label the probeID, the estimated regression coefficient, gene symbol and NCBI Identifier.

Supplementary Table 8. The optimal Elastic Net predictor representing the hepatocyte-specific clock.

Probe	β	Gene symbol	NCBI ID
cg02918910	-6.017436881	MLLT11	10962
cg05675373	-5.459241272	KCNC4	3749
cg06085726	-2.176126012	PLEKHG5	57449
cg06850099	-1.163595542	LAMC1	3915
cg10483534	-11.27967918	DISC2	27184
cg15458155	-2.432514762	ODF2L	57489
cg26883161	0.549543803	RGL1	23179
cg21211367	34.7997767		
cg24996440	0.556654931		
cg03553587	5.580119066		
cg04461197	-7.79656029	SPATA16	83893
cg10719144	-7.169039318		
cg11320271	1.542575768	CCDC36	339834
cg14098708	15.73487273	SLC6A6	6533
cg16978797	0.359053738	ARHGEF3	50650

cg11956467	6.515843914	SORBS2	8470
cg04670168	19.46881394	MYOZ3	91977
cg08271852	1.41841243		
cg09457131	0.032538949		
cg11801489	3.521123817		
cg14129473	-0.357563612	DAP	1611
cg16016281	3.376959884		
cg26919818	21.54003737	N4BP3	23138
cg05441830	-7.915169358		
cg13705956	-2.34826709	SMOC2	64094
cg14111380	-1.25046115	GABBR1	2550
cg09803262	-3.796130414	DLX6AS	285987
cg10542223	13.85943901	ZNF815	401303
cg10741603	4.520953706	SDK1	221935
cg14646653	0.223822725	GLI3	2737
cg21329649	-0.084885825	MAD1L1	8379
cg25644224	5.864462497	PTPRN2	5799
cg26836793	-2.691338723	RP9P	441212
cg00935717	-0.150122949	ZMAT4	79698
cg07350076	4.765699061	FAM66A	100133172
cg03020810	0.76432236	TLX1	3195
cg05719612	9.687510549	CACNB2	783
cg11478495	5.014419362	PRKG1	5592
cg14824052	2.470626826		
cg20475550	-0.230176091	PPP1R3C	5507
cg20530855	-2.132530095		
cg24202448	9.199788218		
cg24353740	36.27907046	KCNK18	338567
cg09036621	-4.604553593	PRDM10	56980
cg11782387	-4.345922791		
cg22204050	-6.92582377	FAM90A1	55138
cg24418063	-2.75996521	ATN1	1822
cg06170065	-1.035475913		
cg08112616	25.35109122	MEG8	79104
cg13222081	2.22617245	MRPL52	122704
cg02038233	-6.204072378	DUOX1	53905
cg02736313	-45.91000831	MEGF11	84465
cg09207718	40.84797516	CYP1A2	1544
cg22242148	8.916645472		
cg22492020	0.000210105	TIPIN	54962
cg23203997	-0.425526553		
cg24799460	8.864890338	TARSL2	123283
cg27057329	1.793435599		
cg14696348	-1.63049215	MC1R	4157
cg00735218	-2.253242355	DNAH17	8632

cg19909658	1.643780317	CDC6	990
cg26749306	-3.784572463	METRNL	284207
cg17192381	10.38975753	BCL2	596
cg02412123	-20.87621835	SIGLEC9	27180
cg07517893	28.27230223	PPP1R13L	10848
cg21940568	3.440818072	NCRNA00085	147650
cg00633740	-4.469064195	EDN3	1908
cg07474022	4.649140961	SRC	6714
cg15333674	5.226801949		
cg18170680	1.471564905		

Columns label the probeID, the estimated regression coefficient, gene symbol and NCBI Identifier.

Supplementary Table 9. The optimal Elastic Net predictor representing the liver-clock.

Probe	β	Gene symbol	NCBI ID
cg00147160	0.066771418	CNKSRI	10256
cg04055049	-0.469147506	GPR153	387509
cg07014742	8.956233592	DEDD	9191
cg08639523	-7.425961333	HIST3H2BB	128312
cg09084829	-0.379237029	KCNH1	3756
cg10501085	-0.967451996	SLC30A2	7780
cg11037787	-0.216347627	PLA2G2A	5320
cg11837451	-0.666694294	KCTD3	51133
cg19860717	-0.063381391	PTPN14	5784
cg22792674	-11.50674235	VPS72	6944
cg27552960	-4.447482513	ARF1	375
cg01138171	-0.779451213	ARHGEF4	50649
cg03602297	1.741921083	ALK	238
cg03956823	3.411298693	CHRND	1144
cg11796910	1.9510978	ROCK2	9475
cg12476443	4.327233425	C2orf39	92749
cg23059725	-7.400779735	WDR35	57539
cg26469081	-0.847409586		
cg00480331	1.458519337	BCL6	604
cg01631215	-2.943168366	SUCLG2	8801
cg17631042	-0.044251663	GOLGA4	2803
cg20668838	-2.712245058	PCCB	5096
cg24572577	-3.016548007	SLC41A3	54946
cg02647878	1.045446946	RNF175	285533
cg15848095	0.705701448		
cg27650726	-5.014070673	MRFAP1L2	93622
cg04671082	2.063865623		
cg16049600	0.187949221	PCDHB11	56125
cg01015199	1.635787049	GCLC	2729
cg16866653	51.19827934	BRP44L	105216101

cg22735834	17.66787787	PPP1R14C	81706
cg23528708	-1.501442393	CCDC90A	63933
cg25589039	0.41920236	C6orf70	55780
cg02190873	-3.651450381	PDIA4	9601
cg04727332	-0.698193261	IGFBP3	3486
cg04841607	2.510801346	C7orf33	202865
cg07023501	-0.382460827		
cg08097290	14.04520054	MGC87042	256227
cg09642951	28.95375853	EPHA1	2041
cg10548292	1.556617492		
cg21577626	-1.052262676	EPHB6	2051
cg02396865	1.267908237	DSCC1	79075
cg02605601	-0.127802997		
cg03069902	3.330232605	C10orf12	84458
cg15322963	-5.642289186	IDE	3416
cg22576165	5.496227518		
cg24402267	10.10386528		
cg07200877	-0.391044839	SPRYD5	84767
cg12139707	-2.247439968	ATM	472
cg13654391	0.27780663	DKFZp779M0652	374387
cg15248242	1.967578063		
cg23030863	-4.468072544	JAM3	83700
cg24742649	-1.315776892	OR51G1	79324
cg14120112	-1.819998161	VWF	7450
cg19535609	4.742022158	LRRC43	254050
cg22221074	-0.373301431	EFCAB4B	84766
cg27632050	4.193695173	RASA3	22821
cg04963651	-3.046651002	ERH	2079
cg22831425	31.60062472		
cg25487438	-1.253946288	DCAF11	80344
cg06486467	0.128498105		
cg18548762	-10.21521945		
cg19642394	-4.217757262	UBL7	84993
cg20901246	1.293243827		
cg22929280	-0.856017399	MIR549	69313
cg02741291	0.666219074	UBN1	29855
cg05304184	-3.141257729		
cg09285525	8.136402387	SLC7A5	8140
cg10150962	-2.324888102		
cg27106356	-1.512703347	TNP2	7142
cg01574134	-1.648422529	MIR636	693221
cg04124526	-0.360735134	FAM171A2	284069
cg06144469	7.583048847	OTOP3	347741
cg07142886	6.768928912		
cg13868393	-1.535866415	SEC14L1	6397

cg15138339	-0.07961536	COASY	80347
cg22352029	1.924760193	PYY	5697
cg02423030	-0.805027326	MAST1	22983
cg06456376	8.581426017		
cg07543967	8.03627842	SULT2B1	6820
cg10604333	0.332748993	ZNF222	7673
cg13551088	-0.045109103	ZNF573	126231
cg15429854	-7.110702776	SLC27A1	376497
cg20118823	0.400570278	TCF3	6929
cg11399508	-0.704090071	DSCR9	257203
cg01133262	-4.790660983	KDEL3	11015
cg01585852	6.523374051	MIF	4282
cg07917473	-12.28794972		
cg10648908	-2.631696356	ARFGAP3	26286
cg15508776	0.276760274	CACNA1I	8911

Columns label the probeID, the estimated regression coefficient, gene symbol and NCBI identifier.

Supplementary Table 10. The overlap between brain cell-type specific (CTS) clocks and DamAge clock.

Brain CTS clock	Brain CTS clock coef	DamAge coef	Gene	Regulatory region
Neu-In	-0.004419393	0.670384139	SESN2	TSS1500
Neu-In	-0.138237554	0.469640644	SGMS1	5'UTR
Neu-In	-0.386904875	0.436183395	EPS8L2	5'UTR
Neu-In	0.083751079	0.501032631		
Neu-Sin	-0.001363156	0.670384139	SESN2	TSS1500
Neu-Sin	0.081413726	0.239157373	ZNF642	TSS1500
Neu-Sin	-3.016294236	0.469640644	SGMS1	5'UTR
Neu-Sin	-15.09060534	0.436183395	EPS8L2	5'UTR
Neu-Sin	9.088271948	0.501032631		
Glia-In	0.132534035	0.203634862		
Glia-In	-0.153640725	0.716232962		
Glia-In	-0.019645819	0.794712929	ST3GAL4	1stExon; 5'UTR
Glia-In	-0.061458766	-13.49218833	CBX7	Body
Glia-Sin	2.821787117	0.203634862		
Glia-Sin	-6.85227349	0.716232962		
Glia-Sin	0.001254957	0.390334572	TOMM40L	TSS1500
Glia-Sin	-2.467746175	0.794712929	ST3GAL4	1stExon; 5'UTR
Glia-Sin	-4.937783763	-13.49218833	CBX7	Body

Columns label the probeID, the brain CTS clock, the brain CTS clock estimated regression coefficient, the DamAge clock estimated regression coefficient, the mapped gene, the mapped regulatory region.

Supplementary Table 11. Two Neu-In clock CpGs that define cis-mQTLs with SNPs that have been associated with Alzheimer's Disease (AD).

CpG	SNP	PvalInADgwas	Chromosome	SNP_location	CpG_location	CpGmappedGene	mQTLtype
cg24519157	rs6024860	1.21E-09	20	56408730	56412605	CASS4	cis
cg24519157	rs1884913	3.55E-10	20	56409008	56412605	CASS4	cis