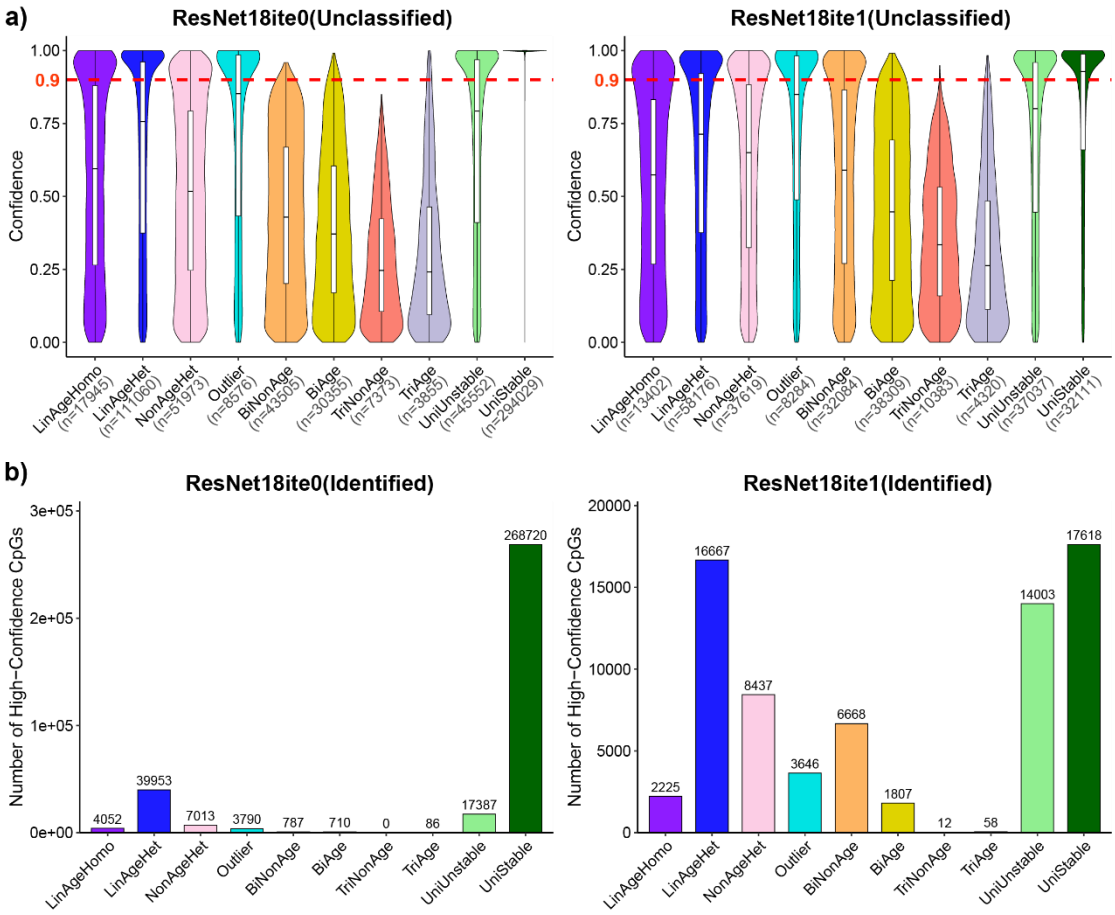
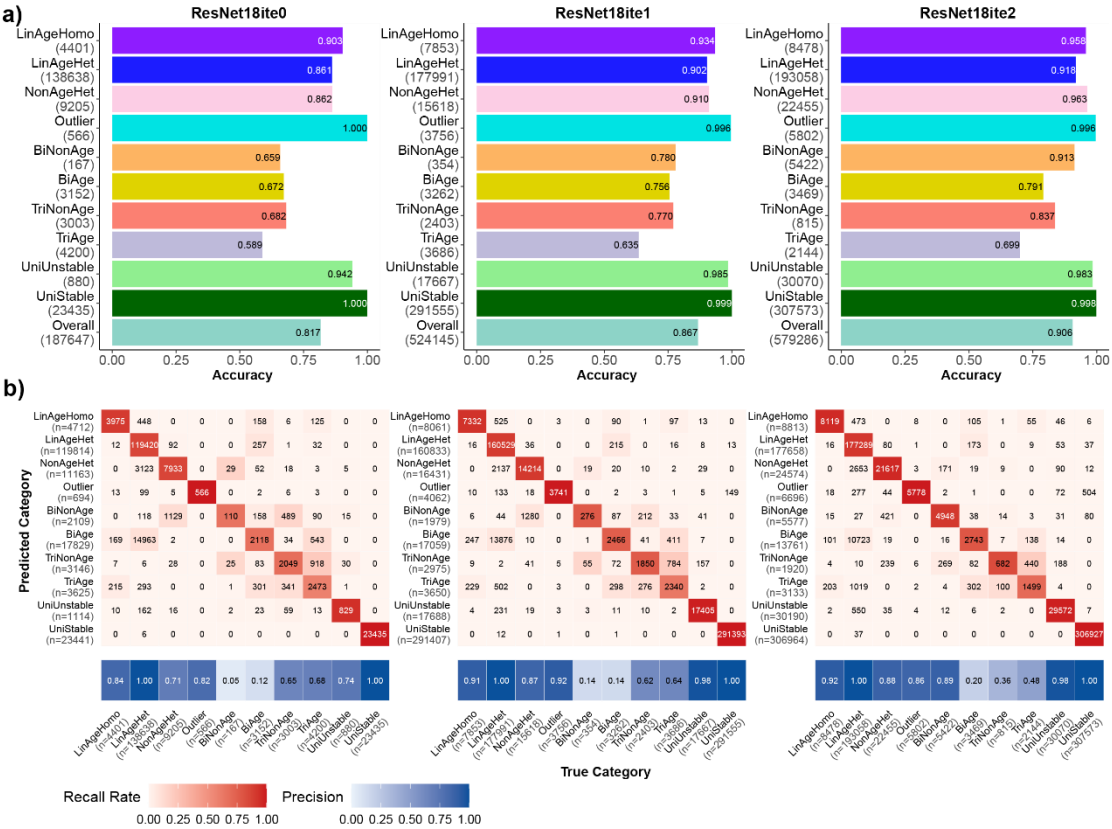
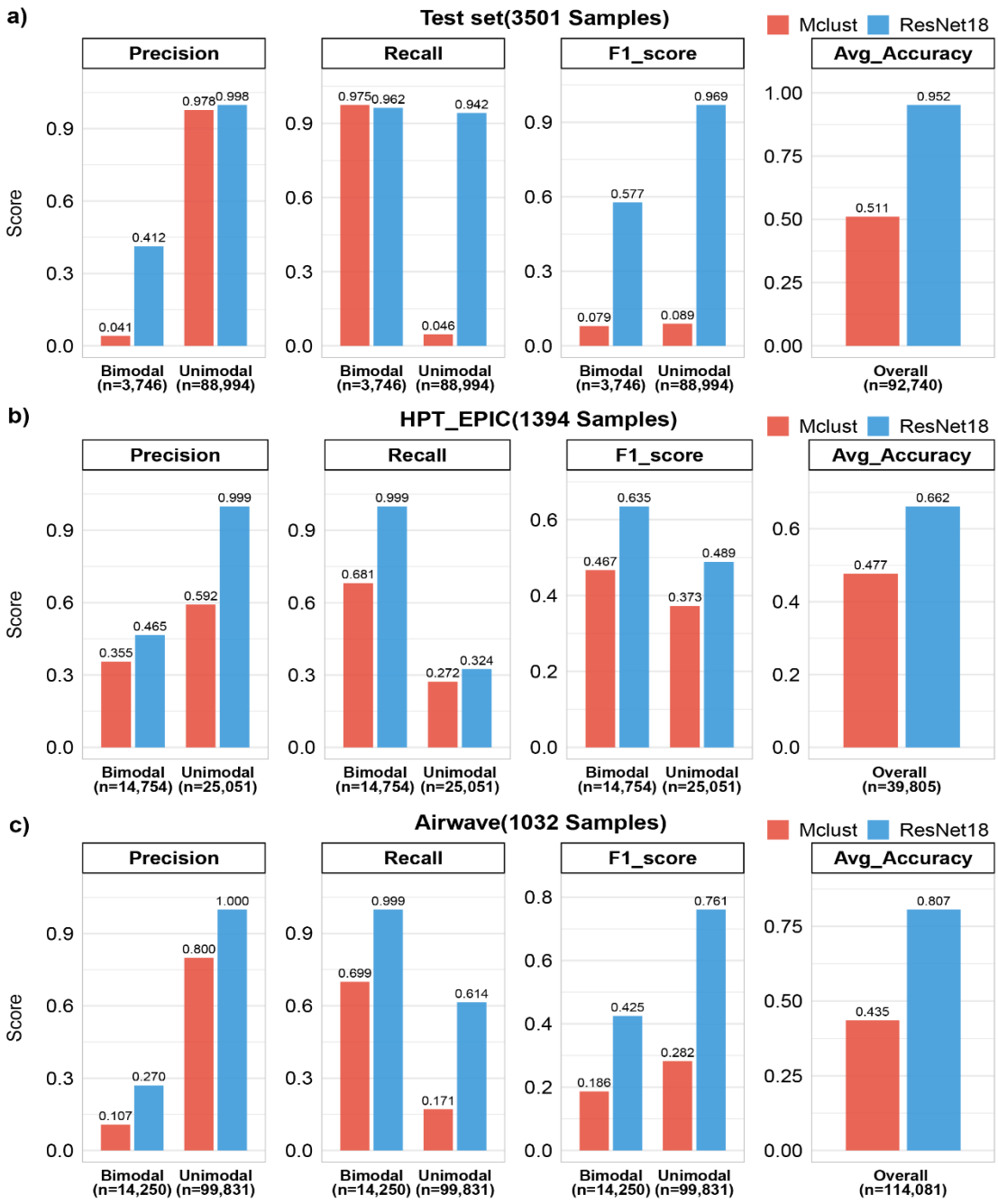
Supplementary Figures



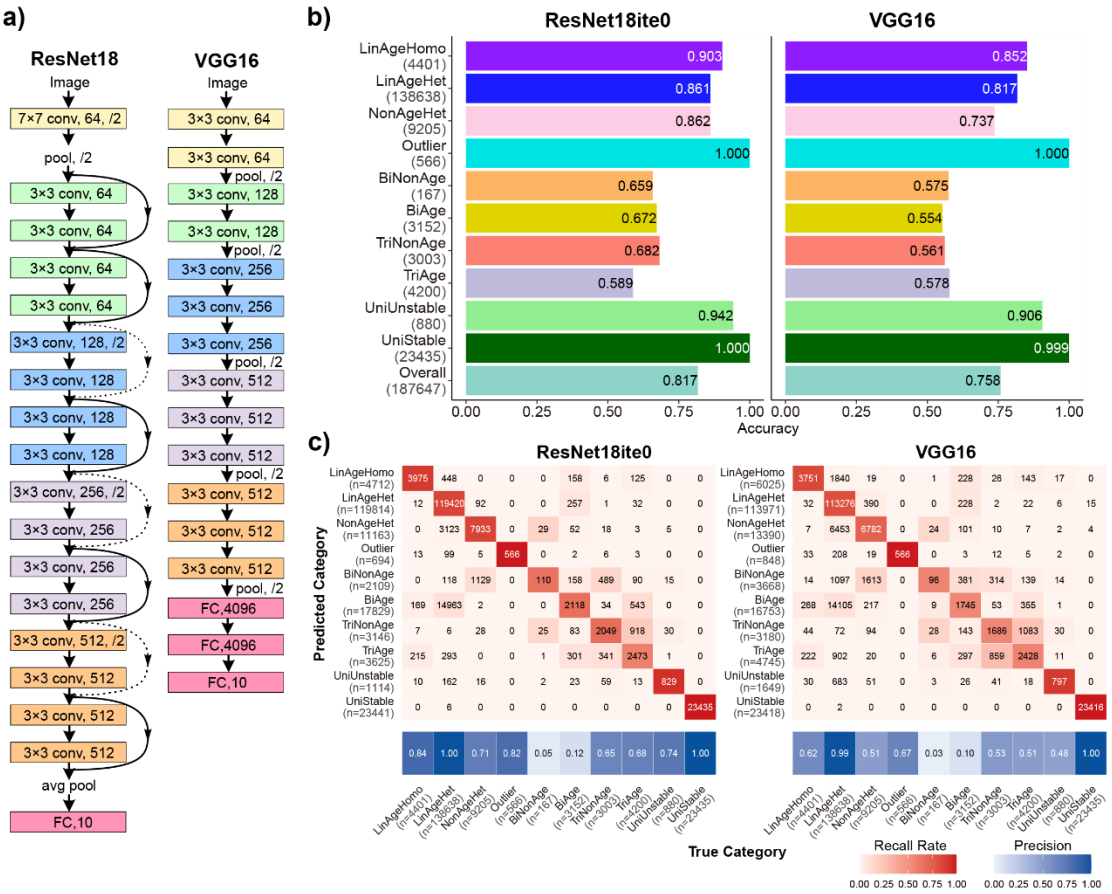
SI fig.S1. Iterative identification and expansion of categorized CpGs using high-confidence filtering. a) Confidence score distributions for unclassified CpGs predicted by the initial (ResNet18ite0) and iterative (ResNet18ite1) models. Here, confidence is calculated as the margin between the top two predicted category probabilities. A strict threshold (red dashed line, confidence > 0.9) was applied to filter out ambiguous predictions. X-axis labels indicate the total number of candidate CpGs for each predicted category. **b)** Counts of high-confidence CpGs retained after filtering. The exact number of newly classified CpGs is indicated above each bar.



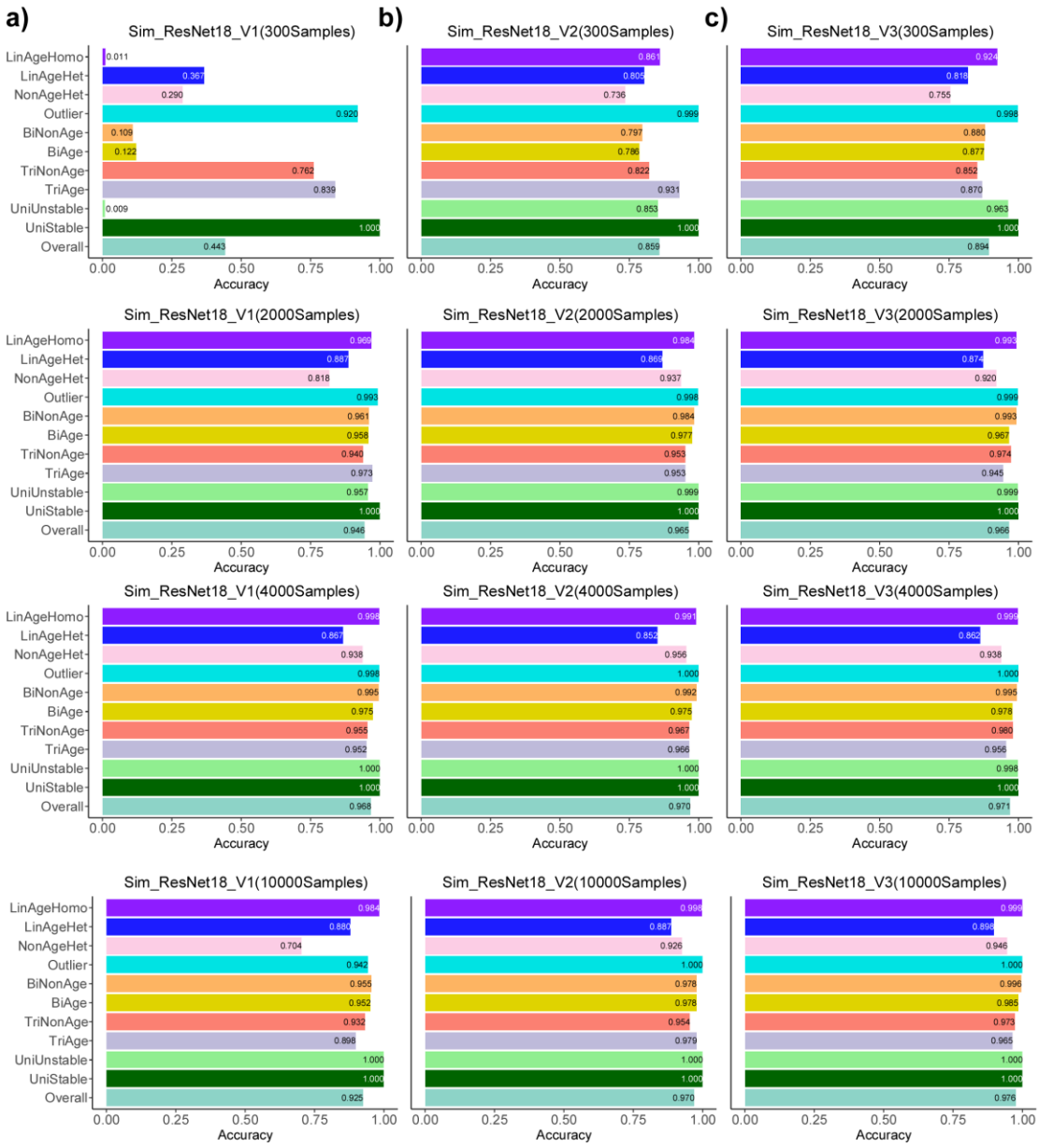
SI fig.S2: Validation accuracies and confusion matrices in NSPT cohort. a) Validation accuracies per CpG category and iterative version of ResNet18. **b)** Confusion matrices over CpG categories and for each iterative version of ResNet18.



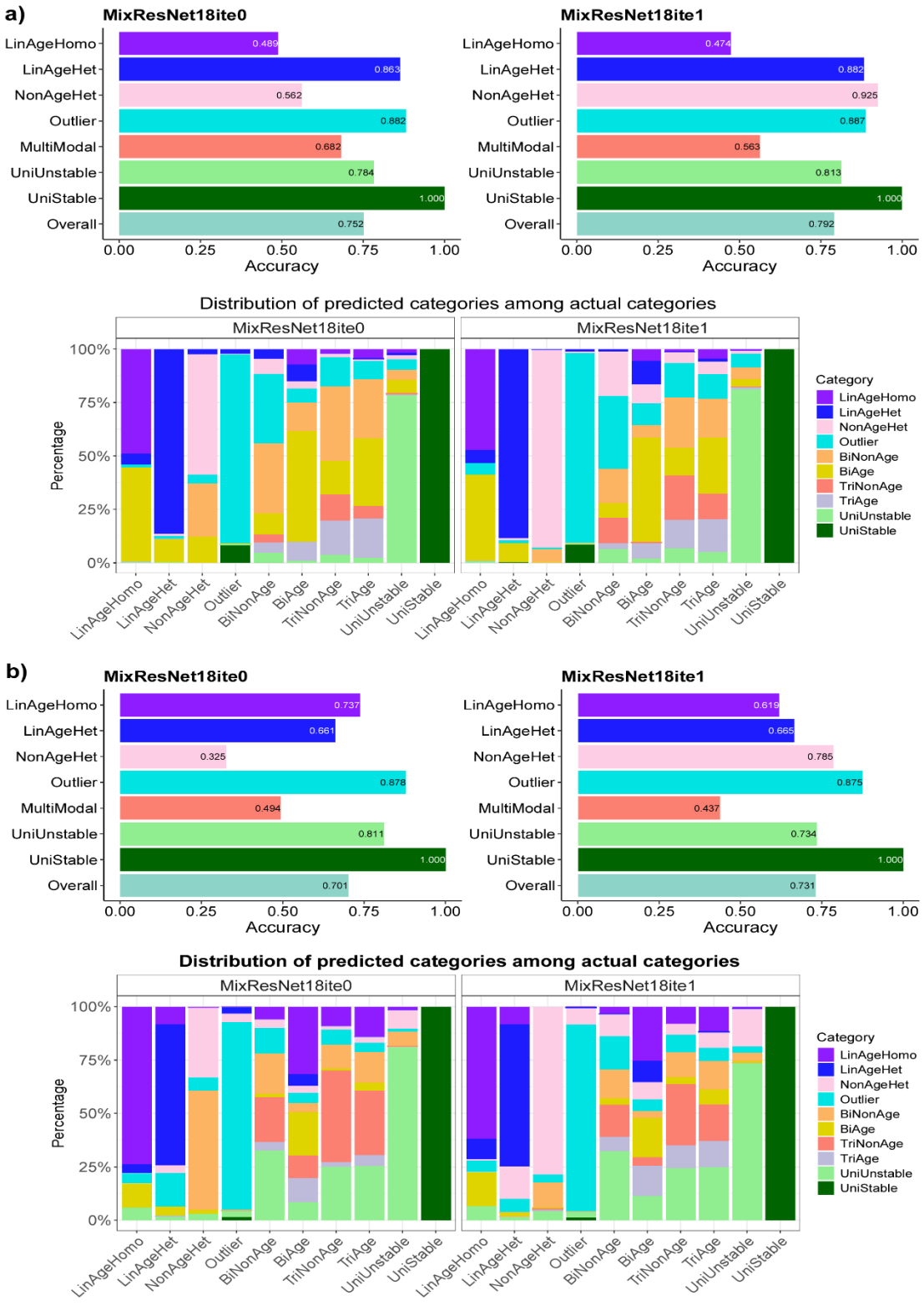
SI fig.S3: Comparison of ResNet18 to Mclust in predicting bi-modality in DNAm data. **a)** Evaluation of ResNet18 and Mclust in distinguishing gold-standard sets of bi and uni-modal CpGs in the NSPT cohort (n=3,501). Of note, ResNet18 was trained using a training set of CpGs, with both methods then evaluated on blinded test set CpGs (not used in training) containing gold-standard bi and unimodal CpGs. Bar charts display Precision, Recall, and F1-scores stratified by category (Bimodal and Unimodal), alongside the overall Average Accuracy. Precision = TP / (TP + FP); Recall = TP / (TP + FN); F1-score = 2 * (Precision * Recall) / (Precision + Recall). **b)** Comparative analysis of the ResNet18 classifier trained on NSPT with Mclust in the independent HPT_EPIC dataset (n=1,394). **c)** As b) but on the independent Airwave EPIC dataset (n=1,032).



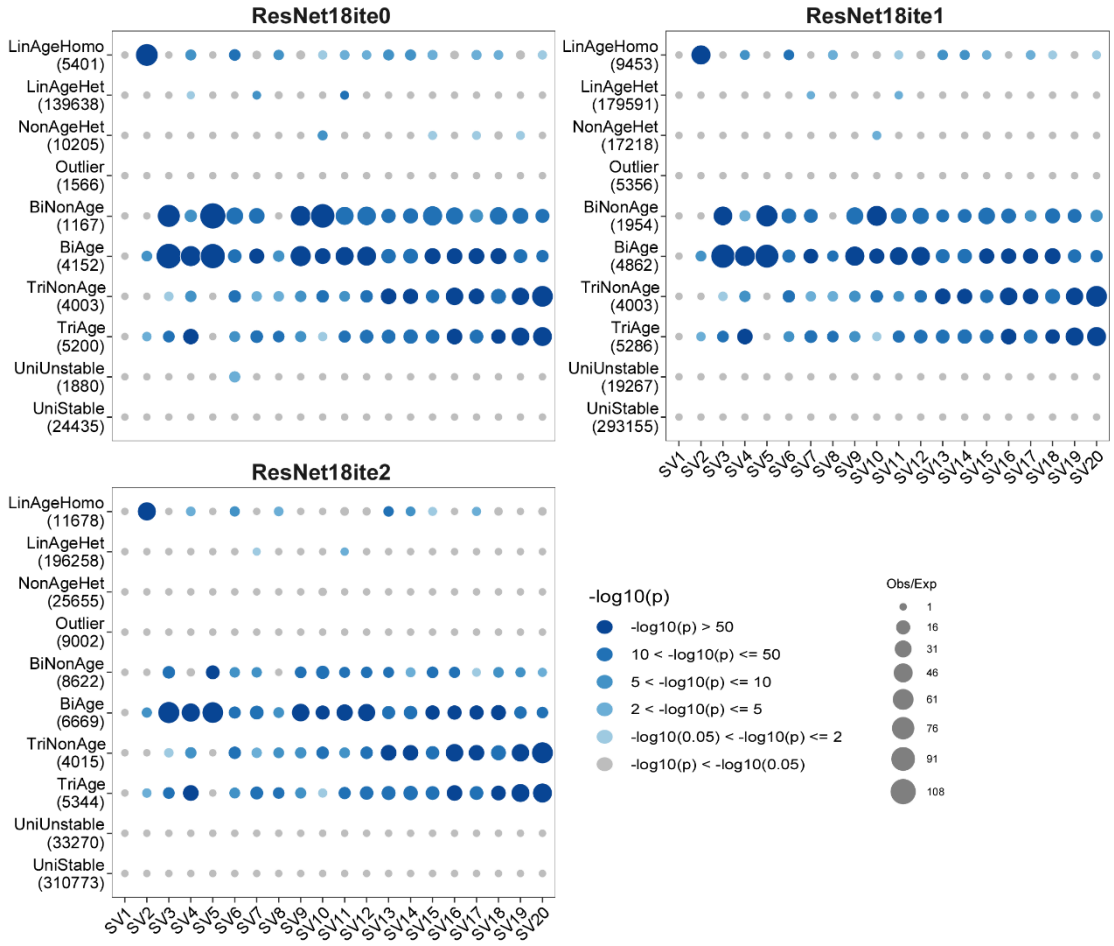
SI fig.S4. Benchmarking of ResNet18 against VGG16 on CpGs classification. a) Schematic representation of ResNet18(Left) and VGG16(Right) architectures. Notably, ResNet18 incorporates residual skip connections (curved arrows) to facilitate gradient flow. **b)** Bar plots displaying the classification accuracy of ResNet18-iteration0 and VGG16 models across ten distinct CpG categories. The y-axis lists the specific categories along with the number of CpGs in the test set for each category. The overall accuracy for each model is indicated at the bottom of the respective plot. **c),** Confusion matrices and precision metrics for the each model. The central heatmaps illustrate the concordance between the Predicted Category (y-axis) and the True Category (x-axis). Numbers within the tiles represent the count of CpGs. The color intensity of the main heatmap (red gradient) denotes the Recall, calculated as the proportion of correctly identified CpGs relative to the total number of CpGs in that true category. The bottom marginal heatmaps (blue gradient) display the Precision (positive predictive value) for each predicted category.



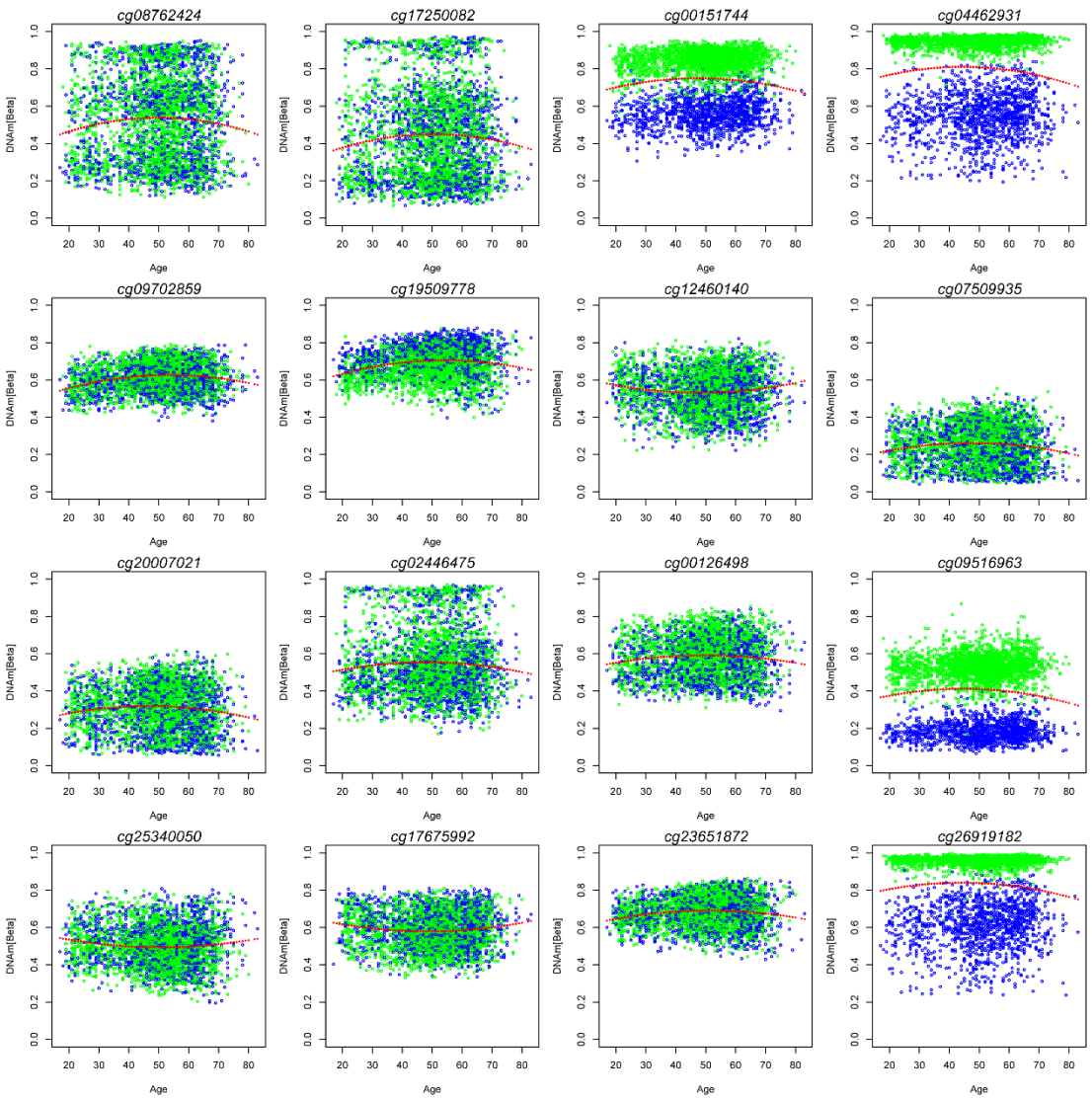
SI fig.S5: Influence of sample size on validation accuracy in a simulation model. a) Barplots represent the validation accuracy of a ResNet18 model (V1) trained on simulated data at a fixed sample size of n=3500, in 4 independently generated DNAm datasets of variable sample size (n=300, 2000, 4000 and 10000). **b)** As a) but for a ResNet18 model (V2) trained on simulated data at two sample sizes (n=800 and n=3500). **c)** As a) but for a ResNet18 model (V3) trained on simulated data at variable sample sizes (n=500, 1,500, 3,000, 5,000, and 8,000).



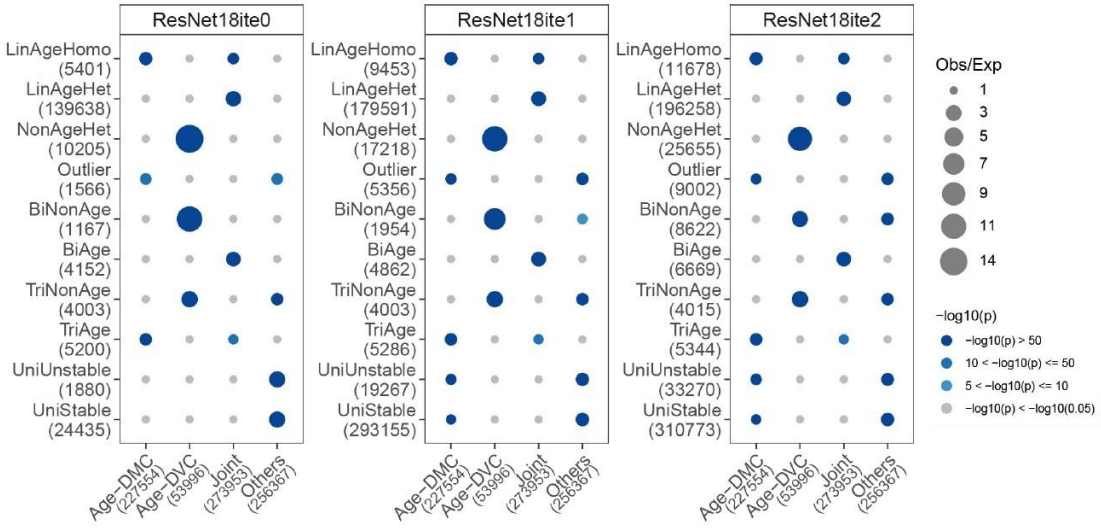
SI fig.S6: Validation accuracy of MixedResNet18 model in independent EPIC DNAm datasets. a) Barplots display the validation accuracy of the mixed ResNet18 models (zero iterations left-ite0, 1 iteration right-ite1) in the independent EPIC HPT DNAm dataset. Bi and tri-modal categories have been merged. Barplots displaying the actual distribution of predicted categories among actual CpG categories. **b)** As a) but for the Airwave EPIC DNAm dataset.



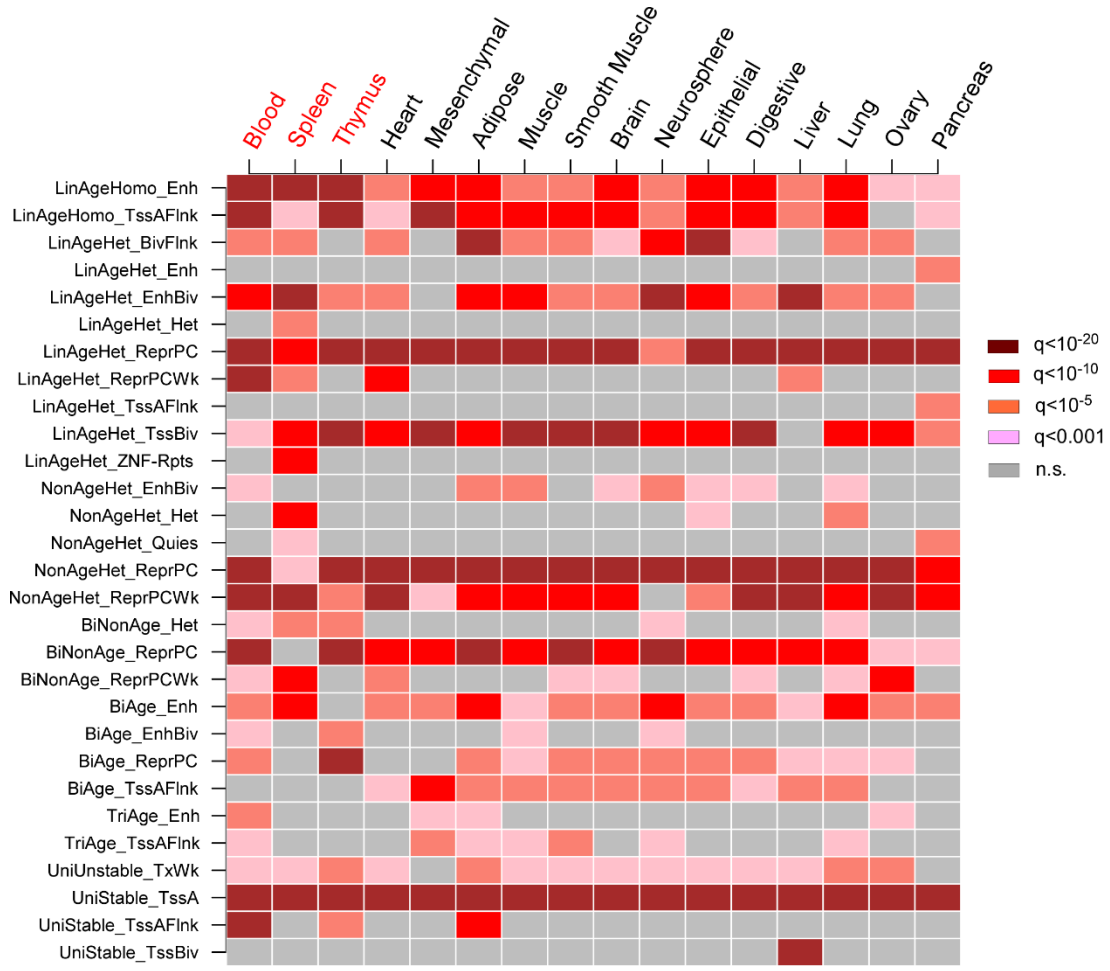
SI fig.S7: Enrichment of singular vectors (principal components) for the 10 age-related CpG categories. Balloon plots display the observed to expected ratio of CpG overlaps between the 10 different age CpG taxa (rows) and the 500 top-ranked CpGs driving each of the singular vectors (SVs) (equivalent to principal components-PCs, columns), for each of 3 iterations of ResNet18. The size of the balloon indicates the observed/expected ratio, color indicates the level of statistical significance under a one-tailed Binomial test. The number of CpGs in each age-category are given. The top-ranked 500 CpGs for each SV are those with the highest absolute weights.



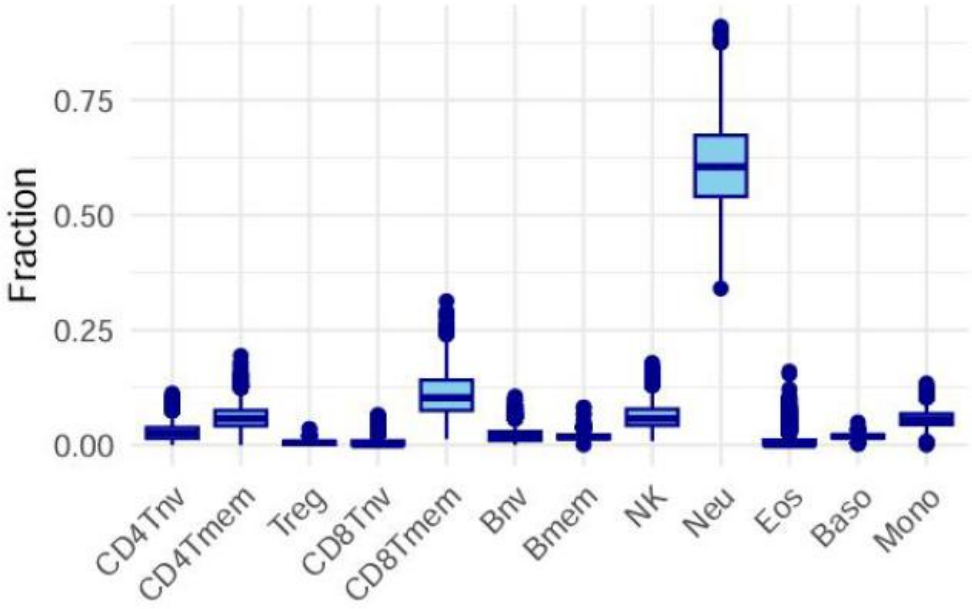
SI fig.S8: The non-monotonicity of top-ranked non-monotonic CpGs is of low effect-sise. Scatterplots display the DNAm values (y-axis) against chronological age (x-axis) for the top 16 CpGs displaying non-monotonic quadratic patterns. The best quadratic fit is shown in red. Color indicates sex: green=female, blue=male. Note how the effect size associated with the age-related non-monotonicity is small compared to the overall DNAm variance.



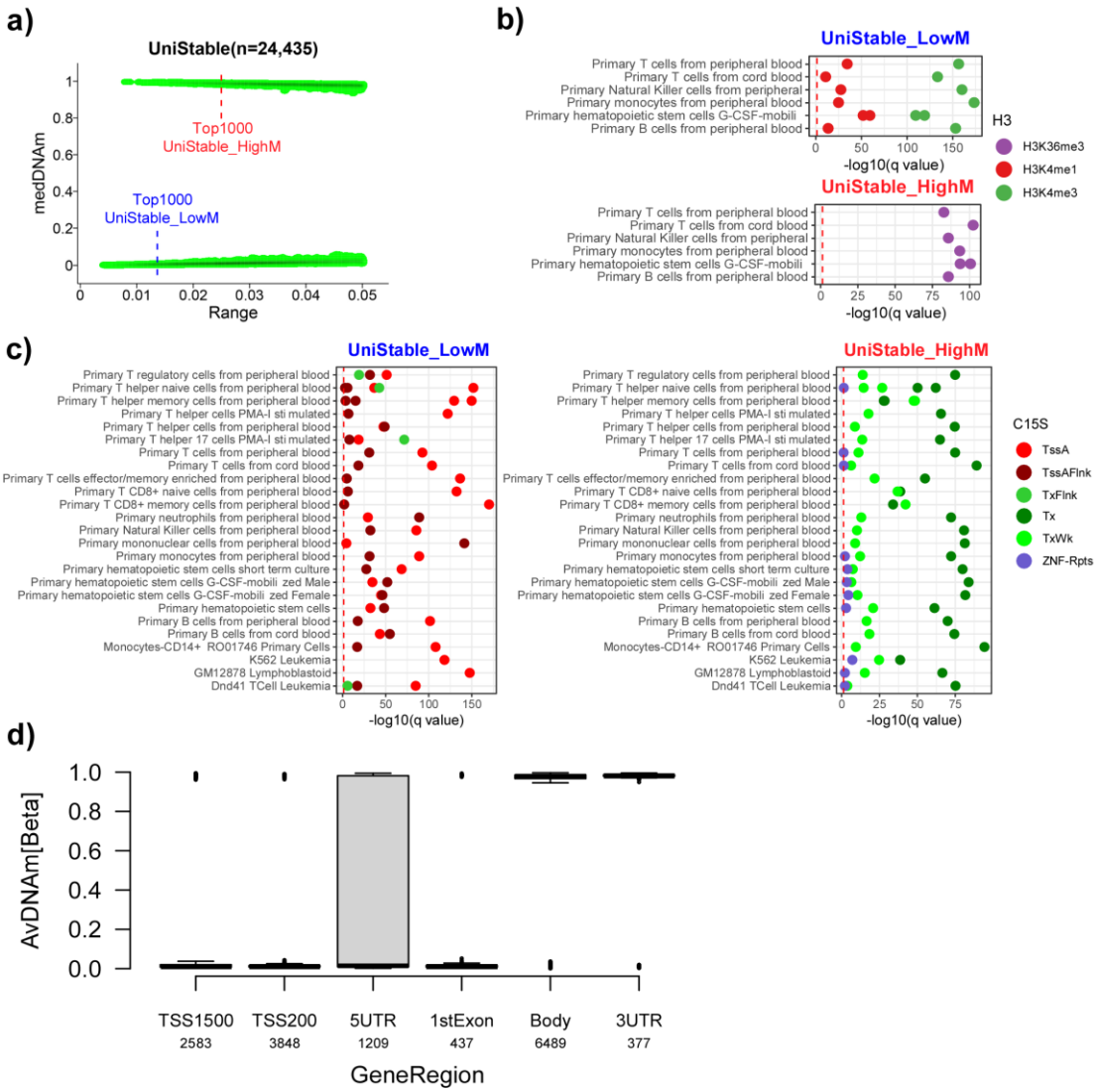
SI fig.S9: Overlap enrichment analysis with Seale's 3 to 4 CpG categorization scheme. For each iteration of the ResNet18 classifier, the balloon plot displays the enrichment (observed to expected ratio) of overlap between our 10 CpG-classification scheme and the 4 categories from Seale et al: age-DMCs: CpGs displaying change in mean DNAm but not change in DNAm variance, age-DVCs: CpGs displaying change in DNAm variance but no change in mean, Joint: CpGs displaying change in DNAm variance and DNAm mean, Other=all other CpGs. Colors indicate statistical significance of overlap according to a one-tailed Fisher-test.



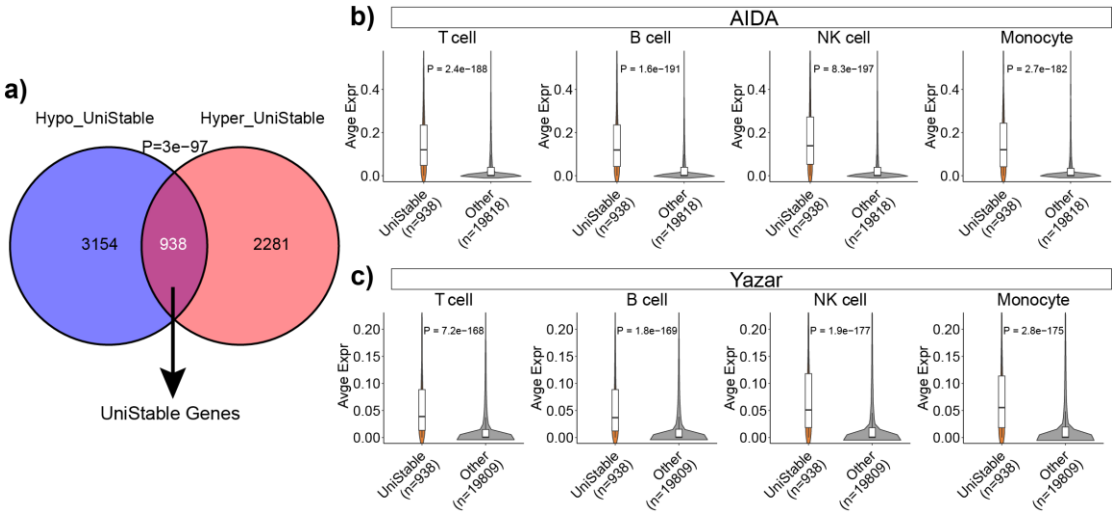
SI fig.S10: eFORGE2 enrichment analysis result for CpG taxa chromatin-state pairs across tissues. We only display CpG taxa chromatin state pairs where there was significant enrichment of that chromatin state in at least one tissue-type, where enrichment was defined as having an average $-\log_{10}$(q-value) larger than 3, and where the average $-\log_{10}$(q-value) was taken over all cell-types mapping to a given tissue-type. Fetal tissue types, stem-cells and carcinoma samples were removed. Q-values as estimated with eFORGE2 using the top 1000 CpGs in each taxa.



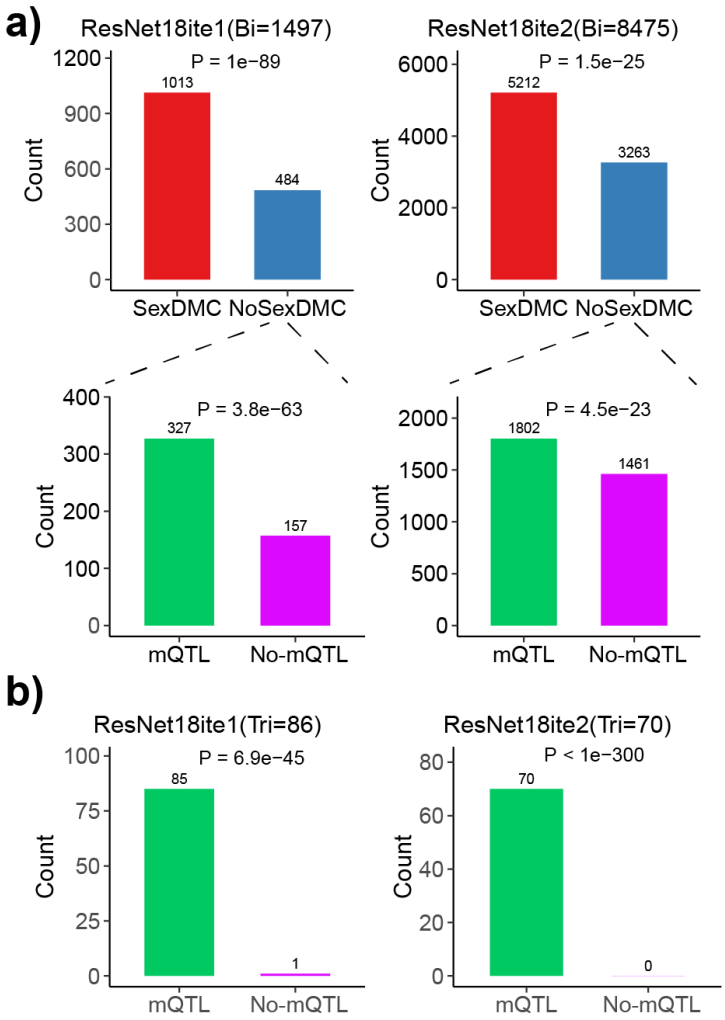
SI fig.S11: Immune-cell fraction distribution in NSPT cohort. Boxplots displaying the estimated fractions of 12 adult immune cell-types in the NSPT cohort (n=3501 samples).



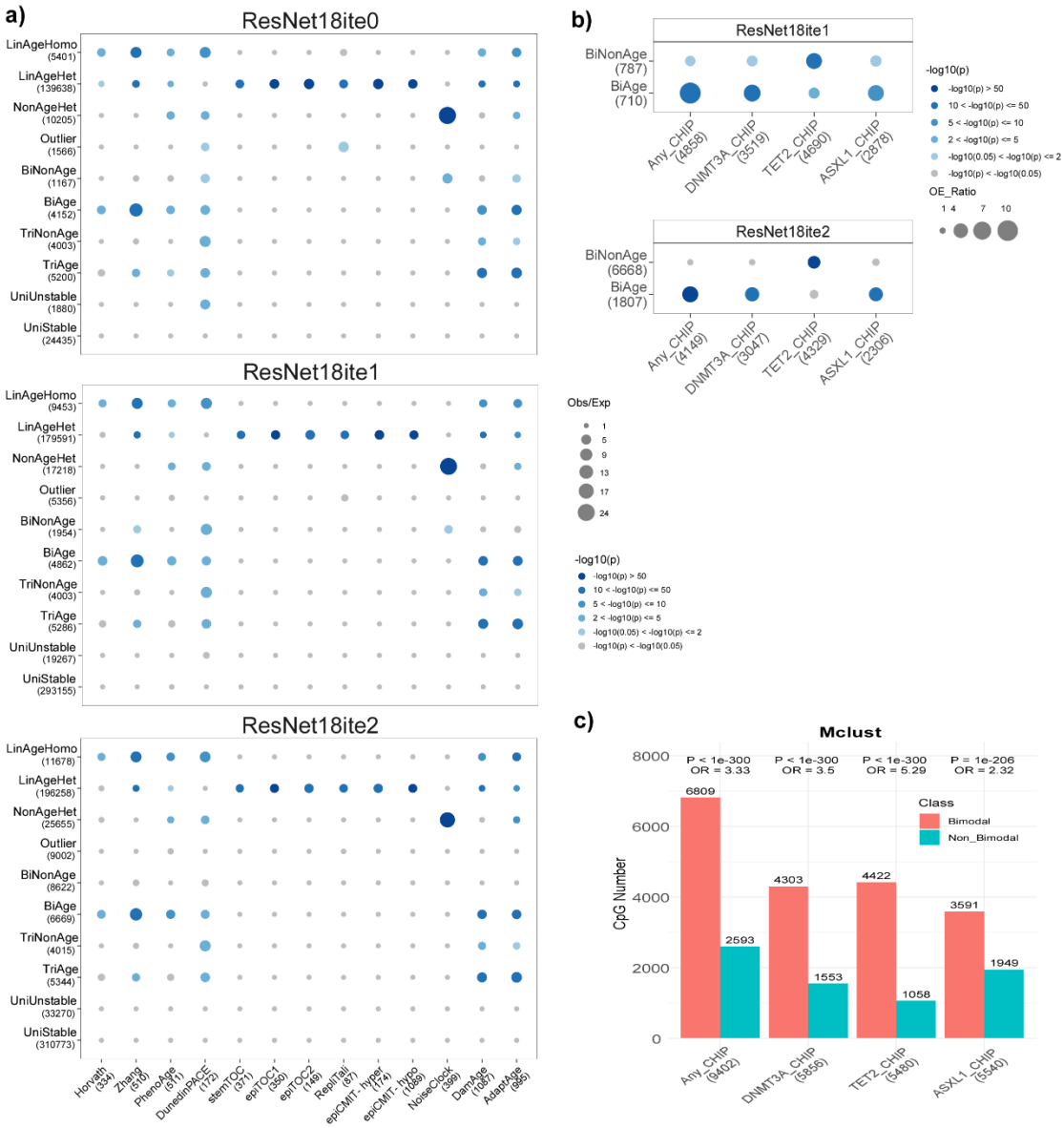
SI fig.S12: Enrichment pattern of UniStable CpG subtypes. a) Smoothed scatterplot of the 24,435 UniStable CpGs, with the median DNAm-level displayed on y-axis and the range in DNAm (max-min) displayed on the x-axis. The vertical dashed lines indicate the top-1000 UniStable CpGs according to low or high median DNAm-levels, as they have been ranked in increasing values of the range. **b)** eFORGE2 enrichment analysis of UniStable low DNAm and UniStable high DNAm against histone (H3) marks, as indicated. Significance is displayed at the -$\log_{10}$(q-value) level, where the P-value is derived from a one-tailed Binomial test. **c)** As b) but testing for enrichment against 15 chromatin-states. **d)** Boxplot displays the average DNAm of UniStable CpGs in the NSPT cohort against the particular gene-region it maps too. Of note, UniStable CpGs with ambiguous mappings were excluded. Number of CpGs in each region is indicated. TSS1500 and TSS200 means 1500bp and 200bp upstream of the transcription start-site. Body means all exons except 1st exon.



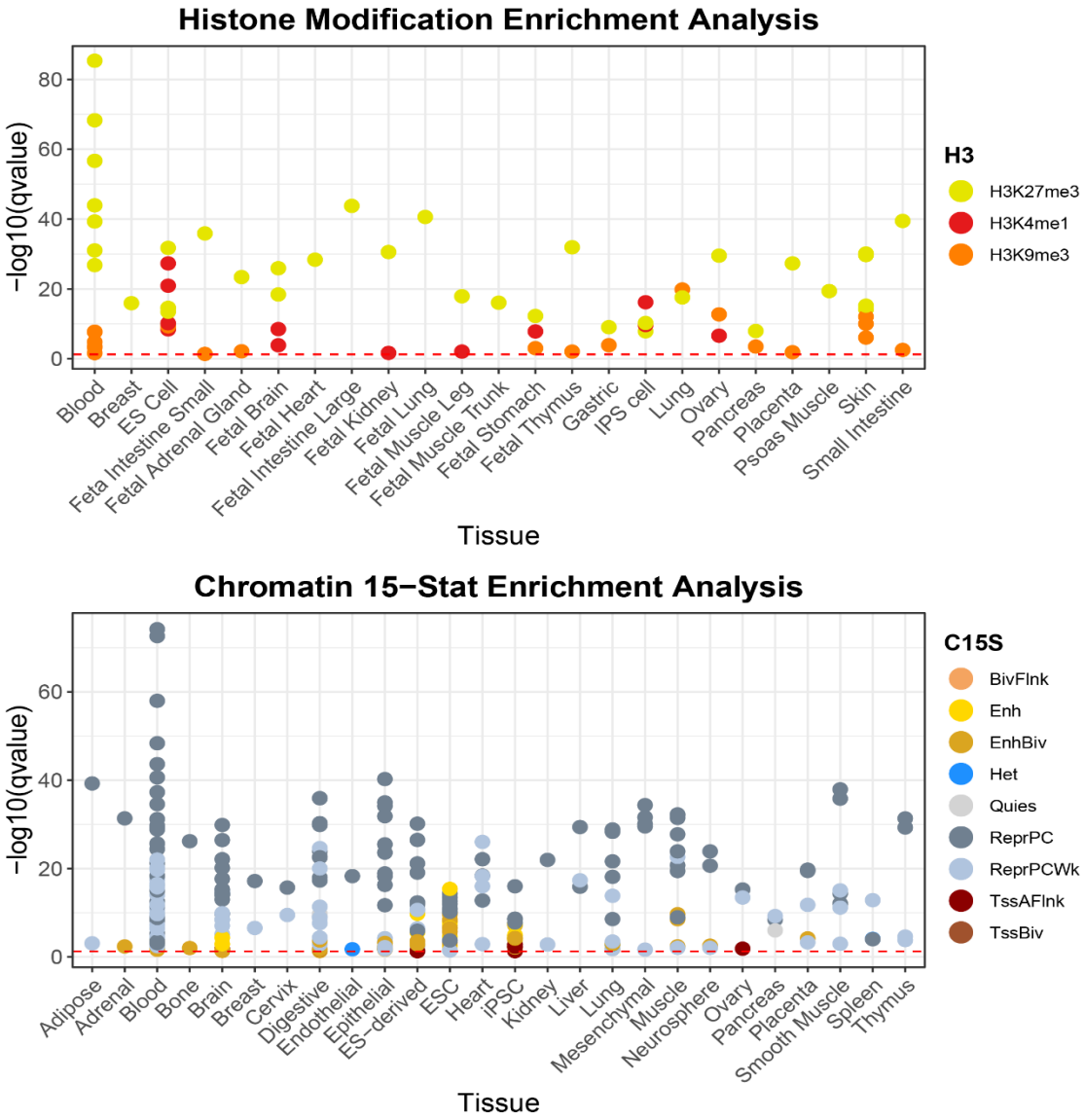
SI fig.S13: UniStable genes are more actively transcribed than non-UniStable genes. **a)** Venn diagram of genes containing a lowly methylated (hypomethylated) UniStable CpG with genes containing a highly methylated (hypermethylated) UniStable CpG. Since the former map to the promoter of 1st exon regions whilst the latter map to gene-body or 3'UTR, we define the overlap to be UniStable genes. P-value was obtained using a one-tailed hypergeometric test. **b)** Boxplots comparing the average expression level of UniStable genes to genes not containing any UniStable CpGs in four of the main immune cell types of the AIDA scRNA-Seq dataset. P-values were obtained using a one-tailed Wilcoxon rank-sum test. **c)** Similar to b), but in the Yazar scRNA-seq data set.



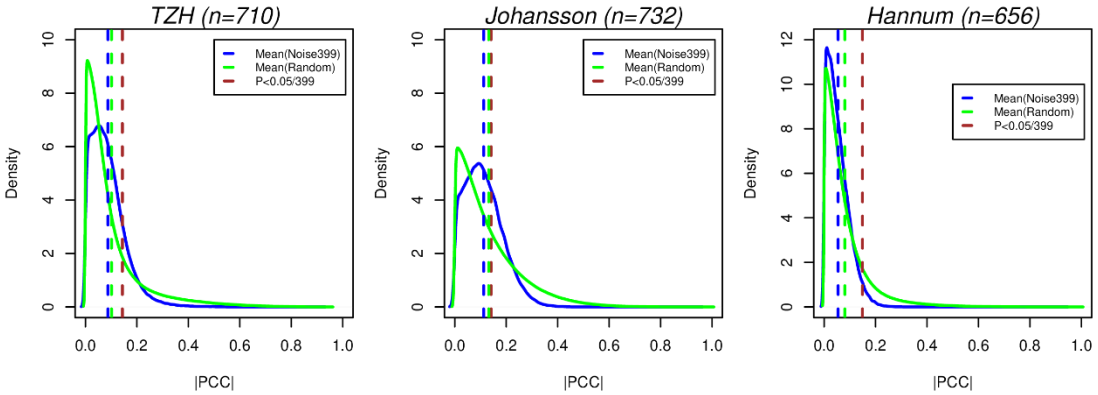
SI fig.S14: Newly assigned bi-modal and tri-modal CpGs map to sex-DMCs and mQTLs. **a)** For two iterations of ResNet18, we display the numbers of newly and confidently (maximum probability >0.9) assigned bi-modal CpGs that define sex-DMCs (nominal limma t-test P-value < 0.05) or not. P-value of enrichment is derived from a one-tailed Fisher-exact test. The lower panel in a) then asks if these bi-modal non-sex DMCs are enriched for low effect size mQTLs. **b)** For two iterations of ResNet18, we display the numbers of newly and confidently (maximum probability >0.9) assigned tri-modal CpGs that define mQTLs, using the full list of mQTLs as defined by Peng Q et al Nat Genet 2024.



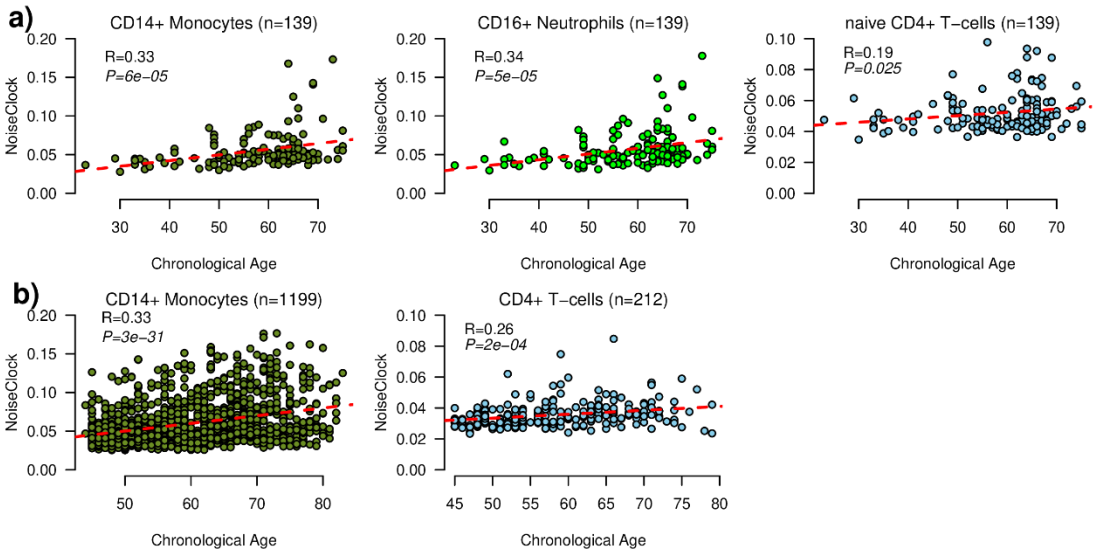
SI fig.S15: Robustness of clock and CHIP-CpG enrichments for different iterative versions of ResNet18. a) Enrichment balloon plot of the 10 different age CpG taxa for epigenetic clocks and for the 3 iterations of ResNet18. The size of the balloons indicate the observed/expected ratio, color indicates the level of statistical significance under a one-tailed Binomial test. The number of CpGs in each age-category and clock are given. **b)** Enrichment balloon plots of CHIP-associated CpG categories, focusing on CpGs that were initially unclassified, but which the ResNet18ite1 or ite2 algorithm newly assigned to the BiNonAge or BiAge category. **c)** Barplots indicate the number of CpGs in each CHIP-related set that display bi-modality (or non-bi-modality) according to the Mclust algorithm, which fits a Gaussian mixture to the DNAm profiles. Odds Ratio and one-tailed Fisher-test P-value is given.



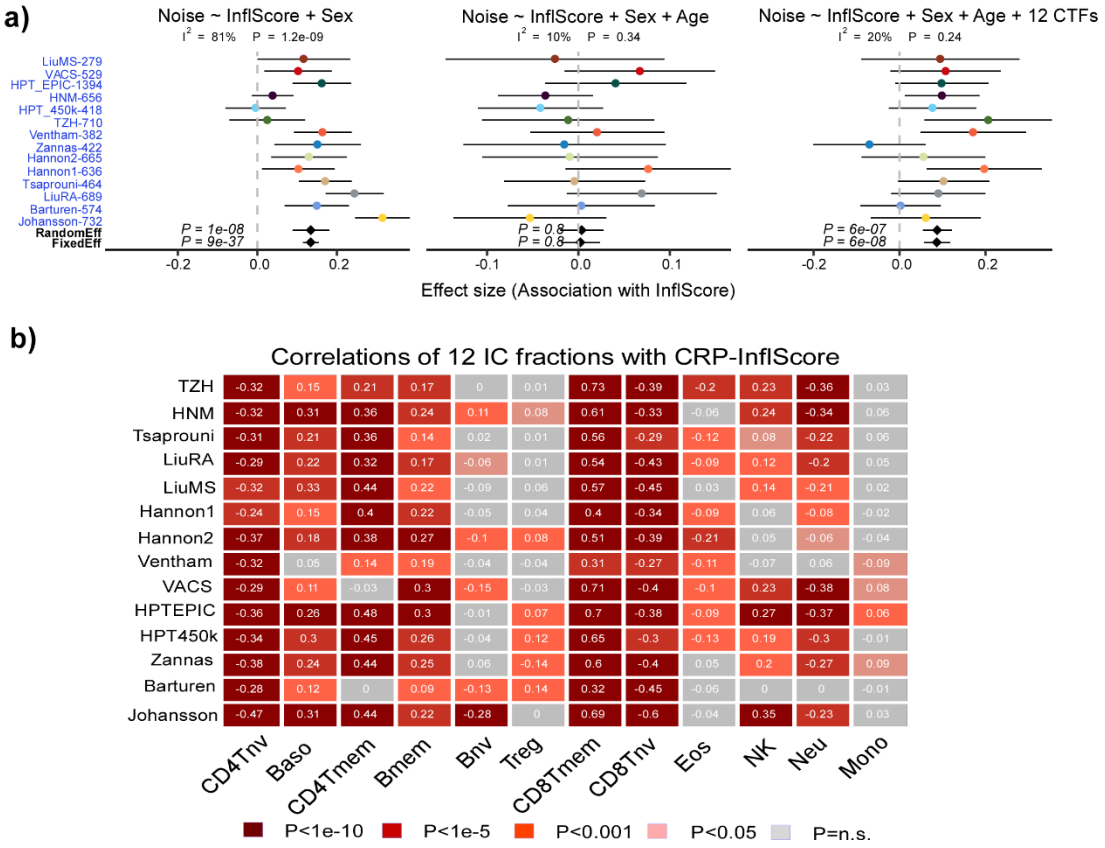
SI fig.S16: eFORGE2 enrichment analysis of 399 noise clock CpGs. Enrichment analysis of the 399 NonAgeHet noise-clock CpGs for histone marks (top) and chromatin states (bottom), as assessed using eFORGE2. Y-axis labels the -$\log_{10}$[q-value] where the q-value is the FDR associated with a one-tailed binomial test P-value.



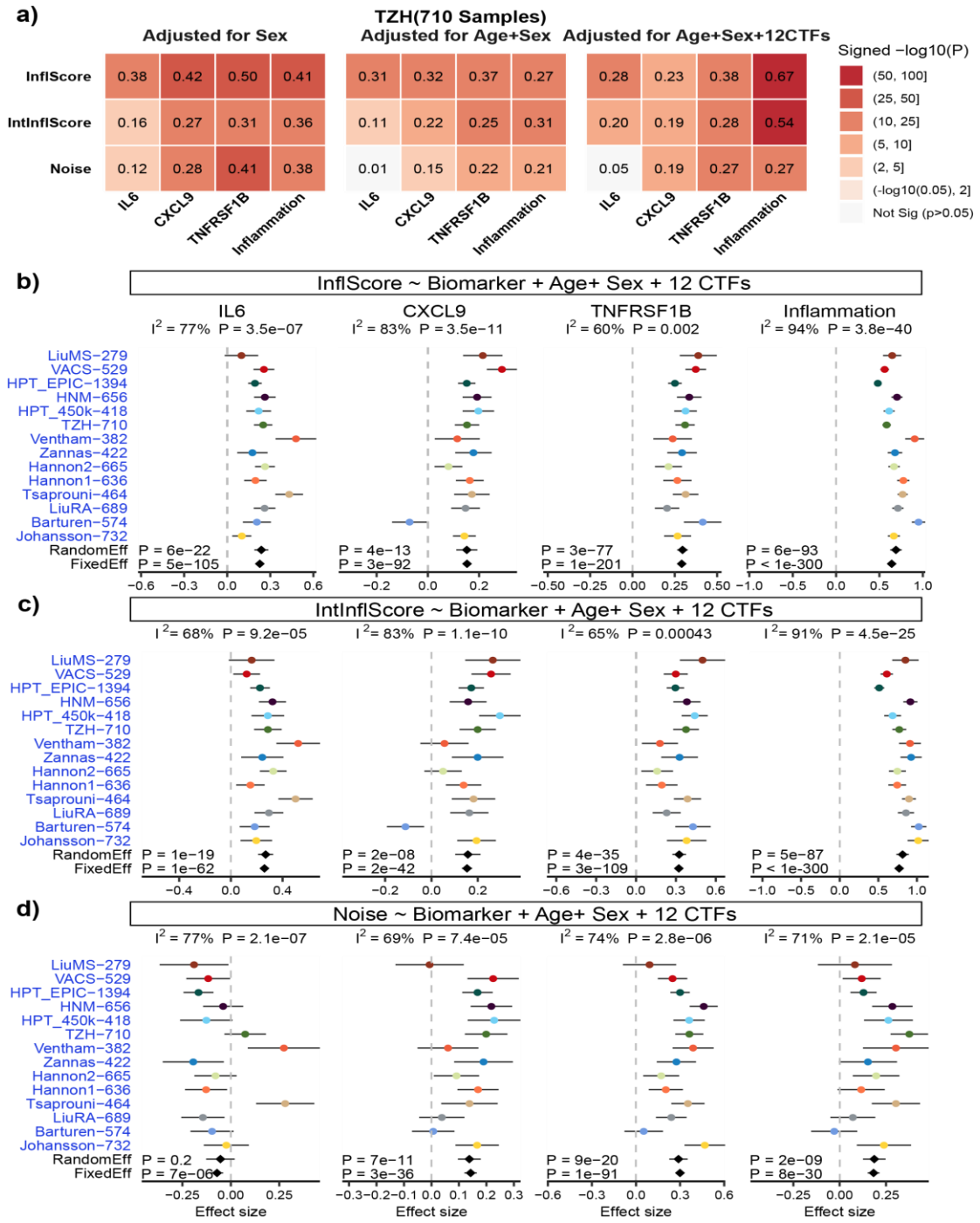
SI fig.S17: Correlation strength of 399 noise clock CpGs. For each of the 3 DNAm datasets that were used to identify the 399 noise clock CpGs, we compare the density distribution of absolute PCC values displayed by pairs of the 399 noise clock CpGs (blue) to those of 5000 randomly selected CpGs (green). The average values are indicated by vertical dashed lines. Also shown is a brown vertical dashed line which indicates the line of statistical significance (Bonferroni P < 0.05/399~0.000125).



SI fig.S18: Validation of NoiseClock in sorted immune cell datasets. a) NoiseClock predictor vs chronological age for CD14+ monocytes (n=139), CD16+ neutrophils (n=139) and naïve CD4+ T-cells (n=139) from BLUEPRINT (Chen L et al Cell 2016). R-value and linear regression P-value are given. **b)** As a) but for 2 CD14+ monocytes (n=1199) and CD4+ T-cells (n=212) from the MESA study (Reynolds LM et al Nat Commun 2014).

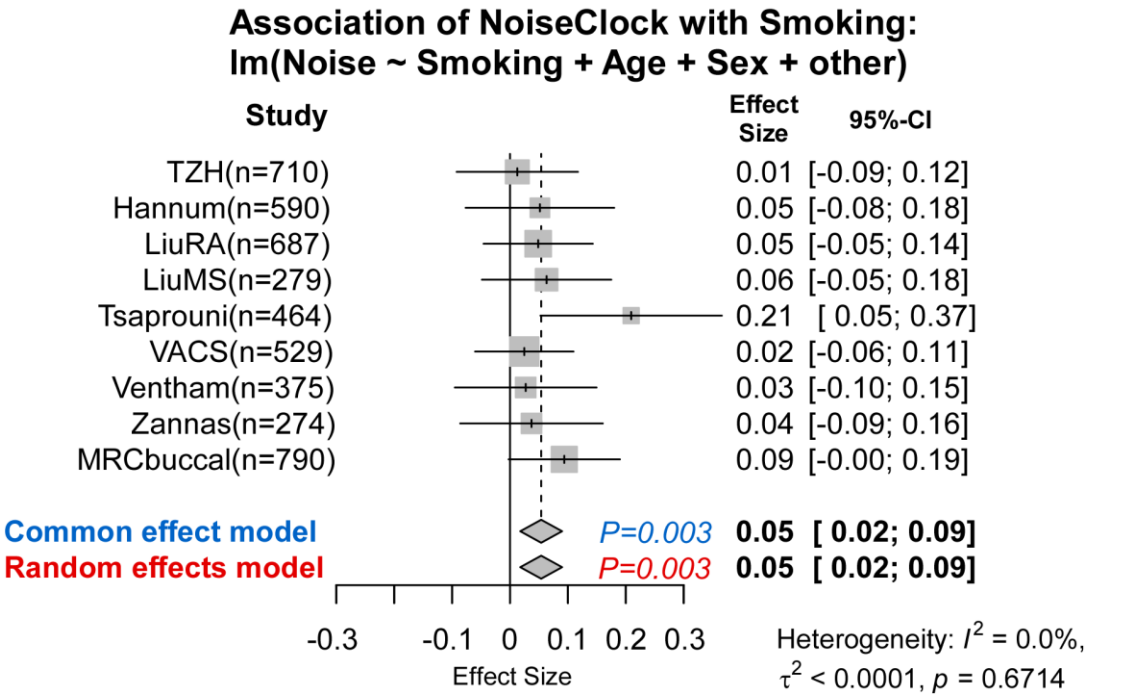


SI fig.S19. Correlation of NoiseClock with a 1765 CpG DNAm-based proxy of CRP levels. a) Forest plots of association of the NoiseClock with the InflScore (1765 CpG DNAm-based proxy of CRP levels) as assessed across 14 independent blood DNAm datasets, adjusted for sex, age and sex and age+sex+ 12 immune-cell fractions. We give the P-values of meta-analysis fixed and random effect models. Bars represent 95% confidence intervals. All variables have been standardized to unit variance before running the meta-analyses. At the top of each panel we given the I^2 heterogeneity index and P-value. **b)** Heatmap of signed Pearson correlation coefficients between the InflScore and the 12 immune-cell fractions across the 14 cohorts.

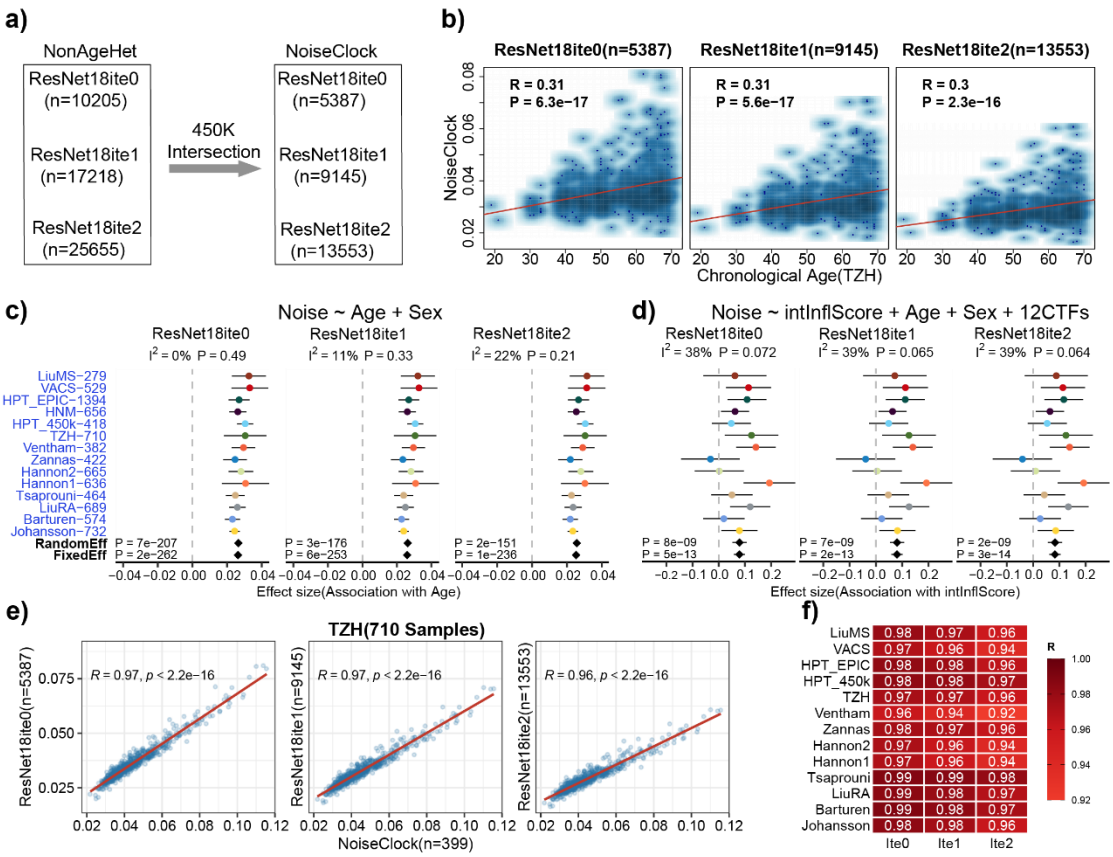


SI fig.S20: Correlation of NoiseClock and hsCRP-based DNAm proxies to other inflammatory biomarkers. a) Heatmap displaying the partial correlation analysis between the derived scores (InflScore, IntInflScore, and NoiseClock) and four DNA methylation-based inflammatory proxies (IL6, CXCL9, TNFRSF1B, and SystemsAge Inflammation) in the TZH cohort (n = 710). The analysis was performed using three adjustment models: adjusted for Sex (left), Age and Sex (middle), and Age, Sex, plus 12 cell type fractions (right). The numbers within the cells represent the Pearson correlation coefficients (PCC), while the background color intensity corresponds to the signed -log10(P-value). **b–d)** Forest plots summarizing the meta-analysis of the associations between the inflammatory biomarkers and **b)** InflScore**, c)** IntInflScore,

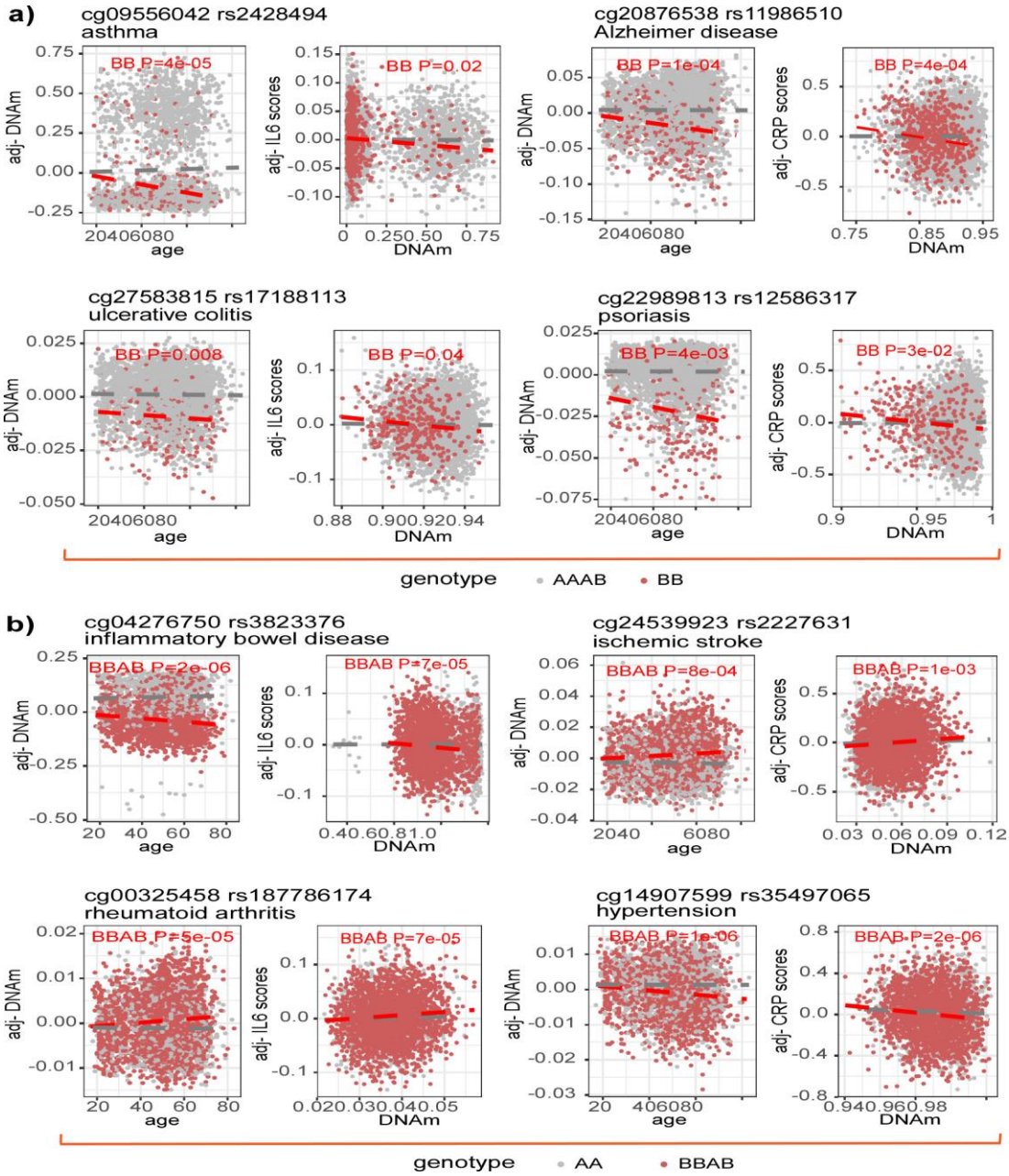
and **d)** NoiseClock across 14 independent blood datasets. In each cohort, associations were assessed using multivariate linear regression adjusted for age, sex and immune cell fractions. Error bars represent 95% confidence intervals (CI). Results from fixed- and random-effects meta-analyses are shown at the bottom.



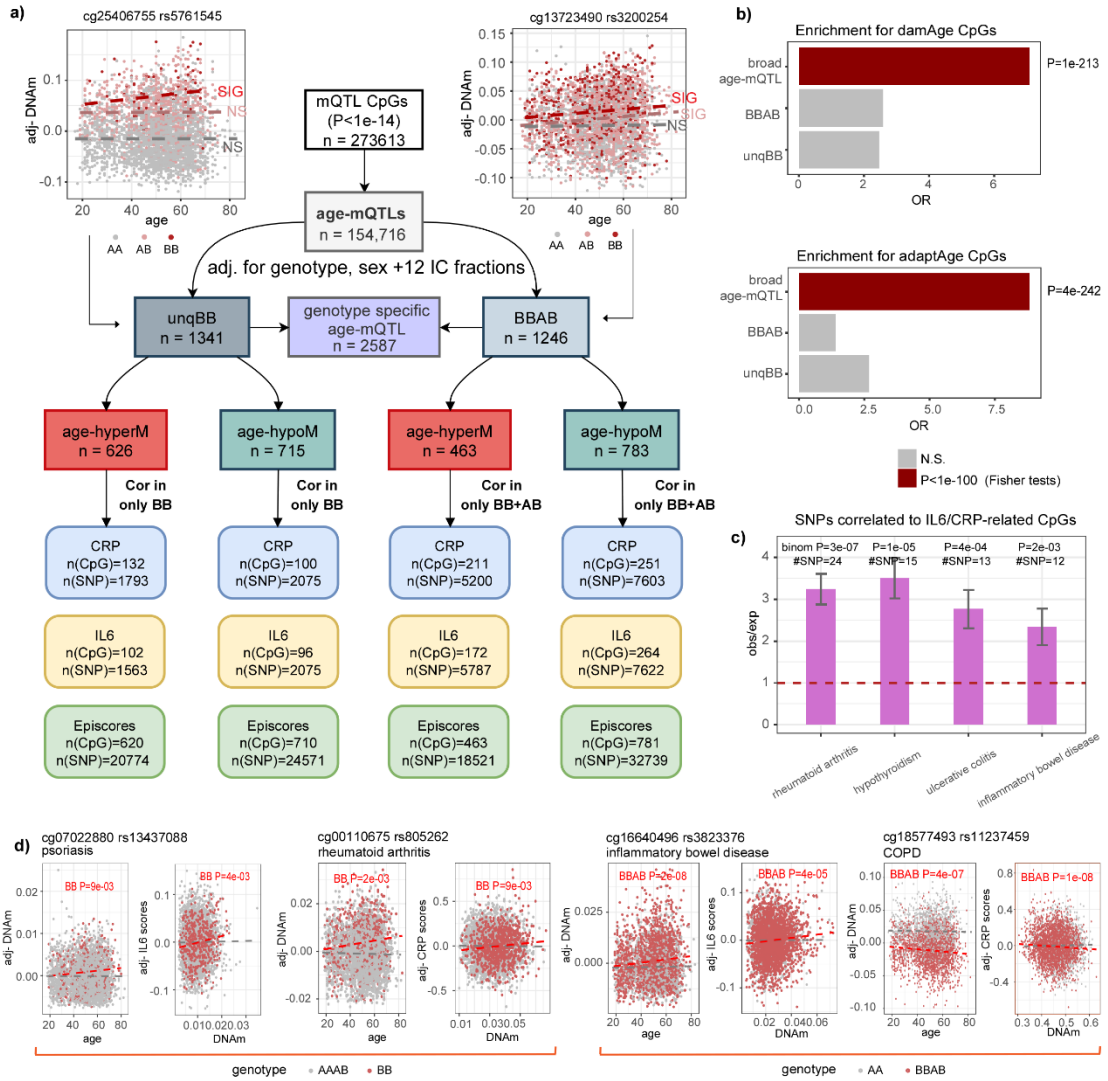
SI fig.S21: NoiseClocks is weakly associated with smoking. Forest plot of the NoiseClock predictor (normalized to unit variance) vs smoking status (0=never, 1=ex-smoker, 2=current-smoker) adjusted for age, sex, race and disease status (if applicable). The number of samples with smoking information is given to the left. P-values from a Fixed and Random Effect inverse variance models are given. We also show the heterogeneity indices and P-value.



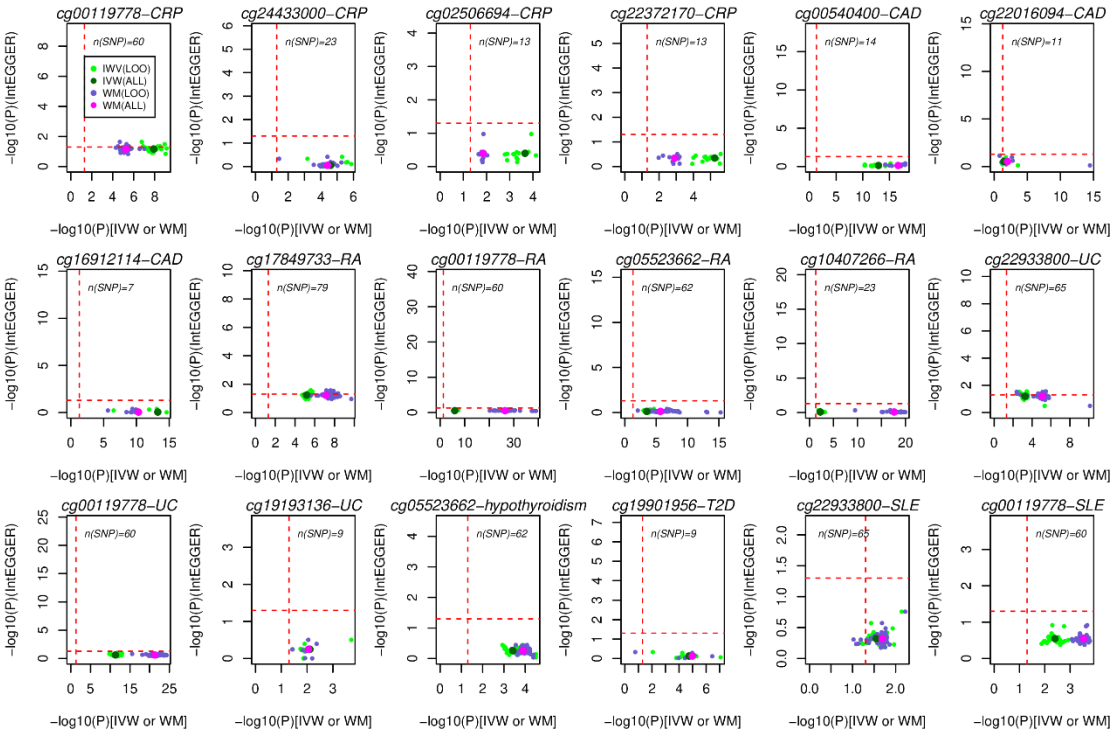
SI fig.S22: NoiseClocks built exclusively from nonAgeHet CpGs in NSPT cohort. a) Flowchart depicting the number of nonAgeHet EPIC CpGs at each iteration level of the ResNet18 classifier and the number of overlapping 450k probes for building the NoiseClocks. **b)** NoiseClock predictive score vs chronological age in the TZH subcohort for the 3 sets of nonAgeHet CpGs (one for each iterative version of ResNet18). R-values and P-values of linear regression are given. **c)** Forest plot of NoiseClock against age, adjusted for sex for each of the 3 NoiseClocks (labelled by ResNet18 iteration version), across 14 whole blood cohorts. Random and Fixed Effect meta-analysis P-values are given at the bottom. At the top of each panel, we give the heterogeneity indices and P-values. **d)** As c), but now a Forest plot of NoiseClock against the intrinsic inflammation score (IntInflScore) adjusted for age, sex and 12 immune cell-type fractions. **e)** Scatterplot between the ResNet18-derived nonAgeHet-based CpG NoiseClocks (y-axis) with the 399 CpG NoiseClock derived using statistical criteria (x-axis), in the TZH cohort. R-values and P-values from linear regression are given. **f)** Heatmap of R-values between the various ResNet18-derived NoiseClocks with the 399 CpG NoiseClock across the 14 whole blood cohorts.



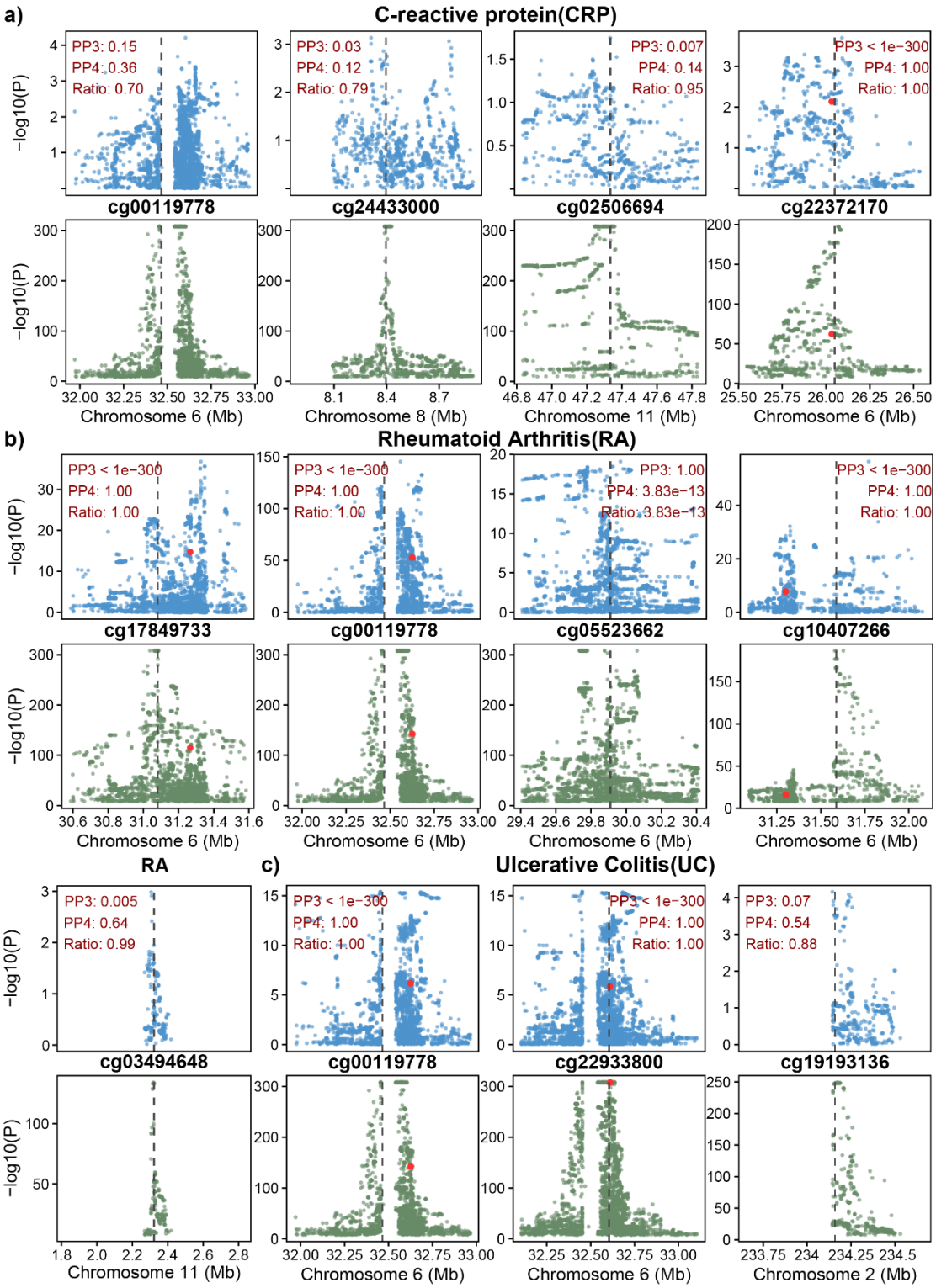
SI fig.S23: Examples of BB and ABBB age-mQTLs with SNPs associated with inflammation-related diseases. a) Four examples of BB age-mQTLs. For each case, left scatterplot displays the adjusted DNAm value against age, with samples stratified by relevant genotype, with the right scatterplot displaying the adjusted CRP or IL6 score vs DNAm. P-values are from a linear regression. **b)** As a) but for BBAB age-mQTLs.



SI fig.S24: Analysis of age-mQTLs adjusting for 12 immune cell fractions. a) Identification flowchart of genotype specific age-mQTLs, which are mQTLs displaying an association with age independently of variations in genotype and 12 IC fractions, and where the association with age is present in only the BB or AB+BB genotypes (B=minor allele). Flowchart indicates the numbers that are hypermethylated and hypomethylated with age and how many of these also display associations with CRP, IL6 or DNAm-based surrogates of other serum protein levels (EpiScores). At the top, two examples are given. **b)** Enrichment barplot displaying odds ratio (OR) of DamAge and AdaptAge causal clock CpGs among all age-mQTLs and the genotype specific ones. P-values are from one-tailed Fisher-tests. **c)** Observed to expected enrichment ratios of SNPs correlated to genotype specific age-mQTLs that also display associations with CRP/IL6 levels, for various inflammatory diseases, as indicated. **d)** Four examples of genotype specific age-mQTLs correlating with IL6/CRP. P-values are from a linear regression.

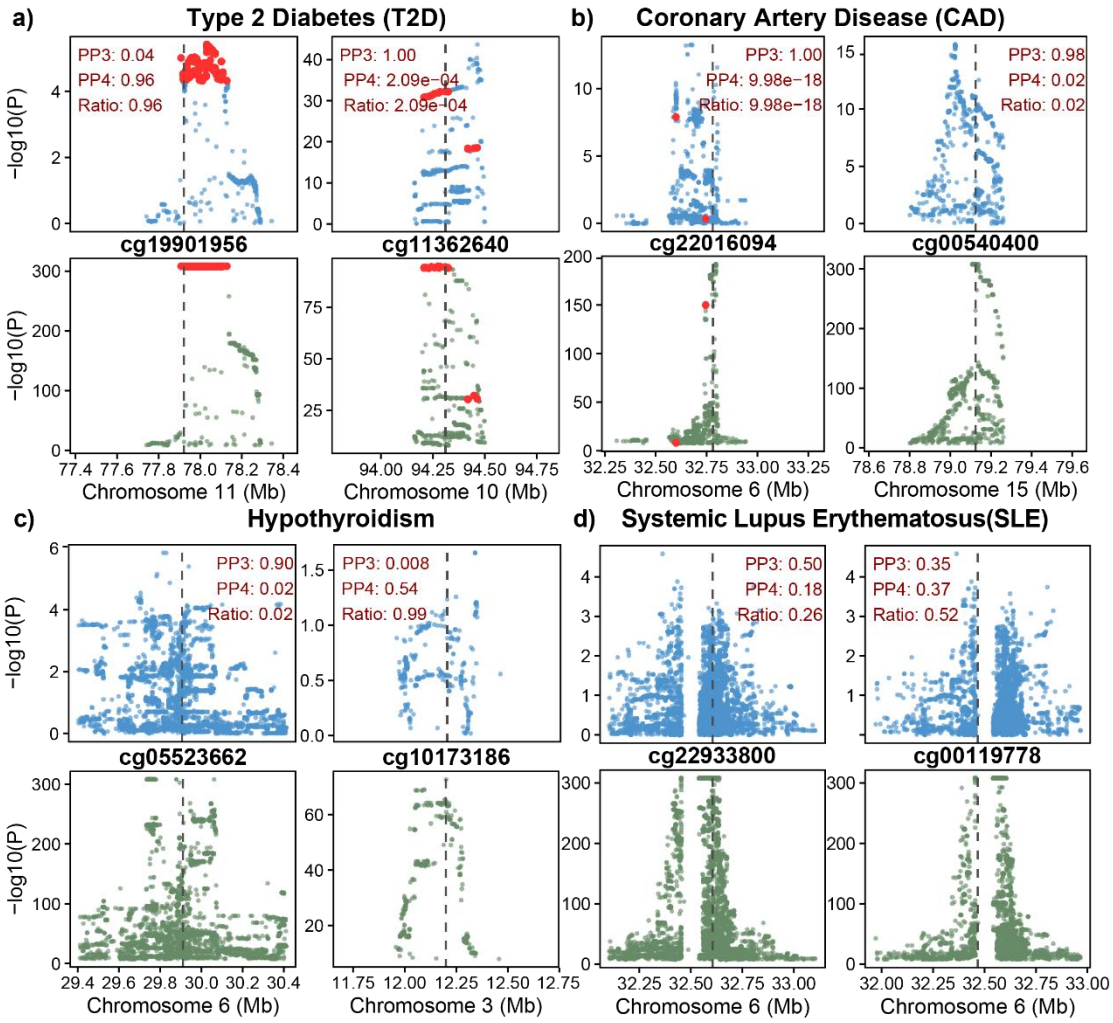


SI fig.S25: MR Leave-One-Out (LOO) analysis. Scatterplots of -log10(P) for the intercept term in MR-Egger vs -log10(P) (slope) for either IVW or WM (shown in different colors, see 1st panel) for all CpG-outcome pairs with more than 3 SNPs. Number of SNPs is indicated in the panel. Red-dashed lines indicate P=0.05. CRP=C-reactive protein, RA=rheumatoid arthritis, CAD=cardiovascular disease, SLE=systemic lupus, T2D=type-2 diabetes, UC=ulcerative colitis.



SI fig.S26: Colocalization analysis of GWAS signals and mQTLs. a) Locus comparison plots for C-reactive protein (CRP) GWAS (upper panels, blue) and mQTL (lower panels, green). Each point represents a single nucleotide polymorphism (SNP), plotted by its genomic position(Mb) against the -log10(P-value) of the association. The vertical dashed grey line indicates the physical position of the index CpG site. Colocalization was evaluated using the Bayesian coloc framework. PP3 denotes the posterior

probability of a model with two distinct causal variants (one per trait), whereas PP4 represents the probability of a single shared causal variant for both DNA methylation and the complex trait. Red points highlight SNPs belonging to the SuSiE-derived credible sets (CS). In instances where SuSiE failed to identify valid credible sets, results from the coloc.abf analysis are displayed instead, characterized by the absence of red-highlighted SNPs. **b-c)** Same as a), but for rheumatoid arthritis (RA) and ulcerative colitis (UC).



SI fig.S27: Colocalization analysis of GWAS signals and mQTLs (part2). a) Locus comparison plots for type-2 diabetes (T2D) GWAS (upper panels, blue) and mQTL (lower panels, green). Each point represents a single nucleotide polymorphism (SNP), plotted by its genomic position(Mb) against the -log10(P-value) of the association. The vertical dashed grey line indicates the physical position of the index CpG site. Colocalization was evaluated using the Bayesian coloc framework. PP3 denotes the posterior probability of a model with two distinct causal variants (one per trait), whereas PP4 represents the probability of a single shared causal variant for both DNA methylation and the complex trait. Red points highlight SNPs belonging to the SuSiE-derived credible sets (CS). In instances where SuSiE failed to identify valid credible sets, results from the coloc.abf analysis are displayed instead, characterized by the absence of red-highlighted SNPs. **b-d)** Same as a), but for coronary artery disease (CAD), hypothyroidism and systemic lupus erythematosus (SLE).