Precise and interpretable neural networks reveal epigenetic signatures of aging in youth across health and disease

David Martínez-Enguita
Linköping University https://orcid.org/0000-0002-6363-6298
Thomas Hillerton
Linköping University
Julia Åkesson
Linköping University
Daniel Kling
Oslo University Hospital
Maria Lerm
Linköping University
Mika Gustafsson
mika.gustafsson@liu.se

Linköping University

Research Article

Keywords: DNA methylation, neural networks, age clock, epigenetic age, aging, youth

Posted Date: August 7th, 2024

DOI: https://doi.org/10.21203/rs.3.rs-4863132/v1

Additional Declarations: The authors declare potential competing interests as follows: DM-E, ML, and MG are founders of PredictMe, a company that provides DNA methylation analysis services. The other authors declare no competing interests

TITLE

Precise and interpretable neural networks reveal epigenetic signatures of aging in youth across health and disease

AUTHORS

David Martínez-Enguita[1], Thomas Hillerton[1], Julia Åkesson[1], Daniel Kling[2], Maria Lerm[3], Mika Gustafsson[1]

[1]Division of Bioinformatics, Department of Physics, Chemistry and Biology, Linköping University, Linköping, Sweden

[2]Department of Forensic Genetics and Toxicology, National Board of Forensic Medicine, Linköping, Sweden

[3]Division of Inflammation and Infection, Department of Biomedical and Clinical Sciences, Linköping University, Linköping, Sweden

KEYWORDS

DNA methylation, neural networks, age clock, epigenetic age, aging, youth

ABSTRACT

DNA methylation age clocks are powerful tools for measuring biological age, providing insights into aging risks and outcomes beyond chronological age. While traditional clocks are effective, their interpretability is often limited due to their dependence on small, potentially stochastic sets of CpG sites. We propose that the reliability of epigenetic age estimation models should stem from their ability to detect comprehensive and targeted aging signatures. Here we introduce NCAE-CombClock, a neural network regressor that improves age prediction accuracy in large datasets by integrating methylation embeddings with CpG sites. Additionally, we developed explainable neural networks for robust age classification across adolescence and young adulthood. Epigenetic aging signatures identified at single-year resolution from these models were enriched in developmental, immune, and metabolic processes. We showcase the utility of our approach by exploring the biological mechanisms underlying the altered pace of aging observed in pediatric Crohn's disease. Our models offer broad applications in personalized medicine and aging research, providing a valuable resource for interpreting aging trajectories in health and disease.

MAIN

Age estimation through DNA methylation (DNAm) age clocks has been a considerably successful application of epigenetic data for modeling purposes. A wide variety of these clocks have been developed, differing in their training approaches, sources of tissue, and study populations[1 2 3 4 5 6 7], even including non-human species[8]. First-generation DNAm clocks aim to accurately track chronological age by leveraging variations in the methylation levels of specific genomic locations (CpG sites) that occur throughout lifetime[9], whereas second-generation age clocks incorporate other biological and clinical parameters to predict all-cause mortality risk[10 11 4]. The divergence between estimated and true chronological age is called epigenetic age acceleration, with a positive value indicating a faster rate of biological aging. An accelerated aging pace has been linked to an increased risk of cardiovascular, musculoskeletal, and neurodegenerative disorders [12 13 4], as well as general adverse health outcomes, such as frailty, cognitive decline, or immune dysregulation[14 15]. However, while changes that slow epigenetic aging often correlate with increased longevity, recent research demonstrated that interventions extending cellular lifespan do not consistently affect epigenetic aging rates[16]. Altogether, this emphasizes the premise that aging is the distinct, deterministic, and malleable consequence of genetic, environmental, and stochastic mechanisms operating and interacting in parallel[17 18].

The importance of precise age estimation is particularly pronounced in adolescents and young adults due to the profound biological changes occurring during these stages, which shape the future health trajectory of an individual. During childhood and adolescence, DNAm patterns undergo significant modifications, accelerating epigenetic

aging as part of natural development[19 20]. These changes can be further influenced by adverse conditions such as disease, physical or emotional abuse, familial instability, and economic hardship, which can expedite the aging process[21 22]. Consequently, providing accurate age estimates for individuals in their adolescence and young adulthood is highly relevant in areas such as developmental healthcare, forensic science, criminology, and the study of population dynamics[23 24 25]. DNAm age clocks can help identify individuals with developmental disorders or those at risk of accelerated aging due to environmental stressors, enabling timely interventions to mitigate these effects. They can also provide assistance in legal and criminal proceedings, and inform public health strategies and policies about the aging patterns of populations exposed to different socioeconomic conditions. This information may allow for the adequate allocation of resources, infrastructure, and services to improve the well-being of these groups.

Despite the high accuracy of existing DNAm age clocks, their role as predictors of biological aging or lifespan is hindered by their limited interpretability, even when trained using extensive clinical and omics data[26]. Current clocks rely on disparate sets of CpG sites with minimal overlap and ambiguous association with risk factors of aging, which has led to concerns about their signal-to-noise ratios[27 28 29]. The question has recently been raised of whether the precise age estimates generated by these clocks primarily reflect stochastic DNAm changes rather than true biological aging processes, partly since these models are often trained on relatively small and homogeneous populations[18 5]. Furthermore, traditional methods for CpG site selection like elastic net omit large portions of the epigenome, potentially missing critical aging markers. While approaches

using principal components of DNAm data have been shown to enhance reliability[26 29], they are inherently limited to modeling linear effects and may overlook the complex, interconnected nature of aging mechanisms. Instead, incorporating data from diverse, well-annotated datasets within a non-linear framework, such as autoencoder embeddings, may enhance the generalizability and robustness of age clocks by better capturing subtle patterns and interactions in DNAm data.

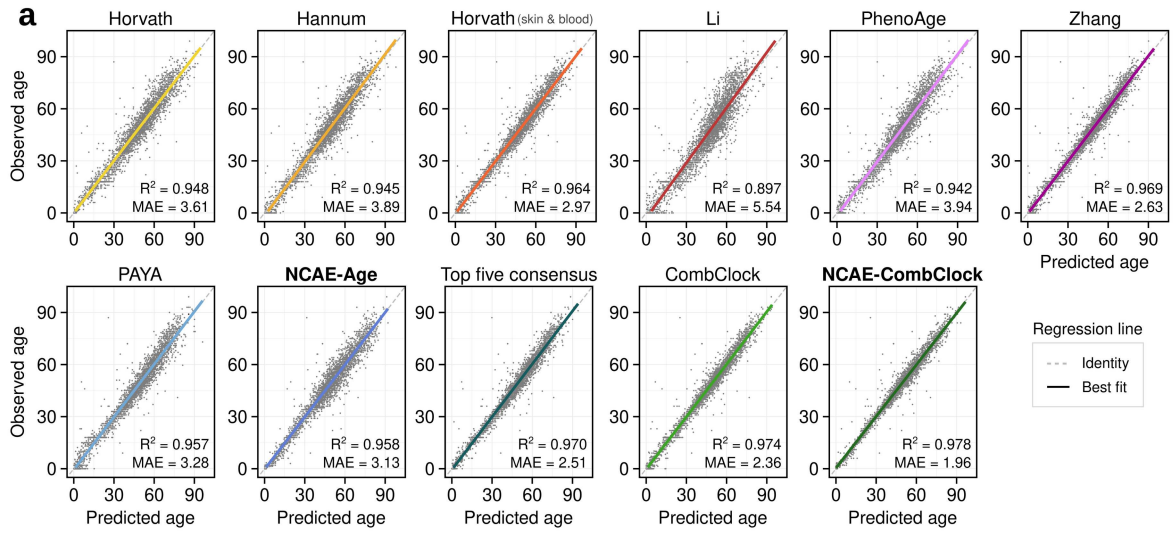
Here, we hypothesized that the reliability and accuracy of an epigenetic age estimation model should be grounded in its ability to effectively detect comprehensive and target-specific DNAm signatures of aging. We propose that leveraging a compressed representation of DNAm data, in the form of autoencoder embeddings, as input for explainable neural network models may optimally balance thoroughness and efficiency. By employing interpretable, data-driven methods that utilize markers across the entire measurable human methylome, we ensure that the model serves not only as a predictive tool but also as a valuable informative resource. Our approach aims to improve the robustness of age estimation by capturing fine-grained signatures of age-related DNAm variation throughout the human lifespan, with particular focus on critical developmental stages such as adolescence and young adulthood, and even at single-year resolution. In doing so, we seek to uncover the underlying biological processes driving the pace of aging and age-related health trajectories and vulnerabilities, thereby advancing our understanding and facilitating applications in healthcare and beyond.

RESULTS

A deep neural network leveraging DNA methylation embeddings and CpG sites yields highly accurate age estimates

Initially, we evaluated the regression performance and generalizability of existing DNAm age clocks using large-scale datasets. We processed 17,726 publicly available whole blood DNAm samples from individuals aged 0 to 112 years, with a mean (SD) age of 41.0 (22.9) years (**Supplementary File 1**), and divided them into training and test sets (80:20). We then conducted a literature search for DNAm age clocks trained on whole blood samples with a known set of CpG sites. We selected the clocks Horvath pan-tissue[1], Horvath skin and blood[2], Hannum[3], PhenoAge[4], Zhang (elastic net)[5], Li[6], and PAYA[7], which use elastic net for feature selection and include between 71 and 514 CpGs (**Supplementary File 2**). Additionally, we included our NCAE-Age[30] model, a supervised deep neural network (DNN) age predictor that uses a compressed representation of the human methylome consisting of 128 variables (embeddings) from an autoencoder with a biologically relevant latent space, named network-coherent autoencoder (NCAE)[30]. NCAE-Age clocks can be developed for multiple tissues or cell types, including whole blood, lymphocytes, buccal epithelial cells, and saliva (**Supplementary Table 1**).

We observed that the selected DNAm age clocks were able to deliver accurate estimates of chronological age (**Fig. 1a-b, Supplementary File 3**) for the test set samples (n = 3,502). Elastic-net-based clocks achieved coefficient of determination (R^2) values between R^2 = 0.897 for Li's clock (mean absolute error, MAE = 5.54 years; root mean square error, RMSE = 7.34 years), optimized for age prediction in children and



b

Age estimation performance of DNAm age clocks on whole blood samples
(test set, n = 3,502 controls, 0 to 112 years, mean age = 41.0 ± 22.9 years).

Model	R^2	MAE	MedAE	RMSE	Pearson *r*	Spearman ρ
NCAE-CombClock	0.978	1.96	1.21	3.37	0.989	0.986
CombClock	0.974	2.36	1.75	3.70	0.987	0.984
Top five consensus	0.970	2.51	1.75	3.94	0.985	0.982
Zhang et al.	0.969	2.63	1.95	4.05	0.984	0.981
Horvath (skin & blood)	0.964	2.97	2.20	4.37	0.982	0.977
NCAE-Age	0.958	3.13	2.08	4.68	0.979	0.971
PAYA	0.957	3.28	2.50	4.74	0.978	0.976
Horvath	0.948	3.61	2.71	5.21	0.974	0.969
Hannum	0.945	3.89	3.07	5.37	0.972	0.968
PhenoAge	0.942	3.94	3.00	5.52	0.970	0.967
Li et al.	0.897	5.54	4.44	7.34	0.947	0.942

MAE: mean absolute error. MedAE: median absolute error. RMSE: root mean square error.

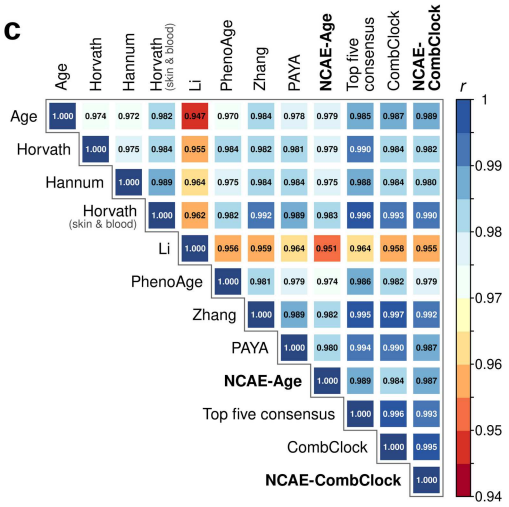


Fig. 1: Regression performance benchmark of DNA methylation age clocks trained on whole blood samples from healthy individuals. a Scatter plots comparing observed ages versus predicted ages for eleven different DNA methylation age estimation models on regression test set samples (n = 3,502 healthy controls, ages 0 to 112 years). The coefficient of determination (R^2) and mean absolute error (MAE) values are indicated for each model. **b** Summary table of age estimation performance metrics for DNAm age clocks evaluated on the regression test set. **c** Correlation matrix of Pearson correlation coefficients (PCC) between observed ages and predicted ages

across various DNA methylation age estimation models. Higher values, shown in dark blue, indicate stronger correlations.

adolescents, to R^2 = 0.969 for Zhang's clock (MAE = 2.63 years, RMSE = 4.05 years), designed as a "near-perfect" age estimator across the entire age range. Horvath's skin and blood clock performed better (R^2 = 0.964, MAE = 2.97 years, RMSE = 4.37 years) than its pan-tissue counterpart (R^2 = 0.948, MAE = 3.61 years, RMSE = 5.21 years). The embedding-based NCAE-Age clock produced an R^2 of 0.958 (MAE = 3.13 years, RMSE = 4.68 years), akin to PAYA (R^2 = 0.957, MAE = 3.28 years, RMSE = 4.74 years). Notably, we found that taking the median of the predictions from the five top-performing clocks resulted in a consensus estimate that outperformed every other DNAm age clock until that point, with an R^2 of 0.970 (MAE = 2.51 years, RMSE = 3.94 years). Therefore, we proposed to identify the upper boundary of our benchmark data in terms of age prediction performance. Building on this wisdom-of-the-crowds approach, we combined the 1,743 unique CpG sites used by the selected elastic-net-based clocks to train a novel elastic net regressor model, named CombClock. Reassuringly, CombClock generated age predictions with a top test R^2 of 0.974, lowering the MAE to 2.36 years and the RMSE to 3.70 years. Thus, we inquired whether additional performance could be gained by leveraging these 1,743 CpG sites plus the 128 NCAE embeddings used by NCAE-Age. For this purpose, we developed a high-precision hybrid model, which we refer to as NCAE-CombClock, trained as a supervised DNN for age prediction with both CpG sites and embeddings as input features. NCAE-CombClock achieved the highest test set performance of the benchmarked models (R^2

= 0.978), with an MAE below two years (MAE = 1.96 years), median absolute error close to one year (MedAE = 1.21 years) and RMSE of 3.37 years.

Comparing the ages estimated by the DNAm age clocks, we obtained pairwise Pearson correlations coefficients (PCC) between 0.947 and 0.997 (**Fig. 1c**). CombClock and Zhang predictions were the most similar (PCC = 0.997) due to Zhang being the largest contributor of CpG sites to CombClock, with 514 shared CpGs. Estimated ages from elastic-net-based DNAm clocks were less correlated to NCAE-CombClock predictions than to CombClock ones. For example, the NCAE-CombClock model had slightly lower correlation with Zhang's clock (PCC = 0.992) than CombClock. Likewise, predictions from the DNN-based NCAE-Age showed a higher correlation with those from the DNN-based NCAE-CombClock (PCC = 0.987). Estimates from NCAE-CombClock were the closest to the true ages (PCC= 0.989), above those from CombClock, Zhang, and Horvath skin and blood (PCC = 0.987, 0.984, and 0.982, respectively). In summary, we showed that the use of large-scale training sample sets coupled with the combination of knowledge-driven CpG sites and data-driven autoencoder DNAm embeddings enabled the development of a highly precise neural network age regressor, named NCAE-CombClock, that outperformed all other tested whole blood DNAm age clocks.

A consensus of DNAm age clocks optimizes threshold-based age determination of adolescents and young adults

Next, we investigated whether a high performance in predicting chronological age from DNAm data translates into an accurate determination of an individual being below or

above specific age thresholds. We established a classification benchmark for our DNAm age clocks, focusing on three critical points of adolescence and young adulthood: 15, 18, and 21 years. These thresholds mark key transitional stages of youth development with significant implications in legal, health, and socioeconomic contexts. Importantly, they allow to assess the practical applicability of DNAm age clocks in real-world scenarios where precise age classification is crucial. We thus aimed to identify the most reliable model for accurately categorizing samples by these age thresholds.

Hence, we applied a five-fold cross-validation approach to generate age predictions for a young-age subset of 1,404 whole blood samples from healthy individuals with ages from 10 to 30 years, with mean (SD) age of 21.8 (5.3) years (**Supplementary File 1**). To mitigate the impact of the varied age ranges in the original training data for our DNAm age clocks, we retrained them, if applicable, in two different settings: young-age (10 to 30 years) and complete-age (1 to 101 years). Within the young-age setting, DNAm age clocks were retrained using four out of five folds of the young-age subset. Conversely, a cohort of 8,846 whole blood samples from healthy individuals with ages from 1 to 101 years (complete-age, mean (SD) age of=8.6 (19.4) years) was included besides the four young-age folds (**Supplementary File 1**). We generated age predictions for the left-out young-age fold until the entire set had been processed, for every clock and training setting. Age estimates were binarized depending on whether they were lower (0), or equal to or higher (1) than the thresholds of 15, 18, and 21 years. To address class imbalance and since this method produces class predictions rather than probability estimates, we opted to evaluate the classification performance using the average F1-score across folds.

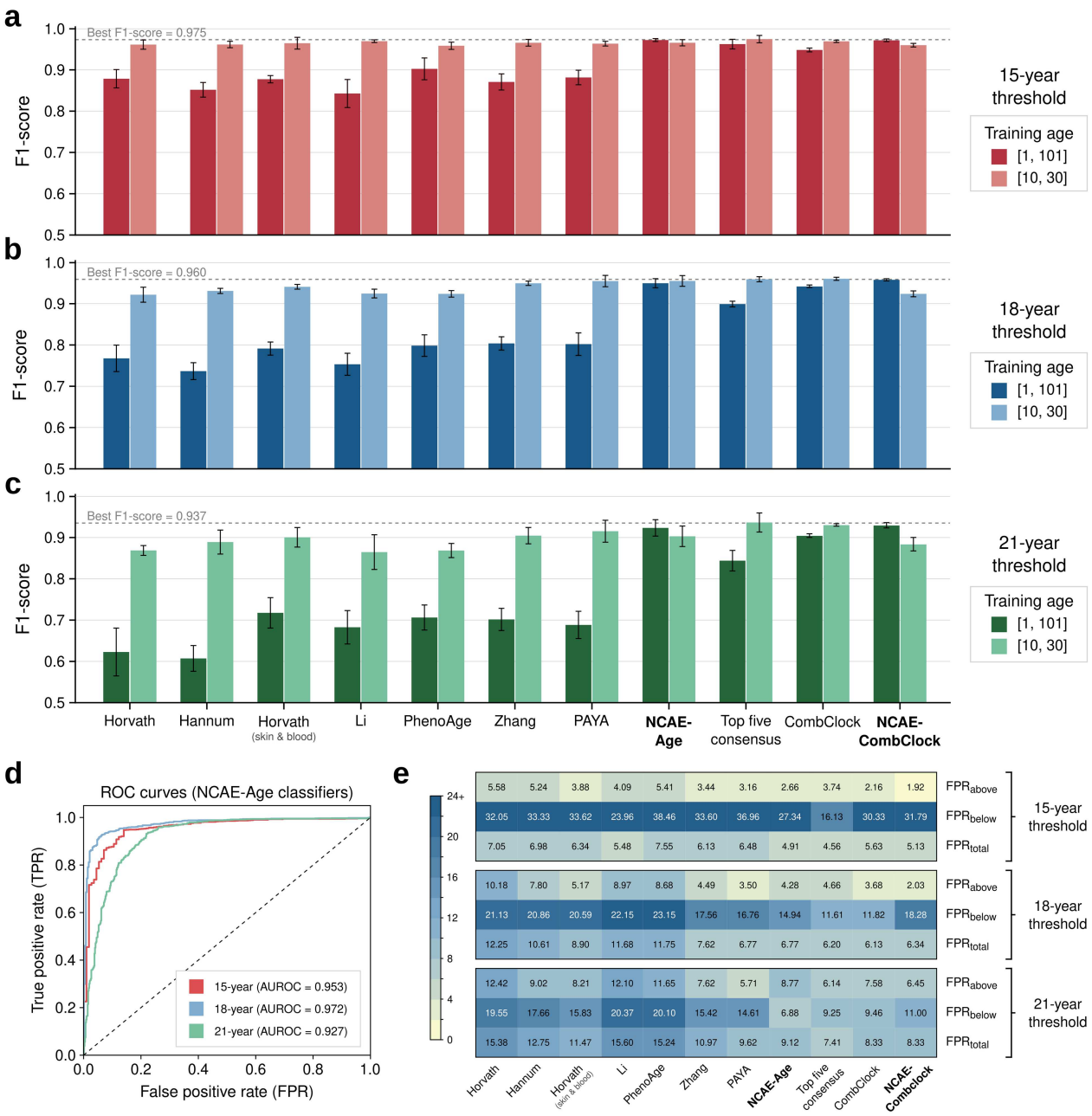


Fig. 2: Classification performance benchmark of DNA methylation age clocks. a-c Bar plots of classification performance (F1-score) of DNAm age clocks based on their capacity to accurately determine if an individual's chronological age is equal to or above the target age cutoffs of 15, 18, and 21 years. DNAm age clocks were evaluated on classification test set samples (n = 1,404 healthy controls, ages 10 to 30 years) divided

into five folds. Age clocks were retrained, if applicable, in complete-age (dark color, ages 1 to 101 years) or young-age (light color, ages 10 to 30 years) training settings. The best average F1-score per target age is indicated and marked as a grey dotted line. **d** Receiver operating characteristic (ROC) curves for NCAE-Age classifiers with cutoffs of 15, 18, and 21 years, evaluated on the classification test set. **e** Heatmaps of false positive rates (FPR) for DNAm age clock estimations above or below the target age cutoffs of 15, 18, and 21 years, and in total, considering the best-performing training setting in each case. Values indicating a higher FPR are shown in dark blue.

We found that the benchmarked DNAm age clocks can accurately classify samples based on our specific young-age thresholds, when trained under an adequate age setting (**Fig. 2a-c**, **Supplementary File 3**). Published elastic-net-based clocks were the most sensitive to the age range of their training data, underperforming on complete-age (mean F1-scores between 0.607 and 0.903) compared to young-age settings (F1-score between 0.865 and 0.970). CombClock was only marginally affected when retrained on complete-age, decreasing in performance but still above F1-score = 0.900 in all thresholds. On the contrary, the DNN-based clocks NCAE-Age and NCAE-CombClock achieved better results in general on complete-age, with F1-scores between 0.924 (21-year, NCAE-Age) and 0.973 (15-year, both). Overall, classification performance decreased with higher threshold ages. The best clock was the top five consensus of young-age models, with the highest F1-scores for the 15-year (F1-score = 0.975) and 21-year (F1-score = 0.937) thresholds, while the latter and CombClock (F1-score = 0.960) were tied at the 18-year category. We found significant correlations between the

R^2 of DNAm age clocks in the regression benchmark and F1-scores on best-performing classification settings at the 18-year (PCC = 0.753, p = 7.43e-3) and 21-year (PCC = 0.793, p = 3.61e-3) thresholds, but not at the 15-year one (PCC = 0.229, p-value = 0.499). We observed that by modifying the training objective and loss function of NCAE-Age clocks, they can also work as binary classifiers and generate probability estimates for a sample to be equal to or older than a certain age cutoff. NCAE-Age classifiers trained under young-age settings performed with area under the receiver-operating curve (AUROC) values of 0.953 for 15-year, 0.972 for 18-year, and 0.927 for 21-year thresholds (**Fig. 2d**). Regarding misclassification ratios measured at best-performing settings (**Fig. 2e**), the top five consensus had the lowest total false positive rate (FPR) at the 15-year (FPR_{total} = 4.56%) and 21-year (FPR_{total} = 7.41%) categories, while CombClock was marginally better at the 18-year category (FPR_{total} = 6.13% versus 6.20%). Most misclassified samples were due to clocks estimating a lower-than-true age, leading to high FPRs below the thresholds: 16.13 to 38.46% (15-year), 11.61 to 23.15% (18-year), and 6.88 to 20.37% (21-year).

An array of single-year NCAE-Age classifiers provides high-resolution young age probability estimation

Our findings up to this point showed that DNAm age clocks can be optimized to serve as quasi-perfect regressors of chronological age, enabling accurate classifications of individuals based on thresholds of interest. However, even the top age regressor (NCAE-CombClock) and the top age classifier (consensus of clocks) exhibited non-negligible misclassification rates. Given the critical importance of precise identifications

in various real-world contexts, a more refined approach is required. We argued that an array of classifiers, each trained at a single-year cutoff, could collectively provide an ensemble of age probability estimates that might improve the robustness and reliability of age assessments at single-year categories.

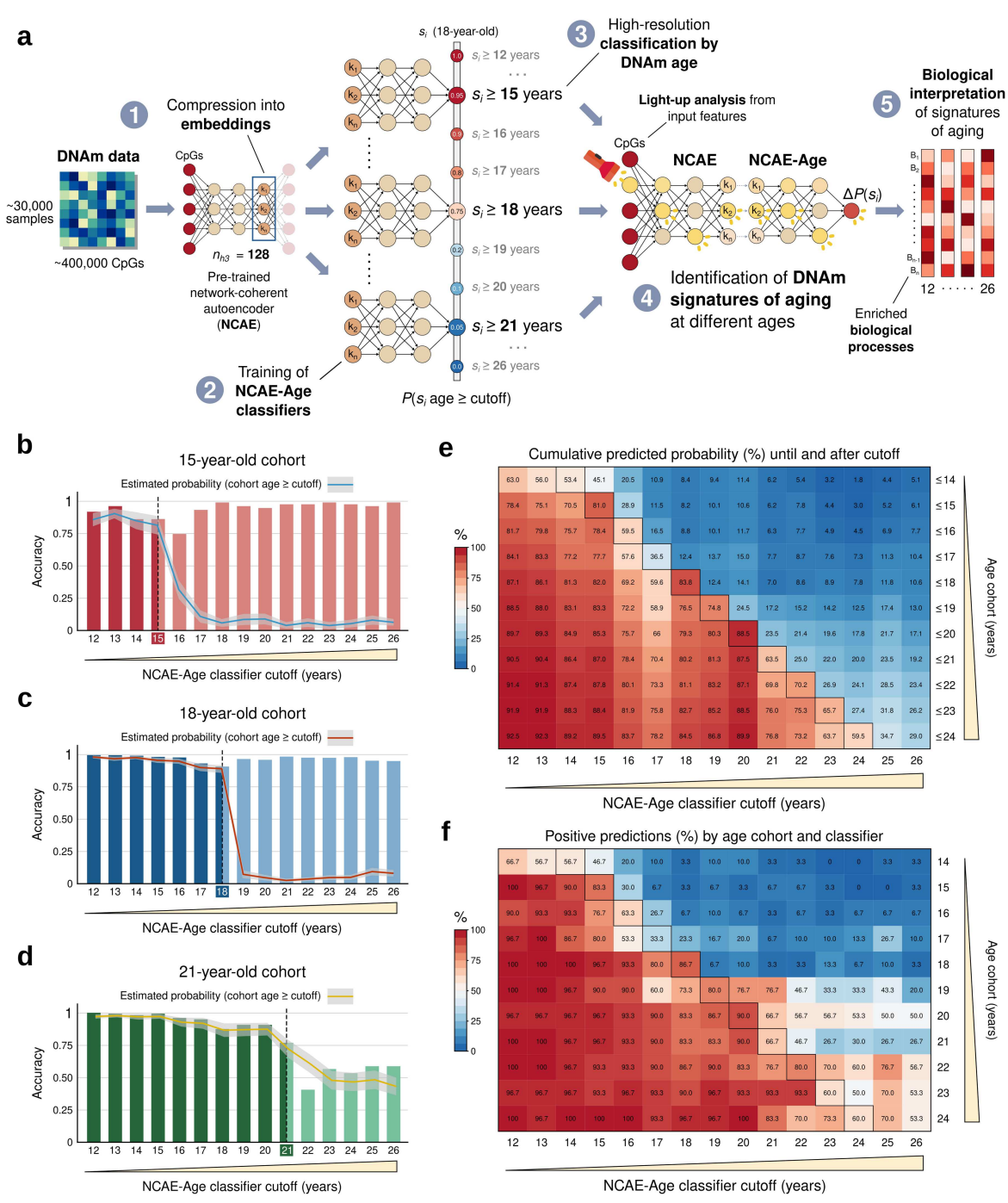


Fig. 3: Workflow summary and performance evaluation of NCAE-Age classifiers by model cutoff and target age. a Graphical summary of the training, evaluation, and biological interpretation of an array of NCAE-Age neural network classifiers from 128-dimensional network-coherent autoencoder (NCAE) embeddings of DNAm data. **b-d** Bar plots of binary classification accuracy of NCAE-Age classifiers for the target age cutoffs of 15, 18, and 21 years. The overlaying lines indicate the probabilities estimated by each model for the cohort samples to be equal to or above the model cutoff, within a 95% confidence interval. The cutoff corresponding to the true age of the cohort is marked with a dark square. **e** Heatmap of cumulative predicted probabilities (%) for the balanced test set cohorts (n = 330, ages 14 to 24 years) to be equal to or above each NCAE-Age classifier cutoff, calculated until or after the cutoff. **f** Heatmap of positive predictions (%, estimated probability > 0.5) per NCAE-Age classifier and balanced test set cohort. Outlined cells indicate that the true age of the single-age test set cohort is equal to the model cutoff. Higher percentages are shown in dark red, lower percentages are shown in dark blue. Age cohorts are sex-balanced groups of 30 healthy controls of the same age, randomly selected from left-out data.

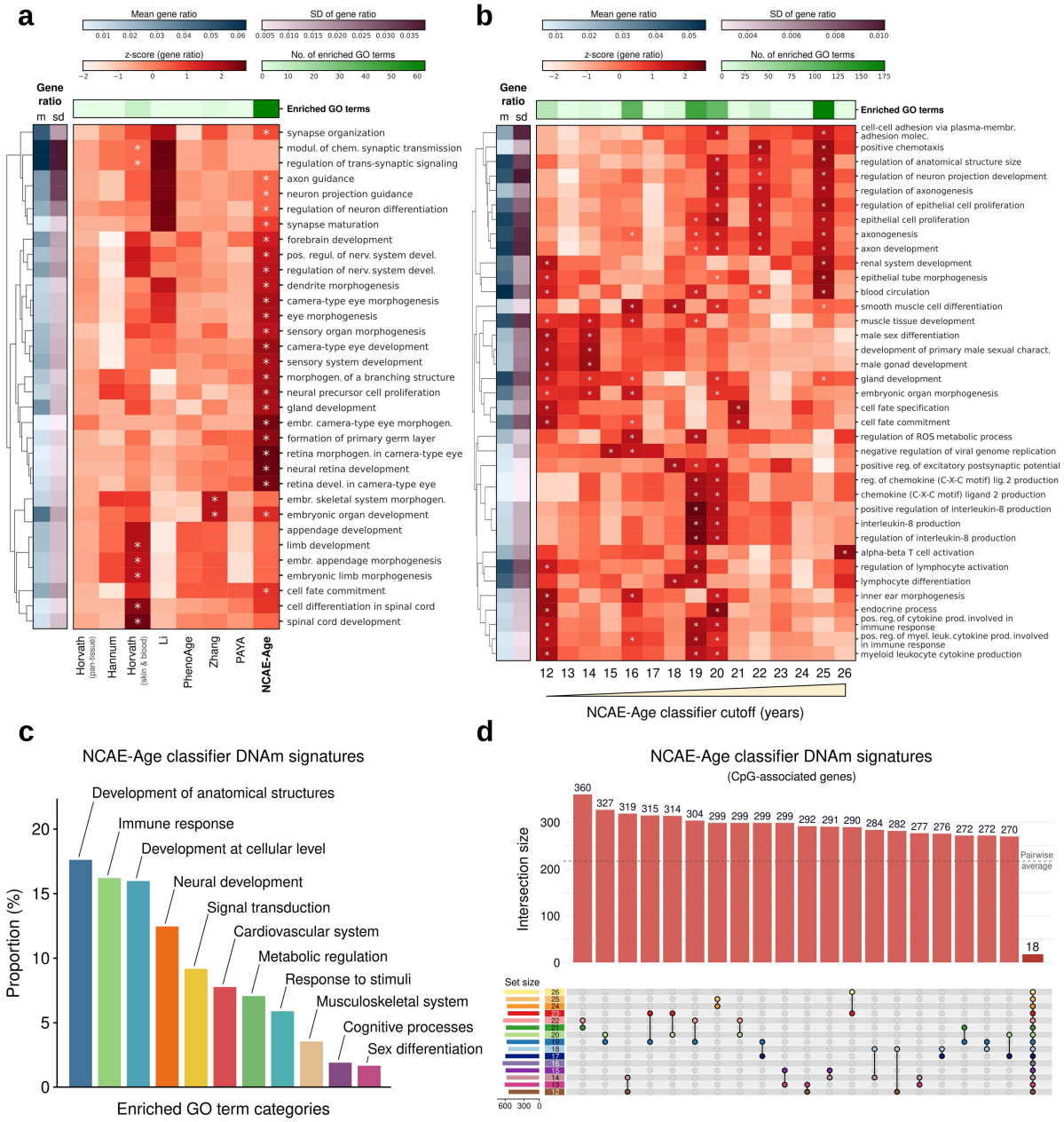
As seen in the classification benchmark, NCAE-Age clocks can be trained as classifiers with specific age cutoffs. Thus, we trained fifteen NCAE-Age classifiers with single-year cutoffs spanning through adolescence and young adulthood (12 to 26 years), using the young-age sample set (**Fig. 3a**). We evaluated their classification performance using an age- and sex-balanced test set composed by eleven single-age cohorts of 30 healthy controls each (total n = 330, mean age (SD) = 19.1 (3.1) years, 48.7% female) with an

age range of 14 to 24 years. First, we applied the NCAE-Age array to generate probability estimates for DNAm samples from cohorts of the same age as the previous thresholds (15, 18, and 21 years) (**Fig. 3b-d, Supplementary File 4**). We determined a prediction to be positive if the classifier estimated a probability over 0.5 for a sample's age to be equal to or above its target cutoff and calculated the accuracy per cutoff and cohort. For example, an accurate classifier with a cutoff of 14 years evaluating an 18-year-old sample should estimate a probability of at least over 0.5, indicating that it predicts that the sample's age is above its cutoff. We found that all single classifiers achieved highly accurate classifications for the 18-year (0.90-0.99) and 15-year (0.74-0.99) cohorts, and for 21-year cohorts until the 21-year cutoff (0.77-1.00). The accuracy dropped to 0.40-0.59 when determining whether 21-year-olds were older than 22-26 years. Predicted probabilities sharply decreased, as expected, when the cohort age became larger than the NCAE-Age cutoff. This can also be seen when considering the accumulated probabilities until or after each cutoff (**Fig. 3e**), showcasing that the use of the complete classifier array leads to optimal classifications. By percentage of positive predictions (**Fig. 3f**), 18-year-cutoff classifiers predicted 86.7% of 18-year samples to be equal to or above 18, down to 23.3% for 17-year, 6.7% for 16-year, and 3.3% on 15- and 14-year cohorts. Other NCAE-Age classifiers showing an above-average performance were the 15-year (83.3% for 15-year), 19-year (80.0% for 19-year, only 6.7% for 18-year cohort), and the 22-year (80.0% for 22-year cohort). Cohorts between 20 and 24 years were harder to classify at the same granularity. In sum, NCAE-Age classifiers enabled the precise categorization of individuals at single-year resolution, with particularly high accuracy for cohorts from 14 to 19 years.

Interpretable NCAE-Age clocks predict age by capturing developmental, immune, and metabolic mechanisms

We then analyzed the potential of the NCAE-Age classifier array as a data-driven tool capable of extracting meaningful epigenetic signatures of aging. We proposed that for the estimates of a DNAm age clock to be a reliable reflection of the true epigenetic aging status of an individual, the model must effectively capture a wide range of DNAm markers linked to age-associated biological mechanisms. By using interpretable models with specific age cutoffs and that leverage a larger portion of the methylome, rather than clocks with a fixed set of CpG sites, we could identify novel DNAm signatures that describe aging processes throughout key development stages, such as adolescence and young adulthood, on a year-to-year basis.

Thus, we inspected the functional associations of the CpG sites used by DNAm age clocks and the epigenetic signatures from NCAE-Age models. First, we retrieved and annotated the lists of CpGs selected by each age clock (ranging from 71 CpGs for Hannum's clock to 514 CpGs for Zhang's model, **Supplementary File 2**) to genes. While small overlaps existed, no single site was common across all the sets of CpGs from the studied clocks. For NCAE-Age models, we applied the light-up technique (Methods) for the interpretation of pre-trained supervised DNNs in terms of their input feature importance. This allowed to retrieve the contribution of original CpG sites to the training objective (age estimation), selecting the top 1,000 sites as the DNAm signature of aging for every NCAE-based model (**Supplementary File 5**). We examined the biological context of the age clock signatures by calculating the significance of their



359 **Fig. 4: Functional enrichment analysis of CpG sites from DNAm age clocks and**
360 **NCAE-Age DNAm signature genes. a** Heatmap of gene ratio z-scores of the top
361 significantly enriched Gene Ontology (GO) terms for CpG-associated genes from DNAm
362 age clocks and NCAE-Age DNAm signatures of aging. **b** Heatmap of gene ratio z-
363 scores of the top enriched GO terms for age-specific DNAm signatures of aging from
364 NCAE-Age classifiers with cutoffs between 12 and 26 years. Significantly enriched

(FDR-adjusted P < 0.05) GO terms are marked with an asterisk (*). Terms are clustered by their gene ratios across enrichments. The mean and standard deviation of the gene ratio (left) and the total number of significantly enriched GO terms (top) are displayed in each panel. **c** Bar plot of the proportion (%) of significantly enriched GO terms (total n = 426 terms) across NCAE-Age DNAm signatures of aging across classifier cutoffs, grouped into eleven common GO term ancestor categories. **d** Upset plot of non-exclusive pairwise overlaps between CpG-associated genes from DNAm signatures of aging from NCAE-Age classifiers (ages 12 to 26 years) in decreasing size order. The size of each signature is displayed as a horizontal bar. Genes overlapping across every signature (n = 18 genes) are shown as the last bar.

enrichments in Gene Ontology (GO) terms from the Biological Process category. We found significant enrichments (FDR-adjusted P < 0.05) for CpGs for Horvath skin and blood (8 GO terms, FDR-adj. P = 1.11e-2 to 4.49e-2) and Zhang (2 GO terms, FDR-adj. P = 2.65e-2 to 4.57e-2) clocks. Other CpG sets were not significantly associated with any term (**Fig. 4a**, **Supplementary File 7**) and even the combined 1,743 CpGs used by CombClock were only enriched in 32 GO terms (**Supplementary File 7**). By contrast, the DNAm signature of the NCAE-Age regressor was significantly enriched in a total of 63 GO terms (FDR-adj. P = 9.37e-4 to 4.93e-2). Out of them, 31 terms (49.2%) were related with the morphogenesis and development of anatomical structures, such as the sensory, hepatic, and urinary systems; and 26 terms (41.3%) described stages of the neural system development, including the regulation of neurogenesis, synapse organization and maturation, and forebrain and telencephalon development. Other

388 enriched GO terms were cell fate commitment and specification, locomotory behavior,
389 response to reactive oxygen species, and regulation of signal transduction via small
390 GTPases (**Supplementary File 7**).

391 To increase the resolution of the analysis to single years, we extracted and examined
392 the GO enrichment of DNAm signatures from every NCAE-Age classifier from 12 to 26
393 years (**Fig. 4b**, **Supplementary File 5**). A total of 426 GO terms were significantly
394 enriched (FDR-adjusted $P < 0.05$) across signatures (**Supplementary File 7**). These
395 included biological mechanisms associated with the development of anatomical
396 structures (75/426, 17.6%) such as organ morphogenesis and growth regulation;
397 immune response processes (69/426, 16.2%) such as leukocyte activation and cytokine
398 production; developmental processes at the cellular level (68/426, 16.0%) such as
399 regulation of cell fate specification, commitment, differentiation, and proliferation; neural
400 development mechanisms (53/426, 12.4%) such as axonogenesis and synapse
401 maturation; signal transduction pathways (39/426, 9.2%); and terms linked with the
402 cardiovascular system (33/426, 7.7%), metabolic regulation (30/426, 7.0%), response to
403 stimuli (25/426, 5.9%), musculoskeletal system (15/426, 3.5%), cognitive and behavioral
404 processes (8/426, 1.9%), and sex differentiation (7/426, 1.6%), among others (**Fig. 4c**).
405 Signatures generally overlapped highly between each other, but only 18 CpG-
406 associated genes were shared among all of them (**Fig. 4d, Supplementary File 5**).
407 Remarkably, we observed that while enrichments in GO terms linked with general
408 development at both organismal and cellular levels were spread throughout the entire
409 range, others were concentrated at specific ages. For example, DNAm signatures
410 between 16 and 20 years were highly enriched in cognitive processes. Similarly, neural

411 development signals were mainly detected at 16 years and above, sex differentiation
412 between 12 and 14 years, immunoregulatory mechanisms between 19 and 20 years,
413 and musculoskeletal development in the range from 12 to 19 years.

414 To further increase the granularity of the analysis, we trained a variation of NCAE-Age
415 models that leveraged DNAm embeddings including an additional 10,619 CpG sites
416 from the X and Y chromosomes, named NCAE-XY-Age models (**Supplementary File**
417 **2**). CpG sites located within sex chromosomes are usually removed from DNAm
418 workflows at the preprocessing stage to avoid sex-dependent biases. Here, we
419 reasoned that developmental changes occurring during youth could be better modeled if
420 these DNAm markers are also considered. NCAE-XY-Age regressors achieved similar
421 benchmark results as NCAE-Age models in whole blood (R^2 = 0.958, MAE = 3.12 years,
422 RMSE = 4.67 years) (**Supplementary File 3**) as well as across various other tissues
423 (**Supplementary Table 2**). Their DNAm signature of aging was enriched for 113 unique
424 GO terms (**Supplementary Fig. 1, Supplementary File 7**). Subsequently, we applied
425 the same approach as before (**Fig. 3a**) to train fifteen NCAE-XY-Age classifiers with
426 single-year cutoffs from 12 to 26 years using the young-age sample set and retrieve
427 their age-specific DNAm signatures of aging (**Fig. 5a-b**, **Supplementary File 6**). We
428 found 521 significantly enriched GO terms across signatures, out of which 172
429 overlapped with the 426 enriched terms from NCAE-Age (**Supplementary File 7**).
430 Compared with NCAE-Age results (**Fig. 5c**), NCAE-XY-Age signatures presented a
431 higher proportion of biological mechanisms associated with the general development of
432 anatomical structures (128/521, 24.6% versus 17.6%), with the musculoskeletal system
433 (36/521, 6.9% versus 3.5%), and with sex differentiation processes (16/521, 3.1%

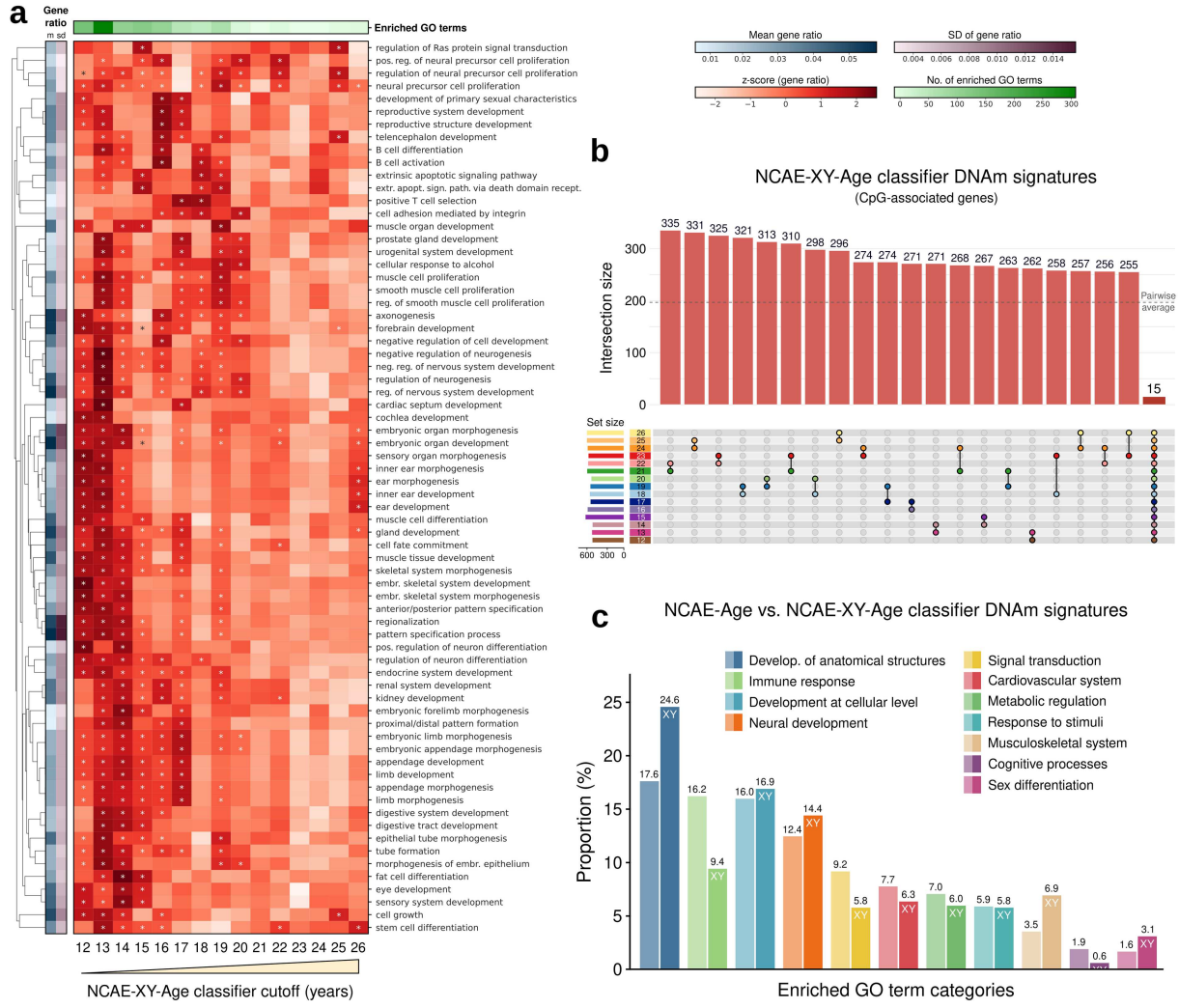


Fig. 5: Functional enrichment analysis of CpG sites from NCAE-XY-Age DNAm signature genes. a Heatmap of gene ratio z-scores of the top enriched GO terms for age-specific DNAm signatures of aging from NCAE-XY-Age classifiers with cutoffs between 12 and 26 years. Significantly enriched (FDR-adjusted P < 0.05) GO terms are marked with an asterisk (*). Terms are clustered by their gene ratios across enrichments. The mean and standard deviation of the gene ratio (left) and the total number of significantly enriched GO terms (top) are displayed. **b** Upset plot of non-exclusive pairwise overlaps between CpG-associated genes from DNAm signatures of

aging from NCAE-XY-Age classifiers (ages 12 to 26 years) in decreasing size order. The size of each signature is displayed as a horizontal bar. Genes overlapping across every signature (n = 15 genes) are shown as the last bar. **c** Bar plot of the proportion (%) of significantly enriched GO terms (total n = 521 terms) across NCAE-XY-Age DNAm signatures of aging across classifier cutoffs grouped into eleven common GO term ancestor categories, compared with NCAE-Age signatures.

versus 1.6%). Notably, XY signatures showed a high concentration of enriched GO terms for organismal and cellular development, and sex differentiation between 12 and 14 years. Immune-related enriched terms were less common (49/521, 9.4%) and their enrichment levels, as well as those for neural (75/521, 14.4%) and musculoskeletal development, followed a decreasing tendency with age (**Supplementary File 7**). Other categories were approximately equally distributed across the age range.

NCAE-Age clocks identify epigenetic signatures of delayed aging in pediatric Crohn's disease

Lastly, we hypothesized that the effects of disease on normal development could be characterized by estimating age probabilities and identifying DNAm signatures of aging at single-year resolution. For that purpose, we chose to study the developmental consequences of Crohn's disease (CD) in children and young adults. CD is a form of inflammatory bowel disease (IBD) that presents as an immunologically mediated, chronic remittent and relapsing inflammation of the gastrointestinal tract. Pediatric CD

has an earlier onset of immunological disruptions and is associated with irregular disease progression[31 32]. Pertinent to our aim, between 65% and 85% of children and adolescents diagnosed with CD suffer from growth deficiency or delay[33 34 35].

We gathered publicly available datasets containing whole blood DNAm samples from persons with CD and healthy controls with ages between 14 and 24 years, resulting in a cohort of 207 persons with CD (mean (SD) age of 17.1 (2.5) years) and 52 healthy controls (mean (SD) age of 18.6 (3.5) years) (**Supplementary File 1**). All CD samples belonged to treatment-naive individuals and were obtained at disease detection time, preventing the introduction of confounding factors due to medical treatment. We generated age probability estimates for every sample using the NCAE-Age classifier array with cutoffs from 12 to 26 years and grouped them by disease status and single-year age. Then, we compared the proportion of positive predictions below and above the true age across every cutoff (**Fig. 6a, b, Supplementary File 4**). As before, we determined a prediction to be positive if a classifier with a certain age cutoff estimated a probability over 0.5 for the sample's age to be equal to or above this cutoff. Strikingly, we observed that persons with CD with ages from 14 to 19 years had their true age consistently underestimated by NCAE-Age classifiers with cutoffs within the same age range (**Fig. 6a**). For example, only 7.9% of the 15-year-old and 20.0% of the 18-year-old CD individuals were predicted to be 15 years or older, and only 15.0% of 18-year-old CD individuals were predicted to be 18 years or older. We found that this delay in the pace of aging detected for adolescents with CD was reversed after the 20-year mark. The ages of 66.7% of the 20-year-old and between 75.0% and all the persons with CD between 21 and 24 years were estimated to be above 26 years. On the

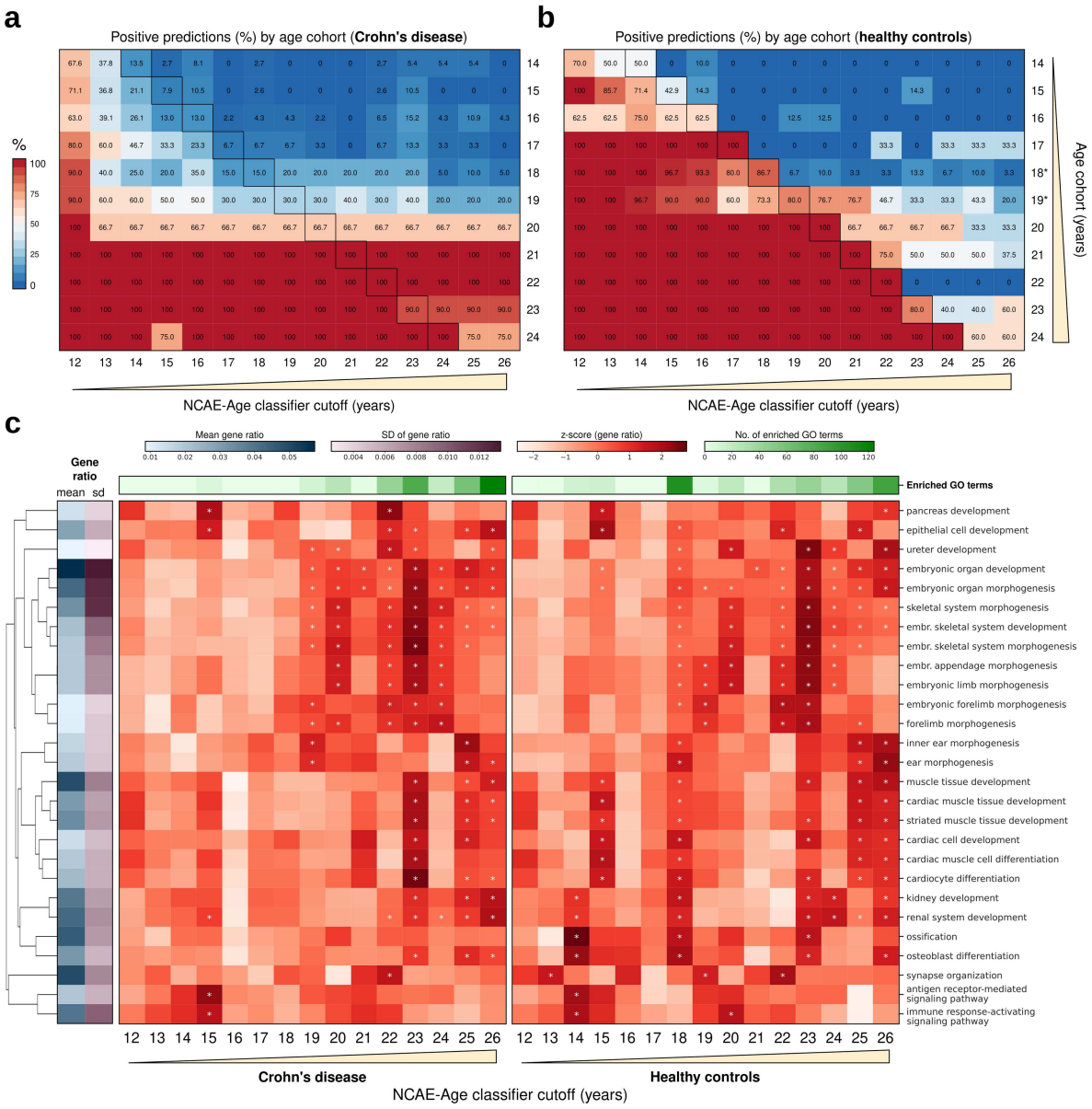


Fig. 6: Epigenetic age and functional enrichment analysis of signatures of aging in Crohn's disease (CD) in treatment-naive adolescents and young adults. a-b Heatmaps of positive predictions (%, estimated probability > 0.5) per NCAE-Age classifier for single-year cohorts (ages 14 to 24 years) of age-matched individuals with CD and healthy controls to be equal to or above the NCAE-Age classifier cutoffs (ages 12 to 26 years). Higher values (dark red) indicate that a higher percentage of samples

493 are classified as being equal to or above the age cutoff, while lower values (dark blue)
494 indicate a lower proportion. Outlined cells indicate that the true age of the single-age
495 cohort is equal to the NCAE-Age classifier cutoff. Cohorts include all available persons
496 with CD and study-matched controls of the same age. Age-matched controls of 18 and
497 19 years old from the classifier test set are displayed to replace non-available control
498 cohorts of those ages from CD datasets. **c** Heatmaps of the gene ratio z-scores of the
499 top enriched GO terms for DNAm signatures of aging from NCAE-Age classifiers with
500 cutoffs between 12 and 26 years for CD patients (left) or healthy controls (right).
501 Significantly enriched (FDR-adjusted $P < 0.05$) GO terms are marked with an asterisk
502 (*). Terms are clustered by their gene ratios across enrichments. The mean and
503 standard deviation of the gene ratio, and the total number of significantly enriched GO
504 terms are shown to the left and top of the figure panel, respectively.

505

506 contrary, none of the study-matched healthy control cohorts presented either an
507 accelerated or decelerated aging pattern at any age (**Fig. 6b**), but rather resembled
508 previous results on control test sets (**Fig. 3f**). The percentages of positive predictions
509 shown by CD individuals did not present significant differences due to sex
510 (**Supplementary Fig. 2**).

511 Next, to further characterize this divergent pace of aging for CD compared to controls,
512 we examined the functional mechanisms captured by the NCAE-Age classifiers. We
513 retrieved age-specific DNAm signatures and explored their enrichment in biological
514 processes represented by GO terms (**Fig. 6c, Supplementary File 7**). We found
515 statistically significant enrichments (FDR-adjusted $P < 0.05$) for DNAm signatures of

aging from 14 to 24 years from CD individuals and healthy controls in up to 271 unique GO biological processes, including every term with at least one significant hit in at least one single-year signature. 27.3% of these enriched GO terms were associated with the general development of anatomical structures (74/271), 17.3% with developmental processes at cellular level (47/271), 13.7% with the development of the neural system (37/271), and 11.1% with the development of the musculoskeletal system (30/271), among other categories (**Supplementary File 7**). The number of significant enrichment hits across all signatures was similar for CD (373 hits) and controls (404 hits), with high correlation between their significance levels (PCC = 0.67). However, restricting the analysis to the interval of interest when delayed aging in CD was observed (14 to 19 years) showed significant enrichments in 121 unique GO terms for control signatures, while CD ones were only enriched in 19 terms. Significant hits for healthy DNAm signatures of aging included multicellular organism growth, muscle tissue development, peripheral nervous system development, renal system development, ossification, bone mineralization, stem cell proliferation, positive regulation of peptidase and endopeptidase activity, among multiple others. None of these terms were enriched in CD signatures until 19 years. The development of anatomical structures remained the largest GO category in both cases (37.2% for controls, 52.6% for CD) but the proportion of terms linked with the immune response was twice as high for CD signatures (5.0% for controls, 10.5% for CD). Other categories coincided, such as the development of the musculoskeletal system (14.0% for controls, 15.8% for CD) or the development of the neural system (7.4% for controls, 5.3% for CD). The most functionally dissimilar DNAm aging signatures were those between CD and controls at 17 years (PCC = 0.37). From

that point onwards, signatures of aging from both groups gradually converged in their term enrichment levels until reaching PCC = 0.67 between CD and control signatures at 26 years. In the interval from 20 to 26 years, CD signatures were enriched in more unique GO terms than control ones (199 versus 145 terms, respectively) (**Supplementary File 7**).

DISCUSSION

In this study, we present the neural network regressor NCAE-CombClock, a high-precision age estimation model trained on a combination of autoencoder embeddings from whole blood DNAm data and established age clock CpG sites. Our model demonstrated exceptional accuracy in predicting chronological age across a large dataset (n > 17,000) encompassing the entire human lifespan (0 to 112 years), achieving a MAE of less than two years. We observed that, while existing DNAm age clocks can accurately determine whether individuals fall below or above crucial young-age milestones (15, 18, and 21 years), their susceptibility to misclassification limits their practical utility. To address this, we developed a suite of neural network classifiers (NCAE-Age array) trained on DNAm embeddings, which provided precise probability estimates of age at various cutoffs, facilitating reliable classification of DNAm samples at the single-year level. Furthermore, these models are interpretable and can thus serve as exploratory tools for identifying cohort-specific DNAm signatures associated with aging. Functional GO term enrichment analyses revealed that the data-driven NCAE-Age signatures of aging identified for healthy adolescent and young adult cohorts between 14 and 24 years were significantly enriched in biological processes that closely

associate with age, such as the development of anatomical structures, maturation of the neural system, immunoregulation, and metabolism, among others. We showcased the applicability of our approach by detecting an altered pace of aging in young individuals with Crohn's disease and investigating the functional pathways captured in their DNAm signatures of aging, compared with those from healthy controls of their same age.

Recent research suggests that DNAm age clocks striving to predict exact chronological age may be modeling increasingly random DNAm variation rather than aging-associated biological mechanisms[18 36 37]. The results of the regression benchmark demonstrated that limited sets of specific CpG sites from clocks could predict chronological age with near-perfect precision, achieving R^2 values on the test set between 0.897 and 0.969 after retraining their coefficients. Additional predictive performance was gained by an elastic net DNAm age clock utilizing the combined sites from clocks (CombClock, $R^2 = 0.974$). Subsequently, the incorporation of information from the broader human methylome jointly with these CpG sites resulted in the NCAE-CombClock regressor, with a top R^2 of 0.978. The fraction of non-stochastic CpG sites selected by clocks built via penalized regression, which could be as low as 10% of their total number of sites[18], likely associates with genes involved in different age-related mechanisms spread throughout the human interactome, rather than functioning as parts of the same biological process [38 28]. This implies that a major part of the epigenetic markers useful for chronological age prediction, regardless of whether stochastic or not, are left aside. An efficient strategy to functionalize this underused DNAm information is to compress it in the form of non-linear autoencoder embeddings, which we here utilized either independently to enhance interpretability, as shown for NCAE-Age classifiers, or

together with previously known sites to optimize age estimation, as in the case of the NCAE-CombClock model.

Adolescence and early adulthood are periods of rapid epigenetic change, influenced by both intrinsic developmental processes and extrinsic environmental factors[19 20]. Some of the DNAm age clocks benchmarked in this study were developed particularly for age prediction in children and adolescents. Li's clock[6], trained using DNAm samples from individuals between 6 and 17 years, performed best when classifying at the 15-year threshold. Likewise, training data for PAYA[7] had an age span of 12 to 25 years, resulting in accurate classifications of 18-year-olds and relatively low FPR. Considering single age clocks, CombClock (in young-age training setting) and NCAE-CombClock (complete-age training) were both top performers for age determination at the 15, 18, and 21-year thresholds, while the consensus of clocks produced the highest F1-score and lowest FPRs across categories. Nevertheless, utilizing an array of clocks such as our introduced NCAE-Age classifiers offers several advantages over a single clock (or consensus of clocks) for ensuring robust predictions in legal, clinical, and other real-life contexts. Probability estimates provide a clearer understanding of uncertainty and confidence levels for decision-makers, enabling a degree of redundancy that decreases the impact of outliers or errors from one model. In the context of health, they may also help guide more tailored preventive or therapeutic interventions and allow for risk stratification. Currently, clinical and lifestyle interventions aimed at mitigating senescence are carried out in older individuals, usually from 50 to 70 years[39 40]. NCAE-Age array estimates of epigenetic aging could feasibly be used alongside long-term, personalized healthcare strategies starting at youth.

DNAm age clocks have been shown to associate with aging factors and outcomes such as mortality, smoking, male sex, body mass index, cardiovascular disease, and cancer[27]. Our goal was to advance beyond merely investigating associations of predicted epigenetic age. Instead, we aimed to implement interpretable clocks that identify data-dependent DNAm signatures relevant to aging, thereby validating their reliability. Within the signatures of NCAE-Age classifiers, we found multiple age-related mechanisms to be significantly enriched at various levels and ages. For example, regulation of organ growth, stem cell division and differentiation[41], regulation of Wnt signaling pathway[42], or Ras signal transduction and positive regulation of the ERK1 and ERK2 cascade[43]. Several of these biological processes, and particularly those from NCAE-XY-Age models, were markedly linked to adolescence and young adulthood, such as the development of primary male sexual characteristics, forebrain development, or mammary gland epithelium development, highlighting the specificity of the detected signatures. With respect to the analysis of pediatric CD, the observed known decelerated epigenetic aging between 14 and 19 years compared to healthy controls[33 34 35] was reflected as a lack of enrichment in developmental mechanisms for CD DNAm signatures of aging at that interval. The aging pace of persons with CD was then accelerated until the last measured cutoff of 26 years, compensating for the delay by a late activation of development-linked processes, visible in the functional enrichment of their DNAm signatures.

While this approach represents a significant advancement in DNAm age estimation and interpretation, it has several limitations. Most age clocks included rely solely on whole blood data, which, although commonly used, may not capture tissue-specific DNAm

variations associated with aging. Additionally, study populations available in public repositories are predominantly Caucasian. Future work could focus on incorporating larger and more demographically diverse datasets, exploring aging-related diseases like neurodegenerative disorders, integrating clinical and lifestyle variables, and examining longitudinal cohorts to better understand the evolution of aging signatures due to intervention or natural progression over time. The use of autoencoders also introduces complexity, requiring prior extensive training and optimization. Finally, although our NCAE-Age array of young-age classifiers provides high-resolution insights, its robustness could be further improved by expanding the studied age interval below 12 and above 26 years, and refining the granularity to even shorter time scales, like semesters or months.

In conclusion, our study underscores that for age estimation models to be reliable, they must effectively capture age-related mechanisms within their DNAm signatures. We developed NCAE-CombClock, a novel and highly precise age prediction model leveraging DNAm embeddings and CpG sites. Additionally, we introduced a neural network classifier ensemble to enhance the precision of age classification in young individuals. By examining the DNAm signatures from these interpretable models, we identified key biological processes associated with age and development during youth. These findings provide valuable insights into the mechanisms of aging in adolescence and young adulthood and their implications for health.

METHODS

Data collection and preprocessing

The data used in this study consisted of publicly available DNA methylation (DNAm) samples from individuals with reported age from the Gene Expression Omnibus (GEO) repository (n = 30,228), from either Illumina Infinium HumanMethylation450K or MethylationEPIC BeadChip arrays. DNAm beta values were normalized using Gaussian Mixture Quantile Normalization (GMQN)[44]. Low-quality samples were filtered out using the ChAMP R package (v2.30.0) and missing beta values were imputed using k-nearest neighbors (k = 10) from the bnstruct R package (v1.0.15). Non-CpG probes, probes related to single nucleotide polymorphisms (SNPs), and multi-hit probes were removed, as well as probes not shared by both Illumina 450K and EPIC arrays. Probes located on the X and Y chromosomes were kept only for models leveraging X and Y chromosome CpG sites, referred to as XY models. In total, 384,629 CpG sites were left after preprocessing, or 395,248 CpG sites for XY models.

DNAm samples were split into training and test sets according to their application. The regression performance of DNAm age clocks was evaluated using a set of 17,726 whole blood samples (0 to 112 years, mean (SD) age = 41.0 (22.9) years, 45.6% female). DNAm samples from various other tissues were used to demonstrate the model's applicability across different contexts, including peripheral blood lymphocytes (PBLs, 0 to 94 years, n = 3,524, mean (SD) age = 54.0 (17.3) years, 68.7% female), peripheral blood mononuclear cells (PBMCs, 0 to 100 years, n = 1,860, mean (SD) age = 33.1 (17.1) years, 46.2% female), buccal epithelium (0 to 112 years, n = 2,174, mean (SD) age = 10.7 (11.9) years, 46.3% female), saliva (3 to 91 years, n = 1,200, mean

(SD) age = 41.5 (26.7) years, 54.4% female), muscle (18 to 112 years, n = 150, mean (SD) age = 38.5 (13.6) years, 35.3% female), and semen (20 to 63 years, n = 132, mean (SD) age = 36.4 (7.1) years). Whole blood samples were divided into a young-age (10 to 30 years, n = 1,404, mean (SD) age = 21.8 (5.3) years, 46.5% female) and a complete-age set (1 to 101 years, n = 8,846, mean (SD) age = 48.6 (19.4) years, 45.2% female) to train and benchmark the classification performance of DNAm age clocks with 15, 18, and 21-year cutoffs. In addition, a sex-balanced set of whole blood DNAm samples (14 to 24 years, n = 330, mean (SD) age = 19.1 (3.1) years, 48.7% female) formed by eleven single-age cohorts of 30 healthy adolescents and young adults each was used to examine the classifier probability estimates at specific ages. To investigate epigenetic aging patterns in Crohn's disease (CD) in adolescents and young adults, we compiled five datasets (GSE112611, GSE32148, GSE81961, GSE87640, GSE87648) and selected whole blood DNAm samples from individuals with an age range of 14 to 24 years, resulting in 207 treatment-naive persons with CD (mean (SD) age = 17.1 (2.5) years, 38.2% female) and 52 healthy controls (mean (SD) age = 18.6 (3.5) years, 50.0% female). A detailed summary of the DNAm samples used throughout the study is available in **Supplementary File 1**.

Design and training of neural network models for age estimation

Artificial neural network models were trained using the Keras 2.4.3 library with TensorFlow 2.4.0 and TensorFlow-GPU 2.2.0 backend, implemented for Python 3.8.10. The training and evaluation of the deep methylation network-coherent autoencoder (NCAE) model was performed as described in Martínez-Enguita et al.[30]. We used the

trained NCAE to compress DNAm samples into low-dimensional latent representations of 128 embeddings. Target age values were min-max scaled from a minimum of 0 and maximum of 114 years to the [0, 1] range to promote numerical stability and enhance model performance. Supervised deep neural network (DNN) models for DNAm age estimation trained on NCAE embeddings (referred to as NCAE-Age models) are feed-forward, fully-connected three-layered DNNs with a 128-feature input layer and either an age regression (prediction of a sample's age in years) or a binary classification (likelihood for the age of a sample to be equal or above a certain cutoff) training objective. An NCAE leveraging CpG sites from the X and Y chromosomes as additional input features (NCAE-XY) was used to obtain DNAm embeddings to train NCAE-XY-Age DNN models, with the same architecture and training strategy as described for the standard NCAE.

NCAE-Age regressors were trained using the Adam (adaptive moment estimation) optimizer to minimize the mean squared error (learning rate = 1e-4, $\beta1 = 0.9$, $\beta2 = 0.999$, $\varepsilon = 1e\text{-}7$, decay = 1e-6), with leaky rectified linear unit (leaky ReLU, $\alpha = 0.3$) as hidden layer activation and output layer function, He uniform initializer, L1 kernel regularization ($\lambda = 0.01$) on the third hidden layer, dropout ($p = 0.1$) and batch normalization (default parameters, momentum = 0.99, $\varepsilon = 1e\text{-}3$) on every hidden layer, and batch sizes of 64 to 1,024. NCAE-Age binary classifiers were trained on the young-age set, with cutoffs of 12 to 26 years, using the Adam optimizer to minimize the binary cross-entropy (learning rate = 1e-4, $\beta1 = 0.9$, $\beta2 = 0.999$, $\varepsilon = 1e\text{-}7$, decay = 1e-6) with an exponential decay learning rate schedule (rate = 0.95, 1e3 steps), leaky ReLU ($\alpha = 0.3$) as hidden layer activation function, sigmoid as output layer function, He uniform

initializer, L2 kernel regularization (λ = 1e-3), dropout (p = 0.1) and batch normalization (default parameters, momentum = 0.99, ε = 1e-3) on every hidden layer, and batch size of 32. Proportional sample weights were introduced during the classifier training phase to compensate for potential class imbalances in each cutoff. NCAE-Age DNNs were trained until early stopping (patience = 1e3 epochs), using an 80:20 ratio of training and validation.

Selection and benchmarking of DNA methylation age clocks

Numerous DNAm age clocks have been developed, differing on training data, methodology, assumptions, and selection of CpG sites. We conducted a literature search to find epigenetic clocks that met the following criteria: the clock should be (a) trained and evaluated using exclusively DNA methylation data (b) from whole blood samples, using a publicly available (c) training methodology and (d) set of CpG sites to (e) estimate chronological age and provide measurements of biological aging (age acceleration). We benchmarked NCAE-Age models and other relevant DNAm age clocks: Horvath pan-tissue[1], Horvath skin and blood[2], Hannum[3], PhenoAge[4], Zhang (elastic net)[5], Li[6], and PAYA (Predictor for Adolescents and Young Adults)[7]. The CpG sites and reported coefficients of the selected DNAm age clocks are available in **Supplementary File 2**. Besides them, we included a consensus model ("top-five consensus") based on the median prediction of the five top-performing models from each evaluation; CombClock, a novel age regressor trained using elastic net (α = 0.5) and leveraging the combined unique 1,743 CpG sites of the selected clocks; and NCAE-CombClock, a novel supervised DNN utilizing the combined 1,743 CpG sites and

128 NCAE embeddings, with the same architecture and training strategy as described for NCAE-Age regressors except for the number of input features.

The regression benchmark was conducted using 17,726 samples from healthy individuals with ages from 0 to 112 years, divided into training and test sets (80:20). Retraining of elastic-net-based age clocks was performed following the same approach as described for the classification benchmark. Regression performance was measured using the coefficient of determination (R^2), mean absolute error (MAE), median absolute error (MedAE), root mean square error (RMSE), Pearson correlation coefficient (PCC, *r*), and Spearman correlation coefficient (ρ). Performance metrics were calculated using the MLmetrics R package (v1.1.1).

The classification benchmark was conducted using the young-age (10 to 30 years, n = 1,404) and complete-age sample sets (1 to 101 years, n = 8,846). Both sets were used to retrain the DNAm age clocks to account for differences in data and preprocessing between the current and original study settings. The retraining was performed by reimplementing the elastic net regression (α = 0.5) using the glmnet R package (v4.1-8), selecting the reported CpG probes for every clock and including an intercept term if it was present originally. In the young-age setting, young-age samples were used to train and evaluate the performance of the age clocks by applying a five-fold cross-validation approach. Each model was trained in four folds while the left-out fold was used as test set, until every fold was used once as test set. In the complete-age setting, models were trained in four folds of young-age samples plus all complete-age samples. They were evaluated on the left-out young-age fold, again until every fold was evaluated. DNAm age clocks were tested on their capacity to predict whether a sample belonged to an

individual equal to or over, or under the cutoffs of 15, 18, and 21 years old. To obtain a comparable metric between DNAm age clocks, estimated ages were converted to a binary variable based on the cutoffs and the classification performance was measured using the F1-score and false positive rate (FPR) above and below the cutoff, and total, for each fold. Since NCAE-Age classifiers generate probability estimates, unlike other DNAm age clocks, their performance was also evaluated using the area under the receiver operating characteristic curve (AUROC). Classification accuracy and probability estimates are reported when evaluating the classifiers on single-age cohorts from the balanced test set.

Identification of DNA methylation signatures of aging

Light-up analyses were performed to retrieve importance rankings of the input features (CpGs) from the trained NCAE-Age model. The light-up technique[45 30] for the interpretation of supervised DNNs is based on the recursive forward propagation of perturbations on the model's input (in this case, CpG sites, either towards hypermethylation, $\beta = 1$, or hypomethylation, $\beta = 0$) through a pre-trained model (in this case, the concatenated layers of the NCAE and NCAE-Age). Input samples for light-up may consist of DNAm profiles representative of a specific cohort of interest, e.g. 21-year-old healthy controls. The perturbation signal is iteratively added to every input CpG and the altered DNAm profiles are propagated through the pre-trained NCAE and DNN models. The contribution of a CpG to the training objective of the pre-trained DNN can be measured by the observed changes (delta values, δ) in the model outcome caused by each perturbation, i.e. changes in the predicted epigenetic age of the sample. Then,

the list of δ-values returned by the light-up analysis can be used to rank the CpG input list by relevance to the model's training objective. We selected the top 1,000 CpGs in increasing order of negative light-up δ-value, which we refer to as the NCAE-Age or NCAE-XY-Age DNAm signature of aging for a particular age estimation DNN and a particular condition and age, for further analysis.

CpG annotation and functional enrichment analysis

Genome-wide mapping of CpG probes to genes was performed using Infinium MethylationEPIC probe annotation files from ChAMP (v2.30.0) and ChAMPdata (v2.32.0) R packages. Gene Symbols were annotated to NCBI (National Center for Biotechnology Information) gene IDs using AnnotationDbi (v1.62.2) and org.Hs.eg.db (v3.17.0) R packages. Functional gene enrichment analysis of CpG-associated DNAm signatures in Gene Ontology (GO) terms from the sub-ontology category "Biological Process" was performed using clusterProfiler R package (v4.8.3) with default parameters and FDR (false discovery rate) P-value adjustment. GO terms with FDR-adjusted $P < 0.05$ were considered significantly enriched and were grouped into manually curated ancestor term categories using annotations from the QuickGO browser (https://www.ebi.ac.uk/QuickGO/) from the European Molecular Biology Laboratory - European Bioinformatics Institute (EMBL-EMBI). GO terms displayed in heatmaps are clustered by the z-score of their gene ratios across significant enrichments for each case, including the mean and standard deviation of the gene ratios for comparison.

Statistical analyses

Statistical analyses and data processing were conducted in R 4.4.0, within RStudio 2023.12.1, and Python 3.8.10. Figures were created using ggplot2 (v3.4.4), gplots (v3.1.3.1), corrplot (v0.92), and UpSetR (v1.4.0) R packages, and seaborn (v0.11.1) and matplotlib (v3.4.2) Python libraries. Subsequent figure editing for enhanced visualization was done in Inkscape v0.92.

ACKNOWLEDGEMENTS

Computational resources were granted by the National Academic Infrastructure for Supercomputing in Sweden (NAISS), including the National Supercomputer Centre resources Berzelius (Berzelius-2022-156, Berzelius-2024-5) and Sigma (LiU-compute-2023-38, NAISS 2023/5-303).

DATA AVAILABILITY

The DNA methylation samples used in this study are publicly available at the Gene Expression Omnibus repository (https://www.ncbi.nlm.nih.gov/geo/). The retrained elastic-net-based DNAm age clocks and the trained NCAE-based neural network models for age estimation, including the tissue-specific regressors and classifier array, are available at https://figshare.com/projects/821420_NCAE_based_interpretable_DNAm_age_clocks.

The authors declare that all other data supporting the findings of this study are available within the paper and its supplementary information files.

CODE AVAILABILITY

The code used for data processing, model training, and signature analysis is available in Zenodo with the identifier https://doi.org/10.5281/zenodo.12906211.

CONFLICTS OF INTEREST

DM-E, ML, and MG are founders of PredictMe, a company that provides DNA methylation analysis services. The other authors declare no competing interests.

FUNDING

This work is supported by the Swedish Research Council (Grant 2019-04193) and the Wallenberg AI, Autonomous Systems and Software Program (WASP) and SciLifeLab and Wallenberg National Program for Data-Driven Life Science (DDLS) (WASPDDLS21-040/KAW 2020.0239).

REFERENCES

1. Horvath, S. DNA methylation age of human tissues and cell types. *Genome Biol.* **14**, (2013).

2. Horvath, S. *et al.* Epigenetic clock for skin and blood cells applied to Hutchinson Gilford Progeria Syndrome and ex vivo studies. *Aging (Albany. NY).* **10**, 1758–1775 (2018).

3. Hannum, G. *et al.* Genome-wide Methylation Profiles Reveal Quantitative Views of Human Aging Rates. *Mol. Cell* **49**, 359–367 (2013).

4. Levine, M. E. *et al.* An epigenetic biomarker of aging for lifespan and healthspan. *Aging (Albany. NY).* **10**, 573–591 (2018).

5. Zhang, Q. *et al.* Improved precision of epigenetic clock estimates across tissues and its implication for biological ageing. *Genome Med.* **11**, (2019).

6. Li, C. *et al.* Age prediction of children and adolescents aged 6-17 years: An epigenome-wide analysis of DNA methylation. *Aging (Albany. NY).* **10**, 1015–1026 (2018).

7. Aanes, H. *et al.* A new blood based epigenetic age predictor for adolescents and young adults. *Sci. Rep.* **13**, (2023).

8. Lu, A. T. *et al.* Universal DNA methylation age across mammalian tissues. *Nat. Aging* **3**, 1144–1166 (2023).

9. Zavala, D. V. *et al.* Epigenetic Age Acceleration and Chronological Age: Associations With Cognitive Performance in Daily Life. *Journals Gerontol. - Ser. A Biol. Sci. Med. Sci.* **79**, (2024).

10. McCrory, C. *et al.* GrimAge Outperforms Other Epigenetic Clocks in the Prediction of Age-Related Clinical Phenotypes and All-Cause Mortality. *Journals Gerontol. - Ser. A Biol. Sci. Med. Sci.* **76**, 741–749 (2021).

11. Liu, Z. *et al.* A new aging measure captures morbidity and mortality risk across dIVerse subpopulations from nhanes IV: A cohort study. *PLoS Med.* **15**, (2018).

12. Chen, B. H. *et al.* DNA methylation-based measures of biological age: Meta-analysis predicting time to death. *Aging (Albany. NY).* **8**, 1844–1865 (2016).

13. Levine, M. E. Modeling the rate of senescence: Can estimated biological age predict mortality more accurately than chronological age? *Journals Gerontol. - Ser. A Biol. Sci. Med. Sci.* **68**, 667–674 (2013).

14. Belsky, D. W. *et al.* Quantification of biological aging in young adults. *Proc. Natl. Acad. Sci. U. S. A.* **112**, E4104–E4110 (2015).

15. Elliott, M. L. *et al.* Disparities in the pace of biological aging among midlife adults of the same chronological age have implications for future frailty risk and policy. *Nat. Aging* **1**, 295–308 (2021).

16. Kabacik, S. *et al.* The relationship between epigenetic age and the hallmarks of ageing in human cells. *Nat. Aging* **2**, 484–493 (2022).

17. Rando, T. A. & Wyss-Coray, T. Asynchronous, contagious and digital aging. *Nat. Aging* **1**, 29–35 (2021).

18. Tong, H. *et al.* Quantifying the stochastic component of epigenetic aging. *Nat. Aging* **4**, 886–901 (2024).

19. Simpkin, A. J. *et al.* The epigenetic clock and physical development during childhood and adolescence: Longitudinal analysis from a UK birth cohort. *Int. J. Epidemiol.* **46**, 549–558 (2017).

20. Jones, M. J., Goodman, S. J. & Kobor, M. S. DNA methylation and healthy human aging. *Aging Cell* **14**, 924–932 (2015).

21. Marini, S. *et al.* Adversity exposure during sensitive periods predicts accelerated epigenetic aging in children. *Psychoneuroendocrinology* **113**, (2020).

22. Sumner, J. A. *et al.* Stressful life events and accelerated biological aging over time in youths. *Psychoneuroendocrinology* **151**, (2023).

23. Bozack, A. K. *et al.* DNA methylation age at birth and childhood: performance of epigenetic clocks and characteristics associated with epigenetic age acceleration in the Project Viva cohort. *Clin. Epigenetics* **15**, (2023).

24. Vidaki, A. & Kayser, M. From forensic epigenetics to forensic epigenomics: Broadening DNA investigative intelligence. *Genome Biol.* **18**, (2017).

25. Hjern, A., Brendler-Lindqvist, M. & Norredam, M. Age assessment of young asylum seekers. *Acta Paediatr. Int. J. Paediatr.* **101**, 4–7 (2012).

26. Fong, S. *et al.* Principal component-based clinical aging clocks identify signatures of healthy aging and targets for clinical intervention. *Nat. Aging* (2024) doi:10.1038/s43587-024-00646-8.

27. Oblak, L., van der Zaag, J., Higgins-Chen, A. T., Levine, M. E. & Boks, M. P. A systematic review of biological, social and environmental factors associated with epigenetic clock acceleration. *Ageing Res. Rev.* **69**, (2021).

28. Liu, Z. *et al.* Underlying features of epigenetic aging clocks in vivo and in vitro. *Aging Cell* **19**, (2020).

29. Higgins-Chen, A. T. *et al.* A computational solution for bolstering reliability of epigenetic clocks: implications for clinical trials and longitudinal tracking. *Nat. Aging* **2**, 644–661 (2022).

30. Martínez-Enguita, D., Dwivedi, S. K., Jörnsten, R. & Gustafsson, M. NCAE: data-driven representations using a deep network-coherent DNA methylation autoencoder identify robust disease and risk factor signatures. *Brief. Bioinform.* **24**, (2023).

31. Van Limbergen, J. *et al.* Definition of Phenotypic Characteristics of Childhood-Onset Inflammatory Bowel Disease. *Gastroenterology* **135**, 1114–1122 (2008).

32. Ruemmele, F. M. *et al.* Consensus guidelines of ECCO/ESPGHAN on the medical management of pediatric Crohn's disease. *J. Crohn's Colitis* **8**, 1179–1207 (2014).

33. Gasparetto, M. & Guariso, G. Crohn's disease and growth deficiency in children and adolescents. *World J. Gastroenterol.* **20**, 13219–13233 (2014).

34. Wong, K., Isaac, D. M. & Wine, E. Growth Delay in Inflammatory Bowel Diseases: Significance, Causes, and Management. *Dig. Dis. Sci.* **66**, 954–964 (2021).

35. Rosen, M. J., Dhawan, A. & Saeed, S. A. Inflammatory bowel disease in children and adolescents. *JAMA Pediatr.* **169**, 1053–1060 (2015).

36. Tarkhov, A. E. *et al.* Nature of epigenetic aging from a single-cell perspective. *Nat. Aging* **4**, 854–870 (2024).

37. Seale, K., Horvath, S., Teschendorff, A., Eynon, N. & Voisin, S. Making sense of the ageing methylome. *Nat. Rev. Genet.* **23**, 585–605 (2022).

38. Levine, M. E., Higgins-Chen, A., Thrush, K., Minteer, C. & Niimi, P. Clock Work: Deconstructing the Epigenetic Clock Signals in Aging, Disease, and Reprogramming. *bioRxiv* (2022).

39. Fahy, G. M. *et al.* Reversal of epigenetic aging and immunosenescent trends in humans. *Aging Cell* **18**, (2019).

40. Fitzgerald, K. N. *et al.* Potential reversal of epigenetic age using a diet and lifestyle intervention: a pilot randomized clinical trial. *Aging (Albany. NY).* **13**, 9419–9432 (2021).

41. Brunet, A., Goodell, M. A. & Rando, T. A. Ageing and rejuvenation of tissue stem cells and their niches. *Nat. Rev. Mol. Cell Biol.* **24**, 45–62 (2023).

42. Nusse, R. & Clevers, H. Wnt/β-Catenin Signaling, Disease, and Emerging Therapeutic Modalities. *Cell* **169**, 985–999 (2017).

43. Slack, C. *et al.* The Ras-Erk-ETS-Signaling Pathway Is a Drug Target for Longevity. *Cell* **162**, 72–83 (2015).

44. Xiong, Z., Li, M., Ma, Y., Li, R. & Bao, Y. GMQN: A Reference-Based Method for Correcting Batch Effects and Probe Bias in HumanMethylation BeadChip. *Front. Genet.* **12**, (2022).

45. Dwivedi, S. K., Tjärnberg, A., Tegnér, J. & Gustafsson, M. Deriving disease modules from the compressed transcriptional space embedded in a deep autoencoder. *Nat. Commun.* **11**, (2020).

SUPPLEMENTARY FILES

Supplementary File 1: Summary of DNA methylation sample sets for DNAm age clocks training and evaluation.

Supplementary File 2: CpG sites and coefficients of benchmarked DNAm age clocks.

Supplementary File 3: Regression and classification benchmark performance of DNAm age clocks.

Supplementary File 4: Estimated probabilities from NCAE-Age classifier array.

Supplementary File 5: DNAm signatures of aging from NCAE-Age clock and classifier array.

Supplementary File 6: DNAm signatures of aging from NCAE-XY-Age clock and classifier array.

Supplementary File 7: Functional enrichment analyses of CpG sets and DNAm signatures of aging from DNAm age clocks and NCAE-Age classifier arrays.

Supplementary Files

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

- suppltablefigures.pdf
- supplfile1.xlsx
- supplfile2.xlsx
- supplfile3.xlsx
- supplfile4.xlsx
- supplfile5.xlsx
- supplfile6.xlsx
- supplfile7.xlsx