

DNA methylation profile is a quantitative measure of biological aging in children

Xiaohui Wu^{1,2,3,4,*}, Weidan Chen^{5,*}, Fangqin Lin¹, Qingsheng Huang¹, Jiayong Zhong¹, Huan Gao¹, Yanyan Song⁶, Huiying Liang¹

¹Institute of Pediatrics, Guangzhou Women and Children's Medical Center, Guangzhou Medical University, Guangzhou, Guangdong, China

²Department of Medical Genetics, School of Basic Medical Sciences, Southern Medical University, Guangzhou, Guangdong, China

³Guangdong Technology and Engineering Research Center for Molecular Diagnostics of Human Genetic Diseases, Guangzhou, Guangdong, China

⁴Guangdong Province Key Laboratory of Psychiatric Disorders, Guangzhou, Guangdong, China

⁵Department of Cardiovascular Surgery, Guangzhou Women and Children's Medical Center, Guangzhou Medical University, Guangzhou, Guangdong, China

⁶The Guangdong Early Childhood Development Applied Engineering and Technology Research Center, Guangzhou Women and Children's Medical Center, Guangzhou Medical University, Guangdong, China

*Equal contribution

Correspondence to: Huiying Liang, Yanyan Song; **email:** lianghuiying@hotmail.com, yansong84@126.com

Keywords: children, biological age, prediction, methylation, age-related diseases

Received: April 12, 2019

Accepted: October 26, 2019

Published: November 22, 2019

Copyright: Wu et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY 3.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

ABSTRACT

DNA methylation changes within the genome can be used to predict human age. However, the existing biological age prediction models based on DNA methylation are predominantly adult-oriented. We established a methylation-based age prediction model for children (9-212 months old) using data from 716 blood samples in 11 DNA methylation datasets. Our elastic net model includes 111 CpG sites, mostly in genes associated with development and aging. The model performed well and exhibited high precision, yielding a 98% correlation between the DNA methylation age and the chronological age, with an error of only 6.7 months. When we used the model to assess age acceleration in children based on their methylation data, we observed the following: first, the aging rate appears to be fastest in mid-childhood, and this acceleration is more pronounced in autistic children; second, lead exposure early in life increases the aging rate in boys, but not in girls; third, short-term recombinant human growth hormone treatment has little effect on the aging rate of children. Our child-specific methylation-based age prediction model can effectively detect epigenetic changes and health imbalances early in life. This may thus be a useful model for future studies of epigenetic interventions for age-related diseases.

INTRODUCTION

Due to the declining fertility rate and the increasing life expectancy, the world is on the brink of a demographic milestone: adults above the age of 65 will soon outnumber children under the age of 5 [1–3]. Population aging is accompanied by increases in illness, disability

and dependency [4]. Consequently, noncommunicable diseases that more commonly occur in adults and older people are imposing the greatest burden on global health. Extending the period of life free of disability and disease is the key to limiting health and social costs. The most promising approach to this end is to identify age-related biological changes in body function or

structure that are more accurate than the chronological age in predicting the future onset of age-related diseases or the remaining years of life [5].

Biomarkers of biological aging can be classified as molecular markers (based on DNA, RNA, etc.) or phenotypic biomarkers (based on anthropometric data such as bone age assessment, blood pressure, lipid levels, etc.) [5]. However, most studies of such biomarkers have been conducted in animals or older individuals [6–9]. Animals with short lifespans are typically not accurate models of the complex multifactorial exposures during human aging. Furthermore, most elderly study participants already suffer from age-related diseases. The theory of the fetal origins of adult disorders [10–13] proposes that many health problems in adults or the elderly are rooted in early life experiences and living conditions. Thus, interventions to reverse or delay age-related diseases and aging itself must be performed in childhood. The lack of tools to quantify aging in children is a significant obstacle to this goal.

Thus far, the most remarkable biological age predictor has been the epigenetic clock. Hannum et al. built a quantitative model of aging by measuring over 450,000 CpG markers in whole blood samples from 656 human subjects aged 19 to 101 years [14]. Horvath et al. developed a biomarker of aging called the multi-tissue predictor based on DNA methylation levels [15]. Using only three CpG sites, Weidner et al. constructed an age prediction model that was more precise than techniques based on telomere length [16]. The above studies demonstrated the feasibility of biological age prediction based on DNA methylation, but these models were mostly focused on adults. Although some of these predictive models included samples from children, the large age range (0–101 years old) and age unit (years) of these models reduced their accuracy both in predicting the biological ages of children (0 - 18 years old) and in revealing biologically relevant epigenetic abnormalities.

There has been some progress in DNA methylation research in children, but there are still many problems to be solved. For instance, Alisch et al. found 2078 age-associated CpG sites in boys (3–17 years old), but did not propose operational quantitative tools [17]. Almstrup et al. only used the data from 51 healthy children (5–16 years old) before and after pubertal onset to predict adolescent development [18]. Freire-Aradas et al. built a preliminary age prediction model using a dataset of 180 donors (2–18 years old) with the EpiTYPER[®] DNA methylation analysis system [19]. None of these studies completely covered the age range from 0 to 18 years, and each study used a single sample data type with a small number of samples, so the results were not highly accurate or applicable.

To delineate the aging pattern precisely throughout childhood, we analyzed DNA methylation datasets from a large cohort of children to construct a child-specific methylation-based age prediction model covering the whole age period from 0 to 18 years with a small age unit (months). Our model is a new tool for quantifying health imbalances and monitoring predictors of age-related diseases early in life, and thus may facilitate early prevention and intervention.

RESULTS

Establishment of a child-specific methylation-based age prediction model

Characteristics of the DNA methylation datasets

We obtained publicly available DNA methylation datasets that were generated with the Illumina 27K or Illumina 450K array platform. Data from 716 healthy children aged 9 to 212 months from 11 different datasets were used to build the quantitative model (Table 1, Figure 1A). These healthy children included 529 boys and 187 girls (Supplementary Figure 1A). Nearly half of the samples (46.5%) were assessed on the Illumina 450K platform (Supplementary Figure 1B). We only studied the 21,979 CpG sites that were present on both Illumina platforms. For simplicity and accuracy, we discarded markers on sex chromosomes and markers with more than 10 missing values across the datasets. DNA methylation levels were recorded as β values between 0 (completely unmethylated) and 1 (completely methylated). To study the link between disease and methylation age in childhood, we also analyzed the datasets of children with diseases (Supplementary Table 1). Details on the above datasets and the data preprocessing steps are provided in the Materials and Methods.

A precise DNA methylation age prediction model in children

By combining sure independence screening [20] and penalized multivariate (elastic net) regression [21], we established a child-specific methylation-based age prediction model in the training cohort, and called the prediction value the DNA methylation age (Figure 2). K-fold cross-validation ($k = 10$) [22] was implemented to divide the training sets and test sets. The optimal model included 111 CpG sites that were accurately predictive of age (regression coefficients in Supplementary Table 2). In the training sets, the model was highly accurate, with a 98% correlation between age and predicted age, and an error of 5.9 months (Figure 1B). The accuracy remained the same when this model was validated on the test sets, as

Table 1. Summary details of the DNA methylation datasets from children.

ID	Availability	Methylation array	n	Age(months)	Gender	Ethnicity	Citation
1	GSE27097	Illumina 27K	334	43.0-212.0	M: 334	white: 266, asian: 14, african-amer: 3, other: 9, more-than-one-race: 35, native-american: 1, native-hawaiian: 1, not-specified: 5	Alisch et al.
2	GSE32148	Illumina 450K	13	60.0-210.0	M: 1 F: 8	null	Harris et al.
3	GSE57484	Illumina 27K	9	120.24-127.8	M: 9	null	Voisin et al.
4	GSE64495	Illumina 450K	18	27.6-129.6	M: 6 F: 12	null	Brunet et al.
5	E-MTAB-4187	Illumina 450K	84	67.0-196.7	M: 51 F: 33	null	Almstrup et al.
6	GSE34257	Illumina 27K	15	9.0	M: 7 F: 8	null	Khulan et al.
7	GSE36054	Illumina 450K	127	12.0-203.0	M: 76 F: 51	Black: 69, White: 8, Other: 42, Asian: 3, Unknown: 5	Alisch et al.
8	GSE23638	Illumina 27K	18	31.1-204.0	M: 9, F: 9	null	Chen et al.
9	GSE41037	Illumina 27K	7	180.0-207.3	M: 6 F: 1	null	Horvath et al.
10	GSE52588	Illumina 450K	3	120.0-180.0	M: 1 F: 2	null	Bacalini et al.
11	GSE73103	Illumina 450K	88	158.2-204	M: 25 F: 63	null	Voisin et al.

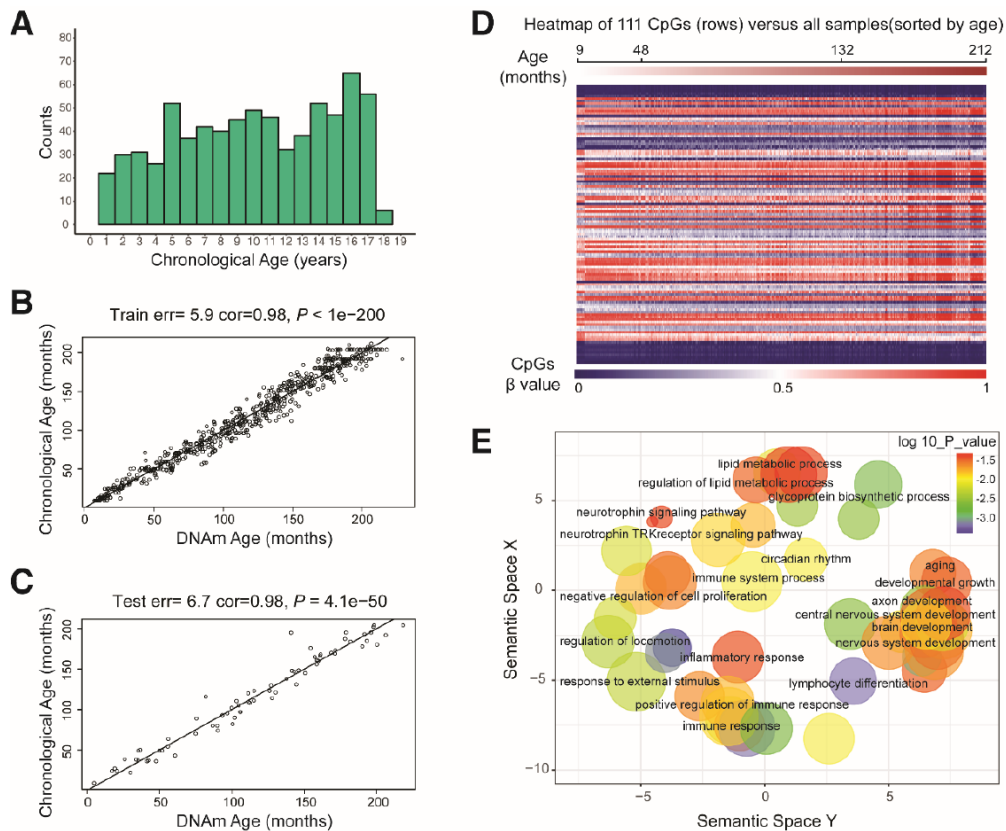


Figure 1. Characteristics of the prediction model. (A) Histogram of the age distribution for healthy children. The x-axis represents the chronological age of the individuals (age unit is years) and the y-axis (counts) represents the number of individuals. (B) Scatterplot of the DNA methylation (DNAm) age (x-axis) against the chronological age (y-axis) for the individuals in the training sets (age unit is months). For the

training data, the correlation between the DNAm age and chronological age was 0.98, and the error (median absolute difference) was 5.9 months. (C) Scatterplot of the DNAm age (x -axis) against the chronological age (y -axis) for individuals in the test sets (age unit is months). For the test data, the correlation was 0.98 and the error was 6.7 months. (D) Heatmap of the DNA methylation levels of 111 CpG sites. Each row represents one CpG site, and the blue to red color spectrum represents β values from 0 to 1. The individuals are sorted by age (9 to 212 months), and it can be seen that the DNA methylation levels change with age. (E) Gene ontology analysis of the 111 CpG sites revealed several ontologies ($P < 0.05$) that may be associated with development and aging. Biological process gene ontologies were plotted in a semantic space with REVIQO, which groups related ontologies together.

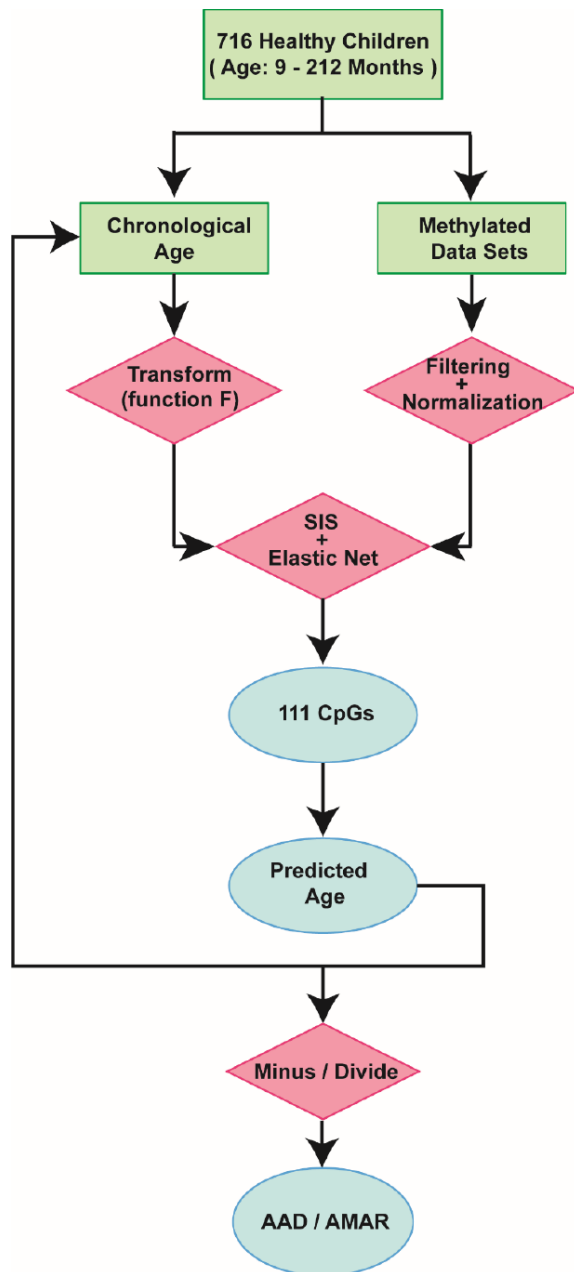


Figure 2. Schematic of the prediction model. A flow diagram of the child-specific methylation-based age prediction model. The green boxes represent the input data, the red diamonds represent the analysis methods and the blue ovals represent the prediction results. AAD: age acceleration difference; AMAR: apparent methylation aging rate; SIS: sure independence screening.

there was a 98% correlation between age and predicted age, and the error was 6.7 months (Figure 1C). The β values of the 111 CpG sites exhibited some certain trends with increasing age, although most of these changes were not very dramatic. The effects of age on the 111 CpG sites were visualized on a heat map, which showed the trends in DNA methylation across subjects (Figure 1D).

Age-related CpG sites associated with development and aging

To determine the biological functions of the 111 age-related CpG sites, we searched for significantly enriched GO terms (biological processes, cellular components and molecular functions, $P < 0.05$) and KEGG signaling pathways among the genes associated with these CpG sites. The top 20 GO terms and KEGG pathways are listed in Supplementary Figure 1. The GO terms were then drawn in a semantic space, and similar terms were combined. The results revealed clusters associated with developmental growth, immune responses, metabolic regulation and age-related diseases such as systemic lupus erythematosus, rheumatoid arthritis and cancer (Figure 1E, Supplementary Figure 1C and 1D, Supplementary Table 3).

Comparing the child-specific age predictor with other age predictors

We then explored the CpG sites selected for different age prediction models. There were only three overlapping sites among Hannum's 71 CpGs [14], Horvath's 353 CpGs [15] and our 111 CpGs (Figure 3A a, b): cg04474832, cg09809672 and cg19722847. These sites are associated with the *ABHD14A*, *EDARADD* and *IPO8* genes, respectively. Of the remaining 108 sites in our study, 59 (54.6%) overlapped with 2,078 previously-identified age-related sites in children [17]; these sites involved 43 genes (Figure 3A c, d). We also compared our model with the model established by Freire-Aradas et al. [19], and obtained only two overlapping genes: *EDARADD* and *PRKG2*.

Since different data and methods were used to construct the above models, it was not possible to identify the most accurate model simply by comparing the error values of the models for the test sets. Thus, to further explore the

accuracy and applicability of our model, we validated it with data from 67 pairs of monozygotic twins in the dataset GSE56105 [23]. We took this approach because the DNA methylation ages of healthy monozygotic twins who share the same genetic background and living environment should theoretically be similar. First, we used our model and the commonly used multi-tissue predictor [15] to calculate the DNA methylation ages of the monozygotic twins. The predictions from our model were closer to the actual ages of the twins, and the distribution of DNA methylation ages was more concentrated than that of the multi-tissue predictor (Figure 3B). Second, we compared the DNA methylation ages of twins 1 and 2 calculated by these two models. The

predicted DNA methylation ages of twins 1 and 2 did not differ significantly when our model was used (Figure 3C, $P = 0.51$, paired t-test), while they did differ significantly when the multi-tissue predictor was used (Figure 3D, $P = 0.025$, paired t-test). Finally, we compared the absolute values of the DNA methylation age differences between twins 1 and 2 calculated by the two models. The absolute values of the two models differed significantly (Figure 3E, $P < 0.01$, t-test), and the values calculated by our model were closer to zero than those calculated by the multi-tissue predictor. Therefore, our prediction model performed better than the multi-tissue predictor in estimating children's DNA methylation ages based on blood samples.

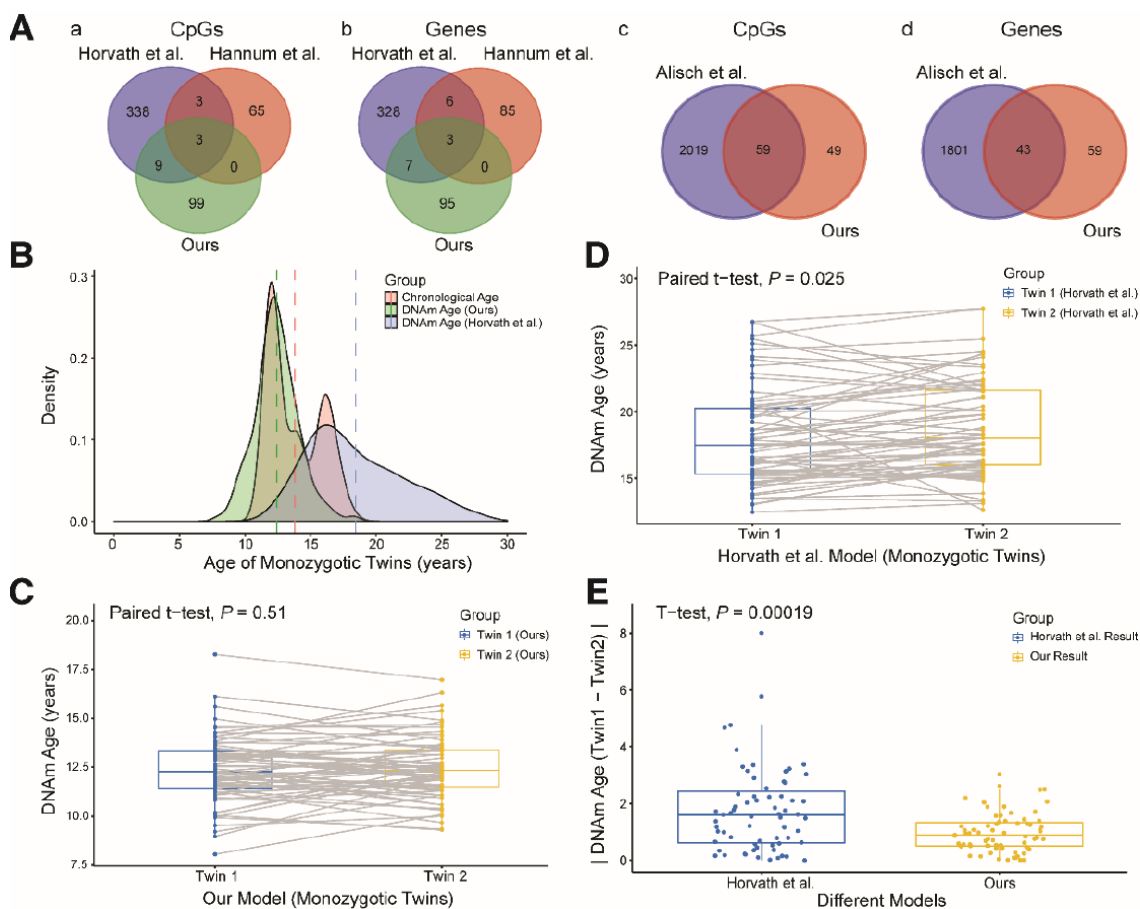


Figure 3. Comparison and verification of our model. (A) (a and b) Venn diagrams of the CpG sites (a) and genes associated with the CpG sites (b) selected from the three models. (c and d) Venn diagrams of the CpG sites (c) and genes (d) associated with age from our study and the study of Alisch et al. (B) Density plot of age and DNA methylation (DNAm) age. The red peak represents the chronological age, the green peak represents the DNAm age predicted by our model, and the blue peak represents the DNAm age predicted by Horvath et al. Dashed lines represent mean values. (C) Boxplot comparing the DNAm ages predicted by our model for monozygotic twins (paired t-test, $n = 67$ each for twins 1 and 2). The blue box indicates twin 1 and the yellow box indicates twin 2. (D) Boxplot comparing the DNAm ages predicted by the model of Horvath et al. for monozygotic twins (paired t-test, $n = 67$ each for twins 1 and 2). The blue box indicates twin 1 and the yellow box indicates twin 2. (E) Boxplot comparing the absolute values of the DNAm age differences of monozygotic twins predicted by the two models (two-sided t-test, $n = 67$). The blue box indicates the results from Horvath et al. and the yellow box indicates our results.

We could not compare the accuracy of our model with that of the model established by Freire-Aradas et al. [19] because the authors did not report their predicted age calculation formula and data; however, the error of 1.25 years (15 months) reported by Freire-Aradas et al. [19] was larger than the error of our model (6.7 months).

Aging patterns in children revealed by our model

Our aging model not only predicted the age of most children with high accuracy, but also revealed individual biological differences and aging trends in the pediatric population [14, 15]. To examine whether these differences were true biological differences (rather than measurement error or intrinsic variability), we used our aging model for two measurements of age acceleration. The first, called the age acceleration difference (AAD), is the DNA methylation age minus the chronological age. The second, called the apparent methylation aging rate (AMAR), is the DNA methylation age divided by the chronological age.

Age acceleration in children seems not to be influenced by gender or ethnicity

We then explored the association of the AAD and AMAR with the potentially clinically relevant factors of gender and ethnicity. In terms of gender, the mean AAD and AMAR values in all the healthy children's samples were -0.01 months and 1.01, respectively. The AAD and AMAR values for boys were 0.003 months and 1.006, respectively, while the values for girls were -0.040 months and 1.020, respectively. Neither the AAD nor the AMAR differed significantly between boys and girls (AAD: $P = 0.92$, Wilcoxon test; Supplementary Figure 2A), although the AMAR was approximately 1.4% faster in girls than in boys (Supplementary Figure 2B). In contrast, in adults, the AMAR was reported to be 4% faster in men than in women [14]. This difference may be due to the fact that girls develop earlier than boys [24–26]. Regarding ethnicity, neither the AAD nor the AMAR differed significantly among children of different ethnicities (AAD: $P = 0.9$, AMAR: $P = 0.26$, ANOVA; Supplementary Figure 2C and 2D).

Age acceleration is the greatest in mid-childhood

We observed a trend in age acceleration in healthy children between the ages of 0 and 18 years. The age acceleration was close to zero before the age of 4, gradually rose after the age of 5, and fell to a negative value after the age of 12 (Figure 4A). To further explore the aging pattern in children, we divided childhood into three periods: toddlerhood (0–4 years), mid-childhood (5–11 years) and adolescence (12–18 years). We found that the AAD and AMAR were significantly greater in

mid-childhood than in toddlerhood, and were significantly lower in adolescence than in mid-childhood (AAD: $P = 7.2 \times 10^{-14}$, AMAR: $P = 4.5 \times 10^{-6}$, ANOVA; Figure 4B and C). The same phenomenon was observed after sex stratification. Moreover, in mid-childhood, the aging rate seemed to be marginally faster in girls than in boys (Figure 4D, Supplementary Figure 3). These differences in the AAD and AMAR at different stages of childhood indicate that the aging rate of children is not completely consistent with the growth curve, which may be related to the development of several major organ systems.

To verify the above results, we analyzed an independent dataset from the study of Almstrup et al. [18]. These authors described longitudinal whole-genome changes in DNA methylation in peripheral blood samples ($n = 84$) before and after adolescence in 42 healthy children. Coincidentally, the pre-puberty (5.6–11.3 years) and post-puberty (12.2–16.4 years) age segments in their study were almost the same as our mid-childhood and adolescent age settings, so we could use this dataset to verify our results. As expected, the rate of aging was significantly higher before puberty than after puberty (AAD: $P = 6.1 \times 10^{-7}$, AMAR: $P = 1.3 \times 10^{-6}$, Wilcoxon test; Figure 4E and 4F). The conclusion of the study by Almstrup et al. also confirmed this finding.

Association of DNA methylation age with disease

To investigate the association of the DNA methylation age with potential health problems in children, we analyzed three datasets, which respectively focused on diseases, short-term interventions and long-term environmental exposures in children.

Age acceleration and autism

Using the dataset GSE27044 (details in Supplementary Table 4), we analyzed the association of the three types of autism (autism, autism spectrum disorder and Asperger syndrome) with the DNA methylation age (Supplementary Figure 4A) [17]. We found no significant difference in the AAD or AMAR between these three types of autistic children and their unaffected siblings (AAD: $P = 0.47$, AMAR: $P = 0.098$, ANOVA; Supplementary Figure 4B and 4C).

We then assessed whether children with the first type of autism conformed to the aforementioned pattern in which the aging rate was significantly greater in mid-childhood. We grouped the samples according to the previous criteria, but there was no toddler group because there was only one child aged less than 48 months. As expected, the rate of aging was significantly higher in mid-childhood than in adolescence in the autistic children (AAD: $P = 2.2 \times 10^{-16}$, Wilcoxon test;

Supplementary Figure 4D). Surprisingly, the aging rate was significantly higher in autistic children than in their unaffected siblings in mid-childhood (AAD: $P = 0.013$, Wilcoxon test; Table 2, Figure 5A and 5B), but this difference was not observed in adolescence. This

finding suggests that autistic children age faster than healthy children in mid-childhood. The above results also indicate that it is worthwhile to examine the significance of mid-childhood and to analyze subgroups throughout childhood (0–18 years).

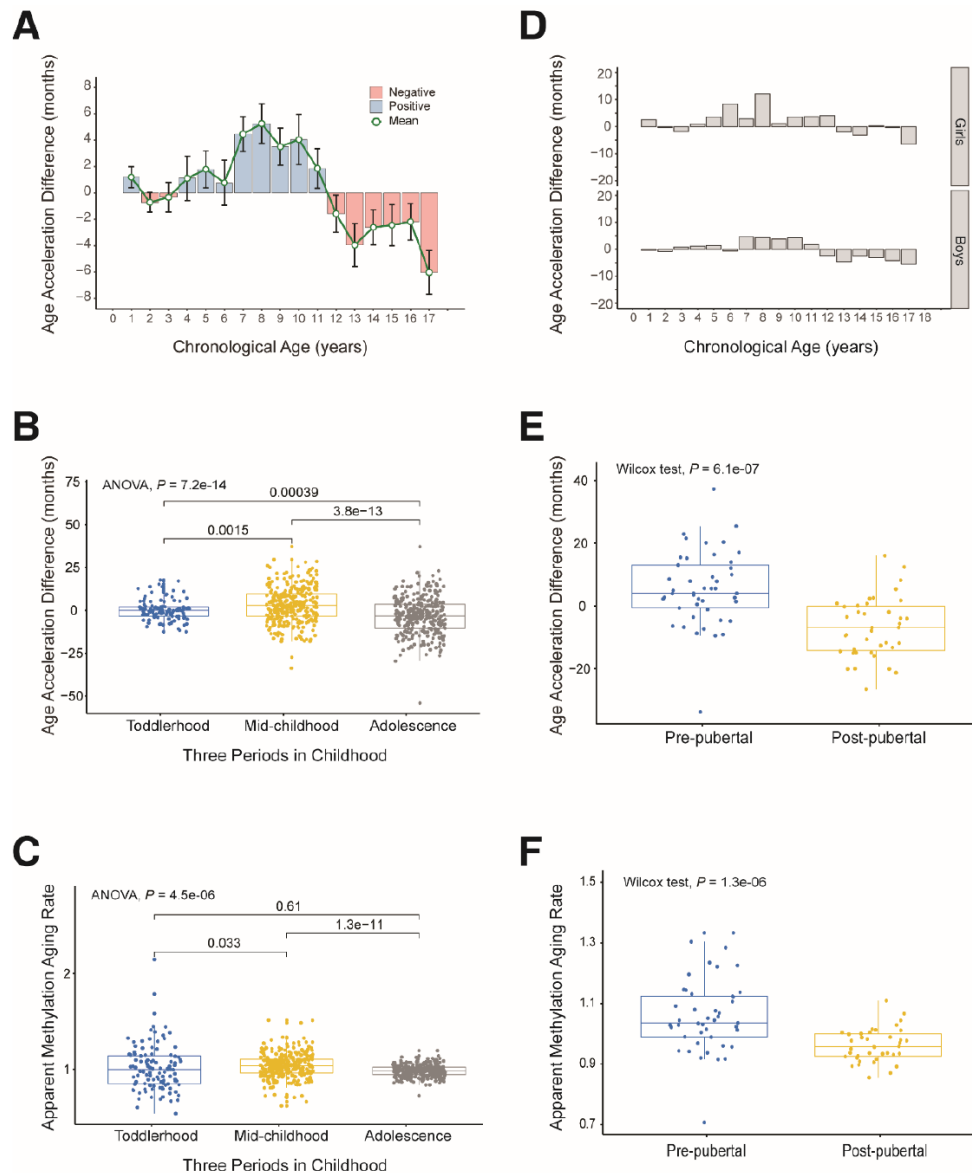


Figure 4. Age acceleration in different periods of childhood. (A) Histogram of the mean value distribution of the age acceleration difference for all individuals. A pink column indicates a negative value, meaning that the average difference between the DNA methylation age and the chronological age is less than zero. A blue column indicates a positive value, meaning that the average difference between the DNA methylation age and the chronological age is greater than zero. The green circle represents the average difference between the DNA methylation age and the chronological age (age unit is years). (B) Boxplot of the age acceleration difference during different periods of childhood. The blue box indicates toddlerhood, the yellow box indicates mid-childhood and the gray box indicates adolescence. (C) Boxplot of the apparent methylation aging rate during different periods of childhood. The box colors have the same meaning as above. (D) Histograms of the mean value distribution of the age acceleration difference for girls and boys, respectively. (E) Boxplot comparing the age acceleration difference between pre-pubertal and post-pubertal individuals. The blue box indicates pre-pubertal individuals and the yellow box indicates post-pubertal individuals. (F) Boxplot comparing the apparent methylation aging rate between pre-pubertal and post-pubertal individuals. The box colors are the same as above.

Table 2. Overview of two measures of age accelerations evaluating the effect of autism.

Group	Sample (n)	Age (months)*	Mean±SD	ΔMean	95%CI	Cohen's d	95%CI	P value	Power (%)
AAD: DNAm age – Age (months)									
Case†	260	48 - 132	6.79±20.67	3.96±1.66	[0.71, 7.21]	0.215	[0.036, 0.394]	0.013	77.0
Control†	226	48 - 132	2.83±15.83						
AMAR: DNAm age / Age									
Case†	260	48 - 132	1.09±0.25	0.1±0.02	[0.06, 0.14]	0.263	[0.084, 0.442]	0.003	91.5
Control†	226	48 - 132	1.03±0.20						

* min – max.

† Case: Autism; Control: Unaffected siblings.

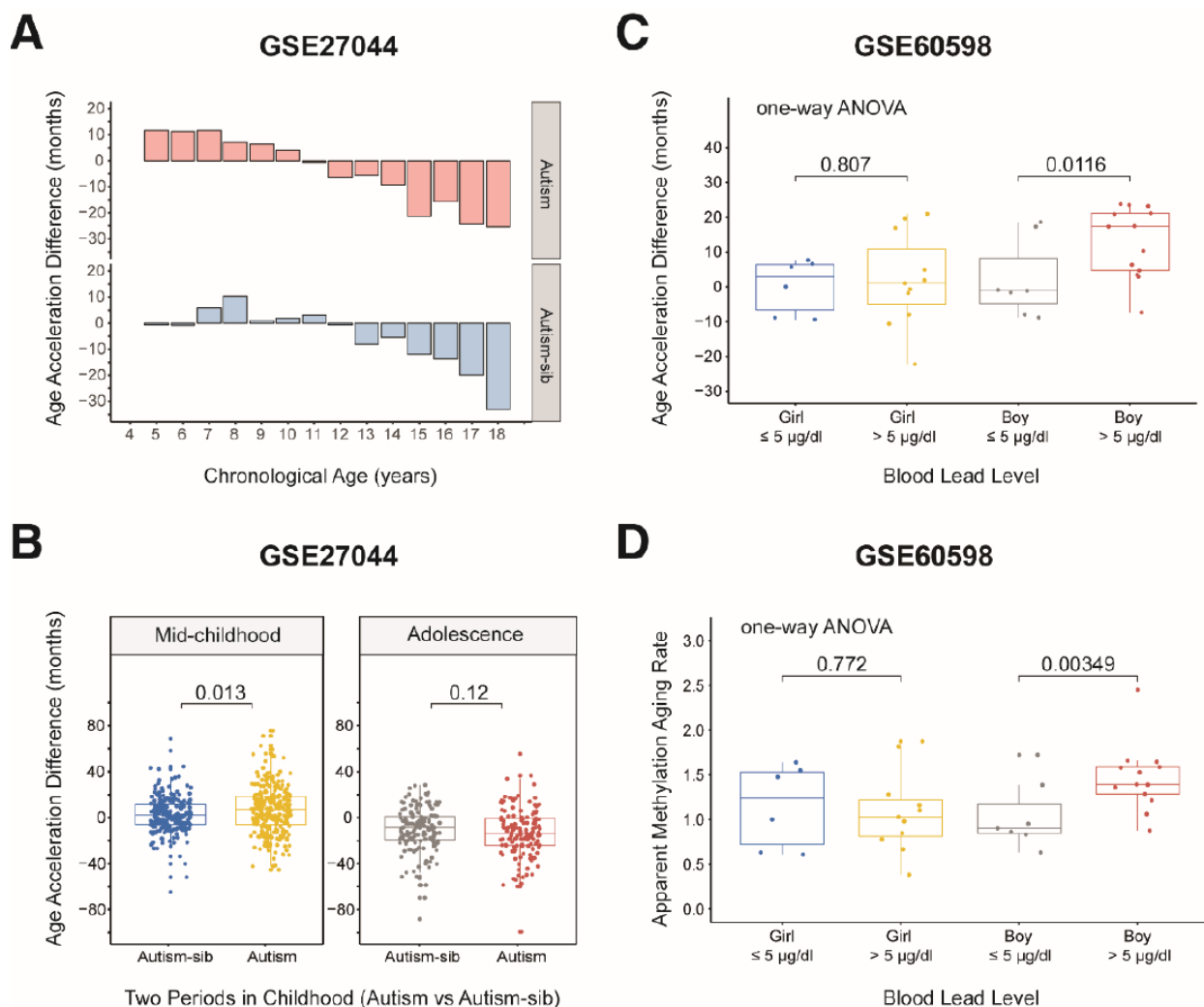


Figure 5. Age acceleration in children with diseases. (A) Histograms of the mean value distribution of the age acceleration difference in autistic children and their unaffected siblings. The pink columns indicate the children with autism, while the blue columns indicate their unaffected siblings ('Autism-sib'). (B) Boxplot comparing the age acceleration difference between autistic children and their unaffected siblings during two periods of childhood. The blue box and the yellow box indicate the autistic children and their unaffected siblings, respectively, in mid-childhood. The gray box and the red box indicate the autistic children and their unaffected siblings, respectively, in adolescence. (C) Boxplot comparing the age acceleration differences of boys and girls with different blood lead levels. A cutoff value of 5 µg/dL was used for the blood lead level. The jitter points represent the age acceleration differences of individual samples. (D) Boxplot comparing the apparent methylation aging rates of boys and girls with different blood lead levels. A cutoff value of 5 µg/dL was used for the blood lead level. The jitter points represent the apparent methylation aging rates of individual samples.

Age acceleration and short-term rhGH treatment

We analyzed data from 48 peripheral blood samples taken from 24 patients prior to the first dose and after four days of rhGH treatment (GSE57205 [27]). The rate of aging in these patients did not differ significantly before and after rhGH treatment ($P > 0.05$, t-test; Supplementary Figure 4E). The diagnoses leading to rhGH treatment in this cohort were classical GH deficiency (classical GH deficiency [STH-D], $n = 7$; panhypopituitarism [PAN], $n = 1$; and small for gestational age [SGA], $n = 1$), neurosecretory dysfunction leading to GH deficiency (NSD, $n = 6$), SGA with a lack of catch-up growth (SGA, $n = 7$), qualitative GH deficiency (Q-STH-D, $n = 2$, Kowarski syndrome), Turner syndrome ($n = 1$) and primary insulin-like growth factor 1 deficiency ($n = 1$). We compared the AAD values of patients with the first three diagnoses before the treatment; the remaining types had too few samples and were not included in the analysis. The AAD was lower in the NSD group than in the other two groups ($P < 0.05$, t-test; Supplementary Figure 4F), which may reflect the underlying neurosecretory dysfunction in NSD patients [28, 29]. These results suggest that short-term rhGH treatment does not significantly influence age acceleration, although different types of GH deficiency may be associated with different rates of age acceleration.

Age acceleration and lead exposure early in life

We then evaluated a dataset of 42 dry blood spots from 24 boys and 18 girls aged three months to five years who were exposed to lead in the environment (from GES60598 [30]). The grouping criterion was a blood lead level (BLL) $> 5 \mu\text{g/dL}$, the maximum safe limit recommended by the Advisory Committee on Childhood Lead Poisoning Prevention in 2012. Boys are more sensitive than girls to lead exposure [31, 32]; thus, we expected to observe significantly greater changes in the AAD and AMAR in boys than in girls if the DNA methylation age was associated with lead exposure. When we grouped the samples by gender, we found that this indeed was the case. The AMAR differed significantly between boys with BLL values $> 5 \mu\text{g/dL}$ and $\leq 5 \mu\text{g/dL}$ in a t-test ($P = 0.033$, Table 3), but this phenomenon was not observed in girls. To control for spurious findings, we performed one-way ANOVA to assess the association between lead exposure and DNA methylation age, controlling for the age of the children and the age, gestational age and smoking status of their mothers. The AAD and AMAR were significantly higher in lead-exposed boys (BLL $> 5 \mu\text{g/dL}$) than in nonexposed boys (BLL $\leq 5 \mu\text{g/dL}$) (AAD: $P = 0.0116$, AMAR: $P = 0.00349$; Figure 5C and 5D).

To sum up, DNA methylation age abnormalities may be associated with certain health problems in children. Our child-specific methylation-based age prediction model can be used to reveal aging trends, to study the relationship of age acceleration with diseases and environmental factors impacting children's growth and development, and to explore the influence of these factors on children's health and longevity.

DISCUSSION

Recently, several age prediction models based on DNA methylation have been published. However, due to their large age range (0–101 years old) and age unit (years), most of these models cannot accurately predict the biological ages of children [14–16]. Although a few studies have examined DNA methylation in childhood, they used small amounts of data detected on infrequently used platforms and provided no explicit quantitative tools [17–19]. Therefore, we employed data from two commonly used methylation chips (Illumina 27K and Illumina 450K array platforms) to construct a child-specific methylation-based age prediction model covering the entire period of childhood (0–18 years old) with a small age unit (months). Our model has the following advantages over other age prediction models: a) it comprehensively reflects the aging patterns in childhood; b) it uses months as the age unit, thus increasing the accuracy of the prediction results; c) it solves the problem of insufficient variable selection methods by using sure independence screening [20] before multiple linear regression (elastic net) [21], as the former performs better when the dimension of the predictor p is much larger than the sample size n ; and d) it is based on data from whole blood samples analyzed with two types of chips (Illumina 27K array and Illumina 450K arrays), and thus can enhance the practical diagnostic design and analysis of samples collected from other studies. Although the multi-tissue age predictor has broader applicability than our model, tissue specificity can influence the accuracy of predictions. We will later validate and optimize our model in other tissue types. Moreover, we will examine how to better balance the accuracy and applicability of the model.

There were only three overlapping sites between the adult-directed age prediction model and our child-specific methylation-based age prediction model. The cg09809672 site is associated with the EDAR-associated death domain gene (*EDARADD*), which has been linked to ectodermal dysplasia, especially hypohidrotic ectodermal dysplasia [16, 33–35]. The cg04474832 and cg19722847 sites are associated with the *ABHD14A* and *IPO8* genes, respectively. *ABHD14A* may be involved in the development of granule neurons, while *IPO8* is involved in a common bone marrow mesenchymal stem cell degenerative joint disease [36]. These three sites are most

Table 3. Overview of two measures of age accelerations evaluating the effect of lead exposure early.

Group	Sample (n)	Age (months)*	Mean±SD	ΔMean	95%CI	Cohen's d	95%CI	P value	Power (%)
AAD: DNAm age – Age (months)									
BLL† > 5 µg/dl	13	12 - 60	12.74±9.97						
BLL† ≤ 5 µg/dl	7	9 - 48	2.07±11.18	10.67±5.05	[0.77, 20.57]	1.027	[0.055, 1.999]	0.058	63.8
AMAR: DNAm age / Age									
BLL† > 5 µg/dl	13	12 - 60	1.46±0.37						
BLL† ≤ 5 µg/dl	7	9 - 48	1.04±0.37	0.42±0.17	[0.09, 0.75]	1.135	[0.151, 2.119]	0.033	76.2

* min – max.

† BLL: Blood lead level.

likely associated with aging throughout the entire life process, but the low number of overlapping sites indicates that children have specific age-related methylation characteristics.

Surprisingly, 54.6% of the remaining 108 sites screened by our model overlapped with previously determined age-related sites in children [17]. This high overlap rate indicates that our model effectively reflects the specific age-related methylation changes in children. Alisch et al. [17] reported that age-related DNA methylation changes in peripheral blood occurred more rapidly during childhood; thus, our choice of peripheral blood samples was appropriate. We identified the gene loci of these 59 CpG sites and annotated them using GO terms. The genes were involved in a concentrated set of developmental processes and immune functions, consistent with the known associations between DNA methylation changes and age-related immune system activities [37, 38].

When we compared our model with that established by Freire-Aradas et al. [19], we identified two overlapping genes. *EDARADD* was introduced in the previous paragraph, and has appeared in most of the relevant studies [15, 16]. The protein encoded by the *PRKG2* gene regulates intestinal fluid balance, and changes in its methylation level correlate highly with age in children [17].

Since our model can effectively reflect the specific age-related methylation characteristics of children aged 0 to 18 years, it can be used to monitor abnormalities in children's growth and development, as well as to predict the occurrence of diseases and the process of aging. Similar to bone age, which can be used to detect precocious puberty [39, 40], the DNA methylation age

can be used to quantify the rate of aging in children. Our results revealed a pattern of changes in epigenetic age acceleration in healthy children. We attempted to explain this phenomenon by using the k-means clustering algorithm to analyze the variation in the β values of the 111 CpG sites (Supplementary Figure 5). The average β values of the sites from some clusters (e.g., cg26227465) were higher in toddlerhood, lower in mid-childhood and higher again in adolescence, whereas those from other clusters (e.g., cg25827666) were lower in toddlerhood, higher in mid-childhood and then slightly lower during adolescence. The cg26227465 site is located near the *IFNG* gene, which encodes a protein secreted by cells of both the innate and adaptive immune systems. This gene is associated with increased susceptibility to viral, bacterial and parasitic infections and to several autoimmune diseases [41–43]. The cg25827666 site is upstream of the *NTRK1* gene, which encodes a member of the neurotrophic tyrosine kinase receptor family. This kinase promotes cellular differentiation and may contribute to sensory neuron subtype specification [44, 45]. DNA methylation is considered an epigenetic marker of expression ability, as decreases in methylation are usually associated with increases in gene expression, and vice versa. Therefore, it was reasonable that the aging patterns of children were reflected in DNA methylation changes with increasing age.

Next, we applied our predictive model to children with autism and early lead exposure, and found that the DNA methylation age was accelerated in autistic children in mid-childhood and in boys exposed to lead. To demonstrate the legitimacy of our findings, we will now report our statistical power. First, the sample of 260

autistic children and 226 unaffected siblings was large enough to detect a significant AMAR difference of 0.1 between the two groups with adequate statistical power (91%). Previous studies have also provided some evidence of the accelerated aging of autistic patients. Autism is a lifelong condition [46] that increases the incidence of nearly all age-related health impairments in adulthood, including immune conditions, gastrointestinal and sleep disorders, seizures, obesity, dyslipidemia, hypertension and diabetes [47]. Autism can also markedly increase premature mortality [48] and reduce the quality of life [49]. Second, although the groups of 13 lead-exposed boys and 7 controls were relatively small, they also achieved 76% power to detect a significant AMAR difference of 0.4. Eid and Zawia also reported that lead induces brain aging and increases susceptibility to adult neurodegenerative diseases, particularly Alzheimer's disease and Parkinson's disease [50]. Given the improvements in our child-specific methylation-based age prediction model, we expect that it will be widely applied in research on pediatric health assessment and disease prevention. This could reveal aging trends with many practical implications.

Although our model can measure the biological ages of children more accurately than previous models, we do not currently have data on the outcomes of the included children later in life (e.g., risk of disease, time of death) to verify whether this new clock accurately measures biological aging caused by pediatric diseases. Therefore, Guangzhou Women and Children's Medical Center is establishing pediatric disease cohorts for long-term follow-up. In this process, multi-omics data at different stages of life (continuing to adulthood and even to death) will be measured to test the ability of children's biological clocks to characterize biological aging. Our pediatric cohort will also provide a larger database for this study, thus addressing the problem of the insufficient sample size and enabling us to explore the association of extensive disease outcomes with biological aging.

In conclusion, childhood (0–18 years) is the fastest period of development of various systems. Age-related DNA methylation changes in the peripheral blood of children occur more rapidly and with greater flexibility than those in adults. We established a methylation-based age prediction model specifically for children, which enabled us to quantify children's biological ages with great accuracy, and to identify several determinants and variation trends of age acceleration in children. In addition to assessing the aging trends that correlated with epigenetic changes in childhood, we also investigated the effects of autism, GH deficiency and lead exposure on biological age in children. In future studies, our model can be used to identify other

factors influencing the AAD and AMAR, including other childhood diseases or environmental factors (such as maternal smoking, alcohol intake or eating habits), and to quantify the impact of these factors on the health and longevity of children. Our biological age prediction model in children could be developed into a quantitative health assessment tool that detects health imbalances early in life, effectively preventing age-related diseases and postponing the aging process.

MATERIALS AND METHODS

Description of the datasets

We collected publicly available genome-wide methylation datasets of healthy children's peripheral blood samples from the Gene Expression Omnibus database and other online resources to build our model. Details about the individual datasets (datasets 1–11) can be found in Table 1, along with the relevant citations. Dataset 1 consisted of leukocyte samples from 334 healthy (entirely male) subjects (mean age 10, range 3–17 years old) [17]. Dataset 2 included 13 unaffected subjects from a DNA methylation study of Crohn's disease and ulcerative colitis [51]. Dataset 3 comprised nine subjects of normal weight from an adolescent dietary fat study [52]. Dataset 4 involved 18 unaffected individuals from a study of age-related diseases [53]. Dataset 5 was obtained from a longitudinal analysis of genome-wide methylation changes in peripheral blood samples ($n = 84$) from healthy children before and after pubertal onset [18]. Dataset 6 consisted of 15 samples from a study on the effects of periconceptional maternal micronutrient supplementation on infant blood methylation patterns [54]. Dataset 7 was generated in the same lab as dataset 1, and contained samples from 127 healthy children measured on the Illumina 450K platform [17]. Dataset 8 included nine healthy boys and nine healthy girls who participated in a sex-specific DNA methylation analysis [55]. Dataset 9 comprised seven healthy control subjects from an analysis of co-methylation modules related to age [56]. Dataset 10 was obtained from the relatives of patients with Down Syndrome (mothers and unaffected siblings) [57]. Dataset 11 included 88 lean individuals aged 14 to 16 years who were recruited by mail and through school visits in Uppsala, Sweden [58]. Five datasets were obtained from Illumina 27K arrays, while six were obtained from Illumina 450K arrays.

The datasets of children with diseases are summarized in Supplementary Table 1 (datasets 1–3). These datasets were also obtained from the Gene Expression Omnibus database. Dataset 1 contained methylation data on 27,578 CpG dinucleotides in peripheral blood leukocyte DNA samples from autistic children and unaffected siblings

[17]. Dataset 2 included samples from 24 patients at baseline and after four days of recombinant human growth hormone (rhGH) treatment [27]. Dataset 3 consisted of 42 dry blood spots from children exposed to lead [30].

DNA methylation data pre-processing and quality control

The public Illumina DNA data described above were generated with either the Illumina Infinium HumanMethylation27 BeadChip or the Illumina Infinium HumanMethylation450 BeadChip. Both arrays are used to quantify DNA methylation based on β values, which range from 0 (completely unmethylated) to 1 (completely methylated). We merged the data from the two platforms by focusing on the ~26,000 CpG sites that are present in both platforms. The age prediction model was trained on 21,979 probes that were shared between the Illumina 27K and 450K platforms and had ≤ 10 missing values across the datasets. Then, the R ‘impute’ package was used to impute the remaining missing values with the k-nearest-neighbors approach (10 nearest markers) [59]. The BMIQ R function [60] was used to readjust the 21,000 overlapping probes so that their distribution met the gold standard (the mean β value of the largest single dataset (GSE27097) in this article [17]).

We performed a principal component analysis to identify and remove outliers. First, each sample was converted into a z-score statistic based on the squared distance of the first principal component from the population mean. Then, the z-score was converted to the false-discovery rate through the Gaussian cumulative distribution function and the Benjamini-Hochberg procedure [61]. Samples falling below a false-discovery rate of 0.2 were designated as outliers and were removed. This filtering procedure was performed iteratively until no samples were determined to be outliers. The remaining 716 samples were used in the age prediction model. Specific information on these samples is shown in Table 1.

Age conversion and DNA methylation age prediction model

To improve the accuracy of the prediction model, we used months as the age unit. In the included datasets, 84% of the sample ages were recorded as months. The sample ages recorded as years were converted to months for this study. We employed k-fold cross-validation ($k = 10$) in the R ‘caret’ package [22] to randomly cleave the datasets 10 times and build a model for each cohort. During each run, a different cluster was used as the test set, and the remaining clusters were used as the training set, with proportions of 10% and 90%, respectively.

Based on the training set data, we found it advantageous to transform age using function F before building the prediction model. Using the inverse of function F , we transformed the linear part of the regression model into the DNA methylation age. Function F was as follows (toddler.age was set to 48 months):

$$F(\text{age}) = \log(\text{age} + 1) \\ - \log(\text{toddler.age} + 1) \text{ if } \text{age} \leq \text{toddler.age} \\ F(\text{age}) \\ = (\text{age} - \text{toddler.age}) / (\text{toddler.age} + 1) \text{ if } \text{age} \\ > \text{toddler.age}$$

The child-specific biological age prediction model was established through sure independence screening combined with multivariate linear modeling based on the elastic net algorithm. First, we used sure independence screening (implemented in the R package ‘SIS’) [20] to reduce the dimensionality of the ~21,000 β values in the datasets. This step was taken because variable selection methods (e.g., lasso, LARS, SCAD) do not perform well when the dimension of the predictor variable p is much larger than the sample size n . Then, an elastic net regression model (implemented in the glmnet R function) [21] was used to regress a transformed model of age based on 111 β values in the training data. The elastic net approach is a combination of traditional lasso and ridge regression methods, emphasizing model sparsity while appropriately balancing the contributions of correlated variables. The glmnet function requires the user to specify two parameters (alpha and lambda). Since we used an elastic net predictor, alpha was set to 0.48, and lambda was set to 0.000954 based on 10-fold cross-validation of the training data (via the R function cv.glmnet). A heat map was drawn in the ‘pheatmap’ package in RStudio, and Venn diagrams were produced on the Bioinformatics and Evolutionary Genomics website (<http://bioinformatics.psb.ugent.be/webtools/Venn/>).

CpG site annotation and enrichment analysis

The Entrez gene IDs of CpG sites in the HumanMethylation27 and HumanMethylation450 annotation files were used to identify genes. Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) analyses were conducted in R with the ‘clusterProfiler’ package from Bioconductor [62]. Enrichment analyses were performed with Fisher’s exact test. Significant GO terms ($P < 0.05$) were

imported into REVIGO for visualization in a semantic space [63].

Statistical analysis

Paired Student's t-tests were used to compare the DNA methylation ages calculated by our model and the multi-tissue predictor for 67 pairs of monozygotic twins in the dataset GSE56105 [23]. Differences between two groups of samples were assessed with Wilcoxon tests and unpaired Student's t-test, while differences among multiple groups of samples were assessed with analysis of variance (ANOVA). *P* values < 0.05 were considered significant. Statistical analyses were performed with RStudio.

ACKNOWLEDGMENTS

I gratefully acknowledge the many researchers who made their DNA methylation datasets publicly available and responded to my email requests. This study would not have been possible without the valuable data from the NCBI GEO (Gene Expression Omnibus) database.

CONFLICTS OF INTERESTS

The authors declare that they have no conflicts of interests.

FUNDING

National Key R&D Plan (No.2018YFC1315400) Guangdong Science and Technology Project (No.2017A020214002)

REFERENCES

1. Tey NP, Siraj SB, Kamaruzzaman SB, Chin AV, Tan MP, Sinnappan GS, Müller AM. Aging in Multi-ethnic Malaysia. *Gerontologist*. 2016; 56:603–09. <https://doi.org/10.1093/geront/gnv153> PMID:26553738
2. Chen R, Xu P, Li F, Song P. Internal migration and regional differences of population aging: an empirical study of 287 cities in China. *Biosci Trends*. 2018; 12:132–41. <https://doi.org/10.5582/bst.2017.01246> PMID:29607873
3. Mota-Pinto A, Rodrigues V, Botelho A, Veríssimo MT, Morais A, Alves C, Rosa MS, de Oliveira CR. A socio-demographic study of aging in the Portuguese population: the EPEPP study. *Arch Gerontol Geriatr*. 2011; 52:304–08. <https://doi.org/10.1016/j.archger.2010.04.019> PMID:20510469

4. Weir HK, Thompson TD, Soman A, Møller B, Leadbetter S, White MC. Meeting the Healthy People 2020 Objectives to Reduce Cancer Mortality. *Prev Chronic Dis*. 2015; 12:E104–104. <https://doi.org/10.5888/pcd12.140482> PMID:26133647
5. Jylhävä J, Pedersen NL, Hägg S. Biological Age Predictors. *EBioMedicine*. 2017; 21:29–36. <https://doi.org/10.1016/j.ebiom.2017.03.046> PMID:28396265
6. Stubbs TM, Bonder MJ, Stark AK, Krueger F, von Meyenn F, Stegle O, Reik W, Reik W, and BI Ageing Clock Team. Multi-tissue DNA methylation age predictor in mouse. *Genome Biol*. 2017; 18:68–68. <https://doi.org/10.1186/s13059-017-1203-5> PMID:28399939
7. Zhang WG, Zhu SY, Bai XJ, Zhao DL, Jian SM, Li J, Li ZX, Fu B, Cai GY, Sun XF, Chen XM. Select aging biomarkers based on telomere length and chronological age to build a biological age equation. *Age (Dordr)*. 2014; 36:9639–9639. <https://doi.org/10.1007/s11357-014-9639-y> PMID:24659482
8. Yang J, Huang T, Song WM, Petralia F, Mobbs CV, Zhang B, Zhao Y, Schadt EE, Zhu J, Tu Z. Discover the network underlying the connections between aging and age-related diseases. *Sci Rep*. 2016; 6:32566–32566. <https://doi.org/10.1038/srep32566> PMID:27582315
9. Maffei VJ, Kim S, Blanchard E 4th, Luo M, Jazwinski SM, Taylor CM, Welsh DA. Biological Aging and the Human Gut Microbiota. *J Gerontol A Biol Sci Med Sci*. 2017; 72:1474–82. <https://doi.org/10.1093/gerona/glx042> PMID:28444190
10. Calkins K, Devaskar SU. Fetal origins of adult disease. *Curr Probl Pediatr Adolesc Health Care*. 2011; 41:158–76. <https://doi.org/10.1016/j.cppeds.2011.01.001> PMID:21684471
11. Skogen JC, Overland S. The fetal origins of adult disease: a narrative review of the epidemiological literature. *JRSM Short Rep*. 2012; 3:59–59. <https://doi.org/10.1258/shorts.2012.012048> PMID:23301147
12. Anwar MA, Saleh AI, Al Olabi R, Al Shehabi TS, Eid AH. Glucocorticoid-induced fetal origins of adult hypertension: association with epigenetic events. *Vascul Pharmacol*. 2016; 82:41–50. <https://doi.org/10.1016/j.vph.2016.02.002> PMID:26903240
13. Briana DD, Malamitsi-Puchner A. Developmental origins of adult health and disease: the metabolic role

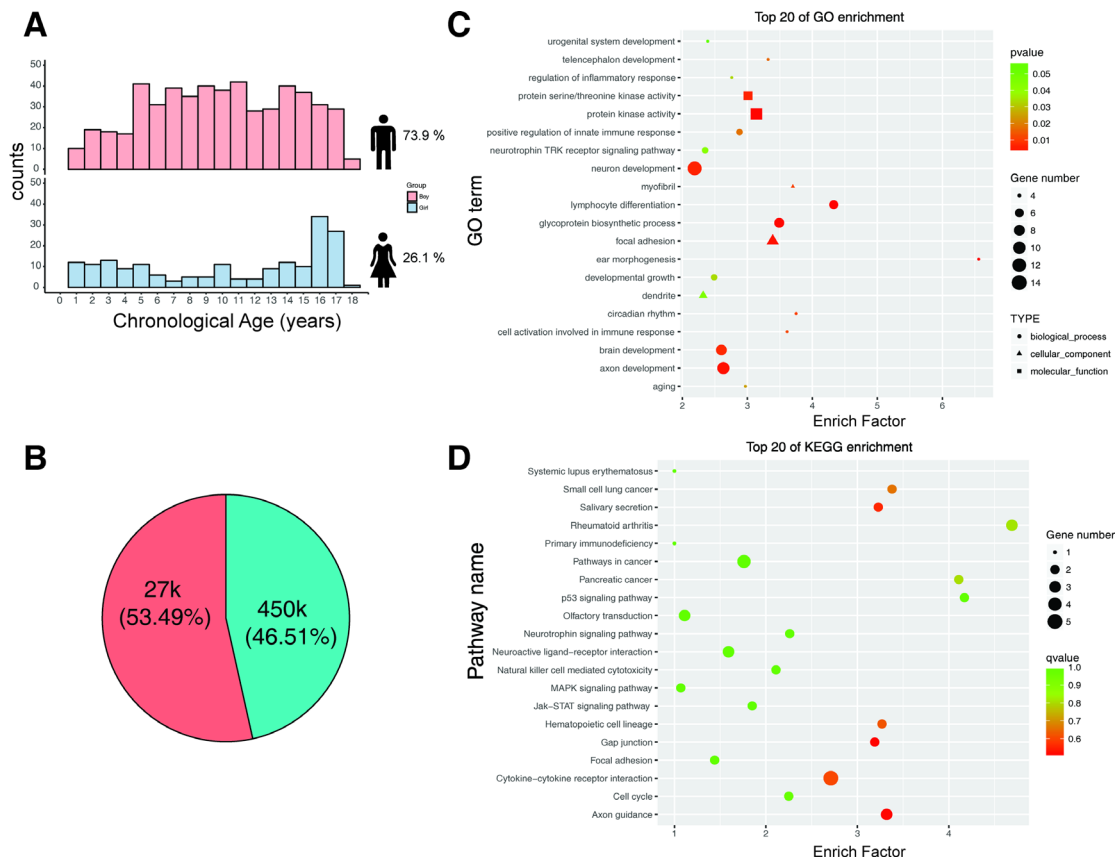
- of BDNF from early life to adulthood. *Metabolism*. 2018; 81:45–51.
<https://doi.org/10.1016/j.metabol.2017.11.019>
PMID:[29217485](https://pubmed.ncbi.nlm.nih.gov/29217485/)
14. Hannum G, Guinney J, Zhao L, Zhang L, Hughes G, Sada S, Klotzle B, Bibikova M, Fan JB, Gao Y, Deconde R, Chen M, Rajapakse I, et al. Genome-wide methylation profiles reveal quantitative views of human aging rates. *Mol Cell*. 2013; 49:359–67.
<https://doi.org/10.1016/j.molcel.2012.10.016>
PMID:[23177740](https://pubmed.ncbi.nlm.nih.gov/23177740/)
 15. Horvath S. DNA methylation age of human tissues and cell types. *Genome Biol*. 2013; 14:R115.
<https://doi.org/10.1186/gb-2013-14-10-r115>
PMID:[24138928](https://pubmed.ncbi.nlm.nih.gov/24138928/)
 16. Weidner CI, Lin Q, Koch CM, Eisele L, Beier F, Ziegler P, Bauerschlag DO, Jöckel KH, Erbel R, Mühleisen TW, Zenke M, Brümmendorf TH, Wagner W. Aging of blood can be tracked by DNA methylation changes at just three CpG sites. *Genome Biol*. 2014; 15:R24.
<https://doi.org/10.1186/gb-2014-15-2-r24>
PMID:[24490752](https://pubmed.ncbi.nlm.nih.gov/24490752/)
 17. Alisch RS, Barwick BG, Chopra P, Myrick LK, Satten GA, Conneely KN, Warren ST. Age-associated DNA methylation in pediatric populations. *Genome Res*. 2012; 22:623–32.
<https://doi.org/10.1101/gr.125187.111>
PMID:[22300631](https://pubmed.ncbi.nlm.nih.gov/22300631/)
 18. Almstrup K, Lindhardt Johansen M, Busch AS, Hagen CP, Nielsen JE, Petersen JH, Juul A. Pubertal development in healthy children is mirrored by DNA methylation patterns in peripheral blood. *Sci Rep*. 2016; 6:28657.
<https://doi.org/10.1038/srep28657> PMID:[27349168](https://pubmed.ncbi.nlm.nih.gov/27349168/)
 19. Freire-Aradas A, Phillips C, Girón-Santamaría L, Mosquera-Miguel A, Gómez-Tato A, Casares de Cal MA, Álvarez-Dios J, Lareu MV. Tracking age-correlated DNA methylation markers in the young. *Forensic Sci Int Genet*. 2018; 36:50–59.
<https://doi.org/10.1016/j.fsigen.2018.06.011>
PMID:[29933125](https://pubmed.ncbi.nlm.nih.gov/29933125/)
 20. Saldana DF, Yang F. SIS : An R Package for Sure Independence Screening in Ultrahigh-Dimensional Statistical Models. *J Stat Softw*. 2018; 83.
<https://doi.org/10.18637/jss.v083.i02>
 21. Friedman J, Hastie T, Tibshirani R. Regularization Paths for Generalized Linear Models via Coordinate Descent. *J Stat Softw*. 2010; 33:1–22.
<https://doi.org/10.18637/jss.v033.i01> PMID:[20808728](https://pubmed.ncbi.nlm.nih.gov/20808728/)
 22. Kuhn M. Building Predictive Models in R Using the caret Package. *J Stat Softw*. 2008; 28:26.
 23. Shah S, McRae AF, Marioni RE, Harris SE, Gibson J, Henders AK, Redmond P, Cox SR, Pattie A, Corley J, Murphy L, Martin NG, Montgomery GW, et al. Genetic and environmental exposures constrain epigenetic drift over the human life course. *Genome Res*. 2014; 24:1725–33. <https://doi.org/10.1101/gr.176933.114>
PMID:[25249537](https://pubmed.ncbi.nlm.nih.gov/25249537/)
 24. Drachler ML, Marshall T, de Carvalho Leite JC. A continuous-scale measure of child development for population-based epidemiological surveys: a preliminary study using Item Response Theory for the Denver Test. *Paediatr Perinat Epidemiol*. 2007; 21:138–53.
<https://doi.org/10.1111/j.1365-3016.2007.00787.x>
PMID:[17302643](https://pubmed.ncbi.nlm.nih.gov/17302643/)
 25. Richter J, Janson H. A validation study of the Norwegian version of the Ages and Stages Questionnaires. *Acta Paediatr*. 2007; 96:748–52.
<https://doi.org/10.1111/j.1651-2227.2007.00246.x>
PMID:[17462065](https://pubmed.ncbi.nlm.nih.gov/17462065/)
 26. Tough SC, Siever JE, Leew S, Johnston DW, Benzie K, Clark D. Maternal mental health predicts risk of developmental problems at 3 years of age: follow up of a community based trial. *BMC Pregnancy Childbirth*. 2008; 8:16. <https://doi.org/10.1186/1471-2393-8-16>
PMID:[18460217](https://pubmed.ncbi.nlm.nih.gov/18460217/)
 27. Kolarova J, Ammerpohl O, Gutwein J, Welzel M, Baus I, Riepe FG, Eggermann T, Caliebe A, Holterhus PM, Siebert R, Bens S. In vivo investigations of the effect of short- and long-term recombinant growth hormone treatment on DNA-methylation in humans. *PLoS One*. 2015; 10:e0120463.
<https://doi.org/10.1371/journal.pone.0120463>
PMID:[25785847](https://pubmed.ncbi.nlm.nih.gov/25785847/)
 28. Aimaretti G, Bellone S, Bellone J, Chiabotto P, Baffoni C, Corneli G, Origlia C, de Sanctis C, Camanni F, Ghigo E. Reduction of the pituitary GH releasable pool in short children with GH neurosecretory dysfunction. *Clin Endocrinol (Oxf)*. 2000; 52:287–93.
<https://doi.org/10.1046/j.1365-2265.2000.00957.x>
PMID:[10718826](https://pubmed.ncbi.nlm.nih.gov/10718826/)
 29. Maghnie M, Lindberg A, Koltowska-Häggström M, Ranke MB. Magnetic resonance imaging of CNS in 15,043 children with GH deficiency in KIGS (Pfizer International Growth Database). *Eur J Endocrinol*. 2013; 168:211–17.
<https://doi.org/10.1530/EJE-12-0801> PMID:[23152438](https://pubmed.ncbi.nlm.nih.gov/23152438/)
 30. Sen A, Heredia N, Senut MC, Hess M, Land S, Qu W, Hollacher K, Dereski MO, Ruden DM. Early life lead exposure causes gender-specific changes in the DNA methylation profile of DNA extracted from dried blood spots. *Epigenomics*. 2015; 7:379–93.
<https://doi.org/10.2217/epi.15.2> PMID:[26077427](https://pubmed.ncbi.nlm.nih.gov/26077427/)
 31. Brubaker CJ, Dietrich KN, Lanphear BP, Cecil KM. The

- influence of age of lead exposure on adult gray matter volume. *Neurotoxicology*. 2010; 31:259–66.
<https://doi.org/10.1016/j.neuro.2010.03.004>
 PMID:20226811
32. Jedrychowski W, Perera F, Jankowski J, Mrozek-Budzyn D, Mroz E, Flak E, Edwards S, Skarupa A, Lisowska-Miszczyk I. Gender specific differences in neurodevelopmental effects of prenatal exposure to very low-lead levels: the prospective cohort study in three-year olds. *Early Hum Dev*. 2009; 85:503–10.
<https://doi.org/10.1016/j.earlhumdev.2009.04.006>
 PMID:19450938
 33. Thesleff I, Mikkola ML. Death receptor signaling giving life to ectodermal organs. *Sci STKE*. 2002; 2002:pe22.
<https://doi.org/10.1126/stke.2002.131.pe22>
 PMID:11997580
 34. Bocklandt S, Lin W, Sehl ME, Sánchez FJ, Sinsheimer JS, Horvath S, Vilain E. Epigenetic predictor of age. *PLoS One*. 2011; 6:e14821.
<https://doi.org/10.1371/journal.pone.0014821>
 PMID:21731603
 35. Bekaert B, Kamalandua A, Zapico SC, Van de Voorde W, Decorte R. Improved age determination of blood and teeth samples using a selected set of DNA methylation markers. *Epigenetics*. 2015; 10:922–30.
<https://doi.org/10.1080/15592294.2015.1080413>
 PMID:26280308
 36. Schildberg T, Rauh J, Bretschneider H, Stiehler M. Identification of suitable reference genes in bone marrow stromal cells from osteoarthritic donors. *Stem Cell Res*. 2013; 11:1288–98.
<https://doi.org/10.1016/j.scr.2013.08.015>
 PMID:24080205
 37. Richardson BC. Role of DNA methylation in the regulation of cell function: autoimmunity, aging and cancer. *J Nutr*. 2002 (8 Suppl); 132:2401S–05S.
<https://doi.org/10.1093/in/132.8.2401S>
 PMID:12163700
 38. Zhang Z, Deng C, Lu Q, Richardson B. Age-dependent DNA methylation changes in the ITGAL (CD11a) promoter. *Mech Ageing Dev*. 2002; 123:1257–68.
[https://doi.org/10.1016/S0047-6374\(02\)00014-3](https://doi.org/10.1016/S0047-6374(02)00014-3)
 PMID:12020947
 39. Xu YQ, Li GM, Li Y. Advanced bone age as an indicator facilitates the diagnosis of precocious puberty. *J Pediatr (Rio J)*. 2018; 94:69–75.
<https://doi.org/10.1016/j.jped.2017.03.010>
 PMID:28866322
 40. Alessandri SB, Pereira FA, Villela RA, Antonini SR, Elias PC, Martinelli CE Jr, Castro M, Moreira AC, Paula FJ. Bone mineral density and body composition in girls with idiopathic central precocious puberty before and after treatment with a gonadotropin-releasing hormone agonist. *Clinics (São Paulo)*. 2012; 67:591–96.
[https://doi.org/10.6061/clinics/2012\(06\)08](https://doi.org/10.6061/clinics/2012(06)08)
 PMID:22760897
 41. Leng RX, Pan HF, Liu J, Yang XK, Zhang C, Tao SS, Wang DG, Li XM, Li XP, Yang W, Ye DQ. Evidence for genetic association of TBX21 and IFNG with systemic lupus erythematosus in a Chinese Han population. *Sci Rep*. 2016; 6:22081.
<https://doi.org/10.1038/srep22081>
 PMID:26916970
 42. Nakao F, Ihara K, Kusahara K, Sasaki Y, Kinukawa N, Takabayashi A, Nishima S, Hara T. Association of IFN-gamma and IFN regulatory factor 1 polymorphisms with childhood atopic asthma. *J Allergy Clin Immunol*. 2001; 107:499–504.
<https://doi.org/10.1067/mai.2001.113051>
 PMID:11240951
 43. Kantarci OH, Goris A, Hebrink DD, Heggarty S, Cunningham S, Alloza I, Atkinson EJ, de Andrade M, McMurray CT, Graham CA, Hawkins SA, Billiau A, Dubois B, et al. IFNG polymorphisms are associated with gender differences in susceptibility to multiple sclerosis. *Genes Immun*. 2005; 6:153–61.
<https://doi.org/10.1038/sj.gene.6364164>
 PMID:15674394
 44. Wang L, He F, Zhong Z, Lv R, Xiao S, Liu Z. Overexpression of NTRK1 Promotes Differentiation of Neural Stem Cells into Cholinergic Neurons. *Biomed Res Int*. 2015; 2015:857202.
<https://doi.org/10.1155/2015/857202>
 PMID:26509167
 45. Braskie MN, Jahanshad N, Toga AW, McMahon KL, de Zubicaray GI, Martin NG, Wright MJ, Thompson PM. How a common variant in the growth factor receptor gene, NTRK1, affects white matter. *Bioarchitecture*. 2012; 2:181–84.
<https://doi.org/10.4161/bioa.22190>
 PMID:22986407
 46. Wright SD, Wright CA, D'Astous V, Wadsworth AM. Autism aging. *Gerontol Geriatr Educ*. 2019; 40:322–338.
<https://doi.org/10.1080/02701960.2016.1247073>
 PMID:27749163
 47. Croen LA, Zerbo O, Qian Y, Massolo ML, Rich S, Sidney S, Kripke C. The health status of adults on the autism spectrum. *Autism*. 2015; 19:814–23.
<https://doi.org/10.1177/1362361315577517>
 PMID:25911091
 48. Hirvikoski T, Mittendorfer-Rutz E, Boman M, Larsson H, Lichtenstein P, Bölte S. Premature mortality in autism spectrum disorder. *Br J Psychiatry*. 2016; 208:232–38.
<https://doi.org/10.1192/bjp.bp.114.160192>
 PMID:26541693

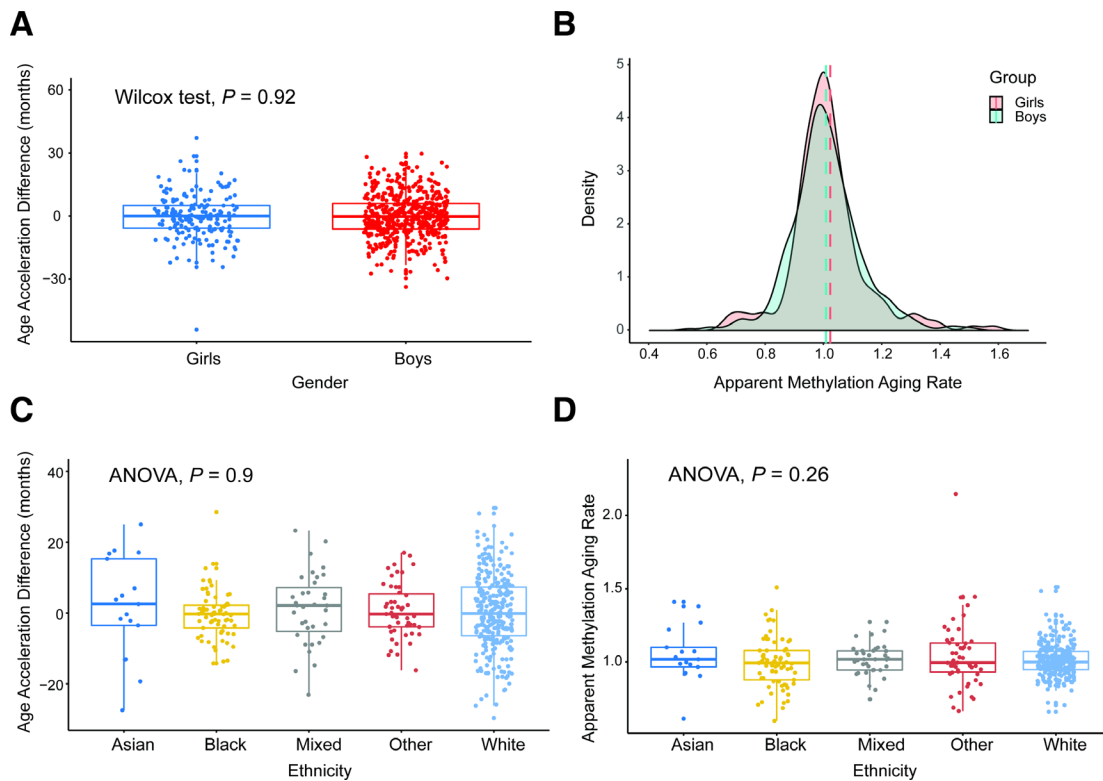
49. van Heijst BF, Geurts HM. Quality of life in autism across the lifespan: a meta-analysis. *Autism*. 2015; 19:158–67.
<https://doi.org/10.1177/1362361313517053>
PMID:24443331
50. Eid A, Zawia N. Consequences of lead exposure, and it's emerging role as an epigenetic modifier in the aging brain. *Neurotoxicology*. 2016; 56:254–61.
<https://doi.org/10.1016/j.neuro.2016.04.006>
PMID:27066759
51. Harris RA, Nagy-Szakal D, Pedersen N, Opekun A, Bronsky J, Munkholm P, Jespersgaard C, Andersen P, Melegh B, Ferry G, Jess T, Kellermayer R. Genome-wide peripheral blood leukocyte DNA methylation microarrays identified a single association with inflammatory bowel diseases. *Inflamm Bowel Dis*. 2012; 18:2334–41. <https://doi.org/10.1002/ibd.22956>
PMID:22467598
52. Voisin S, Almén MS, Moschonis G, Chrousos GP, Manios Y, Schiöth HB. Dietary fat quality impacts genome-wide DNA methylation patterns in a cross-sectional study of Greek preadolescents. *Eur J Hum Genet*. 2015; 23:654–62.
<https://doi.org/10.1038/ejhg.2014.139>
PMID:25074463
53. Brunet A, Berger SL. Epigenetics of aging and aging-related disease. *J Gerontol A Biol Sci Med Sci*. 2014 (Suppl 1); 69:S17–20.
<https://doi.org/10.1093/gerona/glu042>
PMID:24833581
54. Khulan B, Cooper WN, Skinner BM, Bauer J, Owens S, Prentice AM, Belteki G, Constanca M, Dunger D, Affara NA. Periconceptional maternal micronutrient supplementation is associated with widespread gender related changes in the epigenome: a study of a unique resource in the Gambia. *Hum Mol Genet*. 2012; 21:2086–101. <https://doi.org/10.1093/hmg/dds026>
PMID:22307237
55. Chen YA, Choufani S, Ferreira JC, Grafodatskaya D, Butcher DT, Weksberg R. Sequence overlap between autosomal and sex-linked probes on the Illumina HumanMethylation27 microarray. *Genomics*. 2011; 97:214–22.
<https://doi.org/10.1016/j.ygeno.2010.12.004>
PMID:21211562
56. Horvath S, Zhang Y, Langfelder P, Kahn RS, Boks MP, van Eijk K, van den Berg LH, Ophoff RA. Aging effects on DNA methylation modules in human brain and blood tissue. *Genome Biol*. 2012; 13:R97.
<https://doi.org/10.1186/gb-2012-13-10-r97>
PMID:23034122
57. Bacalini MG, Gentilini D, Boattini A, Giampieri E, Pirazzini C, Giuliani C, Fontanesi E, Scurti M, Remondini D, Capri M, Cocchi G, Ghezzi A, Del Rio A, et al. Identification of a DNA methylation signature in blood cells from persons with Down Syndrome. *Aging (Albany NY)*. 2015; 7:82–96.
<https://doi.org/10.18632/aging.100715>
PMID:25701644
58. Voisin S, Almén MS, Zheleznyakova GY, Lundberg L, Zarei S, Castillo S, Eriksson FE, Nilsson EK, Blüher M, Böttcher Y, Kovacs P, Klovins J, Rask-Andersen M, Schiöth HB. Many obesity-associated SNPs strongly associate with DNA methylation changes at proximal promoters and enhancers. *Genome Med*. 2015; 7:103.
<https://doi.org/10.1186/s13073-015-0225-4>
PMID:26449484
59. Troyanskaya O, Cantor M, Sherlock G, Brown P, Hastie T, Tibshirani R, Botstein D, Altman RB. Missing value estimation methods for DNA microarrays. *Bioinformatics*. 2001; 17:520–5.
<https://doi.org/10.1093/bioinformatics/17.6.520>
PMID:11395428
60. Teschendorff AE, Marabita F, Lechner M, Bartlett T, Tegner J, Gomez-Cabrero D, Beck S. A beta-mixture quantile normalization method for correcting probe design bias in Illumina Infinium 450 k DNA methylation data. *Bioinformatics*. 2013; 29:189–96.
<https://doi.org/10.1093/bioinformatics/bts680>
PMID:23175756
61. Benjamini Y, Hochberg Y. Controlling the false discovery rate - a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society*. 1995; 57:289–300.
<https://doi.org/10.1111/j.2517-6161.1995.tb02031.x>
62. Gentleman RC, Carey VJ, Bates DM, Bolstad B, Dettling M, Dudoit S, Ellis B, Gautier L, Ge Y, Gentry J, Hornik K, Hothorn T, Huber W, et al. Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol*. 2004; 5:R80.
<https://doi.org/10.1186/gb-2004-5-10-r80>
PMID:15461798
63. Supek F, Bošnjak M, Škunca N, Šmuc T. REVIGO summarizes and visualizes long lists of gene ontology terms. *PLoS One*. 2011; 6:e21800.
<https://doi.org/10.1371/journal.pone.0021800>
PMID:21789182

SUPPLEMENTARY MATERIALS

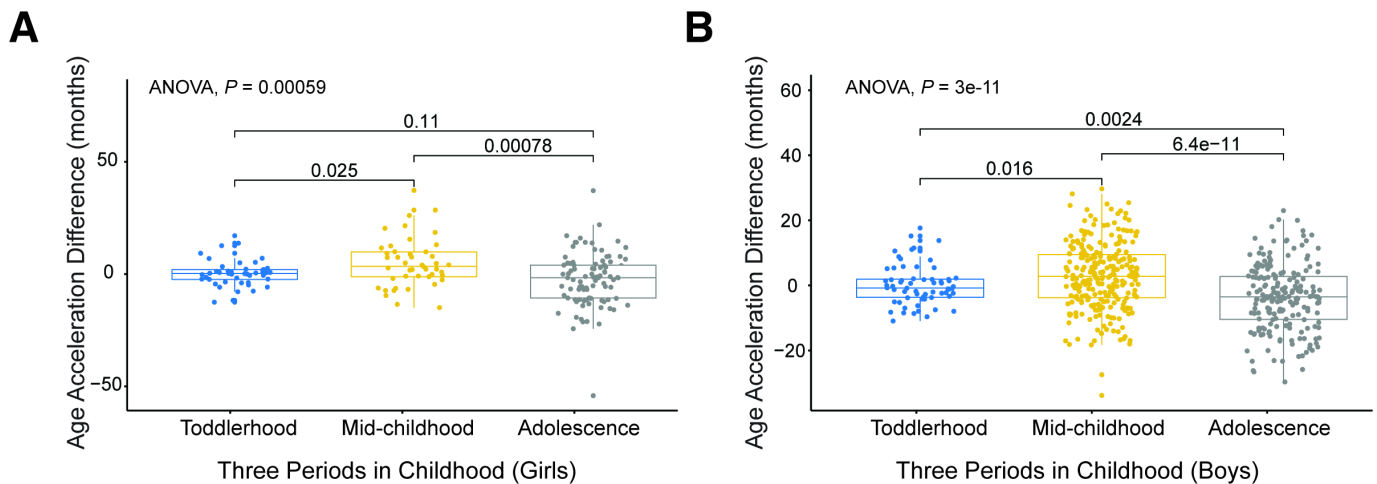
Supplementary Figures



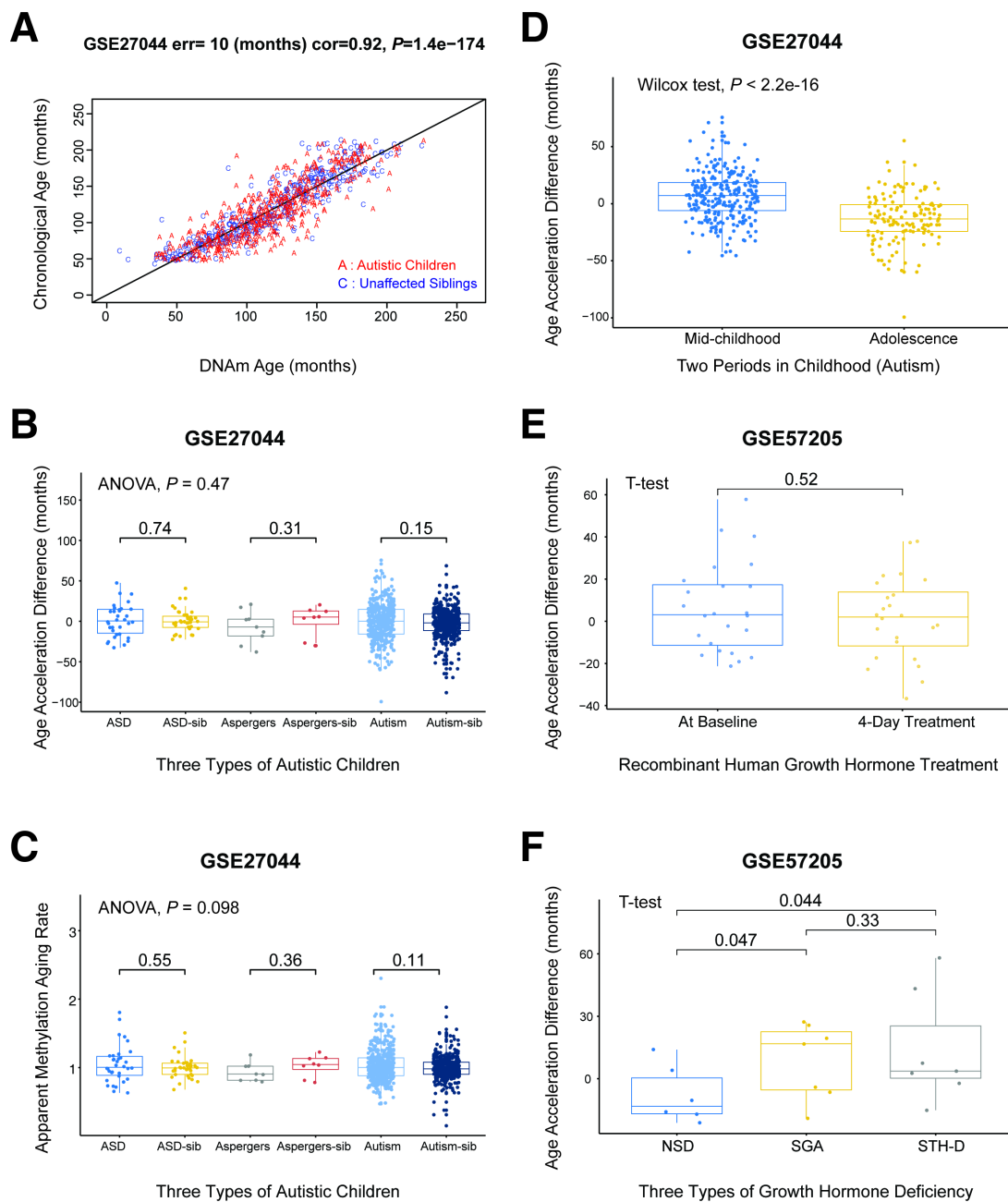
Supplementary Figure 1. Characteristics of the prediction model. (A) Histograms of the age distribution for healthy girls and boys, respectively. The x-axis represents the chronological age of the individuals (age unit is years) and the y-axis (counts) represents the number of individuals. The pink columns indicate boys, and the blue columns indicate girls. (B) Pie chart displaying the proportion of the different types of Illumina Methylation Assays. Red stands for the 27K assay and green stands for the 450K assay. (C) Gene Ontology (GO) analysis of the genes associated with the 111 CpG sites. (D) Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis of the genes associated with the 111 CpG sites.



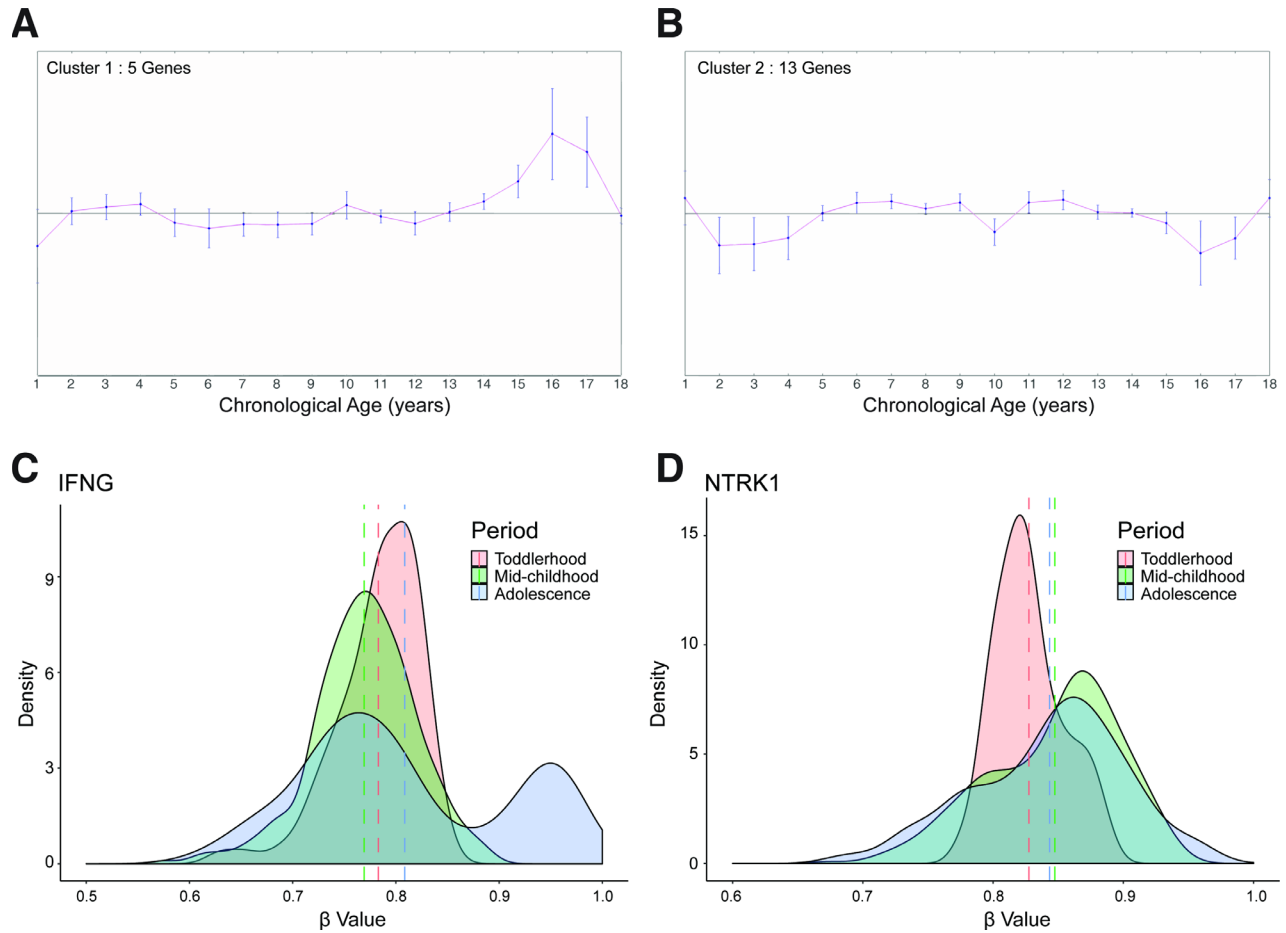
Supplementary Figure 2. Age acceleration in children grouped by gender and ethnicity. (A) Boxplot comparing the age acceleration difference between girls and boys. (B) Density plot of the apparent methylation aging rates of children separated by gender. Dashed lines represent the mean values. (C) Boxplot comparing the age acceleration differences among children of different ethnicities. (D) Boxplot comparing the apparent methylation aging rates among children of different ethnicities.



Supplementary Figure 3. Age acceleration in different periods of childhood by gender. (A) Boxplot of the age acceleration difference during different periods of childhood in girls. (B) Boxplot of the age acceleration difference during different periods of childhood in boys.



Supplementary Figure 4. Age acceleration in children with diseases. (A) Scatterplot of the DNA methylation (DNAm) age (x -axis) against the chronological age (y -axis) in the GSE27044 dataset (age unit is months). In this dataset, the correlation between the DNAm age and the chronological age was 0.92, and the error (median absolute difference) was 10 months. ‘A’ stands for autistic children and ‘C’ stands for their unaffected siblings. (B) Boxplot comparing the age acceleration differences among children with different types of autism and their unaffected siblings (‘-sib’). (C) Boxplot comparing the apparent methylation aging rates among children with different types of autism. (D) Boxplot comparing the age acceleration differences in mid-childhood and adolescence in children with autism. (E) Boxplot comparing the age acceleration differences before and after four days of recombinant human growth hormone treatment. (F) Boxplot comparing the age acceleration differences among children with different types of growth hormone deficiency.



Supplementary Figure 5. K-means clustering analysis of the 111 CpG sites. (A) The trend of DNA methylation levels and the number of genes associated with the CpG sites in Cluster 1. (B) The trend of DNA methylation levels and the number of genes associated with the CpG sites in Cluster 2. (C) Density plot of the β value of cg26227465, located near the *IFNG* gene, during different periods of childhood. This site was in Cluster 1. (D) Density plot of the β value of cg25827666, located near the *NTRK1* gene, during different periods of childhood. This site was in Cluster 2.

Supplementary Tables

Supplementary Table 1. Summary details of the DNA methylation datasets on children in disease.

ID	Availability	Methylation array	n	Age(months)	Gender	Disease	Ethnicity	Citation
1	GSE27044	Illumina 27K	866	43.0-214.0	M: 866	Autism: 397; Autism-sib: 382; ASD: 34; ASD-sib: 36; Aspergers: 9; Aspergers-sib: 8;	White: 698; Asian: 32; African-amer: 14; other: 25; more-than-one-race: 80; Native-american: 2; Native-hawaiian: 2; not-specified: 13;	Alisch et al.
2	GSE57205	Illumina 450K	48(24)	62.0-185.0	M: 14 F: 10	NSD: 6; SGA: 7; STH-D: 7; Q-STH-D: 2; IGF1-D: 1; UTS: 1;	null	Kolarova et al.
3	GSE60598	Illumina 27K	42	3.0-60.0	M: 24 F: 18	BLL \geq 5 μ g/dl: 25; BLL < 5 μ g/dl: 17;	African_american: 41; Caucasion_mixed: 1	Sen et al.

1. NSD: GH deficiency due to neurosecretory dysfunction; SGA: small for gestational age; STH-D: classical GH deficiency; Q-STH-D: qualitative GH deficiency (Kowarski syndrome); IGF1-D: IGF1 deficiency; UTS: Turner-Syndrome;
2. GSE57205 contains 48 blood samples from 24 patients: 24 samples at baseline and 24 samples after 4-day short-term recombinant human growth hormone treatment stimulation;
3. BLL: blood lead level.

Please browse Full Text version to see the data of Supplementary Tables 2 to 4.

Supplementary Table 2. The regression coefficients of child-specific methylation-based age prediction model.

Supplementary Table 3. The result of REVIGO on the significant GO terms.

Supplementary Table 4. The predict details of GSE27044 dataset.