**Supplementary text**

Additional information on the methods, statistical analysis, and results.

**METHODS**

**Material and methods**

Wave 4 of the Finntwin12 study consisted of an invitation to attend an in-person study in Helsinki, targeting twins that had been interviewed at age 14. The response rate was 73% (*n* = 1,347; 21–25-year-old twin individuals). Participants filled several questionnaires about their health-related behaviour and psychosocial factors, including questionnaires about lifestyle factors. They were also interviewed with a semi-structured psychiatric interview. DNA from blood samples was collected. A total of 1,295 twins of the FinnTwin12 study provided a DNA sample and written informed consent. The sample includes opposite-sex pairs who are not analyzed here due to gender differences in physical activity behaviours and experiences.

The older cohort fourth-wave questionnaire on health and risk factors was used to invite and select participants for a detailed in-person study on the epigenetics of hypertension. A total of 447 twins were examined and measured, and blood samples were collected for DNA analyses after written informed consent was signed. A lifestyle interview including items on medications, use of alcohol, smoking, diet, occupation, and physical activity was performed (1,2).

**DNA methylation.** High molecular weight white blood cell DNA was bisulfite-converted using EZ-96 DNA methylation-Gold Kit (Zymo Research, Irvine, CA, USA), according to the manufacturer’s instructions, and the co-twins were always converted on the same plate to minimize potential batch effects. Genome-wide DNAm was measured using Illumina’s Infinium HumanMethylation450 BeadChip and the Infinium MethylationEPIC BeadChip (LTPA discordant twins), according to the manufacturer’s instructions (Illumina, San Diego, CA, USA). The Illumina BeadChips measure single-CpG resolution DNAm levels across the human genome. DNAm data were preprocessed using R package *minfi* (3)*.* Detection *P* values comparing total signal for each probe to the background signal level, were calculated to evaluate the quality of the samples (4). Samples of poor quality (mean detection *P* > 0.01) were excluded from further analysis. Because data included samples from different platforms, data were normalized by using the single-sample Noob normalization method (5). Beta values representing CpG methylation levels were calculated as a ratio of methylated intensities (M) to the overall intensities (Beta value=M/(M+U+100), where U is an unmethylated probe intensity).

**Epigenetic age estimates**

The development of all the three DNAm-based age estimators (DNAm age, DNAm PhenoAge and DNAm GrimAge) have been conductedutilizing the penalized regression model (6–8). The method automatically selects the best predictors for the model from large datasets**. Horvath’s DNAm age** is an estimate for chronological age (6). It was developed by regressing the chronological age on DNAm levels within 21,369 of CpG sites. As a result, 353 optimal CpG sites were selected for the age predictor by the algorithm. **DNAm PhenoAge** was developed by using ‘phenotypic age’ as reference (7). Phenotypic age was developed by regressing the mortality risk on 42 clinical biomarkers and chronological age. Phenotypic age is a combination (weighted sum) of chronological age and the following clinical biomarkers, that best predicted the mortality risk: albumin (with negative weighting coefficient), creatinine, glucose, C-reactive protein, lymphocyte percent (with negative weighting coefficient), mean red cell volume, red cell distribution width, alkaline phosphatase and white blood cell count. Next, phenotypic age was regressed on DNAm levels within 20,169 CpG sites. As a result, 513 CpG sites were selected for the phenotypic age predictor by the algorithm. **DNAm GrimAge** was developed at two stages, as well (8). At the first stage, DNAm based surrogate biomarkers for 88 plasma proteins and smoking pack-years were developed. At the second stage, mortality risk was regressed on the DNAm based surrogate biomarkers, chronological age and sex. As a result, eight DNAm based surrogate biomarkers, chronological age and sex were selected for mortality risk predictor by the algorithm. DNAm GrimAge is a composite biomarker (weighted sum) of chronological age, sex, DNAm based smoking pack-years (PACYRS) and the following surrogate biomarkers for plasma proteins (DNAm based plasma proteins): DNAm adrenomedullin (ADM), DNAm beta-2-microglobulin (B2M), DNAm cystatin-C, DNAm growth differentiation factor 15 (GDF15), DNAm leptin (with negative weighting coefficient), DNAm plasminogen activator inhibitor 1 (PAI-1) and DNAm tissue inhibitor metalloproteinases 1 (TIMP-1).

In this study, the epigenetic age estimates, DNAm age, DNAm PhenoAge, and DNAm GrimAgewere calculated using an online calculator (<https://dnamage.genetics.ucla.edu/new>). We utilized normalization method implemented in the calculator. Epigenetic age acceleration (Acc), which describes the difference between chronological age and epigenetic age estimate, was calculated for all subjects as the residuals from a linear regression model of epigenetic age estimate on chronological age.

**STATISTICAL ANALYSIS**

The correlations and the parameters of the models were estimated using the full information maximum likelihood with robust standard errors. The data were assumed to be missing at random (MAR).

**Univariate twin models**

Univariate modelling was carried out to study the magnitude of genetic and environmental factors affecting epigenetic age Acc measures, and physical activity (PA) indices separately in young adult and older twin pairs. Because the dataset includes monozygotic (MZ) and dizygotic (DZ) twin pairs raised in the same home, it was possible to decompose the variation in the outcome of interest into the latent variables representing additive genetic (A), dominant genetic (D) or shared environmental (C) and non-shared environmental (E) components (ACE model or ADE model) (9). However, C and D components cannot be estimated simultaneously using twin data alone. The sequences of the models were fitted in both cohorts (ACE, ADE, AE, CE and E). The model including D and E components (DE-model), was omitted because it is less plausible biologically (dominance in the absence of additive effects is rare). The univariate models were adjusted for sex, BMI and smoking status.

The Satorra-Bentler scaled chi-squared (χ2)-test, the comparative fit index (CFI), the Tucker–Lewis index (TLI), the root mean square error of approximation (RMSEA), and the standardized root-mean-square residual (SRMR) were used to evaluate the goodness-of-fit of the models. The model fits the data well if the χ2-test is insignificant (*P* > 0.05), CFI and TLI values are close to 0.95, the RMSEA value is below 0.06, and the SRMR value is below 0.08 (10). In addition, Bayesian information criterion (BIC) was used to compare non-nested models. A lower BIC value indicates a better model fit. The parsimonious model with the most appropriate fit to the data was considered optimal. After that, the model was fitted simultaneously in young adult and older twin pairs by using the multiple-group analysis method. The differences in the proportions explained by genetic and environmental factors were tested for significance using Wald test of parameter constraints.

**Bivariate twin models**

First, the full ACE- or ADE-model, including all the components for both latent variables CH1 and CH2 were fit. CH1 represents the PA index, and CH2 represents the residual of epigenetic age Acc after the impact of PA index has been taken into account. After that, a more parsimonious model was considered based on the results of the univariate modeling. The model fit of the nested models was compared by using the Satorra–Bentler scaled chi-squared (χ2) difference test. The more parsimonious model was chosen if the test produced a non-significant loss of fit.

**RESULTS**

**Univariate models**

Mainly, the assumptions of genetic twin modelling were met. There were no systematic differences in the means or variances in the epigenetic age Acc measures or PA indices between MZ and DZ twins. For all the epigenetic age Acc measures, the univariate model including additive genetic and non-shared environmental component (AE-model) was considered optimal in both young adult and older twin pairs (see Tables, SDC 2‒3, the estimation results of the univariate model for epigenetic aging among young adult and older twin pairs, respectively). However, ACE and ADE generally fit the data about as well. Apart from DNAm GrimAge Acc in older twins, CE models (i.e models with no genetic component) provided significantly worse fit.

In young adults, ICCs for PA indices were consistently higher in MZ twins than in DZ twins (Table 2). In the older cohort, the ICCs were quite similar in MZ and DZ twins. However, after adjusting for covariates, AE-model was considered optimal for all the PA indices in young adult and older twin pairs (see Tables, SDC 4‒5, the estimation results of the univariate model for physical activity indices among young adult and older twin pairs, respectively). In young adult twin pairs, additive genetic factors explained 60% of the variation in the sport index, 48% of the variation in the leisure index and 49% of the variation in the work index, while the rest of the variation in each index was explained by non-shared environmental effects (40%, 52% and 51%, respectively). In older twin pairs, additive genetic factors explained 35% of the variation in the sport index, 28% of the variation in the leisure index and 34% of the variation in the work index while the rest of the variation in each index was explained by non-shared environmental effects (65%, 72% and 66%, respectively).

**Bivariate models**

For each PA index and DNAm Grim Acc, the model including all the ACE-model components for both latent variables CH1 and CH2 was fitted. After that, the shared environmental components were omitted based on the results of univariate modelling. In young adult and older twin pairs, the models fitted the data equally well and thus, the more parsimonious model was chosen (see Table, SDC 6, the model fit of the bivariate models).

**REFERENCES**

1. Kaprio J. The Finnish Twin Cohort Study: an update. *Twin Res Hum Genet*. 2013;16(1):157–62.

2. Kaprio J, Bollepalli S, Buchwald J, Iso-Markku P, Korhonen T, Kovanen V, et al. The older Finnish Twin Cohort: 45 years of follow-up. *Twin Res Hum Genet*. 2019;22(4):240–54.

3. Aryee MJ, Jaffe AE, Corrada-Bravo H, Ladd-Acosta C, Feinberg AP, Hansen KD, et al. Minfi: A flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. *Bioinformatics*. 2014;30(10):1363–9.

4. Maksimovic J, Phipson B, Oshlack A. A cross-package Bioconductor workflow for analysing methylation array data [version 3; peer review: 4 approved]. *F1000Res*. 2017;5:1281.

5. Fortin JP, Triche TJ, Hansen KD. Preprocessing, normalization and integration of the Illumina HumanMethylationEPIC array with minfi. *Bioinformatics*. 2017;33(4):558–60.

6. Horvath S. DNA methylation age of human tissues and cell types. *Genome Biol*. 2013;14:R115.

7. Levine ME, Lu AT, Quach A, Chen BH, Assimes TL, Bandinelli S, et al. An epigenetic biomarker of aging for lifespan and healthspan. *Aging (Albany NY)*. 2018;10(4):573–91.

8. Lu AT, Quach A, Wilson JG, Reiner AP, Aviv A, Raj K, et al. DNA methylation GrimAge strongly predicts lifespan and healthspan. *Aging (Albany NY)*. 2019;11(2):303–27.

9. Neale MC, Cardon LR. Methodology for Genetic Studies of Twins and Families. Dordrecht, The Netherlands: Kluver Academic Publisher; 1992.

10. Hu L, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Struct Equ Model A Multidiscip J*. 1999;6(1):1–55.