Research Paper

Different epigenetic signatures of newborn telomere length and telomere attrition rate in early life

Congrong Wang[1], Tim S. Nawrot[1,2], Charlotte Van Der Stukken[1], Dominika Tylus[1], Hanne Sleurs[1], Martien Peusens[1], Rossella Alfano[1], Sabine A.S. Langie[1,3], Michelle Plusquin[1], Dries S. Martens[1]

[1]Centre for Environmental Sciences, Hasselt University, Hasselt, Belgium
[2]Department of Public Health and Primary Care, Leuven University, Leuven, Belgium
[3]Department of Pharmacology and Toxicology, School for Nutrition and Translational Research in Metabolism (NUTRIM), Maastricht University, Maastricht, The Netherlands

Correspondence to: Dries S. Martens; **email:** dries.martens@uhasselt.be
Keywords: telomere length, telomere attrition, DNA methylation, early life, newborn
Received: January 26, 2021 **Accepted:** May 17, 2021 **Published:** June 4, 2021

ABSTRACT

Telomere length (TL) and telomere shortening are biological indicators of aging, and epigenetic associates have been found for TL in adults. However, the role of epigenetic signatures in setting newborn TL and early life telomere dynamics is unknown. In the present study, based on 247 participating newborns from the ENVIR*ON*AGE birth cohort, whole-genome DNA methylation, profiled on the Illumina MethylationEPIC BeadChip microarray, and TL were measured in cord blood. In a follow-up visit at a mean age of 4.58 years, leukocyte TL was evaluated. We combined an epigenome-wide association study and a statistical learning method with re-sampling to select CpGs and their two-way interactions to model baseline (cord blood) TL and early-life telomere attrition rate, where distinct epigenetic signatures were identified for the two outcomes. In addition, a stronger epigenetic regulation was suggested in setting newborn TL than that of telomere dynamics in early life: 47 CpGs and 7 between-CpG interactions explained 76% of the variance in baseline TLs, while 72% of the total variance in telomere attrition rate was explained by 31 CpGs and 5 interactions. Functional enrichment analysis based on the selected CpGs in the two models revealed GLUT4 translocation and immune cell signaling pathways, respectively. These CpGs and interactions, as well as the cellular pathways, are potential novel targets of further investigation of telomere biology and aging.

INTRODUCTION

Telomeres are protective nucleoprotein caps built up from tandem repeats of the hexamer sequence "TTAGGG" at the ends of chromosomes which is crucial for chromosomal stability [1]. Telomeres progressively shorten after each cellular division, owing to the incomplete replication of DNA molecules at the end of chromosomes (referred as the end-replication problem) and telomere maintenance mechanisms that are not capable of preventing telomere attrition. Therefore, telomere length reflects the cycles of mitosis a cell has been through and the extent of biological aging. Telomere length in a somatic cells is the longest at birth and telomere attrition occurs throughout the lifespan [2]. Shorter telomere length is an indicator of a higher risk of age-related diseases and a predictor of mortality [3–7]. In addition to cell division, telomere dynamics has been shown to associate with genetic alterations [8], biochemical factors such as oxidative stress [9], socioeconomic status [10], environmental exposures [11–13] and lifestyle factors such as obesity [14].

Another widely studied molecular marker of the biology of aging is DNA methylation. DNA methylation in human cells is a modification of DNA molecules by the

addition of methyl groups, in the most widely studied context, to cytosine residue in cytosine-phosphate-guanines (CpGs) forming 5-methylcytosines, which plays a pivotal role in the regulation of human gene expression [15]. Differences in DNA methylation, similar to telomere length, exhibit strong correlations with aging, longevity and age-related diseases [16, 17]. Studies have identified various sets of CpGs as predictors of human biological age, which formed the epigenetic age clocks [18–20], for wide-ranged age groups as well as specifically for newborns [21, 22]. The difference between epigenetic age and chronological age measures biological aging status [23].

There have been investigations in the link between these two biomarkers of aging. A population-based study [24] elucidated that DNA methylation at specific sub-telomeric CpGs and imprinted loci were associated with telomere length. A DNA methylation-based predictor of telomere length in adults has been developed and validated in different tissues and cell types which predicted mortality and aging-related diseases [25]. However, despite that telomere length in early life is an important determinant of telomere length in adulthood [26, 27] and shows high variation across newborns [28], currently no study has examined the association between telomere length and DNA methylation in newborns. The regulation of telomere length at early stage of life might involve DNA methylation at different loci compared to those in adulthood, as both telomere length and DNA methylation are dynamic processes over life [2, 29]. Furthermore, no study yet has reported such findings in a longitudinal setting allowing for telomere tracking from birth to young childhood.

We hypothesized that newborn genome-wide DNA methylation is associated with telomere length at birth and telomere attrition rate in early life. The present study was performed on a subset of mother-newborn pairs from the ENVIRonmental influence *ON* AGEing in early life (ENVIR*ON*AGE) birth cohort [30]. We aimed to identify CpGs and CpG-CpG interactions from whole-genome DNA methylation profiles that jointly explained the variation in cord blood telomere lengths and telomere attrition from birth to childhood (age of 4 years), respectively. The identified CpGs were mapped to biological pathways to facilitate our understanding in the potential underlying epigenetic mechanisms of early-life telomere biology.

RESULTS

Study population

Table 1 shows the demographic characteristics of the mother-newborn (mother-child) pairs. The 247 newborns/children had a mean (standard deviation, SD) gestational age of 39.24 (1.37) weeks, and age of 4.58 (0.39) years at follow-up. The study population included 121 (49.0%) girls, and 93.5% of newborns were of European-origin. The mean (SD) cord blood telomere length, as a surrogate for baseline telomere length, was 1.17 (0.27) T/S ratio. The leukocyte telomere length of the children was 1.02 (0.19) T/S and on average 0.14 T/S shorter than cord blood telomere length ($p<0.0001$). The variations in cord blood telomere lengths and leukocyte telomere lengths were significantly different ($p<0.0001$). Telomere attrition corrected for regression-to-the-mean (RTM) effect was on average 0.0025 (0.14) T/S ratio, and telomere attrition rate was on average 0.00072 (0.030) T/S per year.

The DNAm-based explanatory model of baseline telomere length

Selection of CpGs

The study design is schematically shown in Figure 1. The epigenome-wide association study (EWAS) of cord blood telomere length, as a pre-screening for candidate CpG predictors from 787,264 CpG probes, identified 244 individual CpGs to be significant (FDR=5%) (Figure 2A). Two CpGs passed the p-value threshold under Bonferroni correction. A loose criterion of p-value<0.01 was applied to allow a sufficiently large number of CpGs in further selection, also considering that CpGs that jointly associate with baseline telomere length might not be the individually significant ones. Therefore, 22,817 CpGs passed the pre-screening and were entered in a subsequent selection in one and same adaptive lasso [31] model which accounts for the between-CpG correlations (Supplementary Table 1). 208 CpGs were jointly selected by adaptive lasso. In order to account for the CpG-CpG interactions, all possible pairwise interactions (21,528 interactions) were added, making it in total 21,736 terms subject to an elastic net selection [32]. In order not to overfit the data, we repeatedly sampled 1000 random subsets from the whole dataset, each containing 85% of the observations (n=210). Variation was observed among the 1000 subsets in terms of the elastic net regularization parameters and the number of selected predictor terms, as shown in Supplementary Figure 1A. Finally, 36 CpG main effects and 7 interactions involving 47 CpGs in total, which were selected in at least 95% of the subsets, were kept as the most relevant. We added the CpGs which were involved in the selected interactions but not in the selected main effects, and thus the final model contained 47 CpG main effects and 7 interactions (54 terms in total).

The final model and interpretations

The Pearson correlation between baseline telomere length and the weighted sum of the 54 terms was 0.89

Table 1. Characteristics of the study population (n=247).

		N (%) or Mean ± SD [i]
Newborns		
Sex	Female	121 (49.0%)
Ethnicity	European	231 (93.55)
Birth weight (kg)		3.39 ± 0.44
Gestational age (weeks)		39.24 ± 1.37
Age at follow-up (years)		4.58 ± 0.17
Telomere length (T/S ratio)		
	Cord blood	1.17 ± 0.27
	Leukocyte	1.02 ± 0.19
	Telomere attrition [ii]	0.0025 ± 0.14
Telomere attrition rate (T/S per year) [iii]		0.00072 ± 0.030
Mothers		
Educational level [iv]		
	No diploma	24 (9.75%)
	High school diploma	67 (27.1%)
	A 3-year college	124 (50.2%)
	A 4-year college or university	32 (13.0%)
Smoking status		
	Never smoker	167(67.6%)
	Former smoker	48 (19.4%)
	Smoker	32 (13.0%)
Parity		
	Primiparous	129 (52.2%)
	Secundiparous	95 (38.5%)
	Multiparous	23 (9.3%)
With pregnancy complications		37 (15.0%)
Age at delivery (years)		30.21 ± 4.38
Pre-pregnancy BMI (kg/m^2)		24.28 ± 4.56
Fathers		
Age at child birth (years)		32.87 ± 5.60

[i]Categorical variables were summarized with frequency (N) and percentage (%), while continuous variables were summarized as mean and standard deviation (SD).

[ii]Telomere attrition was calculated as the raw difference of cord blood telomere length and leukocyte telomere length corrected for regression-to-the-mean effect.

[iii]Telomere attrition rate was calculated as telomere attrition divided by the children's age at follow-up visit.

[iv]Educational level was defined as the highest diploma obtained by the mother.

and the R^2 was 0.76. Figure 3A and Supplementary Table 2 show the contributing weights (final model coefficients) and the EWAS coefficients of the CpGs and their interactions, as well as the genomic locations. 8 out of the 47 CpG main effects in the final model had been identified significant in EWAS. In the final model, 26 main effect CpGs had a positive coefficient (being positively associated with baseline telomere length) and 21 had a negative coefficient (being inversely associated with baseline telomere length), consistent with their EWAS coefficients. 15 CpGs were promoter-associated and 1 was related to differentially methylated regions (DMRs). The genomic locations of the involved 47 CpGs were distributed across autosomes except chromosomes 5, 9, 18, 20 and 21. All the 7 interaction terms involved CpGs from different chromosomes, among which 3 had positive coefficients indicating that the interactors are mutually promoting their associations with baseline telomere length. The other 4 pairs of interactions with negative coefficients indicated mutual suppression of the associations between their methylation level and baseline telomere length.

40 of the 47 CpGs had gene name annotations, from which the pathway analysis suggested one significantly enriched biological pathway "translocation of SLC2A4 (GLUT4) to the plasma membrane" (Figure 4, left panel). Correlating the selected CpGs to RNA probes from the global gene expression microarray suggested 75 significantly correlated RNAs ($p<0.05$), which had no biological pathways significantly enriched.

The DNAm-based explanatory model of telomere attrition rate in early childhood

Selection of CpGs

Three CpGs were identified to associate with telomere attrition rate based on an FDR of 5% and are highlighted in Figure 2B. Only cg26557617 passed Bonferroni correction. 9,862 CpGs ($p<0.01$) were used

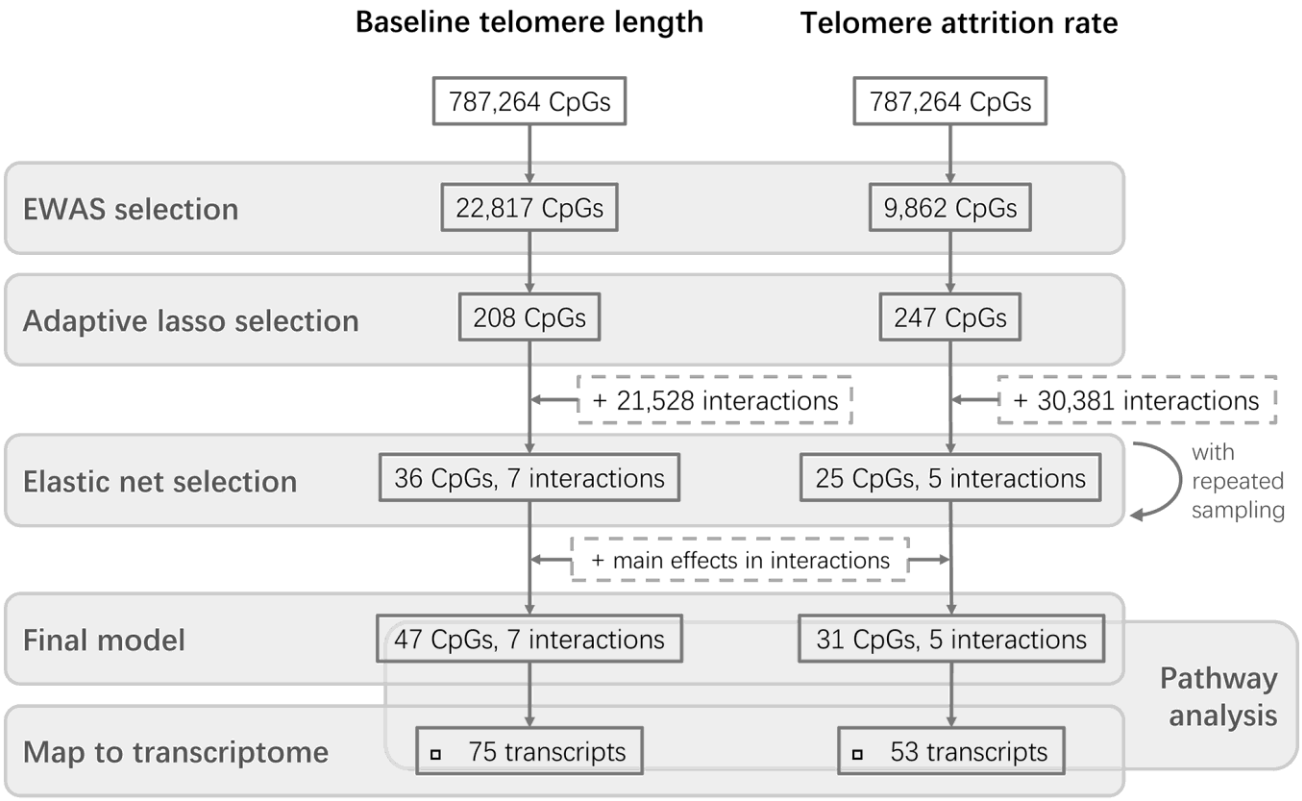


Figure 1. Flowchart of the study design. Two parallel stepwise model selections were performed for baseline telomere length and telomere attrition rate, respectively, followed by further investigations of the CpGs in the final models.

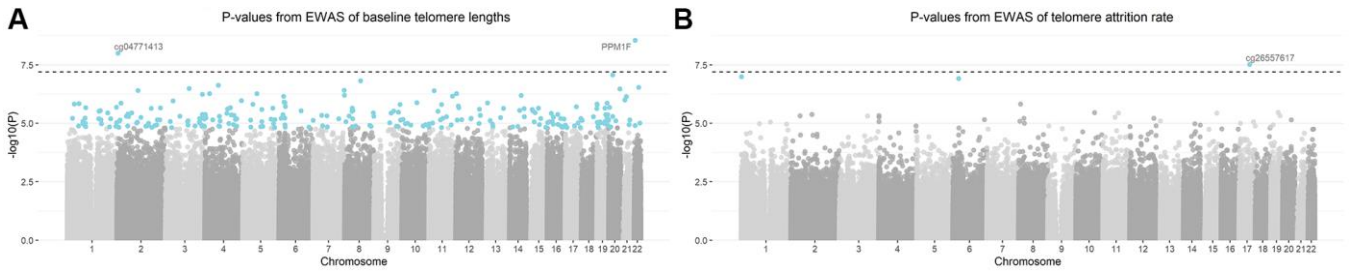


Figure 2. Manhattan plots of -log10 transformed p-values from the epigenome-wide association analysis (EWAS). Models were adjusted for gestational age, newborns' sex, ethnicity and birth weight, maternal age, pre-pregnancy body-mass-index, parity, presence of pregnancy complications, education level and smoking status, paternal age and estimated blood cell counts. CpGs highlighted in blue were significant based on a false discovery rate of 5%. CpG annotated with gene names (or CpG probe names in case that no gene name annotation was available) were the CpGs passing Bonferroni threshold (the dashed horizontal line) of 6.35×10^{-8}. (**A**) Manhattan plot of baseline telomere length; (**B**) Manhattan plot of telomere attrition rate.

for subsequent selection. 247 CpGs were then selected by an adaptive lasso model, based on which there were 30,381 pairwise interactions, summing up to 30,628 terms. The most relevant terms selected using an elastic net in at least 95% of the subsets were 25 CpGs and 5 interactions. Together with 6 CpGs which were involved in the selected interactions but not in the selected main effects, the final model contained 31 CpG main effects and 5 interactions. No overlap was observed between these terms and the model terms of baseline telomere length.

The final model and interpretations

The Pearson correlation between telomere attrition rate and the weighted sum of the 36 terms was 0.87 and the R^2 was 0.72. As is shown in Figure 3B and Supplementary Table 3, in the final model 18 of the 31 CpG main effects had telomere attrition rate-related increase in DNA methylation and 13 had inverse associations, also consistent with their EWAS associations. 7 CpGs were promoter-associated and 1 was related to a reprogramming-specific DMR (RDMR). 4 out of the 5 interactions had negative coefficients and the other had a positive coefficient.

25 CpGs in the final model were annotated with gene names but no pathway was found significantly enriched. Pathway analysis of the 54 RNA transcripts correlated to the selected 31 CpGs resulted in 30 enriched pathways (Figure 4, right panel) which are mostly related to signaling pathways in immune response.

DISCUSSION

To our knowledge, the present study is the first to investigate the linkage of genome-wide DNA methylation profiles to newborn telomere length and early-life telomere attrition rate, incorporating

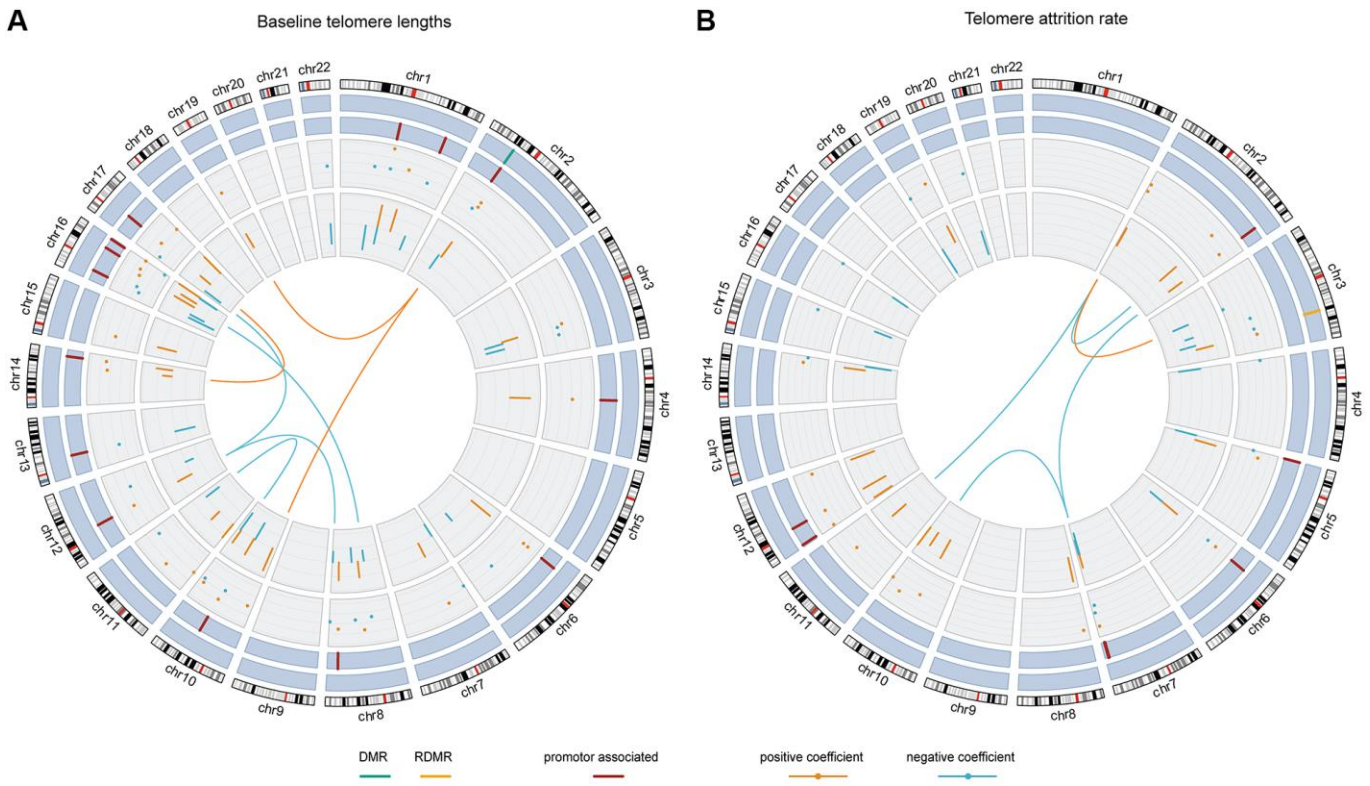


Figure 3. The chord diagrams depicting the contributing weights (final model coefficients) and EWAS associations of the selected CpGs and CpG-CpG interactions with annotations. Chromosomes 1 to 22 are displayed clockwise beginning from the top. Tracks from the outside to the inside represent: chromosome ideograms, whether a CpG is involved in known differentially methylated region (DMR, in green) or reprogramming-specific DMR (RDMR, in yellow), whether a CpG is associated with a promotor region (in dark red), the contributing weights in the final models (dots), the epigenome-wide association analysis (EWAS) associations (segments) and the interaction between CpGs, respectively. Panel (**A**) shows the results for baseline telomere length and panel (**B**) for telomere attrition rate. The EWAS coefficients (dots) were shown with dashed grids: in (**A**) the five grids from inside to outside represent -36, -18, 0, 18, 36; in (**B**) the five grids from inside to outside indicate -3, -2, -1, 0, 1, 2. Similarly, the model coefficients (segments) were also shown on dashed grids, with the inner to outside grids indicating value -0.02, -0.01, 0, 0.01, 0.02 in (**A**) and -0.003, -0.002, -0.001, 0, 0.001 and 0.002 in (**B**), respectively. Coefficients larger than 0 are shown in orange and those smaller than 0 are shown in blue.

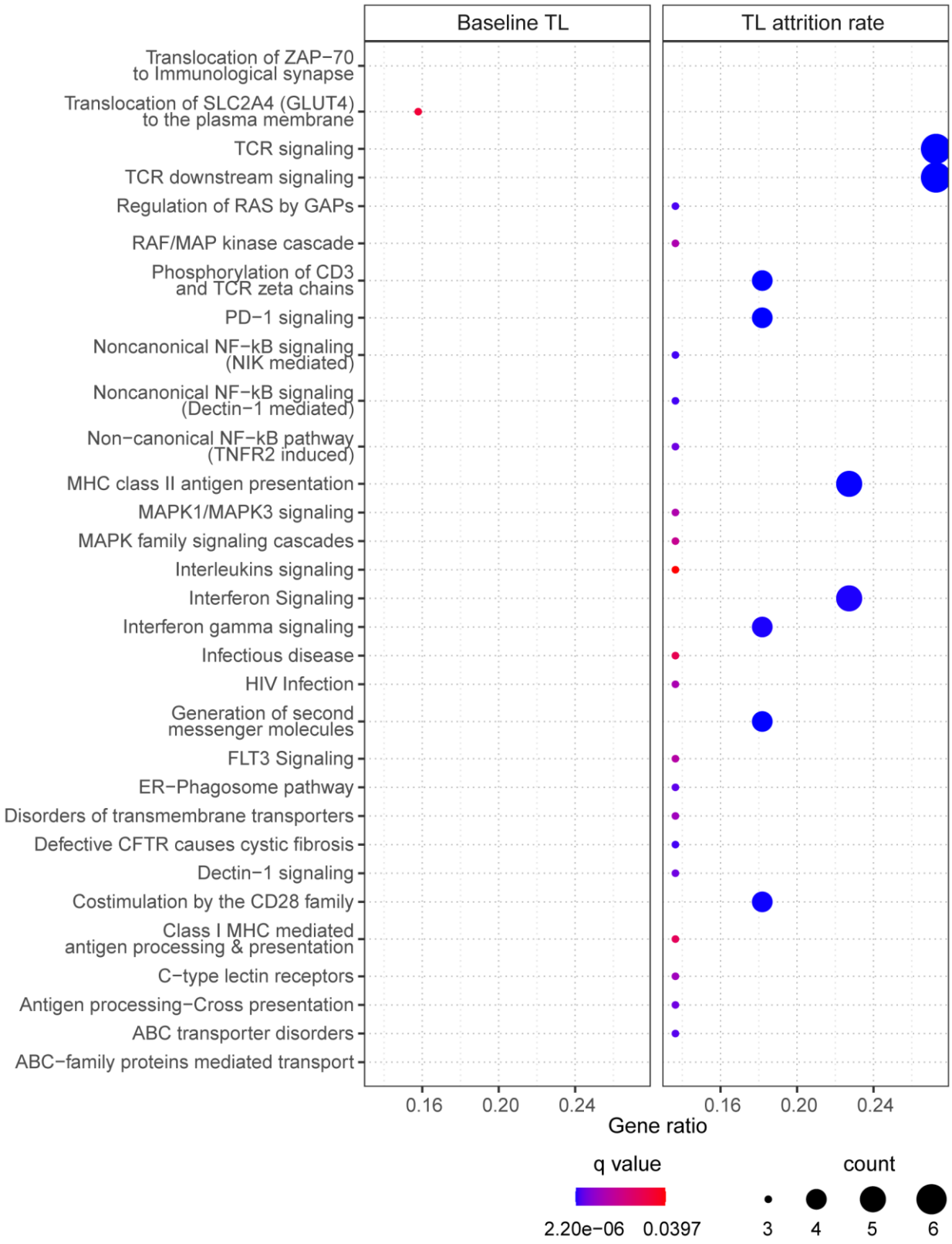


Figure 4. The enriched pathways from the selected CpGs. For baseline telomere length (TL) (left panel), pathways were identified based on CpGs in the final model. For telomere attrition rate (right panel), pathways were identified based on gene transcriptome correlated with the selected CpGs in the final model. Dot size indicates the number of genes (count) in the query list appearing in a pathway. Gene ratio of each pathway was calculated as the count divided by the total number of genes in the pathway. Dot color shows the over-representation analysis q-value.

interactions between epigenomic loci. Starting from whole-genome DNA methylation, a univariable CpG selection using EWAS in combination with a multivariable model construction using statistical learning was adopted to select relevant CpGs and CpG-CpG interactions. The selected DNAm-based explanatory models were distinct between baseline telomere length and telomere attrition rate, and a smaller number of CpGs were selected for telomere attrition rate than for baseline telomere length. Pathway analysis based on the identified CpGs suggested a role for GLUT4 translocation pathway in relation to baseline telomere length. Immune response-related cellular pathways were revealed in relation to telomere attrition rate in early childhood.

Two CpGs were identified strongly associated with baseline telomere length, one of which mapped to the promoter region of gene *PPM1F*, a gene encoding a protein phosphatase that negatively regulates stress response pathways [33]. Identifying this gene might explain the dependence of telomere length on cellular oxidative stress. However, this CpG was not selected in the final model, suggesting that the top CpGs in EWAS were not necessarily of the same importance in a multiple model. Indeed, only 8 of the 224 FDR significant CpGs from EWAS were in the final model of baseline telomere. CpGs in the final model suggested a pathway of translocating GLUT4 to plasma membrane through pathway enrichment analysis. In the meanwhile, two genes involved in this pathway, *TUBA1B* and *SLC2A4*, had their interacting CpGs selected in the final model. GLUT4 is a glucose transporter with a central role in glucose homeostasis and its translocation to plasma membrane is stimulated by insulin [34]. This might be in line with nutrient sensing as one of the hallmarks of aging [6] and supported by study findings that shorter telomere length was associated with type 2 diabetes mellitus [35] and insulin resistance [36]. Moreover, the pathway of GLUT4 translocation has been shown under the regulation of the same poly-ADP-ribosyl-transferase involved in telomere homeostasis [37]. Another CpG, cg13074173 (*RHPN1*), interacting with cg00973724 (*TUBA1B*), might also participate in this pathway since it regulates cytoskeleton activity [38] which is crucial for cellular translocation by vesicular transport.

The 35 CpGs selected in the final model of telomere attrition rate did not show significantly enriched pathways. Nevertheless, the top pathways (with q-value < 0.10) are collagen biosynthesis/formation, extracellular matrix (ECM) formation and N-lined glycosylation, all being related to ECM which plays an important role in tissue homeostasis [39] and is age-dependent [40]. 30 significantly enriched pathways were identified only after mapping to the cord blood transcriptome, with the majority pointing to T cell receptor signaling, NF-κB signaling and mitogen-activated protein (MAP) kinase signaling. Changes in immune system along with aging is termed as immunosenescence [41]. Active T cell signaling is accompanied by a high cell division rate and thus requires the support from telomere maintenance [42]. Transcription factor NF-κB, as a major driver of inflammatory signaling, is involved in several processes regulating telomere homeostasis, such as the modulation of telomerase activity and telomerase reverse transcriptase (TERT) transcription [43]. Additionally, telomere stability is attributed to the regulation by telomeric repeat binding factor 2 and TERT which are both under the control of MAP kinase pathways [44, 45].

Although a number of interaction terms were selected in both models, the number of interactions in the final models was much smaller than the number of main effects. It appears that the main effects play a more important role than their interactions in spite of the dependence between CpGs, which is in line with the idea of most published DNA methylation-based prediction models which only include CpG main effects.

The selected CpGs for the two outcomes showed an overlap only at the EWAS step (687 CpGs), and the final models did not have one predictor in common, possibly partly owing to the sample size and the conservative selection strategy applied. The discrepancy between baseline telomere length and telomere attrition rate in terms of CpG selection mirrors the discrepancy in epigenetic determinants of baseline telomere length and telomere dynamics that were set up at birth. Less CpGs and interactions were found to explain telomere attrition rate than baseline telomere length. This suggests that the initial setting of telomere length is more under epigenetic control compared to the change in telomere length thereafter. Indeed, there might be more external factors than just epigenetic and genetic factors at birth that can contribute to the change in telomere length after the *in utero* life, which may include lifestyle [46], psychological stress [47] and exposure to environmental pollutants [48–50]. Moreover, in the regard that the epigenome is also dynamic throughout life, identifying epigenetic predictors that contribute to telomere attrition might be warranted if DNA methylation is also profiled longitudinally.

Previously a 140-CpG DNAmTL estimator was published for adult telomere length [25] but did not overlap with the 47 CpGs selected for baseline telomere

length in the present study. In addition, the 140-CpG estimator only weakly predicted cord blood telomere length in our study (r=0.088, p-value=0.10). Therefore, certain patterns of DNA methylation that are important at birth and early life may not be as important in later life for the regulation of telomere length, with both telomere length and DNA methylation being dynamic processes over life [2, 29].

A strength of this study is that we based our model on the genome-wide DNA methylation data from EPIC assay which is reliable and reproducible covering a wide range of genome regions [51], and supports a comprehensive search for relevant epigenetic biomarkers. In parallel, we assessed the associations for two outcomes, telomere length at birth and telomere attrition rate, providing a comparison between initial telomere length setting and telomere dynamics early in life. Secondly, in spite of a sample size that may not be sufficiently large and that lasso, as one of the components of elastic net, is inconsistent for variable selection [31], we used a step-wise selection in combination with a repeated sampling to overcome low statistical power and undesired selection bias. In addition, the statistical inference was performed by bootstrapping instead of parametric models to avoid violations to model assumption.

On the other hand, limitations of our study need to be acknowledged. First, real-time PCR was used to determine telomere length in newborns, which shows a higher inter-assay variation compared to the traditionally used terminal restriction fragments (TRF) method or flow-FISH method [52, 53], although an inter-laboratory comparison achieved a coefficient of variation less than 7%. The imperfect correlation between baseline cord blood telomere length and leukocyte telomere length could also be attributed to variation in measurement. This issue often raises an RTM effect in a repeated measurement situation, which was accounted for by applying an RTM-correction [54]. Among different proposed methods exploring telomere attrition in longitudinal settings, the RTM-correction method we applied has been shown not to yield biased effect estimates [55]. Secondly, a pre-screening by EWAS resulted in that not all CpGs in the DNA methylation array could participate in subsequent variable selection. Due to the correction of the test-statistics based on empirical null distribution and correction for multiple testing, EWAS is highly data-driven and unable to detect subtle associations for a quantitative trait such as telomere length. However, EWAS was useful in selecting CpGs that could be candidates for subsequent multivariable model building, being less computationally intense than direct multivariable selection in analyzing high dimensional data. Thirdly, although the influence of a small sample size was coped with as discussed above, our variable selection tended to be conservative since the shrinkage in elastic net often leads to a small magnitude in model parameter estimates. Fourthly, DNA methylation at certain loci could involve genomic imprinting which conveys parental regulation of gene expression, which is an interesting topic for future analyses since allele-specific information of the epigenetic markers are not available in the Infinium MethylationEPIC assay. Finally, we acknowledge that the current findings are explorative, and that external cohorts are needed to further validate our findings and to enable constructing a DNA methylation-based prediction model for newborn telomere length and early-life telomere attrition.

In conclusion, as a first analysis on both telomere length at birth and telomere attrition rate during early life, searching for polygenic explanatory DNA methylation, the present study has added to the evidence that the two aging biomarkers are associated, and identified potential relevant biological pathways. Especially, baseline telomere length and telomere attrition in early-life were identified with epigenetic signatures, and these epigenetic settings at birth contributed more to baseline telomere length than to telomere attrition.

MATERIALS AND METHODS

Study population

The ongoing prospective ENVIR*ON*AGE birth cohort [30] was initiated in 2010. This cohort study has been approved by the ethical committee of Hasselt University and East-Limburg Hospital (Genk, Belgium) (EudraCT B37120107805), and is carried out according to the Declaration of Helsinki.

The mother-newborn pairs were recruited at delivery in the East-Limburg Hospital and were invited for a follow-up visit when the children were around 4 years old. Cord blood samples were collected at birth and blood samples of the children were collected at the follow-up visit. Genome-wide DNA methylation profiled in cord blood was available in 372 samples. Among the 423 children attending the follow-up examination between October 2014 and October 2019, 247 children had matching cord blood (baseline) telomere length, leukocyte telomere length (follow-up) and cord blood DNA methylation. These samples were used to evaluate the relationship between DNA methylation and baseline telomere length or longitudinal change in telomere length from birth to childhood. The selected 247 participants are

representative for the total ENVIR*ON*AGE population in terms of most characteristics as described in Table 1, but slightly differed in maternal education and newborn ethnicity (Supplementary Table 4).

Written informed consent was obtained from the participating mothers and questionnaires were filled out by the mothers to collect information on lifestyles and socioeconomic status. The information on newborns' sex, birth weight and gestational age, and maternal age and parity was collected from medical records in the hospital. The date of conception was estimated based on the first day of the mother's last menstrual period in combination with the first ultrasonographic examination. Maternal body mass index (BMI) was determined as the ratio of the maternal weight to the squared maternal height measured at the first antennal visit (weeks 7-9 of pregnancy). Ethnicity of a newborn was categorized as European if at least 2 grandparents were Europeans and classified as non-European otherwise. Educational level of the mothers was coded as 0 if they did not obtain any diploma, as 1 if they obtained at highest a high school diploma, as 2 if the highest diploma obtained was from a 3-year college and as 3 if obtained a 4-year college or university degree. Maternal smoking status was classified into "never smoker", "former smoker" (when having quit smoking before pregnancy), or "smoker" (in case of continuing smoking during pregnancy). Mothers were regarded as with pregnancy complications if they had any of gestational diabetes, hypertension, hyper- or hypothyroidism, infectious disease, preeclampsia, vaginal bleeding, phenylketonuria and allergy or asthma during pregnancy.

Blood sample collection and DNA extraction

Cord blood was collected immediately after delivery in BD Vacutainer® plastic whole blood tubes with spray-coated K2EDTA (BD, Franklin Lakes, NJ, USA). Samples were centrifuged at 3,200rpm for 15 minutes. Plasma was separated and the remainder with the buffy coats was stored at -80° C until further analysis. At the follow-up visit, blood was collected from the participating children by venipuncture using a winged steel needle in spray-coated K2EDTA tubes. After the sample process as described for cord blood, samples were stored at -80° C until analysis.

Cord blood and child blood DNA was extracted using the QIAamp DNA mini kit (Qiagen, Inc., Venlo, The Netherlands). DNA purity and concentration were verified by Nanodrop 1000 spectrophotometer (Isogen, Life Science, Belgium). DNA was considered pure when the A260/280 was greater than 1.80 and A260/230 greater than 2.0. DNA integrity was assessed with agarose gel electrophoresis.

Average relative telomere length measurement and data processing

Average relative telomere length was measured in triplicates using a previously described quantitative, real-time polymerase chain reaction (qPCR) protocol [14]. Detailed assay specifications are provided in the Supplementary Method. Telomere length were expressed as the ratio of telomere copy number to single-copy gene number (T/S) relative to the average T/S ratio of the entire sample set. Assay-reliability was assessed using an intra-class correlation coefficient (ICC) [56]. The inter-assay ICC was 0.936 (95%CI: 0.808 to 0.969) and the intra-assay ICC was 0.952 (95%CI: 0.947 to 0.956).

The cord blood telomere lengths were measured along with the leukocyte telomere lengths at follow-up in order to ensure that matching samples of individuals were assayed under the same condition: samples from the same individuals were on the same qPCR plate. Having been measured in a single batch, the telomere length data did not undergo batch effect removal. Telomere attrition from birth to follow-up was defined as the difference between telomere length at birth and at follow-up, the latter subtracted from the former. The difference between telomere lengths at birth and at follow-up was assessed using a paired t-test. Whether the children had differential telomere length change (homogeneity) was tested using the test-statistic suggested by Berry et al [54]. To account for the RTM effect which might be a result of the variation in measurement, we adopted a previously published method [54] to obtain the RTM-corrected telomere attrition:

$$\Delta TL = \rho\theta(X_1 - \mu_1) - (X_2 - \mu_2) \qquad (1)$$

where X_1 and X_2 are the notations for telomere length at birth and at follow-up, respectively, $\hat{\rho} = r$ is the Pearson correlation between the telomere lengths measured at the two time points, and $\hat{\theta} = \frac{S_2}{S_1}$ is the ratio of the standard deviation of telomere lengths at follow-up to the standard deviation of baseline telomere lengths. $\hat{\mu}_1 = \overline{X_1}$ and $\hat{\mu}_2 = \overline{X_2}$ denote the mean baseline telomere length and mean leukocyte telomere length at follow-up, respectively. Telomere attrition rate was calculated as telomere attrition divided by age (years) at follow-up.

DNA methylation measurement and data processing

Cord blood DNA samples were bisulphite-converted, amplified and hybridized to the Illumina

HumanMethylationEPIC Bead-Chip array (Illumina, San Diego, CA, USA), at the service lab GenomeScan (Leiden, The Netherlands). The array measurements were scanned using an Illumina iScan and the data quality was assessed using the R script MethylAid. DNA methylation data were preprocessed using the minfi package in R [57]. Briefly, probes with a call rate lower than 95% based on detection p-value > 10e-16 [58], samples with a call rate lower than 98% and samples with discordant sex, predicted using shinyMethyl [59], were removed. Methylation data was normalized using functional normalization [60]. For each CpG locus the methylation level was expressed as beta value calculated by M/(M+U) where M stands for the signal intensity from methylated probes and U stands for the signal intensity of unmethylated probes. The preprocessing resulted in a DNA methylation data with 857,898 CpGs. The missingness in the matrix of beta values was imputed by K-nearest neighbor (KNN) imputation and technical confounding effects (batch and position) were removed from the beta matrix using an empirical Bayes method [61]. Subsequently, we trimmed the data CpG-wisely for outliers which were lower than 3 inter-quarter-range below the 1st quartile, or higher than 3 inter-quarter-range above the 3rd quartile. CpG probes were filtered to exclude CpGs on X and Y chromosomes, those known to be common SNPs and those having cross-reactivity with multiple genomic locations [51]. 787,264 CpGs remained in the DNA methylation data were used in the EWAS of telomere lengths.

Based on the DNA methylation data, blood cell proportions (nucleated red blood cells, granulocytes, monocytes, natural killer cells, B cells, $CD4^+$ T cells, and $CD8^+$ T cells) in the cord blood samples were estimated using Bakulski algorithm [62].

Pre-screening by EWAS

The design of the present study is illustrated in Figure 1. Since variable selection from the whole EPIC array is computationally expensive, we performed EWAS as a pre-screening for candidate CpGs. For the *j*-th CpG (j=1, 2, …, 787264), a multiple linear regression model was fitted to test the association between baseline telomere length (Y_i, i=1, 2, …, 247) and the DNA methylation level at the CpG (X_{ij}), with robust covariance matrix estimated using the R package *sandwich* [63, 64]:

$$Y_i = \beta_0 + \beta_m X_{ij} + C_i^T \beta_c \qquad (2)$$

CpGs entered models in their original scale without extra centering or standardization. Models were adjusted for covariates selected a priori ($\boldsymbol{C_i}$): newborn gestational age, sex, ethnicity and birth weight, maternal pre-pregnancy BMI, age, parity, medical condition during pregnancy and education, paternal age, and cell type heterogeneity. Similarly, an EWAS was performed for telomere attrition rate by fitting the same linear model as equation (2) for each CpG adjusting for the same set of covariates. The test-statistics of β_m were corrected for an inflation factor and bias estimated based on their empirical null distribution [65], hence giving the corrected p-values. The Benjamini-Hochberg adjusted p-values were calculated based on the corrected p-values to identify significant EWAS probes (FDR=5%). In order to select a sufficiently large number of CpGs for the subsequent selection procedure, CpGs whose inflation- and bias-corrected p-value was smaller than a pre-determined threshold of 0.01 were the candidates that become the inputs of the next selection procedure.

Variable selection by statistical learning

The variable selection in this study was within the framework of elastic net where CpG selection and estimation of the contributing weight of each predictor take place simultaneously. We aimed to build explanatory models that express the baseline telomere length or telomere attrition rate as a weighted sum of the methylation level of a set of CpGs. The variable selection method used for both outcomes were identical, therefore we only describe the method in detail for baseline telomere length here.

The aim of this study was to select a number of CpGs (X_k) and CpG-CpG interactions (X_sX_t) that can express the *i*-th newborn's baseline (cord blood) telomere length (Y_i) as:

$$Y_i = w_0 + \sum_k w_k X_{ik} + \sum_{s,t} v_{s,t} X_{is} X_{it} \qquad (3)$$

where w_0 is the model intercept, being the mean of Y_i due to the standardization of CpGs. Parameters w_k and $v_{s,t}$ denote the contributing weights of the CpGs and interaction terms in the model. The model was trained with the matrix of candidate CpGs as the inputs and the vector of baseline telomere lengths as the output. In order to reduce the influence by the difference between CpG's magnitude and possible collinearity, the CpGs were standardized to have a mean of 0 and standard deviation of 1 before entering the model. When interactions are subject to selection, the pairwise interaction of CpGs were made by multiplying pairwisely the standardized CpGs.

The elastic net regression minimizes the objective function (simplified to include only main effects)

$$\frac{1}{2N}\sum_{i=1}^{N}(Y_i - \sum_k w_k X_{ik})^2 + \lambda\left[\frac{1}{2}(1-\alpha)\sum_k w_k{}^2 + \alpha\sum_k |w_k|\right] \quad (4)$$

where N is the number of observations. The second term is the penalty on coefficient estimates of variables (w_k). The regularization parameter λ was determined by 10-fold cross validation based on model deviance. In order to balance between model predictability and model overfitting, the λ value corresponding to a 1σ deviation ($\lambda_{1\sigma}$) from $\lambda=\lambda_{min}$ was chosen, which led to a slightly smaller number of predictors than with the λ_{min}. The tuning parameter α between lasso (α=1) and ridge (α=0) regression was determined by computation over a sequence of α value ranging from 0 to 1 with step of 0.05 and the α with the smallest deviance was selected. The models were fit in R with package *glmnet* [66], where coordinate descent algorithm was used and convergence threshold was set to 10^{-7}.

Adaptive lasso, a consistent penalized model, variable selection, was used to select candidate CpGs in a multiple linear model including all 22,817 CpG's selected from pre-screening by EWAS. The adaptive lasso performs lasso regression (α=1) by shrinking coefficients of less relevant variables to zero, while at the same time assigning unequal penalties to each variable [31]. The penalty term shown in equation (4) for adaptive lasso was $\lambda\sum_k a_k |w_k|$, where a_k was the penalty weight for the k-th CpG and was determined in a data-driven way as the inverse of the absolute value of the ridge regression coefficient of the corresponding CpG, as the ridge regression accounts for correlations among variables. By doing so, the CpGs given more shrinkage in ridge regression were assigned even higher penalty in the adaptive lasso variable selection. 208 CpGs were selected by adaptive lasso, which forms 21,528 pairwise interactions. Thus, a total number of 21,736 terms underwent the elastic net selection.

To encounter the issue that one of the components of elastic net, the lasso penalty, is not statistically consistent, we used 1000 repeatedly randomly sampled subsets from the whole data set for model selection. During each round, 85% of the samples (n=210) were taken and used to fit an elastic net, for which the parameters α, $\lambda_{1\sigma}$ and the number of predictors corresponding to the selected model were recorded. Among all the CpG and interactions, those has been selected in no less than 95% of the 1000 subsets were regarded as active terms and put in the final model. The weights w_k in equation (3) were estimated by fitting a ridge regression model on all the selected CpGs and interactions to the whole data set. Since statistical inference was not available within the ridge regression framework, the p-values of the ridge coefficients were simulated via bootstrap with 10,000 replicates.

Functional annotation of the selected CpGs

The selected CpGs were annotated with genomic regions (promoter region or DMRs) and UCSC reference gene names. Chord diagram visualization was performed using R package *circlize* [67].

To assist the translation of the identified CpGs into biological pathways, the DNA methylation level of the CpGs were correlated to the gene expression data available in the ENVIR*ON*AGE birth cohort. Global gene expression profiling by microarrays was performed in cord blood samples using a method described previously [68]. 44 samples which were available in both DNA methylation and global gene expression profiles were used in searching for RNA probes that are correlated with the selected CpGs. The correlation was assessed by computing the Pearson correlation coefficient within each CpG-RNA pair, which were defined based on gene name annotation or genomic location. Among the 47 CpGs explaining baseline telomere length, 36 CpGs found RNA probes with the same gene name annotated. Additionally, each of the 47 CpGs was correlated to the RNA transcripts on the same chromosome whose starting sites were within 1Mb from the CpG locus. In total, 1,311 correlations were calculated and 75 correlations had p-value<0.05. The gene names corresponding to the 75 correlated RNA transcripts were used for a pathway analysis. Similarly, there were 676 RNA transcripts correlated to the 31 CpGs explaining telomere attrition rate: 22 were found based on gene name and 654 were found within genomic neighborhood. 53 correlations with p-value < 0.05 were used for a pathway analysis.

The gene names of selected CpGs and the correlated transcripts, respectively, were mapped to the Reactome pathway database [69] using the R package *ReactomePA* [70] for an over-representation analysis (ORA). A pathway was considered significantly enriched if its q-value was smaller than FDR of 5% and at least 3 genes were involved in the pathway.

Ethics approval and consent to participate

The study protocol was approved by the Ethical Committee of Hasselt University and East-Limburg Hospital in Genk (Belgium) and has been carried out according to the declaration of Helsinki. Written informed consent was obtained from all participating mothers.

Data availability

Data are available on request.

AUTHOR CONTRIBUTIONS

TSN coordinates the ENVIR*ON*AGE birth cohort. TSN, DSM and CW designed the current study. SL and DT coordinated the DNA methylation array. HS and MP performed sample collection and sample processing. DSM, DT and CVDS performed DNA extractions and telomere length measurements. RA performed DNA methylation data pre-preprocessing. CW conducted all the statistical analyses. CW, DSM and TSN prepared the first draft of the manuscript. All authors contributed in critical revision of the manuscript and approved the final manuscript.

ACKNOWLEDGMENTS

We acknowledge the cohort participants, midwives, the staff of the maternity ward and the clinical laboratory of East-Limburg Hospital in Genk.

CONFLICTS OF INTEREST

The authors declare that they have no conflicts of interest.

FUNDING

The ENVIR*ON*AGE birth cohort is supported by grants from the European Research Council (Grant No. ERC-2012-StG310898), the Flemish Scientific Fund (FWO, Grant No. G073315N) and Kom Op Tegen Kanker DSM (FWO grant 12X9620N) is postdoctoral fellows of the Flanders Research Foundation. SL is the beneficiary of a post-doctoral fellowship (12L5216N) provided by The Research Foundation-Flanders (FWO) and the Flemish Institute of Technological Research (VITO).

REFERENCES

1. Shammas MA. Telomeres, lifestyle, cancer, and aging. Curr Opin Clin Nutr Metab Care. 2011; 14:28–34. https://doi.org/10.1097/MCO.0b013e32834121b1 PMID:21102320

2. Frenck RW Jr, Blackburn EH, Shannon KM. The rate of telomere sequence loss in human leukocytes varies with age. Proc Natl Acad Sci USA. 1998; 95:5607–10. https://doi.org/10.1073/pnas.95.10.5607 PMID:9576930

3. Wang Q, Zhan Y, Pedersen NL, Fang F, Hägg S. Telomere Length and All-Cause Mortality: A Meta-analysis. Ageing Res Rev. 2018; 48:11–20. https://doi.org/10.1016/j.arr.2018.09.002 PMID:30254001

4. Needham BL, Rehkopf D, Adler N, Gregorich S, Lin J, Blackburn EH, Epel ES. Leukocyte telomere length and mortality in the National Health and Nutrition Examination Survey, 1999-2002. Epidemiology. 2015; 26:528–35. https://doi.org/10.1097/EDE.0000000000000299 PMID:26039272

5. Blackburn EH, Epel ES, Lin J. Human telomere biology: A contributory and interactive factor in aging, disease risks, and protection. Science. 2015; 350:1193–98. https://doi.org/10.1126/science.aab3389 PMID:26785477

6. López-Otín C, Blasco MA, Partridge L, Serrano M, Kroemer G. The hallmarks of aging. Cell. 2013; 153:1194–217. https://doi.org/10.1016/j.cell.2013.05.039 PMID:23746838

7. Cawthon RM. Association between telomere length in blood and mortality in people aged 60 years or older. Lancet. 2003; 361:393–95. https://doi.org/10.1016/S0140-6736(03)12384-7

8. Dugdale HL, Richardson DS. Heritability of telomere variation: it is all about the environment!. Philos Trans R Soc Lond B Biol Sci. 2018; 373:20160450. https://doi.org/10.1098/rstb.2016.0450 PMID:29335377

9. Reichert S, Stier A. Does oxidative stress shorten telomeres *in vivo*? A review. Biol Lett. 2017; 13:20170463. https://doi.org/10.1098/rsbl.2017.0463 PMID:29212750

10. Martens DS, Janssen BG, Bijnens EM, Clemente DB, Vineis P, Plusquin M, Nawrot TS. Association of Parental Socioeconomic Status and Newborn Telomere Length. JAMA Netw Open. 2020; 3:e204057. https://doi.org/10.1001/jamanetworkopen.2020.4057 PMID:32364595

11. Martens DS, Nawrot TS. Ageing at the level of telomeres in association to residential landscape and air pollution at home and work: a review of the current evidence. Toxicol Lett. 2018; 298:42–52. https://doi.org/10.1016/j.toxlet.2018.06.1213 PMID:29944903

12. Martens DS, Cox B, Janssen BG, Clemente DB, Gasparrini A, Vanpoucke C, Lefebvre W, Roels HA, Plusquin M, Nawrot TS. Prenatal Air Pollution and Newborns' Predisposition to Accelerated Biological Aging. JAMA Pediatr. 2017; 171:1160–67. https://doi.org/10.1001/jamapediatrics.2017.3024 PMID:29049509

13. Martens DS, Plusquin M, Cox B, Nawrot TS. Early Biological Aging and Fetal Exposure to High and Low Ambient Temperature: A Birth Cohort Study. Environ Health Perspect. 2019; 127:117001. https://doi.org/10.1289/EHP5153 PMID:31691586

14. Martens DS, Plusquin M, Gyselaers W, De Vivo I, Nawrot TS. Maternal pre-pregnancy body mass index and newborn telomere length. BMC Med. 2016; 14:148. https://doi.org/10.1186/s12916-016-0689-0 PMID:27751173

15. Schübeler D. Function and information content of DNA methylation. Nature. 2015; 517:321–26. https://doi.org/10.1038/nature14192 PMID:25592537

16. Ciccarone F, Tagliatesta S, Caiafa P, Zampieri M. DNA methylation dynamics in aging: how far are we from understanding the mechanisms? Mech Ageing Dev. 2018; 174:3–17. https://doi.org/10.1016/j.mad.2017.12.002 PMID:29268958

17. Johnson AA, Akman K, Calimport SR, Wuttke D, Stolzing A, de Magalhães JP. The role of DNA methylation in aging, rejuvenation, and age-related disease. Rejuvenation Res. 2012; 15:483–94. https://doi.org/10.1089/rej.2012.1324 PMID:23098078

18. 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

19. Levine ME, Lu AT, Quach A, Chen BH, Assimes TL, Bandinelli S, Hou L, Baccarelli AA, Stewart JD, Li Y, Whitsel EA, Wilson JG, Reiner AP, et al. An epigenetic biomarker of aging for lifespan and healthspan. Aging (Albany NY). 2018; 10:573–91. https://doi.org/10.18632/aging.101414 PMID:29676998

20. Hannum G, Guinney J, Zhao L, Zhang L, Hughes G, Sadda 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

21. Knight AK, Craig JM, Theda C, Bækvad-Hansen M, Bybjerg-Grauholm J, Hansen CS, Hollegaard MV, Hougaard DM, Mortensen PB, Weinsheimer SM, Werge TM, Brennan PA, Cubells JF, et al. An epigenetic clock for gestational age at birth based on blood methylation data. Genome Biol. 2016; 17:206. https://doi.org/10.1186/s13059-016-1068-z PMID:27717399

22. Bohlin J, Håberg SE, Magnus P, Reese SE, Gjessing HK, Magnus MC, Parr CL, Page CM, London SJ, Nystad W. Prediction of gestational age based on genome-wide differentially methylated regions. Genome Biol. 2016; 17:207. https://doi.org/10.1186/s13059-016-1063-4 PMID:27717397

23. Horvath S, Raj K. DNA methylation-based biomarkers and the epigenetic clock theory of ageing. Nat Rev Genet. 2018; 19:371–84. https://doi.org/10.1038/s41576-018-0004-3 PMID:29643443

24. Buxton JL, Suderman M, Pappas JJ, Borghol N, McArdle W, Blakemore AI, Hertzman C, Power C, Szyf M, Pembrey M. Human leukocyte telomere length is associated with DNA methylation levels in multiple subtelomeric and imprinted loci. Sci Rep. 2014; 4:4954. https://doi.org/10.1038/srep04954 PMID:24828261

25. Lu AT, Seeboth A, Tsai PC, Sun D, Quach A, Reiner AP, Kooperberg C, Ferrucci L, Hou L, Baccarelli AA, Li Y, Harris SE, Corley J, et al. DNA methylation-based estimator of telomere length. Aging (Albany NY). 2019; 11:5895–923. https://doi.org/10.18632/aging.102173 PMID:31422385

26. Bijnens EM, Zeegers MP, Derom C, Martens DS, Gielen M, Hageman GJ, Plusquin M, Thiery E, Vlietinck R, Nawrot TS. Telomere tracking from birth to adulthood and residential traffic exposure. BMC Med. 2017; 15:205. https://doi.org/10.1186/s12916-017-0964-8 PMID:29157235

27. Martens DS, Van Der Stukken C, Derom C, Thiery E, Bijnens EM, Nawrot TS. Newborn telomere length predicts later life telomere length: Tracking telomere length from birth to child- and adulthood. EBioMedicine. 2021; 63:103164. https://doi.org/10.1016/j.ebiom.2020.103164 PMID:33422989

28. Factor-Litvak P, Susser E, Kezios K, McKeague I, Kark JD, Hoffman M, Kimura M, Wapner R, Aviv A. Leukocyte Telomere Length in Newborns: Implications for the Role of Telomeres in Human Disease. Pediatrics. 2016; 137:e20153927. https://doi.org/10.1542/peds.2015-3927 PMID:26969272

29. Jones MJ, Goodman SJ, Kobor MS. DNA methylation and healthy human aging. Aging Cell. 2015; 14:924–32. https://doi.org/10.1111/acel.12349 PMID:25913071

30. Janssen BG, Madhloum N, Gyselaers W, Bijnens E, Clemente DB, Cox B, Hogervorst J, Luyten L, Martens

DS, Peusens M, Plusquin M, Provost EB, Roels HA, et al. Cohort Profile: The ENVIRonmental influence ON early AGEing (ENVIRONAGE): a birth cohort study. Int J Epidemiol. 2017; 46:1386–87m. https://doi.org/10.1093/ije/dyw269 PMID:28089960

31. Zou H. The Adaptive Lasso and Its Oracle Properties. J Am Stat Assoc, 2006. 101:1418–29. https://doi.org/10.1198/016214506000000735

32. Zou H, Hastie T. Regularization and Variable Selection via the Elastic Net. J Royal Stat Soc. Series B (Statistical Methodology). 2005; 67:301–20. https://doi.org/10.1111/j.1467-9868.2005.00503.x

33. Wingo AP, Velasco ER, Florido A, Lori A, Choi DC, Jovanovic T, Ressler KJ, Andero R. Expression of the PPM1F Gene Is Regulated by Stress and Associated With Anxiety and Depression. Biol Psychiatry. 2018; 83:284–95. https://doi.org/10.1016/j.biopsych.2017.08.013 PMID:29054677

34. Stöckli J, Fazakerley DJ, James DE. GLUT4 exocytosis. J Cell Sci. 2011; 124:4147–59. https://doi.org/10.1242/jcs.097063 PMID:22247191

35. Zhao J, Miao K, Wang H, Ding H, Wang DW. Association between telomere length and type 2 diabetes mellitus: a meta-analysis. PLoS One. 2013; 8:e79993. https://doi.org/10.1371/journal.pone.0079993 PMID:24278229

36. Verhulst S, Dalgård C, Labat C, Kark JD, Kimura M, Christensen K, Toupance S, Aviv A, Kyvik KO, Benetos A. A short leucocyte telomere length is associated with development of insulin resistance. Diabetologia. 2016; 59:1258–65. https://doi.org/10.1007/s00125-016-3915-6 PMID:27020448

37. Haikarainen T, Krauss S, Lehtio L. Tankyrases: structure, function and therapeutic implications in cancer. Curr Pharm Des. 2014; 20:6472–88. https://doi.org/10.2174/1381612820666140630101525 PMID:24975604

38. Lal MA, Andersson AC, Katayama K, Xiao Z, Nukui M, Hultenby K, Wernerson A, Tryggvason K. Rhophilin-1 is a key regulator of the podocyte cytoskeleton and is essential for glomerular filtration. J Am Soc Nephrol. 2015; 26:647–62. https://doi.org/10.1681/ASN.2013111195 PMID:25071083

39. Bonnans C, Chou J, Werb Z. Remodelling the extracellular matrix in development and disease. Nat Rev Mol Cell Biol. 2014; 15:786–801. https://doi.org/10.1038/nrm3904 PMID:25415508

40. McCabe MC, Hill RC, Calderone K, Cui Y, Yan Y, Quan T, Fisher GJ, Hansen KC. Alterations in extracellular matrix composition during aging and photoaging of the skin. Matrix Biol Plus. 2020; 8:100041. https://doi.org/10.1016/j.mbplus.2020.100041 PMID:33543036

41. Lewis TJ, Trempe CL. Chapter 7 - Inflammation Friend or Foe?, in The End of Alzheimer's (Second Edition), T.J. Lewis and C.L. Trempe, Editors. 2017, Academic Press. 192–241. https://doi.org/10.1016/B978-0-12-812112-2.00007-0

42. Moro-García MA, Alonso-Arias R, López-Larrea C. Molecular mechanisms involved in the aging of the T-cell immune response. Curr Genomics. 2012; 13:589–602. https://doi.org/10.2174/138920212803759749 PMID:23730199

43. Jose SS, Bendickova K, Kepak T, Krenova Z, Fric J. Chronic Inflammation in Immune Aging: Role of Pattern Recognition Receptor Crosstalk with the Telomere Complex? Front Immunol. 2017; 8:1078. https://doi.org/10.3389/fimmu.2017.01078 PMID:28928745

44. Xu D, Li H, Liu JP. Inhibition of telomerase by targeting MAP kinase signaling. Methods Mol Biol. 2007; 405:147–65. https://doi.org/10.1007/978-1-60327-070-0_12 PMID:18369823

45. Picco V, Coste I, Giraud-Panis MJ, Renno T, Gilson E, Pagès G. ERK1/2/MAPK pathway-dependent regulation of the telomeric factor TRF2. Oncotarget. 2016; 7:46615–27. https://doi.org/10.18632/oncotarget.10316 PMID:27366950

46. Puterman E, Lin J, Krauss J, Blackburn EH, Epel ES. Determinants of telomere attrition over 1 year in healthy older women: stress and health behaviors matter. Mol Psychiatry. 2015; 20:529–35. https://doi.org/10.1038/mp.2014.70 PMID:25070535

47. Kiecolt-Glaser JK, Glaser R. Psychological stress, telomeres, and telomerase. Brain Behav Immun. 2010; 24:529–30. https://doi.org/10.1016/j.bbi.2010.02.002 PMID:20167271

48. Lee EY, Lin J, Noth EM, Hammond SK, Nadeau KC, Eisen EA, Balmes JR. Traffic-Related Air Pollution and Telomere Length in Children and Adolescents Living in Fresno, CA: A Pilot Study. J Occup Environ Med. 2017; 59:446–52. https://doi.org/10.1097/JOM.0000000000000996 PMID:28486341

49. Zhao B, Vo HQ, Johnston FH, Negishi K. Air pollution and telomere length: a systematic review of 12,058 subjects. Cardiovasc Diagn Ther. 2018; 8:480–92. https://doi.org/10.21037/cdt.2018.06.05 PMID:30214863

50. Clemente DB, Vrijheid M, Martens DS, Bustamante M, Chatzi L, Danileviciute A, de Castro M, Grazuleviciene R, Gutzkow KB, Lepeule J, Maitre L, McEachan RR, Robinson O, et al. Prenatal and Childhood Traffic-Related Air Pollution Exposure and Telomere Length in European Children: The HELIX Project. Environ Health Perspect. 2019; 127:87001. https://doi.org/10.1289/EHP4148 PMID:31393792

51. Pidsley R, Zotenko E, Peters TJ, Lawrence MG, Risbridger GP, Molloy P, Van Djik S, Muhlhausler B, Stirzaker C, Clark SJ. Critical evaluation of the Illumina MethylationEPIC BeadChip microarray for whole-genome DNA methylation profiling. Genome Biol. 2016; 17:208. https://doi.org/10.1186/s13059-016-1066-1 PMID:27717381

52. Aviv A, Hunt SC, Lin J, Cao X, Kimura M, Blackburn E. Impartial comparative analysis of measurement of leukocyte telomere length/DNA content by Southern blots and qPCR. Nucleic Acids Res. 2011; 39:e134. https://doi.org/10.1093/nar/gkr634 PMID:21824912

53. Kimura M, Stone RC, Hunt SC, Skurnick J, Lu X, Cao X, Harley CB, Aviv A. Measurement of telomere length by the Southern blot analysis of terminal restriction fragment lengths. Nat Protoc. 2010; 5:1596–607. https://doi.org/10.1038/nprot.2010.124 PMID:21085125

54. Berry DA, Eaton ML, Ekholm BP, Fox TL. Assessing differential drug effect. Biometrics. 1984; 40:1109–15. PMID:6398711

55. Bateson M, Eisenberg DT, Nettle D. Controlling for baseline telomere length biases estimates of the rate of telomere attrition. R Soc Open Sci. 2019; 6:190937. https://doi.org/10.1098/rsos.190937 PMID:31824705

56. Telomere research network: Study Design and Analysis. 2020. https://trn.tulane.edu/resources/study-design-analysis/

57. Aryee MJ, Jaffe AE, Corrada-Bravo H, Ladd-Acosta C, Feinberg AP, Hansen KD, Irizarry RA. Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. Bioinformatics. 2014; 30:1363–69. https://doi.org/10.1093/bioinformatics/btu049 PMID:24478339

58. Lehne B, Drong AW, Loh M, Zhang W, Scott WR, Tan ST, Afzal U, Scott J, Jarvelin MR, Elliott P, McCarthy MI, Kooner JS, Chambers JC. A coherent approach for analysis of the Illumina HumanMethylation450 BeadChip improves data quality and performance in epigenome-wide association studies. Genome Biol. 2015; 16:37. https://doi.org/10.1186/s13059-015-0600-x PMID:25853392

59. Fortin JP, Fertig E, Hansen K. shinyMethyl: interactive quality control of Illumina 450k DNA methylation arrays in R. F1000Res. 2014; 3:175. https://doi.org/10.12688/f1000research.4680.2 PMID:25285208

60. Fortin JP, Labbe A, Lemire M, Zanke BW, Hudson TJ, Fertig EJ, Greenwood CM, Hansen KD. Functional normalization of 450k methylation array data improves replication in large cancer studies. Genome Biol. 2014; 15:503. https://doi.org/10.1186/s13059-014-0503-2 PMID:25599564

61. Johnson WE, Li C, Rabinovic A. Adjusting batch effects in microarray expression data using empirical Bayes methods. Biostatistics. 2007; 8:118–27. https://doi.org/10.1093/biostatistics/kxj037 PMID:16632515

62. Bakulski KM, Feinberg JI, Andrews SV, Yang J, Brown S, McKenney SL, Witter F, Walston J, Feinberg AP, Fallin MD. DNA methylation of cord blood cell types: Applications for mixed cell birth studies. Epigenetics. 2016; 11:354–62. https://doi.org/10.1080/15592294.2016.1161875 PMID:27019159

63. Zeileis A. Econometric Computing with HC and HAC Covariance Matrix Estimators. J Stat Softw. 2004; 11:17. https://doi.org/10.18637/jss.v011.i10

64. Zeileis A. Object-oriented Computation of Sandwich Estimators. J Stat Softw. 2006; 16:16. https://doi.org/10.18637/jss.v016.i09

65. van Iterson M, van Zwet EW, Heijmans BT, and BIOS Consortium. Controlling bias and inflation in epigenome- and transcriptome-wide association studies using the empirical null distribution. Genome Biol. 2017; 18:19. https://doi.org/10.1186/s13059-016-1131-9 PMID:28129774

66. Friedman J, Hastie T, Tibshirani R. Regularization Paths for Generalized Linear Models via Coordinate Descent. J Stat Softw. 2010; 33:1–22. PMID:20808728

67. Gu Z, Gu L, Eils R, Schlesner M, Brors B. circlize

Implements and enhances circular visualization in R. Bioinformatics. 2014; 30:2811–12. https://doi.org/10.1093/bioinformatics/btu393 PMID:24930139

68. Tsamou M, Vrijens K, Wang C, Winckelmans E, Neven KY, Madhloum N, de Kok TM, Nawrot TS. Genome-wide microRNA expression analysis in human placenta reveals sex-specific patterns: an ENVIRONAGE birth cohort study. Epigenetics. 2021; 16:373–88. https://doi.org/10.1080/15592294.2020.1803467 PMID:32892695

69. Jassal B, Matthews L, Viteri G, Gong C, Lorente P, Fabregat A, Sidiropoulos K, Cook J, Gillespie M, Haw R, Loney F, May B, Milacic M, et al. The reactome pathway knowledgebase. Nucleic Acids Res. 2020; 48:D498–503. https://doi.org/10.1093/nar/gkz1031 PMID:31691815

70. Yu G, He QY. ReactomePA: an R/Bioconductor package for reactome pathway analysis and visualization. Mol Biosyst. 2016; 12:477–79. https://doi.org/10.1039/c5mb00663e PMID:26661513

SUPPLEMENTARY MATERIALS

Supplementary Method

Telomere length measurement by qPCR assay

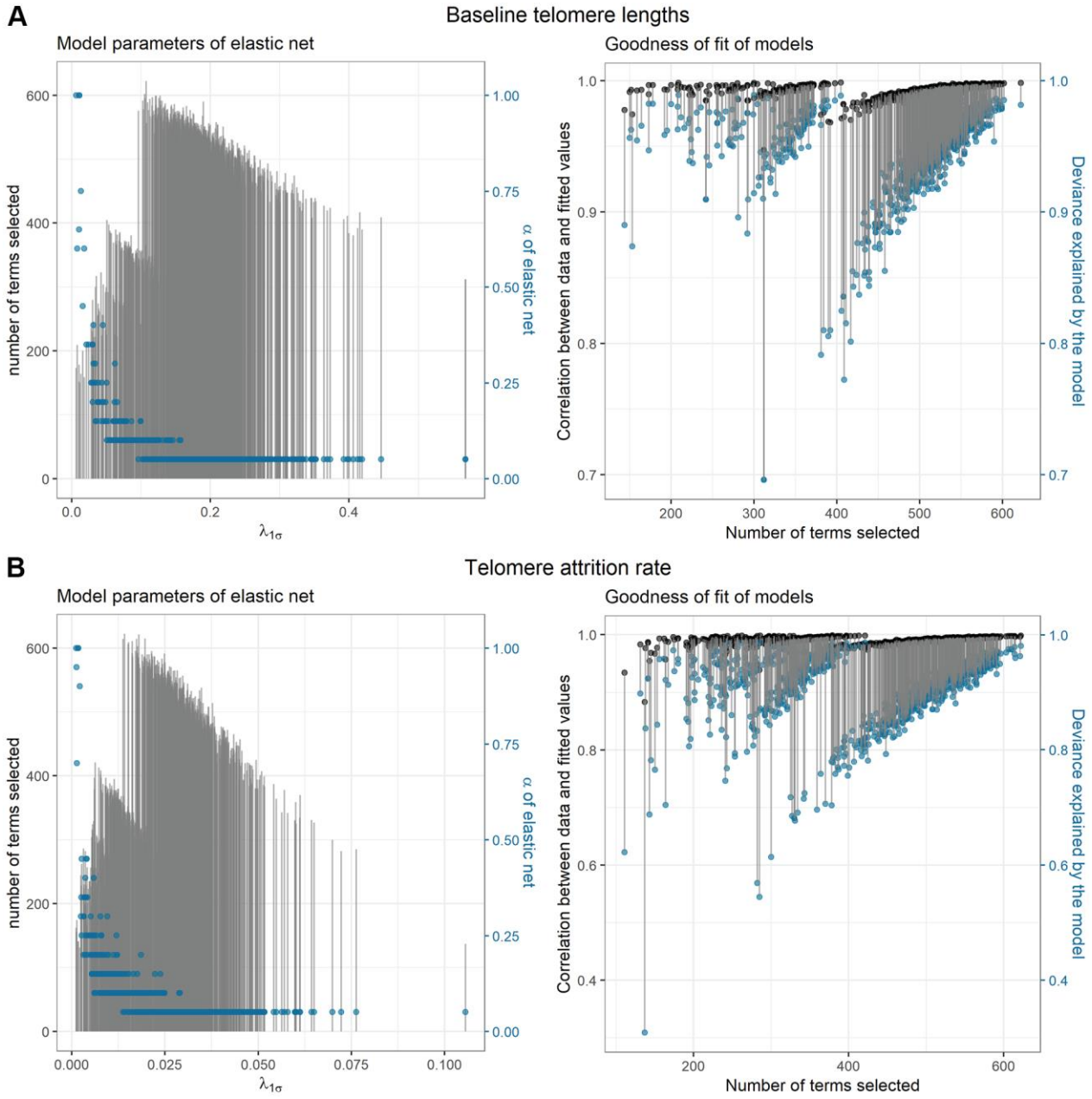
DNA samples were normalized to ensure a uniform DNA input of 5 ng for each qPCR, and this was checked using the Quant-iT™ PicoGreen® dsDNA Assay Kit (Life Technologies, Europe). Cord blood and child blood samples were matched per individual and arranged on the same qPCR plate and were measured in 1 batch. All samples were measured in triplicates on a 7900HT Fast RealTime PCR System (Applied Biosystems) in a 384-well format. The telomere-specific qPCR reaction mixture contained 1x QuantiTect SYBR Green PCR master mix (Qiagen, Inc., Venlo, the Netherlands), 2 mM dithiothreitol (DTT), 300 nM telg primer (ACACTAAGGTTTGGGT TTGGGTTTGGGTTTGGGTTAGTG T) and 900 nM telc primer (TGTTAGGTATCCCTATCCCTATCCCT ATCCCTATCCCTAACA). Used cycling conditions were: 1 cycle at 95° C for 10 min, 2 cycles at 94° C for 15 sec and 49° C for 2 min, and 30 cycles at 94° C for 15 sec, 62° C for 20 sec, and 74° C for 1 min and 40 sec. The single-copy gene (human β globin) qPCR mixture contained 1x QuantiTect SYBR Green PCR master mix, 400 nM HBG1 primer (GCTTCTGACACA ACTGTGTTCACTAGC) and 400 nM HBG2 primer (CACCAACTTCATCCACGTTCACC). Used cycling conditions were: 1 cycle at 95° C for 10 min, 40 cycles at 95° C for 15 sec, and 58° C for 1 min and 20 sec. After each qPCR a melting curve analysis was performed. In each run, a 6-point serial dilution of pooled DNA was run to assess PCR efficiency. PCR-efficiencies ranged from 100-105% for telomere runs and 95-100% for single-copy gene runs. Furthermore 10 inter-run calibrators (IRCs) were run to account for inter-run variability over 16 qPCR plates. qPCR curves for each sample were visually inspected and when technical problems were detected or triplicates showed too high variability, samples were excluded for further analysis. Telomeres were normalized using qBase (Biogazelle, Zwijnaarde, Belgium). The reliability of our assay was assessed by calculating the intraclass coefficient (ICC) with 95% CI of triplicate measures (T/S ratios) [1]. Both the inter-assay (based on 10 IRCs over 16 qPCR plates) and intra-assay ICC (based on all measures) was calculated using the available on-line R script on the Telomere Research Network website [2].

REFERENCES

1. Telomere research network: Study Design and Analysis. 2020. Available from: https://trn.tulane.edu/resources/study-design-analysis/

2. Eisenberg D, Nettle D, Verhulst S. How to calculate the repeatability (ICC) of telomere length measures. 2020. Available from: https://trn.tulane.edu/wp-content/uploads/sites/445/2020/10/How-to-calculate-repeatability.pdf

Supplementary Figure



Supplementary Figure 1. Parameters of elastic net model selection and model goodness-of-fit. Sub-figures in left show that the 1000 randomly sampled subsets are distinct subsets with various regularization parameters: the length of grey segments represents the number of terms selected, and the blue dots show the optimal α values picked for each subset (ranging from 0.05 to 1). Sub-figures in the right column show the fitting performance of each selected model in the corresponding subsets with the black dots depicting the Pearson correlation between telomere length or telomere attrition and the estimation made based on each selected model. The deviance explained by model (equivalent to model R^2) is shown with blue dots. For baseline telomere lengths (**A**), the median of the parameter α was 0.05. $\lambda_{1\sigma}$ ranged from 0.0067 to 0.57 with a median of 0.18. The smallest number of selected terms was 144 (including 51 main effects and 93 interactions) while the largest was 622 (including 154 main effects and 468 interactions). Models including larger numbers of predictors showed better fitting performance. For telomere attrition rate (**B**), the parameter α had a median of 0.05 and $\lambda_{1\sigma}$ ranged from 0.0012 to 0.11 with a median of 0.025. The smallest number of selected terms was 111 (including 47 main effects and 64 interactions) while the largest was 622 (including 173 main effects and 449 interactions).

Supplementary Tables

Supplementary Table 1. The summary of the pairwise Pearson correlations of the CpGs which passed pre-screening by an epigenome-wide association analysis.

	min	5th percentile	Median	95th percentile	max
Cord blood telomere length [1]	-0.95	-0.19	0.05	0.78	0.99
Telomere attrition rate [2]	-0.96	-0.28	0.03	0.49	0.98

[1]260,296,336 pairwise Pearson correlations of 22,817 CpGs.
[2]48,624,591 pairwise Pearson correlations of 9,862 CpGs.

Supplementary Table 2. The 47 CpGs and 7 interactions in the model of baseline telomere length.

Name of CpG or interaction	Coefficient (bootstrap p-value)	EWAS Coefficient (p-value)	Chrom-osome	Position on chromosome	Relation to island	Gene name	DMR	Promoter associated
(intercept)	1.1605 (0.00e+00)							
cg00036011	0.0169 (4.00e-04)	20.68 (9.17e-04)	chr6	26271718	N_Shore	HIST1H3G; HIST1H2BI		
cg00064950	-0.0135 (1.10e-03)	-8.51 (4.32e-03)	chr17	7184216	Island	SLC2A4		
cg00487989	0.0177 (7.00e-04)	7.64 (7.07e-04)	chr16	89688244	OpenSea	DPEP1		
cg00871560	0.0112 (1.12e-02)	1.47 (2.32e-04)	chr3	132756721	N_Shore	TMEM108		
cg00973724	0.0109 (1.18e-02)	7.43 (2.76e-03)	chr12	49524797	Island	TUBA1B		yes
cg00973724:cg00064950	-0.0086 (2.66e-02)					*TUBA1B * SLC2A4*		
cg01567315	0.0156 (1.10e-03)	4.52 (2.60e-04)	chr1	155650264	OpenSea	YY1AP1		
cg02576753	0.0132 (4.40e-03)	1.56 (2.56e-06)	chr15	27772640	OpenSea	GABRG3		
cg02802645	-0.0139 (5.60e-03)	-6.17 (3.64e-03)	chr22	38904757	S_Shore			
cg03891344	-0.0193 (4.00e-04)	-9.98 (1.25e-04)	chr16	19533212	Island	GDE1		yes
cg04372430	-0.0074 (7.54e-02)	-9.06 (6.31e-03)	chr8	48591406	OpenSea	SPIDR		
cg04372430:cg16435922	-0.0171 (2.00e-04)							
cg06105699	0.0162 (4.00e-04)	1.15 (2.53e-03)	chr17	79971436	OpenSea	ASPSCR1		
cg06529600	-0.0066 (1.18e-01)	-2.15 (2.60e-03)	chr6	112652308	OpenSea			
cg06826039	-0.0099 (2.82e-02)	-6.65 (3.07e-04)	chr1	204379650	N_Shore	PPP1R15B		yes
cg07131116	0.0113 (2.57e-02)	9.4 (2.34e-06)	chr14	94546722	N_Shore	DDX24; IFI27L1		yes
cg07417551	0.0108 (1.75e-02)	3.34 (3.30e-03)	chr19	14586295	S_Shore	PTGER1		
cg08878914	-0.0099 (2.53e-02)	-5.9 (4.85e-03)	chr8	86435445	OpenSea			
cg09793256	-0.0112 (9.80e-03)	-7.08 (4.66e-04)	chr2	55646604	N_Shore	CCDC88A	DMR	yes
cg09880724	-0.0142 (1.30e-03)	-34.01 (3.75e-03)	chr16	87425644	Island	FBXO31; MAP1LC3B		yes
cg11300971	0.0143 (4.10e-03)	1.88 (1.33e-04)	chr4	89619038	OpenSea	NAP1L5; HERC3		yes
cg11455582	-0.0129 (3.90e-03)	-8.57 (7.86e-03)	chr3	155272990	OpenSea	PLCH1		
cg12581244	0.0091 (6.35e-02)	4.65 (2.06e-03)	chr17	33700295	N_Shore	SLFN11		yes
cg13074173	-0.0074 (7.85e-02)	-3.27 (4.55e-03)	chr8	144462007	Island	RHPN1		
cg13074173:cg00973724	-0.0146 (2.20e-03)					*RHPN1 * TUBA1B*		
cg14055864	-0.0129 (7.50e-03)	-7.72 (3.72e-03)	chr3	138067120	Island	MRAS		
cg14102251	0.0062 (1.46e-01)	6.08 (7.67e-03)	chr14	75745098	Island	FOS		
cg14102251:cg12581244	0.0101 (2.70e-02)					*FOS * SLFN11*		
cg14264119	0.0169 (0.00e+00)	2.37 (5.41e-05)	chr10	11783794	N_Shore	ECHDC3		
cg14493336	0.0085 (3.73e-02)	0.68 (5.11e-03)	chr2	65295115	OpenSea	CEP68		
cg14493336:cg14264119	0.0091 (1.50e-02)					*CEP68 * ECHDC3*		
cg15380694	0.0125 (1.06e-02)	9.52 (3.49e-04)	chr8	125577065	OpenSea	MTSS1		yes
cg15661580	-0.0132 (6.00e-03)	-3.82 (1.04e-03)	chr10	71896835	OpenSea			
cg15821589	-0.0169 (2.00e-04)	-2.89 (8.26e-06)	chr1	78444904	Island	FUBP1		
cg16435922	0.016 (6.00e-04)	3.84 (4.55e-05)	chr16	55053053	Island			
cg17282982	0.0161 (0.00e+00)	34.56 (4.40e-05)	chr1	114355064	Island	RSBN1		yes
cg17332573	0.0126 (3.10e-03)	10.68 (1.22e-06)	chr6	26250934	N_Shore	HIST1H2BH; IST1H3F		yes
cg19822755	-0.0108 (7.60e-03)	-1.95 (1.97e-04)	chr16	30108259	S_Shore	YPEL3		
cg20249169	0.0128 (6.60e-03)	10.23 (2.72e-03)	chr2	64881356	Island	SERTAD2		

cg20249169:cg07417551	0.0191 (1.00e-03)					*SERTAD2 * PTGER1*	
cg20467667	0.0127 (6.10e-03)	15.65 (1.32e-06)	chr10	75987922	OpenSea	ADK	yes
cg21954346	0.0126 (5.80e-03)	8.03 (2.26e-04)	chr16	68271489	Island	ESRP2	yes
cg22069262	-0.0095 (4.73e-02)	-3.66 (6.60e-03)	chr7	1587402	Island	TMEM184A	
cg22335713	0.0106 (1.36e-02)	4.78 (3.94e-04)	chr10	123357603	Island	FGFR2	
cg22335713:cg22577984	-0.0157 (5.10e-03)					*FGFR2 * C2CD2L*	
cg22577984	-0.011 (1.83e-02)	-2.86 (5.99e-03)	chr11	118981363	S_Shelf	C2CD2L	
cg22749642	0.012 (4.70e-03)	1.29 (6.78e-03)	chr11	857033	N_Shore	TSPAN4	
cg23455212	-0.0056 (2.24e-01)	-1.7 (8.00e-03)	chr12	91505417	OpenSea	LUM	
cg24803876	0.0112 (1.07e-02)	12.81 (7.11e-04)	chr8	67340974	Island	RRS1	
cg26035817	-0.0115 (1.70e-02)	-1.89 (7.30e-04)	chr1	114414802	OpenSea	PTPN22	yes
cg26110653	-0.0139 (3.00e-04)	-6.7 (7.53e-06)	chr13	38924412	Island	UFM1	yes
cg26725552	-0.0143 (2.40e-03)	-2.77 (1.73e-03)	chr10	126314758	OpenSea	FAM53B	
cg26795934	0.0079 (1.01e-01)	23.3 (7.15e-06)	chr11	58421700	OpenSea		
ch.7.1140112F	0.0105 (3.39e-02)	10.2 (1.28e-05)	chr7	49032475	OpenSea		

From left to right column: names of the CpGs or CpG-CpG interactions, coefficients (contributing weights) of each CpG in the final model with bootstrap p-value, the fully adjusted association and corrected p-value obtained in epigenome-wide association analysis (EWAS), the chromosome where the CpGs were located and the genes to which the CpGs were mapped. Gene names of interaction terms (two gene names in italic and separated by "*") indicate that both interacting CpGs were annotated with gene names.

Supplementary Table 3. The 31 CpGs and 5 interactions in the model of telomere attrition rate.

Name of CpG or interaction	Coefficient (bootstrap p-value)	EWAS Coefficient (p-value)	Chromosome	Position on chromosome	Relation to island	Gene name	DMR	Promoter associated
(intercept)	8e-04 (4.62e-01)							
cg02537587	0.002 (1.90e-03)	1.24 (3.32e-05)	chr12	3982809	S_Shore	PARP11		yes
cg02690648	0.0024 (1.00e-04)	1.93 (1.49e-06)	chr8	10683044	OpenSea	PINX1; MIR1322		
cg04043642	-0.0024 (4.00e-04)	-0.17 (6.00e-03)	chr20	55842871	S_Shore	BMP7		
cg04935434	0.0021 (2.10e-03)	1.12 (3.83e-03)	chr2	177024947	Island			
cg05201958	0.002 (7.00e-04)	0.14 (1.34e-03)	chr10	61968731	OpenSea	ANK3		
cg06107116	0.0019 (1.00e-03)	1.31 (6.96e-05)	chr6	33267515	Island	TAPBP; RGL2		yes
cg08134068	-0.0022 (3.00e-04)	-1.37 (9.09e-03)	chr7	151039078	Island	NUB1		yes
cg09366730	-0.0021 (1.90e-03)	-0.2 (4.61e-06)	chr4	2341370	Island	ZFYVE28		
cg10020325	0.0016 (9.90e-03)	0.17 (7.77e-03)	chr2	220362163	Island	GMPPA		yes
cg10020325:cg16798728	-0.0025 (1.00e-04)					*GMPPA * DPP6*		
cg11086312	0.002 (1.60e-03)	0.1 (9.34e-04)	chr14	92720273	OpenSea			
cg11872375	0.0024 (3.00e-04)	0.7 (3.43e-04)	chr12	124709358	OpenSea	ZNF664-FAM101A		
cg12267637	0.0018 (2.90e-03)	0.15 (7.96e-04)	chr2	4102455	OpenSea			
cg12267637:cg23423086	-0.0017 (4.20e-03)							
cg12782249	-0.0017 (4.50e-03)	-0.07 (6.00e-04)	chr15	91458455	OpenSea	MAN2A2		
cg13960352	0.0017 (3.10e-03)	0.59 (4.79e-06)	chr19	45826863	Island	CKM		
cg15072796	0.0015 (1.40e-02)	1.31 (9.86e-04)	chr10	105727772	S_Shore	SLK		
cg16467775	-0.0017 (4.30e-03)	-0.2 (7.31e-04)	chr16	90085110	Island	DBNDD1		
cg16718624	-0.0016 (3.70e-03)	-3.55 (1.38e-03)	chr3	77089518	S_Shore	ROBO2		
cg16798728	-0.0015 (7.50e-03)	-0.22 (7.03e-03)	chr7	153984755	OpenSea	DPP6		
cg16798728:cg15072796	-0.0022 (3.00e-04)					*DPP6 * SLK*		
cg16975599	0.0017 (7.40e-03)	0.15 (4.88e-05)	chr3	158962761	OpenSea	IQCJ		
cg17751153	-0.0021 (1.00e-03)	-0.39 (3.51e-04)	chr5	447880	OpenSea	EXOC3		yes
cg19027504	0.0025 (0.00e+00)	1.23 (1.39e-04)	chr12	40013700	OpenSea	ABCD2		yes
cg19291087	0.002 (1.00e-03)	1.27 (1.67e-04)	chr2	3502760	N_Shore	ADI1		
cg19291087:cg04935434	-0.0022 (4.00e-04)							
cg19291087:cg16718624	0.0021 (5.00e-04)					*ADI1 * ROBO2*		
cg19887293	-0.0015 (8.80e-03)	-0.2 (1.19e-03)	chr6	33132693	S_Shelf	COL11A2		
cg20054536	-0.0027 (7.00e-04)	-0.27 (1.40e-05)	chr19	376069	Island	THEG		

cg23155965	0.0021 (5.00e-04)	0.11 (2.70e-03)	chr10	134778176	N_Shore			
cg23423086	0.0017 (6.50e-03)	0.58 (1.24e-03)	chr11	85856245	OpenSea			
cg23965459	-0.0025 (0.00e+00)	-0.52 (3.71e-04)	chr14	104647094	S_Shore	KIF26A		
cg24101053	0.0021 (1.00e-03)	0.56 (1.03e-03)	chr5	14419738	OpenSea	TRIO		
cg24790788	-0.0015 (8.60e-03)	-0.1 (1.05e-03)	chr3	145879626	S_Shore	PLOD2	RDMR	
cg24942922	-8e-04 (1.28e-01)	-0.17 (2.41e-04)	chr3	120003547	N_Shore			
cg25615010	0.0013 (3.78e-02)	1.9 (7.57e-03)	chr7	152456741	Island	ACTR3B		yes

From left to right column: names of the CpGs or CpG-CpG interactions, coefficients (contributing weights) of each CpG in the final model with bootstrap p-value, the fully adjusted association and corrected p-value obtained in epigenome-wide association analysis (EWAS), the chromosome where the CpGs were located and the genes to which the CpGs were mapped. DMR: differentially methylated region. RDMR: reprogramming-specific DMR. Gene names of interaction terms (two gene names in italic and separated by "*") indicate that both interacting CpGs were annotated with gene names.

Supplementary Table 4. Comparison of characteristics of the selected with the total data in ENVIR*ON*AGE cohort.

		Selected data (N=247)	Total data (N=916)
Newborns			
Sex	Female	121 (49.0%)	459 (50.1%)
Ethnicity	European	231 (93.5%)	808 (88.2%)
Birth weight (kg)		3.39 ± 0.44	3.39 ± 0.48
Gestational age (weeks)		39.24 ± 1.37	39.13 ± 1.58
Mothers			
Education			
	No diploma	24 (9.7%)	115 (12.6%)
	High school diploma	67 (27.1%)	333 (36.4%)
	A 3-year college	124 (50.2%)	354 (38.6%)
	A 4-year college or university	32 (13.0%)	114 (12.4%)
Smoking status			
	Never smoker	167(67.6%)	582 (63.5%)
	Former smoker	48 (19.4%)	210 (22.9%)
	Smoker	32 (13.0%)	124 (13.5%)
Parity			
	Primiparous	129 (52.2%)	503 (54.9%)
	Secundiparous	95 (38.5%)	313 (34.2%)
	Multiparous	23 (9.3%)	100 (10.9%)
With pregnancy complications		37 (15.0%)	166 (18.1%)
Age at delivery (years)		30.21 ± 4.38	29.29 ± 4.59
Pre-pregnancy BMI (kg/m2)		24.28 ± 4.56	24.26 ± 4.54
Fathers			
Age at the mother's delivery (years)		32.87 ± 5.60	31.88 ± 5.52