BMC Genomics

RESEARCH

Open Access

The multi-omics signatures of telomere length in childhood

Congrong Wang[1], Dries S. Martens[1], Mariona Bustamante[2,3,4], Rossella Alfano[1], Michelle Plusquin[1], Lea Maitre[2,3,4], John Wright[5], Rosemary R. C. McEachan[5], Johanna Lepeule[6], Remy Slama[7], Marina Vafeiadi[8], Leda Chatzi[9], Regina Grazuleviciene[10], Kristine B. Gutzkow[11], Hector Keun[12,13], Eva Borràs[3,14], Eduard Sabidó[3,14], Angel Carracedo[15,16], Georgia Escarami[17,18], Augusto Anguita-Ruiz[2,19], Dolors Pelegrí-Sisó[2], Juan R. Gonzalez[2,3,4], Martine Vrijheid[2,3,4] and Tim S. Nawrot[1,20*]

Abstract

Background Telomere length is an important indicator of biological age and a complex multi-factor trait. To date, the telomere interactome for comprehending the high-dimensional biological aspects linked to telomere regulation during childhood remains unexplored. Here we describe the multi-omics signatures associated with childhood telomere length.

Methods This study included 1001 children aged 6 to 11 years from the Human Early-life Exposome (HELIX) project. Telomere length was quantified via qPCR in peripheral blood of the children. Blood DNA methylation, gene expression, miRNA expression, plasma proteins and serum and urinary metabolites were measured through microarrays or (semi-) targeted assays. The association between each individual omics feature and telomere length was assessed in omics-wide association analyses. In addition, a literature-guided, sparse supervised integration method was applied to multiple omics, and latent components were extracted as predictors of child telomere length. The association of these latent components with early-life aging risk factors (child lifestyle, body mass index (BMI), exposure to smoking, etc.), were interrogated.

Results After multiple-testing correction, only two CpGs (cg23686403 and cg16238918 at PARD6G gene) out of all the omics features were significantly associated with child telomere length. The supervised multi-omics integration approach revealed robust associations between latent components and child BMI, with metabolites and proteins emerging as the primary contributing features. In these latent components, the contributing molecular features were known as involved in metabolism and immune regulation-related pathways.

Conclusions Findings of this multi-omics study suggested an intricate interplay between telomere length, metabolism and immune responses, providing valuable insights into the molecular underpinnings of the early-life biological aging.

Keywords Telomere length, Biological aging, Early-life, Multi-omics

*Correspondence:
Tim S. Nawrot
tim.nawrot@uhasselt.be

Full list of author information is available at the end of the article

Introduction

Telomeres are protective nucleoprotein caps at the ends of chromosomes which is crucial for chromosomal stability [1], and their shortening with chronological age is an important contributor to biological aging [2]. Therefore, telomere length is recognized as a marker of biological age and aging-related diseases [3]. Several studies have evaluated the molecular basis of telomere length in genomics [4], epigenomics [5], transcriptomics [6], and metabolomics [7]. In addition, omics-based aging clocks have been associated with telomere length in early life [8, 9]. Currently, studies have not incorporated an integrative approach across multiple omics for telomere length. With the advances in high-throughput omics methods, combining information from multi-omics has become feasible and would contribute to unraveling the complexities of telomere biology by capturing a more comprehensive and, especially, integrative view of the biological processes underlying telomere interactome.

Most studies evaluated biological aging in adulthood, however in early life such as childhood, differences in aging and developmental trajectories involving distinct biological processes are described and this may set the stage for vulnerability differences to diseases [10]. Therefore, children living free of diseases may have underlying biological aging differences that might persist and cumulate overtime. Considering the evidence that telomere length tracks over the lifespan [11, 12], studying telomere biology in early-life allows us to gain more insights into the underlying mechanism of biological aging and enhance the perception of extending a healthy lifespan.

Biological aging in children, as a multi-faceted process, has been reported to be accelerated by risk factors such as exposure to passive smoking [13], adiposity [14, 15], unhealthy lifestyle [9, 16], and lower socio-economic status [17]. Specifically for telomere shortening, the potential underlying mechanism of the effect of these risk factors could be increased cellular and genomic damage, elevated psychological stress [18] and insufficient intake of antioxidants from diet [19], which leads to chronic inflammation [20] and higher oxidative stresses [21].

In this study, we investigated the multi-omics signature of telomere length in children aged 6–11 years from the Human Early Life Exposome (HELIX) project. Our hypothesis posits that the multi-omics signatures of childhood telomere length differ depending on risk factors of early-life aging. Leveraging molecular measurements from various omics assays, we first assessed the association of individual omics features with telomere length. Subsequently, we conduct an integrative analysis of molecular features from multiple omics (DNA methylome, transcriptome, metabolome, and proteome), using a multi-block sparse partial least squares (sPLS) regression, to identify the multi-omics signatures of childhood telomere length. Finally, we evaluated the association of the multi-omics signatures to a priori selected early-life aging risk factors, including lifestyle factors, exposure to tobacco smoke, socioeconomic status, and body mass index (BMI).

Results

The study population

Within the HELIX project, a collaborative project of six established longitudinal birth cohorts in Europe (Spain, UK, France, Lithuania, Norway, and Greece) (Maitre et al., 2018; Vrijheid et al., 2014), multi-omics molecular profiles were assessed in a subcohort of children, aged between 6 and 11 years, including the average relative telomere length (quantitative real-time PCR (qPCR)), genome-wide genotyping (Infinium Global Screening Array, Illumina), blood DNA methylation (450 K, Illumina), blood gene expression (HTAv2.0,Affymetrix), blood miRNA expression (SurePrint Human miRNA rel 21, Agilent), plasma proteins (3 Luminex multiplex assays), serum metabolites (targeted LC-MS/MS metabolomic assay, Biocrates AbsoluteIDQ p180 kit), and urinary metabolites (^{1}H nuclear magnetic resonance (NMR) spectroscopy) (Supplementary Figure S1). The number of molecular probes and sample size in each omics platform are summarized in Supplementary Table S1. Lists of biomarkers in plasma protein, serum and urinary metabolites assays are available in Supplementary Data S1 – S3, respectively. The current study was based on 1001 children, who were of European ancestry defined based on the genome-wide genotype data, and had complete data on relative telomere length measurements and blood cell proportions estimated based on DNA methylation profiles available (Fig. 1).

The general characteristics of the HELIX children involved in the current study ($N = 1001$) are shown in Table 1. As a comparison, the characteristics of all HELIX subcohort children ($N = 1301$), which have been described previously [9, 22], and are listed in parallel in Supplementary Table S2. The children had a mean age of 7.9 years (range 5.4–12.0 years) and 45.2% were girls. The six cohorts comprised 9.2–19.8% of the total sample size of the current study. Compared with all the HELIX subcohort children, this study included less children from BiB and EDEN who were not of European ancestry. There was no difference observed in the other characteristics. All children in the study population had the genome-wide genetic data available. Around 74% of the children had data available in all the other six omics layers and 99% had data available in at least 4 omics (Supplementary Table S3).

To assist the learning of multi-omics signatures of telomere length, we calculated additional estimators of telomere length, namely, the DNA methylation-based

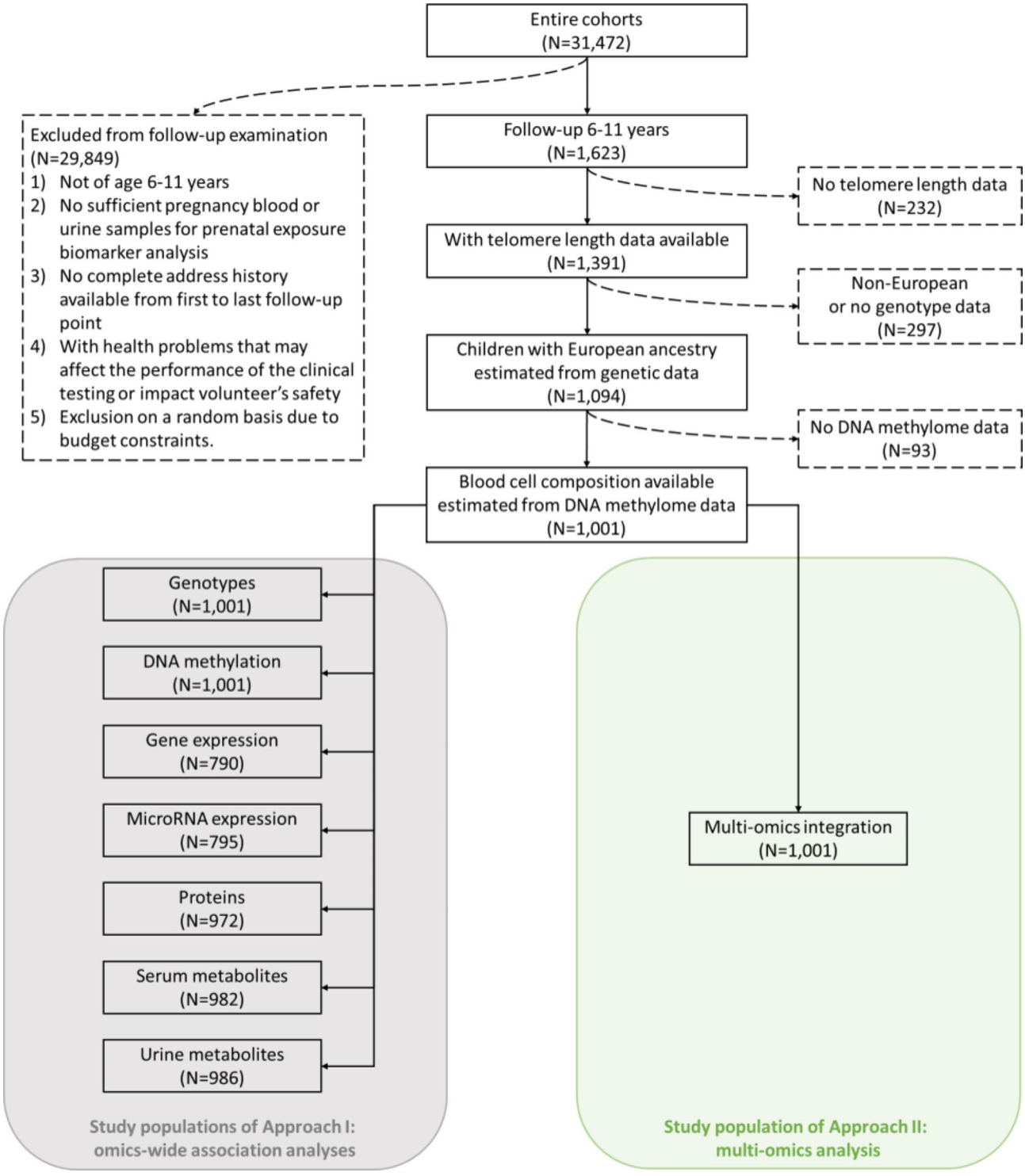

Fig. 1 Flowchart of the participant inclusion. Sample sizes and inclusion/exclusion criteria of the Human Early Life Exposome (HELIX) project and the study populations of the current study

telomere length estimator (DNAmTL) [5] based on the DNA methylation data, and two polygenic scores (PRS) of telomere length, hereafter referred to as Li's PRS [23] and Codd's PRS [24], based on the whole-genome genetic data. The measured telomere length showed a correlation with DNAmTL ($r = 0.25$; $p < 0.001$) and two polygenic scores ($r = 0.21$, $p < 0.001$ and $r = 0.23$, $p < 0.001$, respectively) (Supplementary Figure S2). No correlation was detected between DNAmTL and the two PRSs ($r = 0.052$; $p = 0.10$ and $r = 0.047$; $p = 0.13$, respectively), while

Table 1 Characteristics of the study population ($N = 1001$)

Characteristics		*n* (%) or mean ± SD
Cohort	BiB	92 (9.2%)
	EDEN	137 (13.7%)
	INMA	199 (19.8%)
	KANC	193 (19.3%)
	MoBa	194 (19.4%)
	RHEA	186 (18.6%)
Sex	Male	549 (54.8%)
	Female	452 (45.2%)
Birth weight	< 2500 g	28 (2.8%)
	[2500 g, 3500 g)	537 (53.6%)
	[3500 g, 4500 g)	313 (31.3%)
	≥ 4500 g	99 (9.9%)
Child BMI[a]	Underweight	62 (6.2%)
	Normal	709 (70.8%)
	Overweight	150 (15.0%)
	Obese	59 (5.9%)
Maternal education level	Primary school	113 (11.3%)
	Secondary school	340 (34.0%)
	University or higher	510 (50.9%)
Maternal pre-pregnancy BMI[b]	Underweight	37 (3.7%)
	Normal	557 (55.6%)
	Overweight	239 (23.9%)
	Obese	14 (14.0%)
Maternal smoking status	Never smoker	778 (77.8%)
	Smoked before pregnancy	0 (0)
	Sustained smoker during pregnancy	99 (9.9%)
Childhood parental smoking	Neither	587 (58.6%)
	One	275 (27.5%)
	Both	112 (11.2%)
Family affluence score[c]	Low	89 (8.9%)
	Middle	378 (37.8%)
	High	513 (51.3%)
Age (years)		7.9 ± 1.6
Gestational age (weeks)		39.6 ± 1.6
Child moderate-to-vigorous physical activity (min/day)		40.0 ± 25.3
KIDMED index in children[d]		2.8 ± 1.7

[a]. Child body-mass-index (BMI) categories were defined according to the CDC growth charts of sex-specific BMI-for-age percentile curves. Children with BMI less than the 5th percentile were "underweight", from the 5th to 85th percentiles were "normal", from the 85th to the 95th were "overweight", and those greater than the 95th percentile were "obese"

[b]. Maternal pre-pregnancy BMI was grouped according to WHO categories for underweight (< 18.5 kg/m2), normal (18.5–24.9 kg/m2), overweight (25–29.9 kg/m2) and obese (≥ 30 kg/m2)

[c]. Family affluence score was categorized as "low" for scores 0, 1 and 2, as "medium" for scores 3, 4, and 5, and as "high" for scores 6, 7, 8, and 9

[d]. The Mediterranean Diet Quality Index (KIDMED index) was used to quantify the children's Mediterranean diet patterns. Ranging from −4 to 11, a higher KIDMED index reflects greater adherence to a Mediterranean diet

the latter were moderately correlated with each other ($r = 0.56$; $p < 0.001$). Child age was slightly correlated with the measured telomere length (r=-0.07; $p = 0.04$) and showed a stronger correlation with DNAmTL (r=-0.37; $p < 0.001$), but was not found to correlate with the two PRSs ($r = 0.003$; $p = 0.92$ and r=-0.04; $p = 0.16$, respectively). Compared to girls, boys had shorter telomere lengths and shorter DNAmTL (both with $p < 0.001$), while no difference in the two PRSs was observed between girls and boys ($p = 0.79$ and $p = 0.37$, respectively). Children's BMI z-scores (zBMI) showed a weak but significant correlation with the measured telomere length (r=-0.10, $p = 0.002$) and DNAmTL (r=-0.07, $p = 0.03$).

Omics-wide association analysis showed diverse relevance to childhood telomere length

We first conducted omics-wide association analyses, where the association between the child's relative telomere length and each single molecular feature was evaluated using multiple linear regression models (Fig. 2A, Approach I). All models were adjusted for child age, sex, the blood cell compositions estimated from the DNA methylation data, and the first four principal components (PCs) of the whole-genome genetic data which corrected the population heterogeneity due to cohort and ethnicity. Nominal p-values ($-\log_{10}$-transformed) from the omics-wide association analyses are plotted in Fig. 3 and summarized in Supplementary Figure S3. Manhattan plots with additional information on the omics features' genomic locations and metabolite compound classes can be found in Supplementary Figure S4 and S5, respectively. After omics-specific Bonferroni correction for multiple testing, the only two significant hits in all omics were from DNA methylation. In general, more signals, suggested by the p-values, were observed in the genetic variants, DNA methylation and gene expression than in other omics layers.

For each omics layer, the top 20 molecular features with the lowest p-values are listed in Supplementary Data S4 – S10, respectively. The top SNPs (with p-value $< 10^{-6}$) were annotated to gene loci *AC139768.1*, *DAZAP2*/*SMAGP*, and *RP11-351O2.1*. In DNA methylation, two out of 386,518 CpGs were significant after Bonferroni correction and showed a positive association with telomere length, both with genomic location near the *PRD6G* gene (cg23686043 in the promoter and cg16238918 in gene body). The top transcript from the gene expression profile belonged to the *DEXI* homolog gene cluster (inversely associated with telomere length, $p = 1.28 \times 10^{-6}$), and the top miRNA was *MIR6752* (positively associated with telomere length, $p = 0.003$). The associations of metabolites and plasma proteins to telomere length were relatively weak. The top plasma protein was IL-1beta ($p = 0.013$), inversely associated with telomere length.

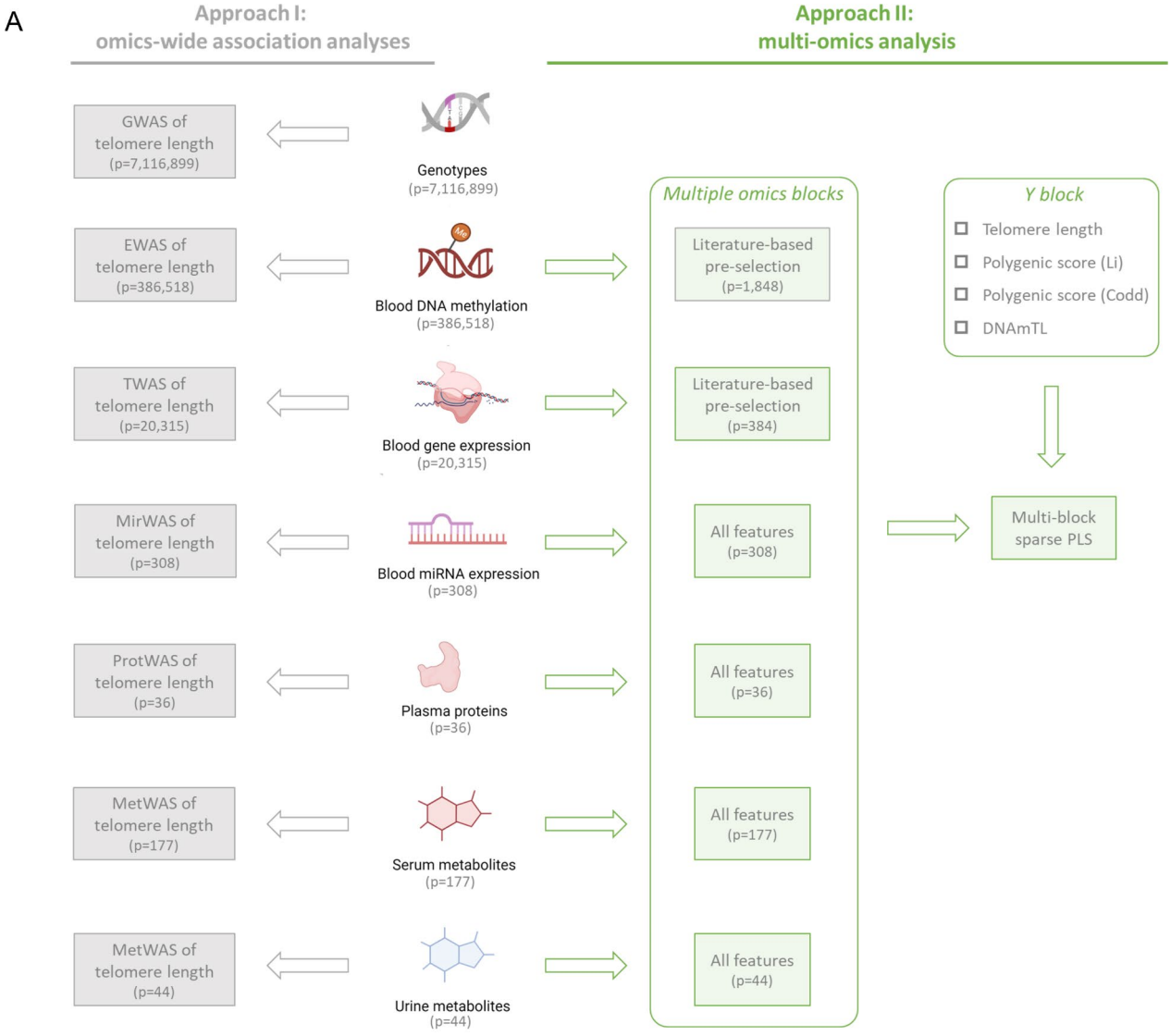

Fig. 2 (See legend on next page.)

(See figure on previous page.)

Fig. 2 Data analysis procedures of the current study. (**A**) The statistical analyses using two approaches. In Approach I (grey color in the left half), omics-wide association analyses were conducted within each omics to assess the association between each individual feature and telomere length. In Approach II (green color in the right half), multiple (pre-selected) omics were analyzed via a supervised method, multi-block sparse Partial Least Squares (sPLS), against four telomere length measures. (**B**) The literature-based omics feature pre-selection in genome-wide CpG methylation and gene expression. All green boxes represent procedures based on literature or databases, while the blue box stands for a data-driven filtering of gene transcripts where the transcript with the highest variance within the same gene was selected. Stage ①: significant SNPs from published genome-wide association studies (GWAS's) of telomere length (TL) were used to extract DNA methylation quantitative trait loci (mQTL) and gene expression quantitative trait loci (eQTL) from publicly accessible databases, which were in turn used to select a first set of CpGs in the DNA methylation data and a first set of transcripts in the gene expression data. Stage ②: genes involved in telomere regulation and two cellular aging-related signaling pathways, mTOR and AMPK pathways, were used to select a second set of gene transcripts, and to extract gene expression quantitative trait methylation (eQTM) from a published study which were then used to select a second set of CpGs. Stage ③: an epigenome-wide association study (EWAS) of TL was used to select a third set of CpGs. All selected CpGs were further filtered considering the probe reliability in the Illumina 450 K array

The top metabolites were serum acylcarnitine C10:1 ($p = 0.023$) which was inversely associated with telomere length, and urinary formate ($p = 0.033$) which was positively associated with telomere length.

Literature-based preselection of molecular features for multi-omics integration analysis

Prior to the integrative analysis of multiple omics, we pre-selected CpGs and gene transcripts, where the data dimension was relatively higher, based on literature and publicly accessible databases (Fig. 2A, Approach II). The purposes of this pre-selection were to denoise the omics data, to improve between-omics intercorrelation in the two omics layers of the highest dimensions, and to reduce model overfitting to the data. The workflow of the literature-based feature pre-selection is illustrated in Fig. 2B and a detailed description can be found in Supplementary Methods. As a result of the pre-selection, 1848 out of 386,518 CpGs and 384 out of 20,315 gene expression transcript clusters remained. For the other omics (miRNA expression, plasma proteins, serum metabolites and urine metabolites), all available features were used for the integrative analysis. We show a comparison between the associations in the reference studies and the associations estimated in the HELIX children for the same omics features. The SNPs and CpGs that were selected from published genome-wide association studies (GWAS) and epigenome-wide association studies (EWAS) are listed in Supplementary Data S11 and S12, respectively, with the feature annotation and model estimates from HELIX and the reference studies. All the selected SNPs and CpGs were genome-wide significant in the corresponding reference study, among which there were only two SNPs with a p-value $< 10^{-5}$ in the HELIX children (both related to *TERC* gene) and one CpG with a p-value $< 10^{-4}$ (in the gene body of *CHL1*). The direction of the estimated associations, on the other hand, showed consistency between the reference studies and HELIX children. Similarly, among the 384 selected gene transcripts, only four transcripts had a p-value < 0.01 (Supplementary Data S13). We did not detect an overrepresentation of the selected CpGs and gene transcripts in the top signals from EWAS and TWAS, respectively (Supplementary Table S4).

Latent components from multi-omics were associated with child BMI

Features across six omics (DNA methylation, gene expression, miRNA expression, plasma proteins, serum metabolites and urine metabolites) were analyzed to identify the multi-omics signature of childhood telomere length (Fig. 2A, Approach II), using a multi-block sparse partial least squares (multi-block sPLS) method implemented in the mixOmics package (Rohart et al., 2017). In addition to the relative telomere length, we included Li's PRS and Codd's PRS, and DNAmTL in the outcome block (Y-block) as well, to increase the variation in the phenotype, assisting the supervised learning from the high-dimension omics blocks.

The numbers of latent components, which maximized the covariance between multiple omics blocks and the Y-block, are listed in Supplementary Table S5. A six-component model was selected across all omics layers that balanced the proportion of covariance explained and model simplicity (Supplementary Figure S6). This model defined a data space spanned by six axes. The features from each omics layer were projected onto these axes, generating one omics-specific component for each axis. Within each omics block, the proportions of variance explained by the corresponding six components are depicted in Fig. 4. Comparing between omics layers, the higher the data dimension, the lower the proportion of variance explained by the components. In addition, there was a clear decreasing trend of variance explained from component 1 to component 6 in the lower-dimension omics that could not be observed in DNA methylation. Based on the specification of keeping six components, the number of omics features to keep was tuned by sparse PLS within each omics block with a five-fold cross validation, from which the number of features was obtained as shown in Supplementary Table S6. The component loadings of the selected omics features are listed in Supplementary Data S14.

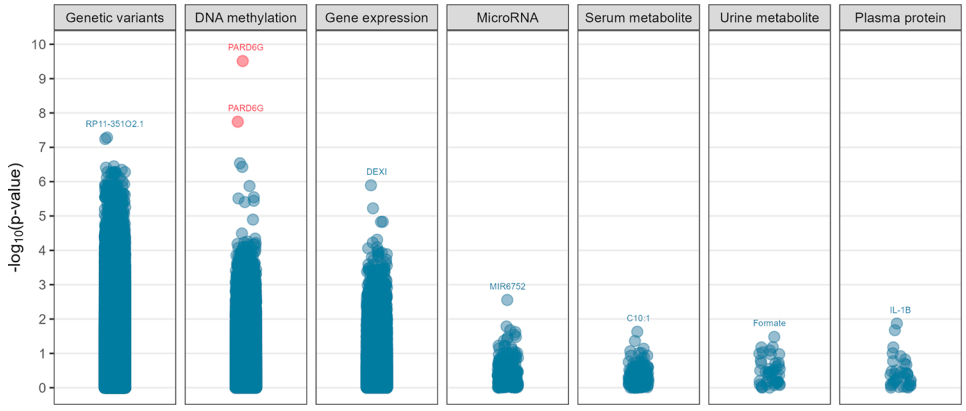

Fig. 3 Summary of significance of the omics-wide association analyses with telomere length in HELIX children. All omics features are shown as points by -$\log_{10}$-transformed nominal p-value versus the omics and the type of model. Omics-wide significant features under Bonferroni correction are colored red. The top molecular features were labeled with the corresponding gene/metabolite names. Models were adjusted for key covariates: child age, sex, the first four genetic PCs and the estimated blood cell compositions

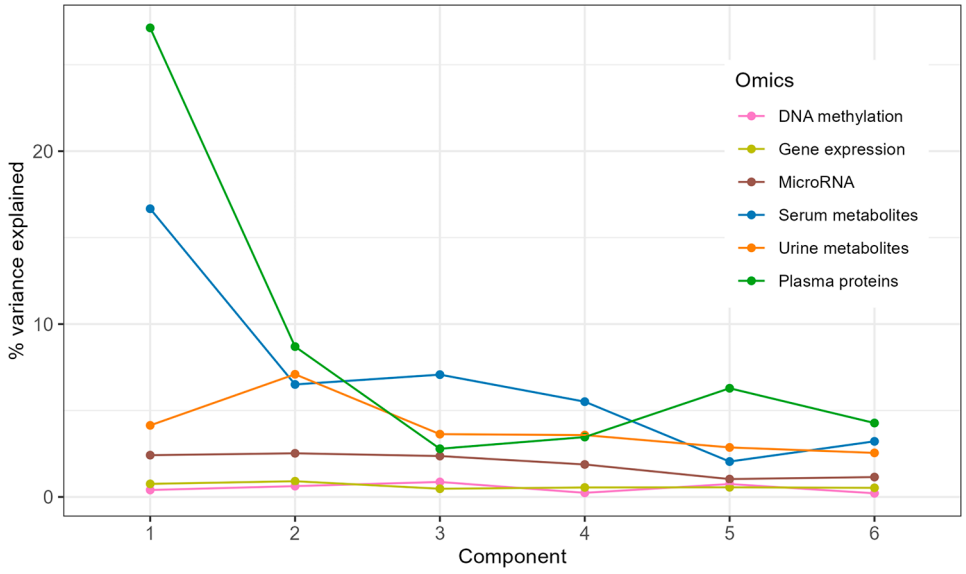

Fig. 4 The proportion of variance explained by the six-component multi-block sPLS model. The model derived six components in each omics block. Displayed are the proportion of variance of each omics explained by each of the components

A correlation heatmap between components of the six omics blocks and the four telomere length measures is shown in Supplementary Figure S7. For each component, the most correlated components in other omics were along the same axis in the component space. Relatively stronger inter-correlations were found between serum metabolite components, urine metabolite components and plasma protein components, between DNA methylation and gene expression, and between gene expression and serum metabolites. The four telomere length measures showed only weak correlations with the omics components, among which five had an absolute correlation higher than 0.10: $r = 0.122$ ($p < 0.001$) between DNA methylation component 3 and Li's polygenic score, r=-0.104 ($p < 0.001$) between miRNA component 4 and Li's polygenic score, $r = 0.102$ ($p = 0.001$) between DNA methylation component 2 and DNAmTL, r=-0.101 ($p = 0.001$) between miRNA component 4 and measured telomere length, and $r = 0.100$ ($p = 0.001$) between gene expression component 2 and DNAmTL. For the Y block components, DNA methylation components 2 and 3, gene expression components 2 and 5, and miRNA component 4 showed absolute correlations higher than 0.10 (all with $p < 0.05$).

Pathway enrichment analyses were performed on genes suggested by the components that were most correlated with telomere length measures or Y components described above (DNA methylation component 2 and component 3, gene expression component 2 and component 5, and miRNA component 4), where omics features

(CpGs, gene transcripts, and miRNAs) with an absolute loading > 0.10 on the corresponding components were used. In total, a list of nine, three, five, six, and nine unique genes were annotated for each of the four components, respectively, which formed a list of 31 unique query genes. The miRNAs also suggested 57 unique target genes. Figure 5 shows the pathways enriched in the annotated genes and target genes, including the regulation of general cellular activities (NOTCH signaling and cell membrane traffic), autophagy and immune responses including SARS-CoV infection-related pathways.

From the final six-component multi-block sPLS model, component scores were calculated for each component in each omics block. Multiple linear regression models were fitted to assess the association between the components in each omics block and early-life aging risk factors: gestational age, birth weight, maternal pre-pregnancy BMI, maternal smoking status, maternal education level, parental smoking in the children's household, family affluence score, child BMI, child physical activity level and child Mediterranean diet score. The association estimates of all risk factors to components scores are listed in Supplementary Tables S7 – S12. Child zBMI was found to be the only significant risk factor under Bonferroni correction, consistent across multiple omics layers and multiple components. Specifically, child zBMI was inversely associated with plasma protein component 2, serum metabolite component 2 and component 5, and urinary metabolite component 2 and component 3. The multi-omics features and their contributions to these components (if having an absolute loading > 0.10), as well as the association between the components and child zBMI, are visualized in Fig. 6. None of the components from DNA methylation, gene expression and miRNA showed an association with child zBMI.

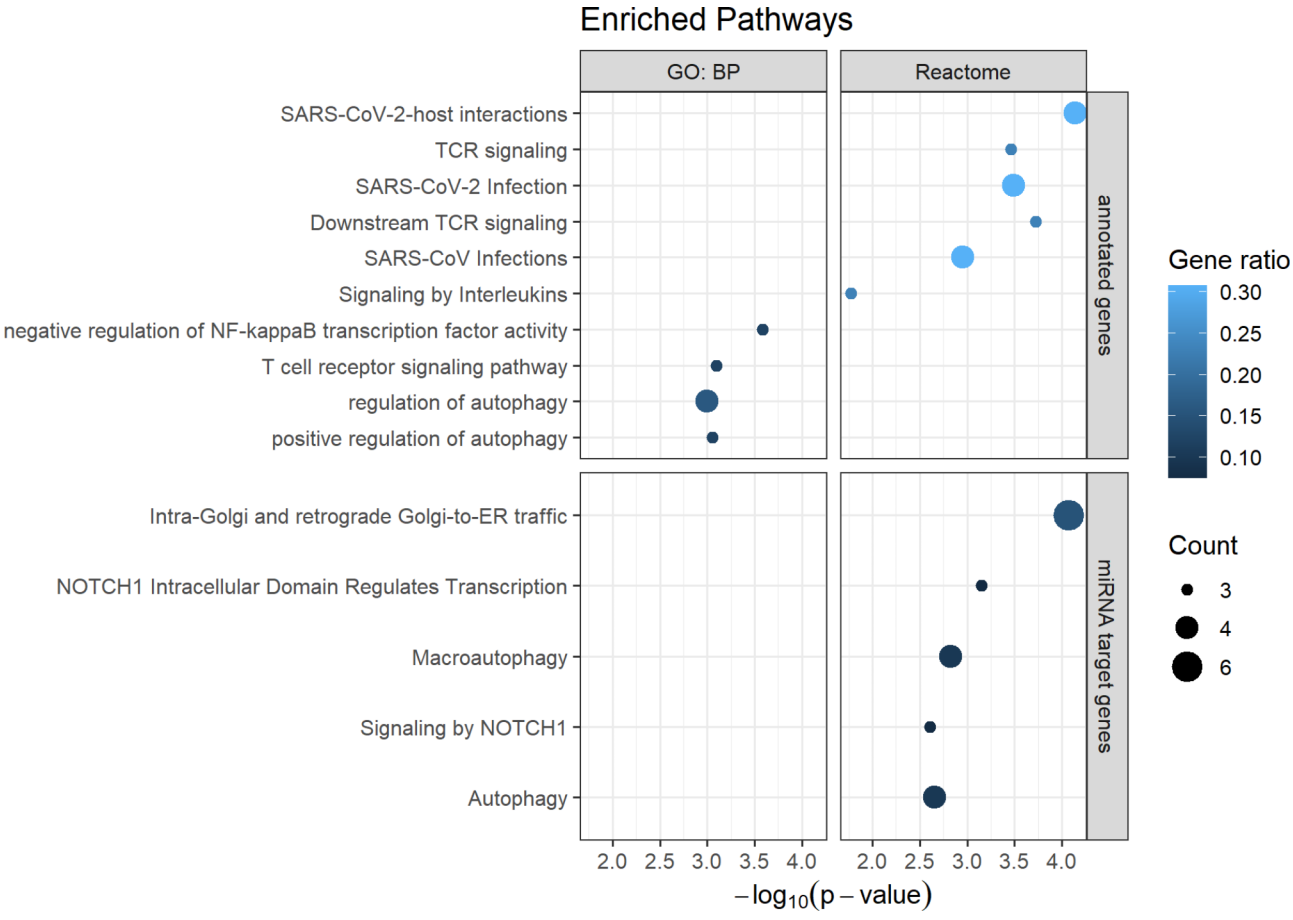

Fig. 5 Pathways enriched in the genes suggested by the components mostly related to telomere length. Genes annotated to features from DNA methylation component 2 and component 3, gene expression component 2 and component 5 and miRNA component 4 (upper panels) and target genes of miRNAs in miRNA component 4 (lower panels) were analyzed separately. Enriched pathways shown are with at least 3 genes from the pathways in the query list, and adjusted p-value < 0.10. Databases used for pathway enrichment analyses were gene ontology (GO) of biological process (BP), cellular component (CC) and molecular function (MF), the Reactome pathway database and KEGG. A database is not shown in the figure if no pathways from the database were identified

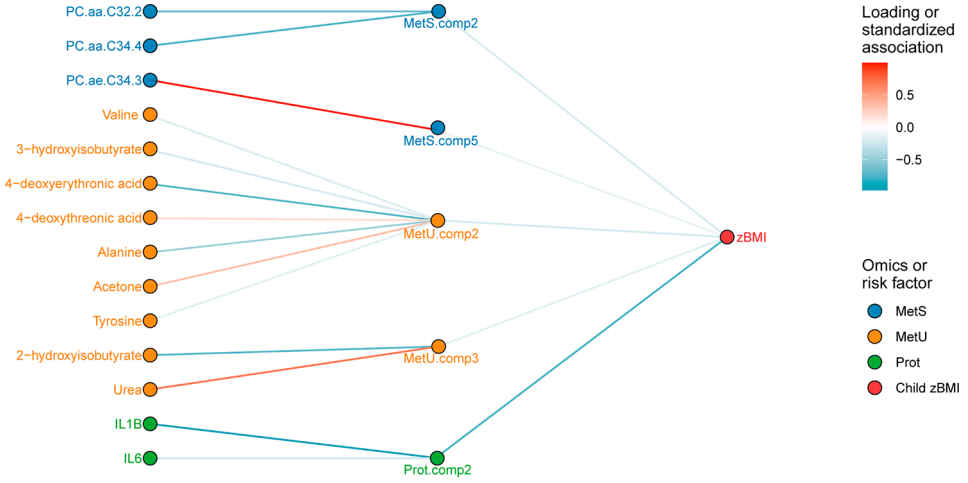

Fig. 6 A summary of the relationships of features, components, and aging-related risk factors. Features shown in the left column are those with absolute loadings higher than 0.1 on the corresponding component. Components are in the middle and the early-life aging risk factors are in the right column. Links between the features and components stand for the loadings. Links between the components and risk factors stand for the standardized associations which were identified as significant under Bonferroni correction. The association with the only risk factor, child body mass index (BMI) z-score (zBMI), was estimated in a multiple regression model adjusted for all the other risk factors (gestational age, birth weight, maternal pre-pregnancy BMI, maternal smoking status, maternal education level, parental smoking in the children's household, family affluence score, child physical activity level and child Mediterranean diet score). MetS: serum metabolites. MetU: urinary metabolites. Prot: plasma proteins

Discussion

Telomere length is an important molecular indicator of disease susceptibility, lifespan and a complex multi-factor trait. Here we present a multi-omics study of child telomere length to investigate the molecular signatures associated with telomere biology in early-life including genomics, DNA methylation, transcriptomics, proteomics and metabolomics. Genes annotated to the selected omics features from DNA methylation, gene expression, and miRNA components showed enrichment in pathways of immune signaling. Child BMI was strongly associated with plasma protein component 2, serum metabolite components 2 and 5, as well as urinary metabolite components 2 and 3. Although the association analyses within single omics layers and the integrative multi-omics analyses did not identify strong signatures of childhood telomere length, our findings may suggest the involvement of immune signaling and metabolic regulation in telomere biology at early stages of life.

The individual omics feature association analysis identified more signals or potential associations in the genome, DNA methylome, and gene transcriptome, but not as likely with the serum and urine metabolites or plasma proteins. The former three omics were profiled using genome-wide assays where a large number of features were measured, while the latter three were quantified in targeted assays that could be missing metabolites and proteins relevant to telomere length. Indeed, the analytical methods of the metabolites guaranteed reliable quantification of omnipresent metabolites, performing well in sensitivity and specificity as well as an explicit metabolite identification, but resulted in only partial coverage of the serum and urine metabolome [25]. Similarly, the candidate plasma proteins were selected a priori based on the literature and on the Luminex kits that were commercially available. In addition, the fact that telomere length, genetic variants and DNA methylation were profiled in the same blood sample extracts might also have increased the interconnections among these traits. On the other hand, our findings might indicate that higher-level biological regulations (i.e. genome and epigenome) are more involved than the lower-level regulations (i.e. gene expression, proteome, and metabolome) of telomere biology.

The relatively stronger signals identified in the omics-wide association analyses showed limited consistency with previous studies of telomere length. However, some of the top omics features have been reported in association with other anthropometric traits or biological processes potentially relevant to early-life aging. One of the top SNP, rs1049467 (*DAZAP2*/*SMAGP*), has been identified in a large-scale GWAS of height in UK Biobank [26]. For DNA methylation, the two epigenome-wide significant CpGs are located in the upstream of transcriptional starting site and within the gene body of *PARD6G* gene, respectivley. This gene encodes a cell polarity regulator involved in cell proliferation activity [27] and centrosomal protein composition [28]. This gene has been found to exhibit an increased expression level associated with higher chronological age in dogs [29]. The top signal from gene expression, the *DEXI* homolog gene, has been shown to function as a potential aetiological gene for type 1 diabetes and to be involved in the activation of

local antiviral immune response in pancreatic beta cells [30, 31].

By comparing the estimated associations with the reference studies that were used for the omics feature pre-selection, there was a large inconsistency where the top signals in HELIX children were mostly not reported in the reference studies. All of the features selected from the reference studies were significant in the corresponding study population with a control for family-wise error rate, but not as significant in the HELIX children. Still, two SNPs near the *TERC* gene showed the same direction of association as in the reference GWAS's, with a p-value $< 10^{-5}$, and a CpG in the *CHL1* gene showed the same direction as the reference EWAS with a p-value $= 1.27X10^{-5}$. The RNA transcript of *TERC* is one of the components of telomerase, a crucial telomere regulator. SNPs in *TERC* have been identified in GWAS's of telomere length also other populations [32, 33]. *CHL1* encodes cell adhesion molecules that are known essential for neural development and to play a role in carcinogenesis [34]. The sample size of the study population could contribute to the difference, as the sample sizes of the reference studies were tens to hundreds of times larger than that of HELIX. This highlights the need of larger-scale molecular studies of telomere length in early life. The inconsistency might also suggest novel findings in children which pointed to a distinction in the biology between early-life aging in children and the aging in adult populations. Indeed, children undergo rapid growth and development, and their biological processes may differ significantly from those of adults.

To increase the statistical power in a study context with high data dimension and limited sample size, we were motivated to avoid a purely data-driven analysis by bringing in prior information from literature for the multi-omics integration approach. Although the SNPs were not used in the integrative analysis with the other six omics due to their distinct data type, the feature pre-selection starting from QTL mapping was a strategy to incorporate the intercorrelation between genetic data and the other omics layers. In addition to the findings from molecular epidemiology studies, well-established biological knowledge was taken into account regarding the biological regulation of telomere length and two key pathways in cellular aging. AMPK (adenosine monophosphate-activated protein kinase) pathway controls cellular metabolism and coordinates cell growth and autophagy [35], and the mTOR (mechanistic target of rapamycin) pathway senses nutrient availability and energy status and is a key regulator of cell cycle and proliferation [36], which are all interconnected with cellular senescence and aging and are therefore relevant for telomere biology. Using the literature-based pre-selection, we filtered out CpGs and gene transcripts that were less likely to be relevant for telomere length. This increased the overall intercorrelation between the remaining CpGs and gene transcripts. Although the intercorrelations between these two pre-selected omics layers and other omics layers, as well as the intercorrelations among the other layers themselves, did not increase directly, the improvement was indirectly transferred to the other omics layers through supervised learning, using the telomere length measures as the outcome variables.

The current study is the first to conduct an analysis of telomere length by integrating multiple omics profiled in children. The biological complexity and the variety of data types in multi-omics data can hold rich information as well as noise. To identify signals relevant to the trait of interest from the high-volume multi-omics data, commonly used algorithms focused on data reduction, which is usually achieved by projecting the original data to a lower-dimension space or by statistical feature selection. In the current study, we were able to combine both methods using the multi-block sPLS. Since the model was supervised, the latent components obtained from the integration maximized the covariance between the telomere length measures and each omics layer, and thus potentially represented the multi-omics signature of telomere length. In addition to the measured telomere length, we added two PRS's and DNAmTL to increase the variability in the Y-block. As a demonstration of the rationale for including additional telomere length measures, we found that in a model using only the measured telomere length as the outcome variable, it was less clear how many components should be extracted across the different omics blocks (Supplementary Figure S8). While using multiple telomere length measures, we assumed that these estimators of telomere length also contained different information about biological aging. As observed in the data, child age was not correlated with the PRS's but showed clearly a medium inverse correlation with DNAmTL. This might suggest that it was the baseline telomere length, rather than telomere attrition, that was reflected by the PRS's. As noted from this supervised learning, however, the weak correlations between these components and telomere length suggested that no strong multi-omics signatures of telomere length could be found. This result is consistent with that of the Approach I analysis where the number of signals was in general low across all omics layers. Especially, the pre-selection has filtered out the CpGs and gene transcripts which showed the strongest associations with telomere length. It was also notable that the correlations between the measured telomere length and the DNAmTL or the PRSs were not high, although they were within the range of correlations previously reported [5, 37, 38]. Telomere length measures predicted based on other omics profiles may capture aspects of telomere maintenance

mechanisms and not necessarily telomere length itself [37, 39]. Therefore, by adding the PRSs and DNAmTL in the Y block, it was possible that the selected omics features were not directly related to the qPCR-based telomere length measurements.

The components in metabolites and plasma proteins were consistently strongly associated with child zBMI with estimated association ranging from −0.11 to -0.76. Since the components were projections that were obtained by maximizing the covariance with the telomere length measures, this result confirmed findings in numerous external studies in adults [40–44] as well as previous studies in the HELIX children [9, 14], where BMI, obesity or BMI genetic risk score has been reported to show an inverse association with telomere length. Besides, studies on diet intervention revealed effects of obesity treatment outcome suggested by longer telomere length in the treatment group compared with the control [45]. Added by the current study, the telomere length-BMI relationship identified indirectly in a multi-omics view has validated our multi-omics analysis approach and exhibited consistency in the molecular measurements in the HELIX project.

From an alternative perspective, the extracted components might not be the signature of telomere length, but the signatures of other facets of biological aging. Given that the measured telomere length and DNAmTL were correlated with child zBMI, and child BMI covaried more strongly with part of the omics [25], the identification of multi-omics signatures of telomere length might have been conveyed to the identification of multi-omics signatures of child BMI. In addition, metabolites and plasma proteins are downstream biological features that result from cumulative regulation at other levels. As a result, these layers may contribute more to explaining the covariance between omics.

It should be noted that the findings in the current study should be interpreted with caution in terms of potential causality. Telomere length was considered as a phenotype and the outcome variable in statistical models, while epigenetic and transcriptional alterations might also regulate the maintenance of telomere [46, 47]. Similarly, the variation in child BMI could reflect long-term cumulative changes in the metabolome and proteome, although shifts in the latter could also be due to BMI change [48]. Reverse causation is thus likely given that all the omics and telomere length were quantified as a snapshot at the same time point for each child. Future longitudinal observational studies or research on biological mechanisms could help elucidate the underlying causal relationship.

Although aging is a multifactorial process, healthy aging is in general believed to be associated with longer telomeres and lower inflammation profiles in adults [49]. From our study, child BMI was connected with telomere length via two plasma protein components where interleukins were identified. Interleukins are a group of cytokines with immunomodulatory functions, some of which also have major roles in the etiology of metabolic diseases. From the plasma proteins component 2, IL-1beta induces insulin resistance [50] which can contribute to obesity [51] and for which obesity can be a triggering factor [52]. Notably, IL-1beta was also the top plasma protein signal from Approach I, despite that the association did not reach Bonferroni significance and the effect size was small. A previous study [53] also reported an association between shorter telomere length and elevated plasma IL-1beta levels in Alzheimer's disease patients. Adipose tissues are the source of IL-6 which is an inflammation stimulus [54] and has been reported to inversely associated with telomere length in chronic obstructive pulmonary disease patients [55]. Taken together, all these findings added to the existing bulk of evidence that a complex inter-relationship exists between cellular aging, metabolic regulation and inflammation. Still, the causal relationship between these role players requires further investigation in subsequent studies.

As for the identified metabolites, serum PC aa C32:2 in component 2 has been reported to be positively associated with obesity in adults [56] and PC aa C34:4 showed reduced expression in the late-onset Alzheimer's disease patients [57], and low levels of both metabolites have been identified as plasma biomarkers of poor muscle quality in an older population [58]. The serum PC ae C34:3 in component 5 has been shown to decrease with childhood obesity [59], consistent with its positive loading on serum metabolite component 5 and the latter's inverse association with child zBMI.

In the urinary metabolites component 2, the highly negatively loaded compound, 4-deoxythronic and 4-deoxyerythronic acid, has been reported as a potential biomarker for type 1 diabetes [60]. The positive associations of valine, alanine and tyrosine to BMI have also been reported in an adult population [61]. Also consistent with our finding, acetone has been shown to display a higher level in people with lower BMI, which might be due to higher metabolic rates and higher fat burning capabilities [62]. 3-hydrooxyisobutyrate, reported to reduced blood glucose level in type 2 diabetes in a clinical trial [63]. To sum up, the urine metabolite component 2 could represent a higher metabolic rate in the children and thus showed an inverse association with child BMI. 2-hydroxyisobutyrate (urine metabolite component 3) has been reported to be elevated in autoimmune diseases [64] and decreased during sleep deprivation [65]. Urinary urea is an established marker of dietary protein intake [66]. Substantial evidence indicated an higher dietary protein intake as a strategy to prevent or treat obesity [67], whereas the effect of dietary protein intake

on telomere length depends on the specific source of protein [19].

Our study has multiple strengths. First, we based the analysis on a large cohort consisting of children from six European countries which increased the generalizability of findings to the population. Second, the multi-omics covered measurements from genome to metabolome, providing an all-round view of the molecular basis. Third, we conducted multiple types of analyses, incorporated knowledge from literature, and applied stringent control for statistical significance. Fourth, our results showed consistencies with other studies regarding the suggested mechanism of the relationship between cellular aging and BMI in children, which proved the effectiveness of the multi-omics method used in the current study.

We also acknowledge the limitations of this work. First, the omics platforms were not completely comparable in their coverage. This has introduced a bias in the type and number of assessed features, and can further bias the biological interpretations. Second, heterogeneity was present in the data due to omics array compatibility, study center difference, technical noises and measurement error. We have conducted a number of normalization and denoising strategies to reduce these variations, which leads to the question regarding signal loss in the processed data. In order to reduce the biological noise in the data, the population of this study was restricted to children with European ancestry, which limited the potential generalizability of the findings to other populations. Third, because of the inconsistency with literature in the association analysis with individual features, we have restricted the multi-omics analysis to the known features. This method has increased the signal-to-noise ratio and reduced overfitting, but might also have limited the likelihood of novel findings. Fourth, given the high dimension of the multi-omics data, the sample size of the study population and the lack of external validation might have limited the statistical power. Finally, as discussed above, causal relationships cannot be inferred from this study.

To conclude, this is the first study of multi-omics signatures of childhood telomere length. We identified multi-omics signatures that showed differentiation with respect to child BMI. These findings contribute to advancing our understanding of, and equip us with a multi-omics toolbox to understand, the intricate relationships between genetic, epigenetic, metabolomic and environmental factors in shaping telomere length during childhood.

Methods

Study population and sample collection

The HELIX study is a collaborative project of six established longitudinal birth cohorts in Europe [68, 69] with singleton deliveries during 2003–2008: the Born in Bradford (BiB) study in the UK [70], the Étude des Déterminants pré et postnatals du développement et de la santé de l'Enfant (EDEN) study in France [71], the INfancia y Medio Ambiente (INMA) cohort in Spain [72], the Kaunus cohort (KANC) in Lithuania [73], the Norwegian Mother, Father and Child Cohort Study (MoBa) [74], and the RHEA Mother Child Cohort study in Crete, Greece [75]. Within all HELIX children, a subset of 1,623 children participated in the follow-up clinical examination in their respective study centers between December 2013 and February 2016 [69]. During the follow-up examinations, mothers were interviewed with questionnaires and children (aged between 6 and 11 years) were examined by trained nurses according to standardized operating procedures. Peripheral blood and urine samples of the children were collected. The children included in the current study were those of European ancestry defined based on the genome-wide genetic data, and with telomere length measurements and DNA methylation profiles available ($N = 1001$). A flowchart of sample inclusion in HELIX and the current study is available in Fig. 1. The biological sample collection procedure is described in detail in Supplementary Method S1.

The HELIX study complies with the Declaration of Helsinki. Each of the six cohorts has received ethical approvals from the corresponding national ethical committees. Informed consent was signed by all participants at recruitment and at the follow-up visit for clinical examinations and biospecimen collection.

Measurements

Telomere lengths

Telomere length measurement in the HELIX children has been described previously [76]. Samples of buffy coat from the child peripheral blood were used for telomere length measurements. DNA extraction and sample quality control are described in detail in Supplementary Method S2. To ensure a uniform DNA input of 5 ng for each qPCR reaction, samples were diluted and checked using the Quant-iT™ PicoGreen® dsDNA Assay Kit (Life Technologies, Europe). Average relative telomere length was measured in the laboratory at the Centre for Environmental Sciences, Hasselt University, Belgium, using a modified quantitative real-time PCR (qPCR) protocol [77]. Telomere and single copy-gene reaction mixture and PCR cycles have been described previously [78] and can be found in Supplementary Method S3. All measurements were performed in triplicate on a 7900HT Fast Real-Time PCR System (Applied Biosystems) in a 384-well format. On each run, a 6-point serial dilution of pooled DNA was run to assess PCR efficiency, which ranged from 90 to 102%; as well as eight inter-run calibrators to account for the inter-run variability. The curves of qPCR for each sample were visually inspected and when technical problems were detected or triplicates showed

too high variability, samples were removed for further analysis. Relative telomere lengths were calculated using qBase software (Biogazelle, Zwijnaarde, Belgium). They were expressed as the ratio of telomere copy number to single-copy gene number (T/S). The coefficients of variation (CV) within triplicates of the telomere runs, single-copy gene runs, and T/S ratios were 0.84%, 0.43%, and 6.4%, respectively. Batch effects were regressed out and alogarithm with base 10 was taken ($\log_{10}$)for the T/S ratios.

Multi-omics profiles

Biological samples used for multi-omics measurements were shown in Supplementary Figure S1. Measurements of the different omics were performed using blood-derived specimens (buffy coat, whole blood or plasma) and urine samples. Details of the blood and urine sample processing can be found in Supplementary Method S1, and DNA and RNA extraction can be found in Supplementary Method S2. Genome-wide genetic variants (Infinium Global Screening Array, Illumina), blood DNA methylation (450 K, Illumina), blood gene expression (HTA v2.0, Affymetrix) and blood miRNA expression (SurePrint Human miRNA rel 21, Agilent) were assessed using micro-array chip-based technologies. Plasma proteins were measured using bead-based sandwich ELISA procedures (Luminex multiplex assays). Serum metabolites were measured using mass spectrometry (targeted LC-MS/MS metabolomic assays, Biocrates AbsoluteIDQ p180 kit), and urinary metabolites were measured using ^{1}H nuclear magnetic resonance (NMR) spectroscopy [25]. The omics assays and data pre-processing are described in detail in Supplementary Methods S4 – S10. The number of probes and sample size in each omics platform are summarized in Supplementary Table S1. Lists of biomarkers in plasma protein, serum and urinary metabolites assays are available in Supplementary Data S1 – S3, respectively. Beyond the general data pre-processing, we made further corrections for batch and sample quality as listed in Supplementary Table S13. Particularly, all blood-cell-based omics measurements were residualized on surrogate variables that captured the variations due to batch and cell type composition.

Using the processed DNA methylation data, we calculated the DNA methylation-based telomere length estimator (DNAmTL) [5], which is an epigenetic aging clock built on the DNA methylation level at 140 CpGs. In addition, based on the whole-genome genetic data, we estimated two polygenic scores (PRS) for telomere length, hereafter referred to as Li's PRS and Codd's PRS, via allelic scoring in PLINK (version 2.0) [79] with the "--score" method [80]. The weights given to the SNPs were the associations estimated by Li et al. [23] (52 SNPs, 43 out of which were available in HELIX data) for the Li's PRS, and that reported by Codd et al. [24] (197 SNPs, 124 out of which were available in HELIX data), for the Codd's PRS. SNPs that were not available in the processed HELIX whole-genome genetic data were weighted with zero. More details of the calculation can be found in Supplementary Method 11.

Key covariates and early-life aging risk factors

Information on child sex, age, and cohort study centers was collected in the HELIX examinations. The cohort- and ancestry-variation was captured by the first four principal components (PCs) of the genome-wide genetic data [81] (Supplementary Table S14). Blood cell type compositions were considered as key covariates, since telomere length was assessed as averaged measurements from multiple blood cell types in the sample. Blood cell proportions were estimated based on the raw methylation data using Houseman algorithm [82] and the Reinius reference panel [83] (Supplementary Method S5). In the current study, the key covariates used for model adjustments were child age, sex, the first four genetic PCs and the estimated blood cell compositions.

We considered the following aspects as potential risk factors for early-life aging in the HELIX children: gestational age, birth weight, maternal pre-pregnancy body-mass-index (BMI), maternal smoking status, maternal education level, parental smoking in the children's household, family affluence score, child BMI, child physical activity level and child Mediterranean Diet Quality Index (KIDMED) score.

Information on birth weight (kg), gestational age (weeks), child height and weight, maternal pre-pregnancy BMI, self-reported maternal smoking status and parental smoking in the household of the children, and self-reported maternal education level was collected during pregnancy or the follow-up examination in childhood. BMI was calculated by dividing weight (in kilos) by the squared height (in meters). For descriptive analysis, child BMI was categorized according to the CDC growth charts of sex-specific BMI-for-age percentile curves (https://www.cdc.gov/growthcharts/). Children with BMI less than the 5th percentile were "underweight", from the 5th to 85th percentiles were "normal", from the 85th to the 95th were "overweight", and those greater than the 95th percentile were "obese". For statistical modeling, child BMI was converted to age- and sex-adjusted z-scores (zBMI) using the international World Health Organization (WHO) reference curves [84]. Maternal pre-pregnancy BMI was grouped according to WHO categories for underweight (< 18.5 kg/m^2), normal (18.5–24.9 kg/m^2), overweight (25–29.9 kg/m^2) and obese (≥ 30 kg/m^2). The maternal smoking status during pregnancy was categorized into "never smoker" if a mother never smoked during pregnancy, "non-sustained smoker"

if a mother smoked only at the beginning of pregnancy, or "sustained smoker during pregnancy". The children's exposure to parental smoking in the household was classified as "neither", "one" or "both". The educational level of the mothers was defined as having the highest education level of primary school, secondary school or having a university degree or higher.

In addition, the physical activity of the children was measured by moderate-to-vigorous physical activity (MVPA) level, based on the physical activity questionnaire developed by the HELIX research group, defined as the amount of time spent by the children in physical activities with intensity above three metabolic equivalent tasks and is expressed in units of min/day.

Family affluence score [85] was defined as a composite score calculated based on the responses to the next four items: [1] Does your family own a car, van, or truck [2]? Do you have your own bedroom for yourself [3]? During the past 12 months, how many times did you travel away on holiday with your family [4]? How many computers does your family own? A score of 0–2 indicates low affluence, a score of 3–5 indicates middle affluence, and a score of 6–9 indicates high affluence.

During the follow-up examination, a semi-quantitative food-frequency questionnaire (FFQ) covering the child's habitual diet was filled in by the parent to collect information on the children's habitual diet in the past years. The FFQ was developed by the HELIX research group, and was translated and applied to all cohorts. The KIDMED index score [86] was calculated as a measure of the adequacy of Mediterranean dietary patterns in children, with higher scores reflecting greater adherence to a Mediterranean diet.

Statistical analyses

The overall statistical workflow is presented in Fig. 2 and was based on two approaches. Approach I was the omics-wide association analyses, where each individual feature within each omics layer was assessed for its association with telomere length through multiple linear regression. In Approach II, an integrative multi-omics analysis was performed with a sparse supervised method against telomere length measures. The two approaches are described in detail below.

Approach I: omics-wide association analyses of individual omics features

All available omics features were first analyzed against the measured telomere length in an association analysis (Fig. 2A, Approach I), where the child telomere length was regressed on each individual omics feature. For genome-wide genetic data, the association analysis was performed in PLINK2 [79, 87] with the "--glm" method, excluding variants with minor allele frequency (MAF) lower than 0.01 and minor allele count (MAC) lower than 5. For the other omics, R version 4.2.2 [88] was used to fit robust linear regression in R with telomere length as the dependent variable. Multiple testing was corrected using Bonferroni method within each omics based on the specific number of hypothesis tests in the omics. All regression models were adjusted for key covariates as described above (child age, sex, the first four genetic PCs, and the estimated blood cell compositions).

Approach II: multi-omics analysis

Multi-omics integrative analysis Features across six omics (DNA methylation, gene expression, miRNA expression, plasma proteins, serum metabolites, and urine metabolites) were analyzed using a supervised multi-block integration method to identify the multi-omics signature of childhood telomere length (Fig. 2A, Approach II). Considering the purpose of feature selection from a high-dimensional collinear feature space within a multi-omics context, a multi-block sparse partial least squares (multi-block sPLS) method implemented in the mixOmics package [89] was applied. The parameters estimation utilizes the Non-linear Iterative PArtial Least Squares (NIPALS) which performs singular value decomposition (SVD) with missing data, without the need to exclude or impute the missing values. A least absolute shrinkage and selection operator (LASSO) penalty was applied simultaneously on the parameters to improve the feature selection for better interpretability [90].

As the input for the predictor blocks in the multi-block sPLS regression, features from each omics were used. A prior pre-selection of the CpGs and gene transcripts, where the data dimension was relatively higher, was conducted based on literature and publicly accessible databases (Fig. 2B). The purposes of this pre-selection were to denoise the omics data, to improve between-omics intercorrelation and to reduce the model's overfitting to the data. A detailed description of the literature-based feature pre-selection can be found in Supplementary Method S12. This literature-based pre-selection resulted in 1,848 selected out of 386,518 CpGs and 384 out of 20,315 gene expression transcript clusters. For the other omics (miRNA expression, plasma proteins, serum metabolites, and urine metabolites), all available features were used as inputs.

On the other side of the model, the measured telomere length, the two polygenic scores estimated based on genetic data, Li's PRS and Codd's PRS, and DNAmTL estimated based on DNA methylation data, as described above, were used to form a four-variable phenotype block (Y block). The aim of adding the estimated scores to the measured telomere length was to increase the variation in the phenotype, assisting the supervised learning from

the high-dimension omics blocks. It was not a concern to have DNA methylation-based information on both sides of the model, since only thirteen out of the 140 CpGs used to estimate DNAmTL were in common with the 1,848 pre-selected CpGs.

As the HELIX project was conducted across six different European countries, between-population differences resulted in data heterogeneity in molecular measurements. In order to minimize the variance or covariance in the omics data that is due to the population characteristics, the omics features were denoised by the most general population characteristics before running the multi-omics analysis, where each of the features in the input of the predictor blocks, as well as the four telomere length measures in the outcome block, were residualized for child sex, age and the first four genetic PCs (representing both cohorts and ethnicity effects). The telomere length measures were additionally residualized for the estimated blood cell compositions (Supplementary Table S13).

The following sequential steps were undertaken for the multi-omics integration. In step 1, we selected the number of components in a multi-omics context without the LASSO penalty that maximized the covariance between the components from the omics blocks and the components from the response (telomere) block. Two measures of covariance were used: the maximum of the squared entries in the covariance matrices (MSC), and the square root of the mean of the squared entries in the covariance matrices (RMSC). In step 2, based on the selected number of components, the number of features was selected within each omics block using the tuning function for sPLS, where a 5-fold cross-validation and 50 repeats of sampling were used for the tuning process, and correlation between the predicted and the actual components was used as the measure of accuracy. In step 3, the final multiblock sPLS model was fitted based on the selected number of components and number of features from the previous two steps. A "regression" mode was specified so that an asymmetric deflation of the response matrix was used. In both steps 1 and 3, the between-block covariance structure was defined using a design matrix. This matrix was constructed as follows: first, for each pair of omics blocks, canonical correlation analysis was performed; second, the first component from each omics block was extracted, and the covariance between these components was calculated; third, this covariance value was then used as the off-diagonal entry in the design matrix for the multi-omics model. As the output of the final model, we obtained the component scores, representing the projection of the original omics features in the latent component space, as well as the corresponding loadings, which measure the correlation between the components and the features.

Downstream analysis of the latent components as multi-omics signatures of telomere length Based on the final multi-block sPLS model, we evaluated the correlations of the latent component scores to telomere length measures to further confirm which components could be used to better interpret the multi-omics signatures of childhood telomere length. From the most correlated components, omics features with an absolute loading > 0.10 were used for the interpretation of the signatures, via pathway enrichment analyses as depicted in Supplementary Figure S9. Briefly, CpGs, transcripts (including gene expression and miRNAs) and proteins were annotated to gene symbols and EntrezIDs. In the meantime, we used the miRWalk (version 3) [91] online database to search for gene targets of the miRNA molecules, while filters were applied to select the targets predicted in all of TargetScan and miRDB with a score no less than 0.95 and validated in miRTarBase. The genes annotated to CpGs, transcripts, and proteins, as well as the target genes, were analyzed through a pathway enrichment analysis, respectively, with reference databases of Reactome, Gene Ontology (GO, including biological process and molecular function) and KEGG, using the R package clusterProfiler [92]. The serum and urinary metabolites were converted into ChEBI IDs, KEGG IDs, and PubChem CIDs, which were used in a metabolite enrichment analysis with reference databases of KEGG, SMPDB, disease signatures in blood and urine, using the MetaboAnalyst 5.0 [93] webtool. Multiple testing of the pathways was controlled with the Benjamini-Hochberg FDR. A pathway was considered significant if the number of query genes or metabolites involved was at least three, and its ratio to the total number of genes or metabolites in the pathway was at least 0.10.

The latent component scores were in parallel used to test the hypothesis that early-life aging risk factors could differentiate the multi-omics signatures of telomere length. This was done by evaluating the association between the component scores and the ten potential risk factors defined above (gestational age, birth weight, maternal pre-pregnancy BMI, maternal smoking status, maternal education level, parental smoking in the children's household, family affluence score, child zBMI, child physical activity level and child Mediterranean diet score). For each component score in each omics block, a multiple regression model was fitted on all ten risk factors, such that the association estimate for each risk factor was conditional on all the others. The global significance level was controlled using Bonferroni correction, where the number of tests was calculated as the number of all combinations of six predictor blocks, K components in each block, and ten risk factors (6 x K x 10 tests in total).

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12864-025-11209-5.

Supplementary Material 1

Supplementary Material 2

Acknowledgements
We would like to thank all the families for their generous contribution.

Author contributions
M. Vrijheid, L.M. and M.B. coordinated the HELIX project. M. Vrijheid, L.M., J.W., R.R.C.M., J.L., R.S., M. Vafeiadi, R.G., K.B.G. are the PIs of the cohorts, or participated in sample and exposure data acquisition. A.C., M.B., G.E., E.S., H.K. participated in omics data acquisition and quality control. C.W., D.S.M., T.S.N. and M.B. conceptualised the current study. C.W. performed data processing and statistical analysis within the current study. D.S.M., T.S.N., M.B., R.A., M.P., A.A.-R., D.P.-S. and J.R.G. suggested on the statistical analyses. C.W. wrote the first draft of the manuscript. D.S.M., T.S.N., M.B., R.A., M.P., AA-R, L.M., R.R.C.M., J.L. and K.B.G. contributed to reviewing and editing the manuscript. All authors read and approved the manuscript. All authors interpreted the results, commented on successive drafts of the manuscript and approved the final version.

Funding
The study has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 874583 (ATHLETE project). Data were collected as part of the European Community's Seventh Framework Programme (FP7/2007 – 206) under grant agreement no 308333 (HELIX project). The genotyping was supported by the projects PI17/01225 and PI17/01935, funded by the Instituto de Salud Carlos III (ISCIII) and co-funded by the European Union (ERDF, "A way to make Europe") and the Centro Nacional de Genotipado-CEGEN (PRB2-ISCIII). BiB received core infrastructure funding from the Wellcome Trust (WT101597MA) and a joint grant from the UK Medical Research Council (MRC) and the Economic and Social Science Research Council (ESRC) (MR/N024397/1). INMA data collections were supported by grants from the Instituto de Salud Carlos III, CIBERESP, and the Generalitat de Catalunya-CIRIT. KANC was funded by a grant from the Lithuanian Agency for Science Innovation and Technology (6-04-2014_31V-66). The Norwegian Mother, Father and Child Cohort Study is supported by the Norwegian Ministry of Health and Care Services and the Ministry of Education and Research. The Rhea project was financially supported by European projects (EU FP6-2003-Food-3-NewGeneris, EU FP6. STREP Hiwate, EU FP7 ENV.2007.1.2.2.2. Project No 211250 Escape, EU FP7-2008-ENV-1.2.1.4 Envirogenomarkers, EU FP7-HEALTH-2009- single stage CHICOS, EU FP7 ENV.2008.1.2.1.6. Proposal No 226285 ENRIECO, EU- FP7-HEALTH-2012 Proposal No 308333 HELIX), and the Greek Ministry of Health (Program of Prevention of obesity and neurodevelopmental disorders in preschool children, in Heraklion district, Crete, Greece: 2011–2014; "Rhea Plus": Primary Prevention Program of Environmental Risk Factors for Reproductive Health, and Child Health: 2012-15). ISGlobal acknowledges support from the Spanish Ministry of Science and Innovation through the "Centro de Excelencia Severo Ochoa 2019–2023" Program (CEX2018-000806-S), and support from the Generalitat de Catalunya through the CERCA Program. L.M. is funded by a Juan de la Cierva-Incorporación fellowship (IJC2018-035394-I) awarded by the Spanish Ministerio de Economía, Industria y Competitividad. The CRG/UPF Proteomics Unit is part of the Spanish Infrastructure for Omics Technologies (ICTS OmicsTech). D.S.M. holds a postdoctoral grant from the Research Foundations Flanders (FWO grants 12×9623 N).

Data availability
The raw data supporting the current study are available from the corresponding author on request subject to ethical and legislative review. The "HELIX Data External Data Request Procedures" are available with the data inventory in this website: http://www.projecthelix.eu/data-inventory. The R code used for all the statistical analysis and data visualization is available upon request.

Declarations

Ethics approval and consent to participate
The HELIX study complies with the Declaration of Helsinki. All six cohorts existed for several years before HELIX started, and had undergone the required evaluation by national ethics committees: EDEN received approval from the ethics committee (CCPPRB) of Kremlin Bicêtre and from CNIL (Commission Nationale Informatique et Liberté), the French data privacy institution; BiB received ethics approval from the Bradford Research Ethics Committee; INMA obtained the approval of the ethics committee of each involved hospital or health center; the research protocol of KANC was approved by the Lithuanian Bioethics Committee; MoBa received approval from a Norwegian regional committee for medical and health research ethics; and the ethics committee of the university hospital at Heraklion approved the study protocols of RHEA. An informed consent has been signed by all participants at recruitment and at the follow-up visit for clinical examinations and biospecimen collection. Each cohort also confirmed that relevant informed consent and approval were in place for the secondary use of data from pre-existing data. The work in HELIX was covered by new ethics approvals in each country. The HELIX project received ethical approvals from the Comité Ético de investigación Clínica Parc de Salut MAR. At follow-up enrolment in the HELIX subcohort and panel studies, participants were asked to sign an informed consent for clinical examination and biospecimen collection and analysis.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Author details
[1]Centre for Environmental Health, Hasselt University, Hasselt, Belgium
[2]ISGlobal, Institute for Global Health, Barcelona, Spain
[3]Universitat Pompeu Fabra (UPF), Barcelona, Spain
[4]Consorcio de Investigación Biomédica en Red de Epidemiología y Salud Pública (CIBERESP), Madrid, Spain
[5]Bradford Institute for Health Research, Bradford Teaching Hospitals NHS Foundation Trust, Bradford, UK
[6]University Grenoble Alpes, Inserm U 1209, CNRS UMR 5309, Team of Environmental Epidemiology applied to Development and Respiratory Health, Institute for Advanced Biosciences, Grenoble, France
[7]Team of Environmental Epidemiology applied to Reproduction and Respiratory Health, Institute for Advanced Biosciences (IAB), Inserm, CNRS, Université Grenoble Alpes, Grenoble, France
[8]Department of Social Medicine, Faculty of Medicine, University of Crete, Heraklion, Greece
[9]Department of Population and Public Health Sciences, University of Southern California, Los Angeles, CA, USA
[10]Department of Environmental Sciences, Vytautas Magnus University, Kaunas, Lithuania
[11]Division of Climate and Environmental Health, Norwegian Institute of Public Health, Oslo, Norway
[12]Division of Systems Medicine, Department of Metabolism, Digestion and Reproduction, Imperial College London, London, UK
[13]Cancer Metabolism & Systems Toxicology Group, Division of Cancer, Department of Surgery & Cancer, Imperial College London, London, UK
[14]Center for Genomic Regulation (CRG), Barcelona Institute of Science and Technology, Barcelona, Spain
[15]Medicine Genomics Group, Centro de Investigación Biomédica en Red Enfermedades Raras (CIBERER-ISCIII), University of Santiago de Compostela, CIMUS, Santiago de Compostela, Spain
[16]Galician Foundation of Genomic Medicine, Instituto de Investigación Sanitaria de Santiago de Compostela (IDIS), Servicio Gallego de Salud (SERGAS), Santiago de Compostela, Galicia, Spain
[17]CIBER in Epidemiology and Public Health (CIBERESP), Barcelona, Spain
[18]Department of Biomedical Sciences, Institute of Neuroscience, University of Barcelona, Barcelona, Spain
[19]CIBEROBN (CIBER Physiopathology of Obesity and Nutrition), Instituto de Salud Carlos III, Madrid, Spain
[20]Department of Public Health and Primary Care, University of Leuven, Leuven, Belgium

Received: 4 July 2024 / Accepted: 2 January 2025
Published online: 27 January 2025

References

1. Shammas MA. Telomeres, lifestyle, cancer, and aging. Curr Opin Clin Nutr Metab Care. 2011;14(1):28–34.
2. McHugh D, Gil J. Senescence and aging: causes, consequences, and therapeutic avenues. J Cell Biol. 2018;217(1):65–77.
3. Fasching CL. Telomere length measurement as a clinical biomarker of aging and disease. Crit Rev Clin Lab Sci. 2018;55(7):443–65.
4. Codd V, Mangino M, van der Harst P, Braund PS, Kaiser M, Beveridge AJ, et al. Common variants near TERC are associated with mean telomere length. Nat Genet. 2010;42(3):197–9.
5. Lu AT, Seeboth A, Tsai PC, Sun D, Quach A, Reiner AP, et al. DNA methylation-based estimator of telomere length. Aging. 2019;11(16):5895–923.
6. Lopizzo N, Tosato S, Begni V, Tomassi S, Cattane N, Barcella M, et al. Transcriptomic analyses and leukocyte telomere length measurement in subjects exposed to severe recent stressful life events. Translational Psychiatry. 2017;7(2):e1042–e.
7. Zierer J, Kastenmüller G, Suhre K, Gieger C, Codd V, Tsai PC, et al. Metabolomics profiling reveals novel markers for leukocyte telomere length. Aging. 2016;8(1):77–94.
8. Reimann B, Martens DS, Wang C, Ghantous A, Herceg Z, Plusquin M, et al. Interrelationships and determinants of aging biomarkers in cord blood. J Transl Med. 2022;20(1):353.
9. Robinson O, Lau CE, Joo S, Andrusaityte S, Borras E, de Prado-Bert P, et al. Associations of four biological age markers with child development: a multi-omic analysis in the European HELIX cohort. eLife. 2023;12:e85104.
10. Etzel RA. The special vulnerability of children. Int J Hyg Environ Health. 2020;227:113516.
11. Bijnens EM, Zeegers MP, Derom C, Martens DS, Gielen M, Hageman GJ, et al. Telomere tracking from birth to adulthood and residential traffic exposure. BMC Med. 2017;15(1):205.
12. 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.
13. de Prado-Bert P, Ruiz-Arenas C, Vives-Usano M, Andrusaityte S, Cadiou S, Carracedo Á, et al. The early-life exposome and epigenetic age acceleration in children. Environ Int. 2021;155:106683.
14. Clemente DBP, Maitre L, Bustamante M, Chatzi L, Roumeliotaki T, Fossati S, et al. Obesity is associated with shorter telomeres in 8 year-old children. Sci Rep. 2019;9(1):18739.
15. Raftopoulou C, Paltoglou G, Charmandari E. Association between Telomere Length and Pediatric Obesity: a systematic review. Nutrients. 2022;14(6).
16. Azcona-Sanjulian MC. Telomere length and Pediatric obesity: a review. Genes (Basel). 2021;12(6).
17. Colich NL, Rosen ML, Williams ES, McLaughlin KA. Biological aging in childhood and adolescence following experiences of threat and deprivation: a systematic review and meta-analysis. Psychol Bull. 2020;146(9):721–64.
18. Alexeeff SE, Schaefer CA, Kvale MN, Shan J, Blackburn EH, Risch N, et al. Telomere length and socioeconomic status at neighborhood and individual levels among 80,000 adults in the Genetic Epidemiology Research on Adult Health and Aging cohort. Environ Epidemiol. 2019;3(3):e049.
19. Balan E, Decottignies A, Deldicque L. Physical activity and Nutrition: two promising strategies for Telomere maintenance? Nutrients. 2018;10(12).
20. Masi S, Nightingale CM, Day IN, Guthrie P, Rumley A, Lowe GD, et al. Inflammation and not cardiovascular risk factors is associated with short leukocyte telomere length in 13- to 16-year-old adolescents. Arterioscler Thromb Vasc Biol. 2012;32(8):2029–34.
21. Barnes RP, Fouquerel E, Opresko PL. The impact of oxidative DNA damage and stress on telomere homeostasis. Mech Ageing Dev. 2019;177:37–45.
22. Maitre L, Bustamante M, Hernández-Ferrer C, Thiel D, Lau C-HE, Siskos AP, et al. Multi-omics signatures of the human early life exposome. Nat Commun. 2022;13(1):7024.
23. Li C, Stoma S, Lotta LA, Warner S, Albrecht E, Allione A, et al. Genome-Wide Association Analysis in humans links Nucleotide metabolism to leukocyte telomere length. Am J Hum Genet. 2020;106(3):389–404.
24. Codd V, Wang Q, Allara E, Musicha C, Kaptoge S, Stoma S, et al. Polygenic basis and biomedical consequences of telomere length variation. Nat Genet. 2021;53(10):1425–33.
25. Lau C-HE, Siskos AP, Maitre L, Robinson O, Athersuch TJ, Want EJ, et al. Determinants of the urinary and serum metabolome in children from six European populations. BMC Med. 2018;16(1):202.
26. Kichaev G, Bhatia G, Loh PR, Gazal S, Burch K, Freund MK, et al. Leveraging Polygenic Functional Enrichment to improve GWAS Power. Am J Hum Genet. 2019;104(1):65–75.
27. Marques E, Englund JI, Tervonen TA, Virkunen E, Laakso M, Myllynen M, et al. Par6G suppresses cell proliferation and is targeted by loss-of-function mutations in multiple cancers. Oncogene. 2016;35(11):1386–98.
28. Dormoy V, Tormanen K, Sütterlin C. Par6γ is at the mother centriole and controls centrosomal protein composition through a Par6α-dependent pathway. J Cell Sci. 2013;126(Pt 3):860–70.
29. Kim HS, Jang S, Kim J. Genome-wide integrative transcriptional profiling identifies Age-Associated signatures in Dogs. Genes (Basel). 2023;14(6).
30. Dos Santos RS, Marroqui L, Velayos T, Olazagoitia-Garmendia A, Jauregi-Miguel A, Castellanos-Rubio A, et al. DEXI, a candidate gene for type 1 diabetes, modulates rat and human pancreatic beta cell inflammation via regulation of the type I IFN/STAT signalling pathway. Diabetologia. 2019;62(3):459–72.
31. Tomlinson MJt, Pitsillides A, Pickin R, Mika M, Keene KL, Hou X, et al. Fine mapping and functional studies of risk variants for type 1 diabetes at chromosome 16p13.13. Diabetes. 2014;63(12):4360–8.
32. Delgado DA, Zhang C, Chen LS, Gao J, Roy S, Shinkle J, et al. Genome-wide association study of telomere length among South asians identifies a second RTEL1 association signal. J Med Genet. 2018;55(1):64–71.
33. Dorajoo R, Chang X, Gurung RL, Li Z, Wang L, Wang R, et al. Loci for human leukocyte telomere length in the Singaporean Chinese population and trans-ethnic genetic studies. Nat Commun. 2019;10(1):2491.
34. Senchenko VN, Krasnov GS, Dmitriev AA, Kudryavtseva AV, Anedchenko EA, Braga EA, et al. Differential expression of CHL1 gene during development of major human cancers. PLoS ONE. 2011;6(3):e15612.
35. Mihaylova MM, Shaw RJ. The AMPK signalling pathway coordinates cell growth, autophagy and metabolism. Nat Cell Biol. 2011;13(9):1016–23.
36. Zou Z, Tao T, Li H, Zhu X. mTOR signaling pathway and mTOR inhibitors in cancer: progress and challenges. Cell Bioscience. 2020;10(1):31.
37. Pearce EE, Horvath S, Katta S, Dagnall C, Aubert G, Hicks BD, et al. DNA-methylation-based telomere length estimator: comparisons with measurements from flow FISH and qPCR. Aging. 2021;13(11):14675–86.
38. Liu D, Aziz NA, Imtiaz MA, Pehlivan G, Breteler MMB. Associations of measured and genetically predicted leukocyte telomere length with vascular phenotypes: a population-based study. GeroScience. 2024;46(2):1947–70.
39. Eisenberg DTA, Ryan CP, Lee NR, Carba DB, MacIsaac JL, Dever K, et al. DNA methylation-based estimators of telomere length show low correspondence with paternal age at conception and other measures of external validity of telomere length. GeroScience. 2024;46(4):3957–69.
40. Gielen M, Hageman GJ, Antoniou EE, Nordfjall K, Mangino M, Balasubramanyam M, et al. Body mass index is negatively associated with telomere length: a collaborative cross-sectional meta-analysis of 87 observational studies. Am J Clin Nutr. 2018;108(3):453–75.
41. Zhang Y, Xu Z, Yang Y, Cao S, Lyu S, Duan W. Association between Weight Change and leukocyte telomere length in U.S. adults. Front Endocrinol (Lausanne). 2021;12.
42. Mundstock E, Sarria EE, Zatti H, Mattos Louzada F, Kich Grun L, Herbert Jones M, et al. Effect of obesity on telomere length: systematic review and meta-analysis. Obesity. 2015;23(11):2165–74.
43. Du M, Prescott J, Cornelis MC, Hankinson SE, Giovannucci E, Kraft P, et al. Genetic predisposition to higher body Mass Index or Type 2 diabetes and leukocyte telomere length in the nurses' Health Study. PLoS ONE. 2013;8(2):e52240.
44. Wulaningsih W, Watkins J, Matsuguchi T, Hardy R. Investigating the associations between adiposity, life course overweight trajectories, and telomere length. Aging. 2016;8(11):2689–701.
45. Welendorf C, Nicoletti CF, Pinhel MAS, Noronha NY, de Paula BMF, Nonino CB. Obesity, weight loss, and influence on telomere length: new insights for personalized nutrition. Nutrition. 2019;66:115–21.
46. Dogan F, Forsyth NR. Telomerase regulation: a role for epigenetics. Cancers (Basel). 2021;13(6).

47. Caslini C. Transcriptional regulation of telomeric non-coding RNA: implications on telomere biology, replicative senescence and cancer. RNA Biol. 2010;7(1):18–22.
48. Rangel-Huerta OD, Pastor-Villaescusa B, Gil A. Are we close to defining a metabolomic signature of human obesity? A systematic review of metabolomics studies. Metabolomics. 2019;15(6):93.
49. Kordinas V, Ioannidis A, Chatzipanagiotou S. The Telomere/Telomerase System in Chronic Inflammatory diseases. Cause or Effect? Genes (Basel). 2016;7(9).
50. Jager J, Grémeaux T, Cormont M, Le Marchand-Brustel Y, Tanti JF. Interleukin-1beta-induced insulin resistance in adipocytes through down-regulation of insulin receptor substrate-1 expression. Endocrinology. 2007;148(1):241–51.
51. Kahn BB, Flier JS. Obesity and insulin resistance. J Clin Invest. 2000;106(4):473–81.
52. Wondmkun YT, Obesity. Insulin resistance, and type 2 diabetes: associations and therapeutic implications. Diabetes Metab Syndr Obes. 2020;13:3611–6.
53. Scarabino D, Peconi M, Broggio E, Gambina G, Maggi E, Armeli F, et al. Relationship between proinflammatory cytokines (Il-1beta, Il-18) and leukocyte telomere length in mild cognitive impairment and Alzheimer's disease. Exp Gerontol. 2020;136:110945.
54. Eder K, Baffy N, Falus A, Fulop AK. The major inflammatory mediator interleukin-6 and obesity. Inflamm Res. 2009;58(11):727–36.
55. Savale L, Chaouat A, Bastuji-Garin S, Marcos E, Boyer L, Maitre B, et al. Shortened telomeres in circulating leukocytes of patients with chronic obstructive pulmonary disease. Am J Respir Crit Care Med. 2009;179(7):566–71.
56. Bagheri M, Djazayery A, Farzadfar F, Qi L, Yekaninejad MS, Aslibekyan S, et al. Plasma metabolomic profiling of amino acids and polar lipids in Iranian obese adults. Lipids Health Dis. 2019;18(1):94.
57. Chang R, Trushina E, Zhu K, Zaidi SSA, Lau BM, Kueider-Paisley A, et al. Predictive metabolic networks reveal sex- and APOE genotype-specific metabolic signatures and drivers for precision medicine in Alzheimer's disease. Alzheimer's Dement. 2023;19(2):518–31.
58. Moaddel R, Fabbri E, Khadeer MA, Carlson OD, Gonzalez-Freire M, Zhang P, et al. Plasma biomarkers of poor muscle quality in older men and women from the Baltimore Longitudinal Study of Aging. Journals Gerontology: Ser A. 2016;71(10):1266–72.
59. Wahl S, Yu Z, Kleber M, Singmann P, Holzapfel C, He Y, et al. Childhood obesity is associated with changes in the serum metabolite profile. Obes Facts. 2012;5(5):660–70.
60. Appiah-Amponsah E, Shanaiah N, Nagana Gowda GA, Owusu-Sarfo K, Ye T, Raftery D. Identification of 4-deoxythreonic acid present in human urine using HPLC and NMR techniques. J Pharm Biomed Anal. 2009;50(5):878–85.
61. Elliott P, Posma JM, Chan Q, Garcia-Perez I, Wijeyesekera A, Bictash M, et al. Urinary metabolic signatures of human adiposity. Sci Transl Med. 2015;7(285):285ra62.
62. Prabhakar A, Quach A, Zhang H, Terrera M, Jackemeyer D, Xian X, et al. Acetone as biomarker for ketosis buildup capability - a study in healthy individuals under combined high fat and starvation diets. Nutr J. 2015;14(1):41.
63. Soto-Mota A, Norwitz NG, Evans R, Clarke K, Barber TM. Exogenous ketosis in patients with type 2 diabetes: Safety, tolerability and effect on glycaemic control. Endocrinol Diabetes Metab. 2021;4(3):e00264.
64. Tsoukalas D, Fragoulakis V, Papakonstantinou E, Antonaki M, Vozikis A, Tsatsakis A et al. Prediction of Autoimmune diseases by targeted metabolomic assay of urinary Organic acids. Metabolites. 2020;10(12).
65. Giskeødegård GF, Davies SK, Revell VL, Keun H, Skene DJ. Diurnal rhythms in the human urine metabolome during sleep and total sleep deprivation. Sci Rep. 2015;5:14843.
66. Bingham SA. Urine Nitrogen as a biomarker for the validation of Dietary protein intake. J Nutr. 2003;133(3):S921–4.
67. Leidy HJ, Clifton PM, Astrup A, Wycherley TP, Westerterp-Plantenga MS, Luscombe-Marsh ND, et al. The role of protein in weight loss and maintenance234. Am J Clin Nutr. 2015;101(6):S1320–9.
68. Vrijheid M, Slama R, Robinson O, Chatzi L, Coen M, van den Hazel P, et al. The human early-life exposome (HELIX): project rationale and design. Environ Health Perspect. 2014;122(6):535–44.
69. Maitre L, de Bont J, Casas M, Robinson O, Aasvang GM, Agier L, et al. Human early life exposome (HELIX) study: a European population-based exposome cohort. BMJ Open. 2018;8(9):e021311.
70. Wright J, Small N, Raynor P, Tuffnell D, Bhopal R, Cameron N, et al. Cohort Profile: the born in Bradford multi-ethnic family cohort study. Int J Epidemiol. 2013;42(4):978–91.
71. Heude B, Forhan A, Slama R, Douhaud L, Bedel S, Saurel-Cubizolles MJ, et al. Cohort Profile: the EDEN mother-child cohort on the prenatal and early postnatal determinants of child health and development. Int J Epidemiol. 2016;45(2):353–63.
72. Guxens M, Ballester F, Espada M, Fernández MF, Grimalt JO, Ibarluzea J, et al. Cohort Profile: the INMA–INfancia y Medio Ambiente–(Environment and Childhood) Project. Int J Epidemiol. 2012;41(4):930–40.
73. Grazuleviciene R, Danileviciute A, Nadisauskiene R, Vencloviene J. Maternal smoking, GSTM1 and GSTT1 polymorphism and susceptibility to adverse pregnancy outcomes. Int J Environ Res Public Health. 2009;6(3):1282–97.
74. Magnus P, Birke C, Vejrup K, Haugan A, Alsaker E, Daltveit AK, et al. Cohort Profile Update: the Norwegian mother and child Cohort Study (MoBa). Int J Epidemiol. 2016;45(2):382–8.
75. Chatzi L, Leventakou V, Vafeiadi M, Koutra K, Roumeliotaki T, Chalkiadaki G, et al. Cohort Profile: the Mother-Child Cohort in Crete, Greece (Rhea Study). Int J Epidemiol. 2017;46(5):1392–k3.
76. Clemente DBP, Vrijheid M, Martens DS, Bustamante M, Chatzi L, Danileviciute A, et al. Prenatal and Childhood Traffic-Related Air Pollution exposure and telomere length in European Children: the HELIX Project. Environ Health Perspect. 2019;127(8):087001.
77. Cawthon RM. Telomere length measurement by a novel monochrome multiplex quantitative PCR method. Nucleic Acids Res. 2009;37(3):e21.
78. 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(1):148.
79. Shaun Purcell CC. PLINK 2.0. 2023.
80. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet. 2007;81(3):559–75.
81. Calvo-Serra B, Maitre L, Lau C-HE, Siskos AP, Gützkow KB, Andrušaitytė S, et al. Urinary metabolite quantitative trait loci in children and their interaction with dietary factors. Hum Mol Genet. 2020;29(23):3830–44.
82. Houseman EA, Accomando WP, Koestler DC, Christensen BC, Marsit CJ, Nelson HH, et al. DNA methylation arrays as surrogate measures of cell mixture distribution. BMC Bioinformatics. 2012;13(1):86.
83. Reinius LE, Acevedo N, Joerink M, Pershagen G, Dahlén SE, Greco D, et al. Differential DNA methylation in purified human blood cells: implications for cell lineage and studies on disease susceptibility. PLoS ONE. 2012;7(7):e41361.
84. de Onis M, Onyango AW, Borghi E, Siyam A, Nishida C, Siekmann J. Development of a WHO growth reference for school-aged children and adolescents. Bull World Health Organ. 2007;85(9):660–7.
85. Boyce W, Torsheim T, Currie C, Zambon A. The Family Affluence Scale as a measure of National Wealth: validation of an adolescent self-report measure. Soc Indic Res. 2006;78(3):473–87.
86. Serra-Majem L, Ribas L, Ngo J, Ortega RM, García A, Pérez-Rodrigo C, et al. Food, youth and the Mediterranean diet in Spain. Development of KIDMED, Mediterranean Diet Quality Index in children and adolescents. Public Health Nutr. 2004;7(7):931–5.
87. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. Gigascience. 2015;4:7.
88. R Core Team. R: a language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2022.
89. Rohart F, Gautier B, Singh A, Lê Cao K-A, mixOmics. An R package for 'omics feature selection and multiple data integration. PLoS Comput Biol. 2017;13(11):e1005752.
90. Lê Cao K, Rossouw D, Robert-Granié C, Besse P. A sparse PLS for variable selection when integrating omics data. Stat Appl Genet Mol Biol. 2008;7(1):35.
91. Sticht C, De La Torre C, Parveen A, Gretz N. miRWalk: an online resource for prediction of microRNA binding sites. PLoS ONE. 2018;13(10):e0206239.
92. Yu G, Wang L-G, Han Y, He Q-Y. clusterProfiler: an R Package for comparing Biological themes among Gene clusters. OMICS. 2012;16(5):284–7.
93. Xia J, Wishart DS. Web-based inference of biological patterns, functions and pathways from metabolomic data using MetaboAnalyst. Nat Protoc. 2011;6(6):743–60.

Publisher's note