Author/s:
Fiorito, G;McCrory, C;Robinson, O;Carmeli, C;Rosales, CO;Zhang, Y;Colicino, E;Dugué, PA;Artaud, F;McKay, GJ;Jeong, A;Mishra, PP;Nøst, TH;Krogh, V;Panico, S;Sacerdote, C;Tumino, R;Palli, D;Matullo, G;Guarrera, S;Gandini, M;Bochud, M;Dermitzakis, E;Muka, T;Schwartz, J;Vokonas, PS;Just, A;Hodge, AM;Giles, GG;Southey, MC;Hurme, MA;Young, I;McKnight, AJ;Kunze, S;Waldenberger, M;Peters, A;Schwettmann, L;Lund, E;Baccarelli, A;Milne, RL;Kenny, RA;Elbaz, A;Brenner, H;Kee, F;Voortman, T;Probst-Hensch, N;Lehtimäki, T;Elliott, P;Stringhini, S;Vineis, P;Polidoro, S;Alberts, J;Alenius, H;Avendano, M;Baltar, V;Bartley, M;Barros, H;Bellone, M;Berger, E;Blane, D;Candiani, G;Carra, L;Castagné, R;Chadeau-Hyam, M;Cima, S;Clavel-Chapelon, F;Costa, G;Courtin, E;Delpierre, C;D'Errico, A;Dermitzakis, M;Elovainio, M;Fagherazzi, G;Fraga, S;Garès, V;Gerbouin-Rerolle, P;Goldberg, M;Greco, D;Guessous, I;Haba-Rubio, J;Heinzer, R;Joost, S;Karimi, M;Kelly-Irving, M;Kähönen, M;Karisola, P;Khenissi, L;Kivimaki, M;Laine, J;Lang, T;Laurent, A;Layte, R;Lepage, B;Lorsch, D;MacGuire, F;Machell, G;Mackenbach, J;Marmot, M;de Mestral, C

Title:
Socioeconomic position, lifestyle habits and biomarkers of epigenetic aging: A multi-cohort analysis

Date:
2019-04-15

Citation:
Fiorito, G., McCrory, C., Robinson, O., Carmeli, C., Rosales, C. O., Zhang, Y., Colicino, E., Dugué, P. A., Artaud, F., McKay, G. J., Jeong, A., Mishra, P. P., Nøst, T. H., Krogh, V., Panico, S., Sacerdote, C., Tumino, R., Palli, D., Matullo, G. ,... de Mestral, C. (2019). Socioeconomic position, lifestyle habits and biomarkers of epigenetic aging: A multi-cohort analysis. Aging, 11 (7), pp.2045-2070. https://doi.org/10.18632/aging.101900.

Persistent Link:
https://hdl.handle.net/11343/273284

Research Paper

Socioeconomic position, lifestyle habits and biomarkers of epigenetic aging: a multi-cohort analysis

Giovanni Fiorito[1,40], Cathal McCrory[2,40], Oliver Robinson[3,40], Cristian Carmeli[4,40], Carolina Ochoa Rosales[5,6,40], Yan Zhang[7,40], Elena Colicino[8,40], Pierre-Antoine Dugué[9,10,11,40], Fanny Artaud[12,40], Gareth J McKay[13,40], Ayoung Jeong[14,15,40], Pashupati P Mishra[16,40], Therese H Nøst[17,18,40], Vittorio Krogh[19], Salvatore Panico[20], Carlotta Sacerdote[21], Rosario Tumino[22], Domenico Palli[23], Giuseppe Matullo[1,24], Simonetta Guarrera[1,24], Martina Gandini[25], Murielle Bochud[4], Emmanouil Dermitzakis[4], Taulant Muka[5,26], Joel Schwartz[27], Pantel S Vokonas[28], Allan Just[8], Allison M Hodge[9,10], Graham G Giles[9,10,11], Melissa C Southey[9,11,29], Mikko A Hurme[30], Ian Young[13], Amy Jayne McKnight[13], Sonja Kunze[31,32], Melanie Waldenberger[31,32,33], Annette Peters[31,32,33,34], Lars Schwettmann[35,36,41], Eiliv Lund[17,41], Andrea Baccarelli[37,41], Roger L Milne[9,10,11,41], Rose A Kenny[2,41], Alexis Elbaz[12,41], Hermann Brenner[7,38,41], Frank Kee[13,41], Trudy Voortman[5,41], Nicole Probst-Hensch[14,15,41], Terho Lehtimäki[16,41], Paul Elliot[3,41], Silvia Stringhini[39,4,41], Paolo Vineis[3,41], Silvia Polidoro[1,41]; and the BIOS Consortium; and the Lifepath consortium[42]

[1]Italian Institute for Genomic Medicine (IIGM), Turin, Italy
[2]The Irish Longitudinal Study on Ageing, Trinity College Dublin, Dublin, Ireland
[3]MRC-PHE Centre for Environment and Health, Imperial College London, London, UK
[4]Institute of Social and Preventive Medicine, Lausanne University Hospital (CHUV), Lausanne, Switzerland
[5]Department of Epidemiology, Erasmus University Medical Center, Rotterdam, the Netherlands
[6]Centro de Vida Saludable de la Universidad de Concepción, Concepción, Chile
[7]Division of Clinical Epidemiology and Aging Research, German Cancer Research Center (DKFZ), Heidelberg, Germany
[8]Icahn School of Medicine, Mount Sinai, New York, NY 10029, USA
[9]Cancer Epidemiology Division, Cancer Council Victoria, Melbourne, Australia
[10]Centre for Epidemiology and Biostatistics, School of Population and Global Health, The University of Melbourne, Victoria, Australia
[11]Precision Medicine, School of Clinical Sciences at Monash Health, Monash University, Clayton, Victoria, Australia
[12]CESP, Faculté de Médecine - Université Paris-Sud, Faculté de Médecine, UVSQ, Institut National de la Santé et de la Recherche Médicale, Université Paris-Saclay, France
[13]Centre for Public Health, Queen's University Belfast, Belfast, Northern Ireland
[14]Swiss Tropical and Public Health Institute, Basel, Switzerland
[15]University of Basel, Basel, Switzerland
[16]Department of Clinical Chemistry, Fimlab Laboratories, and Finnish Cardiovascular Research Center - Tampere, Faculty of Medicine and Health Technology, Tampere University, Tampere 33520, Finland
[17]Department of Community Medicine, Faculty of Health Sciences, UiT-The Arctic University of Norway, Tromsø, Norway
[18]NILU Norwegian Institute for Air Research, The Fram Centre, Tromsø, Norway
[19]Fondazione IRCCS - Istituto Nazionale dei Tumori, Milan, Italy
[20]Department of Clinical Medicine and Surgery, University of Naples Federico II, Naples, Italy
[21]Piedmont Reference Centre for Epidemiology and Cancer Prevention (CPO Piemonte), Turin, Italy
[22]Cancer Registry and Histopathology Department, 'Civic - M. P. Arezzo' Hospital, ASP Ragusa, Ragusa, Italy
[23]Istituto per lo Studio, la Prevenzione e la Rete Oncologica (ISPRO Toscana), Florence, Italy
[24]Department of Medical Sciences, University of Torino, Torino, Italy

[25]Environmental Epidemiological Unit, Regional Environmental Protection Agency, Piedmont Region, Torino, Italy
[26]Institute of Social and Preventive Medicine, University of Bern, Bern, Switzerland
[27]Department of Environmental Health and Department of Epidemiology, Harvard T.H. School of Public Health, Boston, MA 02115, USA
[28]Department of Epidemiology, Boston University School of Public Health, Boston, MA 02115, USA
[29]Department of Clinical Pathology, The University of Melbourne, Melbourne, Australia
[30]Department of Microbiology and Immunology, Faculty of Medicine and Health Technology, Tampere University, Tampere 33014, Finland
[31]Research Unit of Molecular Epidemiology, Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg, Germany
[32]Institute of Epidemiology, Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg, Germany
[33]German Center for Cardiovascular Research (DZHK), Munich, Germany
[34]Ludwig-Maximilians-Universität München, Munich, Germany
[35]Institute of Health Economics and Health Care Management, Helmholtz Zentrum München, German Research Center for Environmental Health (GmbH), D-85764, Neuherberg, Germany
[36]Department of Economics, Martin Luther University Halle-Wittenberg, Halle, Germany
[37]Department of Environmental Health Sciences, Columbia University Mailman School of Public Health, New York, NY 10032, USA
[38]Network Aging Research, University of Heidelberg, Heidelberg, Germany
[39]Unit of Population Epidemiology, Primary Care Division, Geneva University Hospitals, Geneva, Switzerland
[40]Equal contribution
[41]Equal senior researcher
[42]See ACKNOWLEDGMENTS AND FUNDING

Correspondence to: Giovanni Fiorito; **email:** giovanni.fiorito85@gmail.com
Keywords: socioeconomic position, education, biological aging, epigenetic clocks
Received: November 16, 2018 **Accepted:** March 31, 2019 **Published:** April 14, 2019

ABSTRACT

Differences in health status by socioeconomic position (SEP) tend to be more evident at older ages, suggesting the involvement of a biological mechanism responsive to the accumulation of deleterious exposures across the lifespan. DNA methylation (DNAm) has been proposed as a biomarker of biological aging that conserves memory of endogenous and exogenous stress during life.

We examined the association of education level, as an indicator of SEP, and lifestyle-related variables with four biomarkers of age-dependent DNAm dysregulation: the total number of stochastic epigenetic mutations (SEMs) and three epigenetic clocks (Horvath, Hannum and Levine), in 18 cohorts spanning 12 countries.

The four biological aging biomarkers were associated with education and different sets of risk factors independently, and the magnitude of the effects differed depending on the biomarker and the predictor. On average, the effect of low education on epigenetic aging was comparable with those of other lifestyle-related risk factors (obesity, alcohol intake), with the exception of smoking, which had a significantly stronger effect.

Our study shows that low education is an independent predictor of accelerated biological (epigenetic) aging and that epigenetic clocks appear to be good candidates for disentangling the biological pathways underlying social inequalities in healthy aging and longevity.

INTRODUCTION

Aging is characterized by a gradual and constant increase in health inequalities across socioeconomic groups [1, 2], an association based on strong epidemiological evidence known as the 'social gradient in health'. On average, individuals with lower socioeconomic position (SEP) have lower life expectancy, higher risk of age-related diseases, and poorer quality of life at older ages compared with less disadvantaged groups. Although lifestyles differ by SEP, unhealthy habits only partially explain this association [3].

The role of epigenetic mechanisms in response to trauma, and evidence for their involvement in intergenerational transmission of biological impacts of traumatic stress have been proposed to explain how social adversity gets biologically embedded [4], leading to differences in biological functionalities among individuals in different social conditions, especially at older ages. Epigenetics, specifically DNA methylation (DNAm) has been proposed as one of the most powerful biomarkers of biological aging and as one of the plausible biological mechanisms by which social adversities get 'under the skin' and affect physiological and cellular pathways leading to disease susceptibility [5-7].

Two different mechanisms have been proposed to contribute to age-related DNAm changes: 'epigenetic drift' and the 'epigenetic clock' that sometimes are used as synonyms even though describe different molecular mechanisms [8-10]. Although both are related to aging, epigenetic drift represents the trend of increasing DNAm variability over time across the whole genome. On the contrary, the epigenetic clock refers to specific CpG sites identified in specific DNA regions at which DNAm levels constantly increase (or decrease depending on the site) during aging and can be used to predict chronological age with high accuracy [11]. Two measures of epigenetic clocks have gained considerable popularity, Horvath [11] and Hannum [12], and the concept of epigenetic aging acceleration (EAA) has been introduced as the difference between predicted DNAm age and chronological age. EAA has been associated with all-cause mortality, cancer incidence and neurodegenerative disorders, as well as non-communicable disease risk factors such as obesity, poor physical activity, unhealthy diet, cumulative lifetime stress and infections [13, 14]. Recently, Levine and colleagues introduced a 'next-generation epigenetic clock' that is based on a set of CpGs associated with a complex set of clinical measures thought to assess the 'phenotypic age' [15]. Levine EAA was found to outperform other measures with regard to the prediction of a variety of aging outcomes, including all-cause mortality, the incidence of and survival from cancer, and physical functioning [15].

In contrast to EAA, epigenetic drift is a mechanism that involves the whole-genome, where age-related genomic instability and chromatin deterioration lead to increased variability of genome-wide DNAm levels at older ages [16]. Different statistical approaches can be used to evaluate the impact of epigenetic drift on aging and disease susceptibility. For example, Teschendorff and colleagues suggested that methods based on differential DNAm variability could identify risk markers more robustly than statistical measures based on differences in mean DNAm levels [17]. Gentilini and colleagues developed an analytical approach to identify these stochastic epimutations (SEMs) [18] from genome-wide DNAm data, showing that the number of SEMs increases exponentially with age although there is high variability within individuals of the same age. A higher number of SEMs was found to be associated with X chromosome inactivation skewing in women (an age-related condition and risk factor for cancer), hepatocellular carcinoma tumor staging [18, 19], and unhealthy exposure such as cigarette smoking, alcohol intake [20] and exposure to toxicants [21], suggesting SEMs as possible biomarkers of exposure-related accumulation of DNA damage during lifespan.

Given the above, it can be assumed that the various epigenetic clocks (Horvath's, Hannum's, Levine's) and the total number of SEMs describe different aspects of the biological (epigenetic) aging process. We previously showed a dose-response relationship between SEP and EAA. Further, our results suggest that the effect could be partially reversible by improving social conditions during life [5]. In addition, ours and two more recent studies indicate that childhood SEP might have a stronger effect on EAA than adulthood SEP [22, 23].

Despite extensive research in the field, to date no studies have compared the effect of SEP on epigenetic aging biomarkers with those of other lifestyle-related risk factors for age-related diseases. We aimed to systematically investigate the association of education level, as a proxy for SEP, with the total number of SEMs and 'accelerated aging' as assessed using the three epigenetic clocks, and to compare the independent effect of low education with those of the main modifiable risk factors for premature aging: smoking, obesity, alcohol intake and physical inactivity, by conducting a meta-analysis including data for more than 16,000 individuals belonging to 18 cohort studies from 12 different countries worldwide.

RESULTS

After quality control and sample filtering, we analyzed blood DNAm data from 16,245 individuals from 18 cohort studies. The main characteristics of the study sample are shown in Table 1. For each epigenetic outcome, we report the results of a meta-analysis of the association with education, smoking, obesity, alcohol intake and physical activity in Table 2. Model 1 includes age, sex, and cohort-specific covariates as adjustment variables whereas Model 2 is the fully adjusted model (additionally adjusted for smoking, BMI, alcohol intake and physical activity).

For the three epigenetic clocks, the estimated differences presented in Table 2 (βs) represent the change in biological age (in years) compared with the reference group. Accordingly, the estimated effects of risk factors on SEMs were re-scaled to be expressed in years as for the three epigenetic clocks using a two-step approach based on the Cohen's D statistic, described in the supplementary text.

Table 1. Study sample descriptive statistics.

Cohort short name	Cohort full name	Country	Illumina BeadChip	N	Mean age (min - max)	Female N(%)	Reference
AIRWAVE	The Airwave Health Monitoring Study	UK	Illumina EPIC chip (850K)	1,127	41 (13 - 65)	458 (41%)	[46]
EXPOsOMICS 'EPIC CVD'	The European Prospective Investigation into Cancer and Nutrition - EXPOsOMICS subsample	Italy	Illumina 450K BeadChip	313	57 (35 - 75)	167 (53%)	[47]
EPIC	The European Prospective Investigation into Cancer and Nutrition	Italy	Illumina 450K BeadChip	1,803	53 (35 - 75)	1,114 (62%)	[48]
ESTHER 1	Epidemiological investigations on chances of preventing, recognizing early and optimally treating chronic diseases in an elderly population	Germany	Illumina 450K BeadChip	1,000	62 (48 - 75)	500 (50%)	[49]
ESTHER 2	Epidemiological investigations on chances of preventing, recognizing early and optimally treating chronic diseases in an elderly population	Germany	Illumina EPIC chip (850K)	864	63 (48 - 75)	390 (45%)	[49]
KORA	Cooperative Health Research in the Region of Augsburg (KORA-F4)	Germany	Illumina 450K BeadChip	1,727	61 (32 - 81)	882 (51%)	[50]

MCCS	Melbourne Collaborative Cohort Study	Australia	Illumina 450K BeadChip	2,817	59 (40 - 70)	1,095 (39%)	[51]
NAS	Normative aging study	USA	Illumina 450K BeadChip	624	72 (55 - 91)	0 (0%)	[52]
NOWAC	The Norwegian Women and Cancer Study	Norway	Illumina 450K BeadChip	632	56 (47 - 63)	632 (100%)	
NICOLA	Northern Ireland Cohort Longitudinal Study of Ageing	Northern Ireland	Illumina EPIC chip (850K)	1,929	64 (40 - 96)	988 (51%)	[53]
RS-Bios	Rotterdam Study 1,2	Netherlands	Illumina 450K BeadChip	720	68 (52 - 80)	304 (42%)	[54]
RSIII-1	Rotterdam Study 3	Netherlands	Illumina 450K BeadChip	730	60 (46 - 89)	335 (46%)	[54]
SAPALDIA	Swiss Study on Air Pollution and Lung Diseases in Adults	Switzerland	Illumina 450K BeadChip	402	57 (38 - 81)	184 (46%)	[55]
SKIPOGH a	Swiss Kidney Project on Genes in Hypertension	Switzerland	Illumina 450K BeadChip	250	51 (26 - 82)	132 (53%)	[56]
SKIPOGH b	Swiss Kidney Project on Genes in Hypertension	Switzerland	Illumina EPIC chip (850K)	451	54 (25 - 89)	231 (51%)	[56]
TERRE	Case-control study of Parkinson's disease in French farmers (only controls were used)	France	Illumina EPIC chip (850K)	174	67 (41 - 76)	80 (46%)	[57]
TILDA	The Irish Longitudinal Study on aging	Ireland	Illumina EPIC chip (850K)	490	62 (50 - 80)	246 (50%)	[58]
YFS	Young Finns Study	Finland	Illumina 450K BeadChip	186	44 (34 - 49)	72 (39%)	[59]

Education: The level of education was significantly associated with the four biomarkers investigated. In Model 1 (minimally adjusted), lower educated individuals had a higher number of SEMs β = 0.34 (95% CI 0.11; 0.58), higher Horvath EAA β = 0.22 (0.03; 0.41), higher Hannum EAA β = 0.34 (0.17; 0.52), and higher Levine EAA β = 0.84 (0.50; 1.17), compared with the higher educated group who constituted the reference category. The observed associations were still significant after the inclusion of smoking, BMI, alcohol and physical activity in the regression models (Model 2), but the estimated effects were moderately reduced. Comparing the two extreme categories (low vs. high education) the estimated effects were: SEMs β = 0.28 (0.04; 0.51), Horvath EAA β = 0.19 (0.00; 0.39), Hannum EAA β = 0.31 (0.14; 0.48), and Levine EAA β = 0.60 (0.25; 0.94) in the full multivariable adjusted models. Interestingly, the intermediate education group ranked between the high and low education group supporting a dose-response effect (Table 2).

Table 2. Results of linear regressions using epigenetic aging biomarkers as outcomes and lifestyle related risk factors as predictors.

		SEMs		HorvathEAA	
		Model 1	**Model 2**	**Model 1**	**Model 2**
Education (ref: High)	Medium	**0.23 (0.02; 0.44)***	0.17 (-0.07; 0.42)	0.11 (-0.07; 0.28)	0.11 (-0.08; 0.29)
	Low	**0.34 (0.11; 0.58)****	**0.28 (0.04; 0.51)***	**0.22 (0.03; 0.41)***	**0.19 (0.00; 0.39)**$^{+}$
Smoking (ref: Never)	Former	**0.32 (0.14; 0.49)*****	**0.33 (0.16; 0.51)*****	0.13 (-0.04; 0.29)	0.11 (-0.05; 0.26)
	Current	**0.53 (0.32; 0.73)*****	**0.51 (0.30; 0.72)*****	-0.06 (-0.24; 0.13)	-0.08 (-0.27; 0.12)
Obesity (ref: BMI < 25)	BMI < 30	-0.01 (-0.18; 0.16)	-0.01 (-0.18; 0.15)	**0.37 (0.22; 0.52)*****	**0.33 (0.18; 0.48)*****
	BMI ≥ 30	-0.06 (-0.26; 0.15)	-0.07 (-0.27; 0.14)	**0.45 (0.27; 0.63)*****	**0.43 (0.24; 0.61)*****
Alcohol (ref: Abstainer)	Occasional	-0.12 (-0.31; 0.08)	-0.10 (-0.29; 0.08)	-0.02 (-0.19; 0.15)	0.00 (-0.18; 0.18)
	Habitual	0.22 (-0.05; 0.49)	0.15 (-0.11; 0.4)	0.19 (-0.07; 0.44)	**0.25 (0.00; 0.49)***
Physical activity (ref: High)	Medium	0.00 (-0.21; 0.21)	-0.03 (-0.21; 0.15)	0.05 (-0.11; 0.21)	0.08 (-0.09; 0.24)
	Low	0.03 (-0.28; 0.35)	-0.03 (-0.32; 0.26)	**0.22 (0.05; 0.39)***	**0.22 (0.04; 0.40)***
		HannumEAA		**LevineEAA**	
		Model 1	**Model 2**	**Model 1**	**Model 2**
Education (ref: High)	Medium	**0.32 (0.14; 0.49)*****	**0.27 (0.08; 0.46)****	**0.50 (0.22; 0.79)*****	**0.31 (-0.05; 0.67)**$^{+}$
	Low	**0.34 (0.17; 0.52)*****	**0.31 (0.14; 0.48)*****	**0.84 (0.50; 1.17)*****	**0.60 (0.25; 0.94)*****
Smoking (ref: Never)	Former	0.04 (-0.08; 0.16)	0.01 (-0.12; 0.13)	**0.60 (0.37; 0.84)*****	**0.52 (0.28; 0.77)*****
	Current	**0.24 (0.06; 0.42)****	**0.17 (0.00; 0.35)***	**1.57 (1.31; 1.82)*****	**1.41 (1.14; 1.67)*****
Obesity (ref: BMI < 25)	BMI < 30	**0.17 (0.05; 0.28)****	**0.15 (0.03; 0.27)***	**0.37 (0.13; 0.62)****	**0.33 (0.11; 0.55)****
	BMI ≥ 30	**0.22 (0.07; 0.36)****	**0.20 (0.05; 0.34)***	**1.08 (0.79; 1.37)*****	**1.01 (0.74; 1.28)*****

Alcohol (ref: Abstainer)	Occasional	-0.05 (-0.19; 0.09)	0.03 (-0.11; 0.17)	-0.08 (-0.36; 0.20)	0.10 (-0.14; 0.34)
	Habitual	0.14 (-0.03; 0.31)	**0.21 (0.04; 0.39)***	**0.88 (0.49; 1.26)*****	**0.91 (0.57; 1.25)*****
Physical activity (ref: High)	Medium	0.07 (-0.08; 0.22)	0.07 (-0.07; 0.20)	0.16 (-0.17; 0.49)	0.20 (-0.04; 0.44)
	Low	0.08 (-0.15; 0.32)	0.05 (-0.20; 0.30)	0.42 (-0.12; 0.96)	0.31 (-0.13; 0.74)

***** p < 0.001; ** p < 0.01; * p < 0.05; + p < 0.10**

Model 1 includes age, sex, and cohort specific covariates; Model 2 includes additional adjustment for education, smoking, BMI, alcohol and physical activity.

Other modifiable risk factors: Current smokers had a higher number of SEMs, higher Hannum EAA and higher Levine EAA compared with never smokers. The estimated effects were slightly reduced in Model 2 compared with Model 1 when adjusted additionally for other covariates. Further, former smokers had intermediate outcomes between never and current smokers (Table 2). The estimated effect size of the association between smoking and epigenetic aging biomarkers was comparable to those observed for education, except for the magnitude of the association with Levine EAA, which was significantly higher: β = 1.57 (1.31; 1.82) in Model 1; β = 1.41 (1.14; 1.67) in Model 2. A similar pattern of associations was observed looking at the effects of obesity on epigenetic aging biomarkers. Obese individuals (BMI ≥ 30) had higher Horvath EAA, higher Hannum EAA, and higher Levine EAA. As previously described for education and smoking, the effects estimated in Model 2 were slightly lower compared with Model 1, and a dose-response association was observed. The estimated effects of obesity were comparable to those of education except for Levine EAA, which was significantly higher: β = 1.08 (0.79; 1.37) in Model 1; β = 1.01 (0.74; 1.28) in Model 2. Looking at alcohol intake, we did not observe any significant difference comparing abstainers and occasional drinkers, but habitual drinkers had higher Horvath EAA, Hannum EAA and Levine EAA. As observed for the other risk factors, the higher estimated effects were observed for Levine's indicator: β = 0.88 (0.49; 1.26) in Model 1; β = 0.91 (0.57; 1.25) in Model 2. Finally, low physical activity was associated with higher Horvath EAA in both Model 1 β = 0.22 (0.05; 0.39) and Model 2 β = 0.22 (0.04; 0.40).

Figure 1 shows a graphical representation of the results (Model 2) using forest plot which allows one to compare the effect of each risk factor considered in the present paper on the four DNAm outcomes.

In sensitivity analyses, we examined the white blood cell (WBC) adjusted epigenetic aging measures (described in Methods), and found similar associations as the ones described above (Table S1). Further, for each risk factor, we evaluated the interaction with age and sex. Our results indicated no significant differences in associations between men and women, whereas we found a significant interaction with age for the association of SEMs with education, smoking, and obesity, with a significantly stronger effect in older individuals (Table S2).

We examined whether SEMs were randomly distributed across the genome or are enriched in functional genomic regions. We observed overlap between the genomic position of SEMs and regions associated with open chromatin states, and shores (p=0.03, p=0.02 respectively, Table S3). Considering the categories defined by the Encyclopedia of DNA Elements (ENCODE) project with Chromatin Immuno-Precipitation Sequencing (ChIP-Seq) experiments on human embryonic stem cells (hESC), we found enrichment of SEMs in 'inactive/poised promoters' (p<0.0001, Table S4), 'heterochromatin/low signal/CNV' (p<0.0001, Table S4), and 'Polycomb-repressed' regions (p=0.001, Table S4). Furthermore, we found significant overlap with transcription factor binding sites (TFBSs) targeted by two members of the Polycomb repressive complex-2 (PRC2): EZH2 and SUZ12 (p<0.0001, Table S5).

DISCUSSION

Social inequalities in health have been extensively reported and accelerated age-dependent DNAm dysregulation has been proposed as one of the biomolecular mechanisms mediating this association [5, 24, 25]. In this study, we examined the effect on DNAm biomarkers of aging of being in the low education group compared with those of other lifestyle-related risk factors: smoking, obesity, alcohol intake, and low levels of physical activity. We used education as the proxy for SEP as it was the only socioeconomic indicator that was available in all the cohorts and it is usually completed

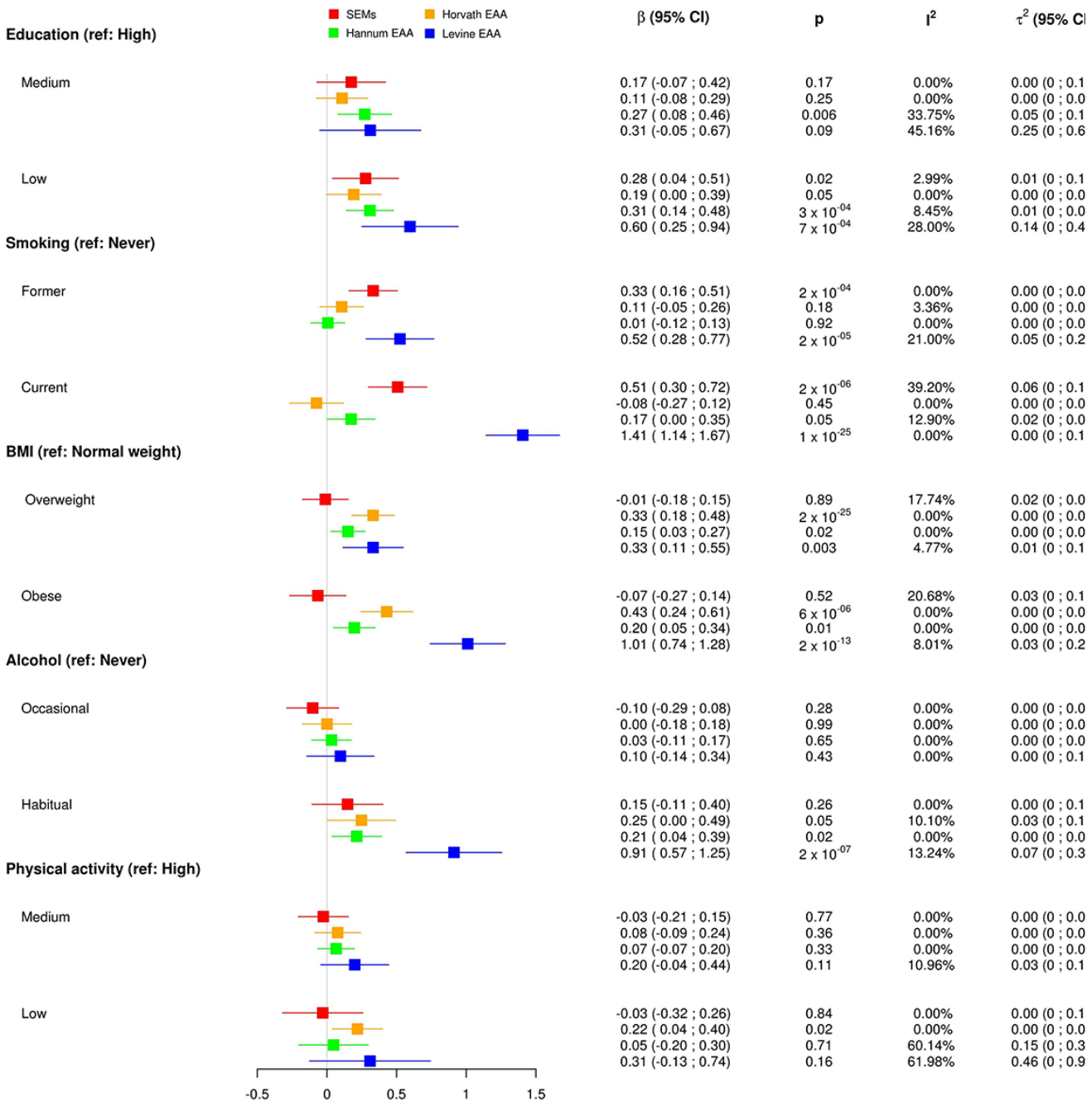

Figure 1. Effect sizes (interpretable as years of increasing/decreasing epigenetic age) of the association between different risk factors and four epigenetic aging biomarkers: total number of stochastic epigenetic mutations (SEMs, red), Horvath epigenetic age acceleration (orange), Hannum epigenetic age acceleration (green) and Levine epigenetic age acceleration next-generation clock (blue).

before the onset of many chronic diseases, therefore reducing the risk of reverse causation [26]. Lower educational attainment was associated with EAA according to the 'first generation' clocks including Horvath's and Hannum's. However, previously it was not clear whether the observed associations depend on other factors associated with low education [6, 27]. For example, Karlsson Linnér and colleagues argue that the association of educational attainment and epigenetic aging is mainly mediated by maternal smoking during pregnancy and smoking during adulthood [6]. To clarify this issue and to increase the epidemiological evidence in the field, we

have examined four biomarkers of age-dependent DNAm dysregulation: the total number of SEMs and three epigenetic clocks (Horvath, Hannum and Levine).

Although all the biomarkers are related to aging, they did not show the same pattern of associations with risk factors, intermediate traits and diseases [28, 29], suggesting that these biological age predictors may reflect different facets of the aging process. The total number of SEMs takes into account whole-genome epigenetic deregulation during aging, a process known as 'epigenetic drift', and has been proposed as a biomarker of exposure-related accumulation of DNA damage during the lifespan [20]. It is necessary to clarify that the word 'epimutation' is sometimes used in a manner that can be misinterpreted. Although some literature uses this term to refer to epigenetic changes driven by genotype differences, the strict definition of epimutation is a heritable change in gene activity that is not associated with a DNA mutation, but rather, with gain or loss of DNA methylation or other heritable modifications of chromatin [30]. Contrary to the definition of genetic mutations, epimutations are defined as potentially (but not necessarily) reversible changes in gene activity not involving DNA mutations, but rather, gain or loss of DNA methyl groups conserved in cells through mitosis [20, 30, 31].

In contrast, Horvath's epigenetic clock is based on DNAm levels at a small subset of CpG sites and is thought to reflect the biological age of different tissues, while Hannum's epigenetic clock is specific to blood samples. Finally, Levine's next-generation clock is computed using a subset of CpGs that were associated with several clinical measures representing the health status of an individual and has been proposed as a biomarker of the individual 'phenotypic age' [15]. Accordingly, Levine's measure of age acceleration tends to be more variable than the first-generation clocks (Horvath and Hannum) as evidenced by the finding that the associations based on this marker showed, in general, a higher degree of heterogeneity in the meta-analysis measured with the I^2 and τ^2 statistics.

Our results from this meta-analysis of more than 16,000 individuals support our working hypothesis. We found that the four aging biomarkers were associated with different sets of risk factors and that the magnitude of the associations differed depending on the epigenetic aging index. We compared the effects of two nested models: the first minimally adjusted model included age, sex and cohort-specific covariates as adjustments; the second (fully adjusted model) was adjusted for all of the risk factors. We did not observe significant differences comparing estimates from the two models (Table 2), supporting the robustness of the results presented. Interestingly, the effect of low education was independent from the other risk factors examined, as it was significant in both the minimally adjusted and fully adjusted models; and the effect sizes were comparable to that of the other risk factors examined, with the exception of smoking, which had an appreciably larger impact on SEMs and Levine's measure. Two previous studies from our group that evaluated the association of low SEP with mortality and physical functioning documented strong patterning by SEP [32, 33]. The current study provides evidence of the potential role of epigenetic modifications as mediators of the association of low SEP and unhealthy lifestyle habits with adverse outcomes at older ages, and further underscores the importance of considering SEP as an important life course risk factor for premature biological aging.

In sensitivity analyses, we also investigated alternative measures of the epigenetic aging biomarkers corrected for the proportion of WBC (estimated from whole-genome DNAm data). Chen and colleagues refer to the WBC-adjusted epigenetic aging as an 'intrinsic' measure of biological aging, which captures cell-intrinsic properties of the aging process, that exhibit some preservation across various cell types and organs [34]. Our results indicate no significant differences in the results using 'extrinsic' (non-WBC-adjusted) vs 'intrinsic' measures. Similarly, stratified analyses by sex indicated no differential effect between men and women.

Finally, we evaluated the potential differential effects of risk factors by (chronological) age group. We found that the effect of education, smoking and BMI on the total number of SEMs, and the effect of smoking on Hannum EAA was significantly greater for older individuals. These results agree with the 'epigenetic memory' hypothesis according to which epigenetic aging biomarkers, particularly SEMs, could reflect the accumulation of deleterious exposures during the lifespan [35].

To elucidate the molecular mechanisms involved in DNAm dysregulation during aging we investigated whether SEMs occurred randomly in the DNA sequence or were enriched in regulatory regions. Our findings confirmed that epimutations preferentially occur in DNA sequences associated with open chromatin (Table S3), as previously observed by Ong et al. [36]. Furthermore, SEMs were enriched in transcriptionally silenced genomic regions such as 'inactive promoters', 'heterochromatin/low signal/copy number variants (CNV)', and 'Polycomb-repressed' regions (Table S4). Specifically, SEMs were more likely to occur in transcription factor binding sites (TFBSs) targeted by two members of Polycomb repressive complex 2

(PRC2): EZH2 and SUZ12 (Table S5). The role of PCR2 proteins and EZH2 specifically in aging and age-related diseases has been extensively investigated in recent years. Targets of PRC2 proteins are enriched for tumor suppressor genes and genes related to mental/neurodegenerative disorders [37-39], providing a link between aging, age-related DNAm modifications and age-related diseases. As an example, downregulations of EZH2 and SUZ12 have been associated with dysregulation of several PRC2 targets including *p53*, a well-known tumor suppressor gene [40]. These results underscore the imperative for further research aimed at identifying disease-specific epimutation signatures.

In conclusion, we have shown that SEP is associated with different biomarkers of biological (epigenetic) aging, which may represent complementary aspects of the aging process. On average, the impact of low education was comparable with that of other lifestyle-related risk factors, with the exception of smoking, for which the effect was more pronounced. Levine's second-generation clock was more strongly associated with education and the other risk factors (smoking, obesity and alcohol intake) compared with the first-generation epigenetic clocks and epimutation biomarkers. This result is not wholly unexpected because Levine's indicator was explicitly designed with a two-step procedure to include both CpG sites associated with aging *per se*, and CpG sites predictive of mortality and associated with biomarkers of chronic diseases. Given the above, Levine's epigenetic clock appeared to be a strong candidate for future studies aiming at disentangling the biological pathways underlying social inequalities in healthy aging and longevity. We confirmed in a very large cross-cohort, cross-country sample, previous observations of 'accelerated aging' in individuals of low SEP [6, 15]. To the best of our knowledge this is also the first study showing that i) socioeconomic status is associated with the total number of stochastic epimutations, which in turn is a biomarker of adverse health outcomes at older ages; and ii) the impact of low socioeconomic status on age-related epigenetic biomarkers is comparable to that of the major lifestyle-related and modifiable risk factors for age related diseases, with the exception of smoking which had a stronger effect on two out of four epigenetic biomarkers investigated.

This study also has limitations: variable collection and classification were not done in a standardized way for all the cohorts. Similarly, for DNA methylation data, each cohort used its own method for data normalization and batch effect removal. To avoid over-estimation of the effects of education and other risk factors on epigenetic aging biomarkers, we used a REML method that incorporates random study effects around the overall mean, to obtain robust global estimates in meta-analysis.

Education was the only SEP indicator used in this study. Critical alternative SEP indicators like income or occupational position were not available in all of the cohorts used in this meta-analysis. We chose to consider all the variables as categorical, with three levels each. On the one hand, this procedure helped us to compare the effects of low education with that of other risk factors, but on the other hand, could lead to a reduction in statistical power to detect significant associations. The lack of association of epigenetic aging measures with low physical activity is not in line with previous literature [27] which might be explained by heterogeneity in the variable definition across cohorts.

MATERIALS AND METHODS

Participating cohorts

We obtained summary-level association statistics from 17 independent cohort studies located in Europe, the United States and Australia (Table 1). The total sample size includes 16,245 individuals of recent European ancestry. All participants provided written informed consent, and all contributing cohorts confirmed compliance with their Local Research Ethics Committees or Institutional Review Boards. Each risk factor was treated as a categorical variable with three categories each; education: 'High' (reference), 'Medium', 'Low'; smoking: 'Never' (reference), 'Former', 'Current'; obesity: 'Normal weight (BMI $\leq$ 25, reference)', 'Overweight (25 < BMI $\leq$ 30)', 'Obese (BMI > 30)'; alcohol consumption: 'Abstainers' (reference), 'Occasional drinkers', 'Habitual drinkers'; physical activity: 'High' (reference), 'Medium', 'Low' (see the supplementary text for the definition of the categories).

Identification of stochastic epigenetic mutations

We identified SEMs based on the procedure described by Gentilini et al. [18]. Briefly, for each CpG, considering the distribution of DNAm beta values across all samples, we computed the interquartile range (IQR) - difference between the third quartile (Q3) and the first quartile (Q1) - and we defined a SEM as a methylation value lower than Q1-(3×IQR) or greater than Q3+(3×IQR). Finally, for each individual, we computed the total number of SEMs across all CpGs. Since the total number of SEMs increased exponentially with age, we used a logarithmic transformation of the outcome for all the analyses. In sensitivity analyses, we computed the same measure corrected for potential confounding by differential WBC proportions among

individuals. Specifically, for each CpG we calculated the residuals from the regression of DNAm beta values on estimated WBC fractions, and then applied the same procedure described above.

Computation of epigenetic clock measures

We computed the epigenetic age acceleration (AA) measures according to the algorithm described by Horvath [11]. Briefly, DNAm age was calculated as a weighted average of 353 age-related CpGs (Horvath DNAm age), 71 blood specific age-related CpGs (Hannum DNAm age), and 513 phenotypic age-related CpGs (Levine DNAm age) respectively. Weights were obtained using a penalized regression model (Elastic-net regularization) [11]. Age acceleration (AA) was defined as the difference between epigenetic and chronological age. Since AA may be correlated with chronological age, the 'extrinsic' EAA is defined as the residuals of AA on age. For sensitivity analyses, we also computed the 'intrinsic' epigenetic age acceleration (IEAA), defined as the residuals from the linear regression of AA on chronological age and WBC percentages [34]. Positive values of EAA (which is by definition independent of age) indicate accelerated aging and negative values decelerated aging.

Association of epigenetic aging measures with risk factors

We investigated the association of four biological (epigenetic) aging biomarkers with education, smoking, obesity, alcohol and physical activity using each epigenetic measure as the outcome and risk factors as predictors in linear regression models. For each outcome and each risk factor, we ran two regression models. Model 1 was adjusted for age, sex and cohort-specific covariates only (see the supplementary text). Model 2 included additional adjustment for education, smoking, BMI, alcohol intake, physical activity, to derive the mutually adjusted estimate for each risk factor considered in the present paper.

Each cohort provided the results of the analyses as point estimates and 95% confidence intervals based on an R script shared with all the data analysts (available in the supplementary material). The pooled estimated association of risk factors with each outcome were obtained by random effect maximum likelihood (REML) meta-analysis using the R package *metafor* [41]. Heterogeneity among studies was assessed using the I^2 and τ^2 statistics.

SEMs enrichment analysis

The genomic locations of SEMs were annotated by merging the Illumina information on the chromosomal position of each probe with ENCODE/NIH Roadmap ChIP-Seq data for chromatin states and transcription factor binding sites (TFBS) in hESC [42-44]. We investigated whether SEMs were randomly distributed across the genome, or were enriched in functional genomic regions using the procedure implemented in the R package *regioneR* [45]. Briefly, the algorithm is specifically designed to test whether a set of genomic *loci* significantly overlap with a set of genomic regions, and includes a permutation procedure that controls for the type I error rate.

CONFLICTS OF INTEREST

Authors declare no conflicts of interest.

ACKNOWLEDGMENTS AND FUNDING

This research was supported by the 'Lifepath' grant to Paolo Vineis at Imperial College, London, Silvia Polidoro at the Italian Institute for Genomic Medicine (IIGM). The Airwave Health Monitoring Study is funded by the Home Office (grant number 780-TETRA) with additional support from the National Institute for Health Research (NIHR) Biomedical Research Centre. The Airwave Study uses the computing resources of the UK MEDical BIOinformatics partnership (UK MED-BIO supported by the Medical Research Council (MR/L01632X/1). The Rotterdam Study is funded by Erasmus Medical Center and Erasmus University, Rotterdam, Netherlands Organization for the Health Research and Development (ZonMw), the Research Institute for Diseases in the Elderly (RIDE), the Ministry of Education, Culture and Science, the Ministry for Health, Welfare and Sports, the European Commission (DG XII), and the Municipality of Rotterdam. The DNA methylation data was funded by the Genetic Laboratory of the Department of Internal Medicine, Erasmus Medical Center, and by the Netherlands Organization for Scientific Research (NWO; project number 184021007) and made available as a Rainbow Project (RP3; BIOS) of the Biobanking and Biomolecular Research Infrastructure Netherlands (BBMRI-NL). TILDA project was supported by the Irish Government, the Atlantic Philanthropies, and Irish Life plc. The SKIPOGH study was supported by a grant from the Swiss National Science Foundation (FN 33CM30-124087); DNA methylation data in the SKIPOGH data was funded by the Swiss national Science Foundation (AmbizioneGrant n° PZ00P3_167732 to Silvia Stringhini) and European commission and the Swiss State Secretariat for Education, Research and Innovation - SERI (Horizon 2020 grant n° 633666). We are grateful to all the participants of NICOLA, Dr Laura Smyth and the NICOLA team, which includes nursing

staff, research scientists, clerical staff, computer and laboratory technicians, managers and receptionists. This work was supported by the following funders who provide core financial support for the NICOLA Study: the Atlantic Philanthropies; the Economic and Social Research Council; the UKCRC Centre of Excellence for Public Health Northern Ireland; the Centre for Ageing Research and Development in Ireland; the Office of the First Minister and Deputy First Minister; the Health and Social Care Research and Development Division of the Public Health Agency; the Wellcome Trust/Wolfson Foundation; and Queen's University Belfast. The MCCS component of this work was supported by the Australian National Health and Medical Research Council (NHMRC, grants 1074383 and 1106016). MCCS cohort recruitment was funded by VicHealth and Cancer Council Victoria. The MCCS was further supported by NHMRC grants 209057 and 396414 and by infrastructure provided by Cancer Council Victoria, and the methylation studies were supported by NHMRC grants 1011618, 1026892, 1027505, 1050198, 1043616, 1074383 and 1088405. The Young Finns Study has been financially supported by the Academy of Finland: grants 286284, 134309 (Eye), 126925, 121584, 124282, 129378 (Salve), 117787 (Gendi), and 41071 (Skidi); the Social Insurance Institution of Finland; Competitive State Research Financing of the Expert Responsibility area of Kuopio, Tampere and Turku University Hospitals (grant X51001); Juho Vainio Foundation; Paavo Nurmi Foundation; Finnish Foundation for Cardiovascular Research; Finnish Cultural Foundation; The Sigrid Juselius Foundation, Wihuri Foundation (for L-PR); Tampere Tuberculosis Foundation; Emil Aaltonen Foundation; Yrjö Jahnsson Foundation; Signe and Ane Gyllenberg Foundation; Diabetes Research Foundation of Finnish Diabetes Association; and EU Horizon 2020 (grant 755320 for TAXINOMISIS); and European Research Council (grant 742927 for MULTIEPIGEN project); Tampere University Hospital Supporting Foundation. DNA methylation data for the TERRE study was funded through the multilateral France-Germany-Canada Call for International proposals 2015 on "Epigenomics of complex diseases" (*Agence nationale de la recherche* grant ANR-15-EPIG-0001) as part of the DECIPHER-PD project. Oliver Robinson was supported by a MRC Early Career Fellowship. Carolina Ochoa-Rosales work was funded by the CONICYT PAI/INDUSTRIA 72170524, Chile. Cathal McCrory is supported by a HRB Emerging Investigator Award (EIA-2017-12). We are grateful to all the participants of NICOLA, and the NICOLA team, which includes nursing staff, research scientists, clerical staff, computer and laboratory technicians, managers and receptionists.

[42]The Lifepath partners include in alphabetic order: Jan Alberts, Harri Alenius, Mauricio Avendano, Valeria Baltar, Mel Bartley, Henrique Barros, Michele Bellone, Eloise Berger, David Blane, Murielle Bochud, Giulia Candiani, Cristian Carmeli, Luca Carra, Raphaele Castagné, Marc Chadeau-Hyam, Sergio Cima, Françoise Clavel-Chapelon, Giuseppe Costa, Emilie Courtin, Cyrille Delpierre, Angelo D'Errico, Manolis Dermitzakis, Pierre-Antoine Dugué, Marko Elovainio, Paul Elliott, Guy Fagherazzi, Giovanni Fiorito, Silvia Fraga, Martina Gandini, Valérie Garès, Pascale Gerbouin-Rerolle, Graham Giles, Marcel Goldberg, Dario Greco, Idris Guessous, Jose Haba-Rubio, Raphael Heinzer, Allison Hodge, Stephane Joost, Maryam Karimi, Michelle Kelly-Irving, Mika Kähönen, Piia Karisola, Leyla Khenissi, Mika Kivimaki, Vittorio Krogh, Jessica Laine, Thierry Lang, Audrey Laurent, Richard Layte, Benoit Lepage, Dori Lorsch, Frances MacGuire, Giles Machell, Johan Mackenbach, Michael Marmot, Carlos de Mestral, Cathal McCrory, Cynthia Miller, Roger Milne, Peter Muennig, Wilma Nusselder, Salvatore Panico, Dusan Petrovic, Lourdes Pilapil, Silvia Polidoro, Martin Preisig, Laura Pulkki-Råback, Olli Raitakari, Ana Isabel Ribeiro, Fulvio Ricceri, Paolo Recalcati, Erica Reinhard, Oliver Robinson, Jose Rubio Valverde, Severine Saba, Carlotta Sacerdote, Frank Santegoets, Roberto Satolli, Terrence Simmons, Gianluca Severi, Martin J Shipley, Silvia Stringhini, Adam Tabak, Vesa Terhi, Joannie Tieulent, Rosario Tumino, Salvatore Vaccarella, Federica Vigna-Taglianti, Paolo Vineis, Peter Vollenweider, Nicolas Vuilleumier, Marie Zins.

REFERENCES

1. Mackenbach JP, Valverde JR, Artnik B, Bopp M, Brønnum-Hansen H, Deboosere P, Kalediene R, Kovács K, Leinsalu M, Martikainen P, Menvielle G, Regidor E, Rychtaříková J, et al. Trends in health inequalities in 27 European countries. Proc Natl Acad Sci USA. 2018; 115:6440–45. https://doi.org/10.1073/pnas.1800028115

2. Artazcoz L, Rueda S. Social inequalities in health among the elderly: a challenge for public health research. J Epidemiol Community Health. 2007; 61:466–67. https://doi.org/10.1136/jech.2006.058081

3. Stringhini S, Sabia S, Shipley M, Brunner E, Nabi H, Kivimaki M, Singh-Manoux A. Association of socioeconomic position with health behaviors and mortality. JAMA. 2010; 303:1159–66. https://doi.org/10.1001/jama.2010.297

4. Cunliffe VT. The epigenetic impacts of social stress: how does social adversity become biologically embedded? Epigenomics. 2016; 8:1653–69. https://doi.org/10.2217/epi-2016-0075

5. Fiorito G, Polidoro S, Dugué PA, Kivimaki M, Ponzi E, Matullo G, Guarrera S, Assumma MB, Georgiadis P, Kyrtopoulos SA, Krogh V, Palli D, Panico S, et al. Social adversity and epigenetic aging: a multi-cohort study on socioeconomic differences in peripheral blood DNA methylation. Sci Rep. 2017; 7:16266. https://doi.org/10.1038/s41598-017-16391-5

6. Karlsson Linnér R, Marioni RE, Rietveld CA, Simpkin AJ, Davies NM, Watanabe K, Armstrong NJ, Auro K, Baumbach C, Bonder MJ, Buchwald J, Fiorito G, Ismail K, et al. An epigenome-wide association study meta-analysis of educational attainment. Mol Psychiatry. 2017; 22:1680–90. https://doi.org/10.1038/mp.2017.210

7. Taylor SE, Repetti RL, Seeman T. Health psychology: what is an unhealthy environment and how does it get under the skin? Annu Rev Psychol. 1997; 48:411–47. https://doi.org/10.1146/annurev.psych.48.1.411

8. 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

9. Zheng SC, Widschwendter M, Teschendorff AE. Epigenetic drift, epigenetic clocks and cancer risk. Epigenomics. 2016; 8:705–19. https://doi.org/10.2217/epi-2015-0017

10. Teschendorff AE, West J, Beck S. Age-associated epigenetic drift: implications, and a case of epigenetic thrift? Hum Mol Genet. 2013; 22:R7–15. https://doi.org/10.1093/hmg/ddt375

11. 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

12. 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

13. Declerck K, Vanden Berghe W. Back to the future: epigenetic clock plasticity towards healthy aging. Mech Ageing Dev. 2018; 174:18–29. https://doi.org/10.1016/j.mad.2018.01.002

14. Dugue PA, Bassett JK, Joo JE, Baglietto L, Jung CH, Ming Wong E, Fiorito G, Schmidt D, Makalic E, Li S, Moreno-Betancur M, Buchanan DD, Vineis P, et al. Association of DNA methylation-based biological age with health risk factors, and overall and cause-specific mortality. Am J Epidemiol. 2018; 187:529–38. https://doi.org/10.1093/aje/kwx291

15. 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

16. Slieker RC, van Iterson M, Luijk R, Beekman M, Zhernakova DV, Moed MH, Mei H, van Galen M, Deelen P, Bonder MJ, Zhernakova A, Uitterlinden AG, Tigchelaar EF, et al, and BIOS consortium. Age-related accrual of methylomic variability is linked to fundamental ageing mechanisms. Genome Biol. 2016; 17:191. https://doi.org/10.1186/s13059-016-1053-6

17. Teschendorff AE, Widschwendter M. Differential variability improves the identification of cancer risk markers in DNA methylation studies profiling precursor cancer lesions. Bioinformatics. 2012; 28:1487–94. https://doi.org/10.1093/bioinformatics/bts170

18. Gentilini D, Garagnani P, Pisoni S, Bacalini MG, Calzari L, Mari D, Vitale G, Franceschi C, Di Blasio AM. Stochastic epigenetic mutations (DNA methylation) increase exponentially in human aging and correlate with X chromosome inactivation skewing in females. Aging (Albany NY). 2015; 7:568–78. https://doi.org/10.18632/aging.100792

19. Gentilini D, Scala S, Gaudenzi G, Garagnani P, Capri M, Cescon M, Grazi GL, Bacalini MG, Pisoni S, Dicitore A, Circelli L, Santagata S, Izzo F, et al. Epigenome-wide association study in hepatocellular carcinoma: identification of stochastic epigenetic mutations through an innovative statistical approach. Oncotarget. 2017; 8:41890–902. https://doi.org/10.18632/oncotarget.17462

20. Yamashita S, Kishino T, Takahashi T, Shimazu T, Charvat H, Kakugawa Y, Nakajima T, Lee YC, Iida N, Maeda M, Hattori N, Takeshima H, Nagano R, et al. Genetic and epigenetic alterations in normal tissues have differential impacts on cancer risk among tissues. Proc Natl Acad Sci USA. 2018; 115:1328–33. https://doi.org/10.1073/pnas.1717340115

21. Haque MM, Nilsson EE, Holder LB, Skinner MK. Genomic Clustering of differential DNA methylated regions (epimutations) associated with the epigenetic transgenerational inheritance of disease and phenotypic variation. BMC Genomics. 2016; 17:418. https://doi.org/10.1186/s12864-016-2748-5

22. Austin MK, Chen E, Ross KM, McEwen LM, MacIsaac JL, Kobor MS, Miller GE. Early-life socioeconomic

disadvantage, not current, predicts accelerated epigenetic aging of monocytes. Psychoneuroendocrinology. 2018; 97:131–34. https://doi.org/10.1016/j.psyneuen.2018.07.007

23. Hughes A, Smart M, Gorrie-Stone T, Hannon E, Mill J, Bao Y, Burrage J, Schalkwyk L, Kumari M. Socioeconomic position and DNA Methylation age acceleration across the lifecourse. Am J Epidemiol. 2018; 187:2346–54. https://doi.org/10.1093/aje/kwy155

24. Miller GE, Yu T, Chen E, Brody GH. Self-control forecasts better psychosocial outcomes but faster epigenetic aging in low-SES youth. Proc Natl Acad Sci USA. 2015; 112:10325–30. https://doi.org/10.1073/pnas.1505063112

25. Simons RL, Lei MK, Beach SR, Philibert RA, Cutrona CE, Gibbons FX, Barr A. Economic hardship and biological weathering: the epigenetics of aging in a U.S. sample of black women. Soc Sci Med. 2016; 150:192–200. https://doi.org/10.1016/j.socscimed.2015.12.001

26. Fox JW. Social class, mental illness, and social mobility: the social selection-drift hypothesis for serious mental illness. J Health Soc Behav. 1990; 31:344–53. https://doi.org/10.2307/2136818

27. Quach A, Levine ME, Tanaka T, Lu AT, Chen BH, Ferrucci L, Ritz B, Bandinelli S, Neuhouser ML, Beasley JM, Snetselaar L, Wallace RB, Tsao PS, et al. Epigenetic clock analysis of diet, exercise, education, and lifestyle factors. Aging (Albany NY). 2017; 9:419–46. https://doi.org/10.18632/aging.101168

28. Jylhävä J, Pedersen NL, Hägg S. Biological Age Predictors. EBioMedicine. 2017; 21:29–36. https://doi.org/10.1016/j.ebiom.2017.03.046

29. 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

30. Oey H, Whitelaw E. On the meaning of the word 'epimutation'. Trends Genet. 2014; 30:519–20. https://doi.org/10.1016/j.tig.2014.08.005

31. Peltomäki P. Mutations and epimutations in the origin of cancer. Exp Cell Res. 2012; 318:299–310. https://doi.org/10.1016/j.yexcr.2011.12.001

32. Stringhini S, Carmeli C, Jokela M, Avendaño M, Muennig P, Guida F, Ricceri F, d'Errico A, Barros H, Bochud M, Chadeau-Hyam M, Clavel-Chapelon F, Costa G, et al, and LIFEPATH consortium. Socioeconomic status and the 25 × 25 risk factors as determinants of premature mortality: a multicohort study and meta-analysis of 1·7 million men and women. Lancet. 2017; 389:1229–37. https://doi.org/10.1016/S0140-6736(16)32380-7

33. Stringhini S, Carmeli C, Jokela M, Avendaño M, McCrory C, d'Errico A, Bochud M, Barros H, Costa G, Chadeau-Hyam M, Delpierre C, Gandini M, Fraga S, et al, and LIFEPATH Consortium. Socioeconomic status, non-communicable disease risk factors, and walking speed in older adults: multi-cohort population based study. BMJ. 2018; 360:k1046. https://doi.org/10.1136/bmj.k1046

34. Chen BH, Marioni RE, Colicino E, Peters MJ, Ward-Caviness CK, Tsai PC, Roetker NS, Just AC, Demerath EW, Guan W, Bressler J, Fornage M, Studenski S, et al. DNA methylation-based measures of biological age: meta-analysis predicting time to death. Aging (Albany NY). 2016; 8:1844–65. https://doi.org/10.18632/aging.101020

35. Vineis P, Chatziioannou A, Cunliffe VT, Flanagan JM, Hanson M, Kirsch-Volders M, Kyrtopoulos S. Epigenetic memory in response to environmental stressors. FASEB J. 2017; 31:2241–51. https://doi.org/10.1096/fj.201601059RR

36. Ong ML, Holbrook JD. Novel region discovery method for Infinium 450K DNA methylation data reveals changes associated with aging in muscle and neuronal pathways. Aging Cell. 2014; 13:142–55. https://doi.org/10.1111/acel.12159

37. Zingg D, Debbache J, Schaefer SM, Tuncer E, Frommel SC, Cheng P, Arenas-Ramirez N, Haeusel J, Zhang Y, Bonalli M, McCabe MT, Creasy CL, Levesque MP, et al. The epigenetic modifier EZH2 controls melanoma growth and metastasis through silencing of distinct tumour suppressors. Nat Commun. 2015; 6:6051. https://doi.org/10.1038/ncomms7051

38. von Schimmelmann M, Feinberg PA, Sullivan JM, Ku SM, Badimon A, Duff MK, Wang Z, Lachmann A, Dewell S, Ma'ayan A, Han MH, Tarakhovsky A, Schaefer A. Polycomb repressive complex 2 (PRC2) silences genes responsible for neurodegeneration. Nat Neurosci. 2016; 19:1321–30. https://doi.org/10.1038/nn.4360

39. Su SK, Li CY, Lei PJ, Wang X, Zhao QY, Cai Y, Wang Z, Li L, Wu M. The EZH1-SUZ12 complex positively regulates the transcription of NF-κB target genes through interaction with UXT. J Cell Sci. 2016; 129:2343–53. https://doi.org/10.1242/jcs.185546

40. Shiogama S, Yoshiba S, Soga D, Motohashi H, Shintani S. Aberrant expression of EZH2 is associated with pathological findings and P53 alteration. Anticancer Res. 2013; 33:4309–17.

41. Viechtbauer W. Conducting meta-analyses in R with the metafor package. J Stat Softw. 2010; 36. https://doi.org/10.18637/jss.v036.i03

42. Ziller MJ, Gu H, Müller F, Donaghey J, Tsai LT, Kohlbacher O, De Jager PL, Rosen ED, Bennett DA, Bernstein BE, Gnirke A, Meissner A. Charting a dynamic DNA methylation landscape of the human genome. Nature. 2013; 500:477–81. https://doi.org/10.1038/nature12433

43. Bernstein BE, Stamatoyannopoulos JA, Costello JF, Ren B, Milosavljevic A, Meissner A, Kellis M, Marra MA, Beaudet AL, Ecker JR, Farnham PJ, Hirst M, Lander ES, et al, and The NIH Roadmap Epigenomics Mapping Consortium. The NIH Roadmap Epigenomics Mapping Consortium. Nat Biotechnol. 2010; 28:1045–48. https://doi.org/10.1038/nbt1010-1045

44. Kundaje A, Meuleman W, Ernst J, Bilenky M, Yen A, Heravi-Moussavi A, Kheradpour P, Zhang Z, Wang J, Ziller MJ, Amin V, Whitaker JW, Schultz MD, et al, and Roadmap Epigenomics Consortium. Integrative analysis of 111 reference human epigenomes. Nature. 2015; 518:317–30. https://doi.org/10.1038/nature14248

45. Gel B, Díez-Villanueva A, Serra E, Buschbeck M, Peinado MA, Malinverni R. regioneR: an R/Bioconductor package for the association analysis of genomic regions based on permutation tests. Bioinformatics. 2016; 32:289–91.

46. Elliott P, Vergnaud AC, Singh D, Neasham D, Spear J, Heard A. The Airwave Health Monitoring Study of police officers and staff in Great Britain: rationale, design and methods. Environ Res. 2014; 134:280–85. https://doi.org/10.1016/j.envres.2014.07.025

47. Fiorito G, Vlaanderen J, Polidoro S, Gulliver J, Galassi C, Ranzi A, Krogh V, Grioni S, Agnoli C, Sacerdote C, Panico S, Tsai MY, Probst-Hensch N, et al. Oxidative stress and inflammation mediate the effect of air pollution on cardio- and cerebrovascular disease: A prospective study in nonsmokers. Environ Mol Mutagen. 2018; 59:234–46. https://doi.org/10.1002/em.22153

48. Riboli E, Kaaks R. The EPIC Project: rationale and study design. European Prospective Investigation into Cancer and Nutrition. Int J Epidemiol. 1997 (Suppl 1); 26:S6–14. https://doi.org/10.1093/ije/26.suppl_1.S6

49. Löw M, Stegmaier C, Ziegler H, Rothenbacher D, Brenner H, and ESTHER study. [Epidemiological investigations of the chances of preventing, recognizing early and optimally treating chronic diseases in an elderly population (ESTHER study)]. Dtsch Med Wochenschr. 2004; 129:2643–47.

50. Herder C, Kannenberg JM, Huth C, Carstensen-Kirberg M, Rathmann W, Koenig W, Heier M, Püttgen S, Thorand B, Peters A, Roden M, Meisinger C, Ziegler D. Proinflammatory cytokines predict the incidence and progression of distal sensorimotor polyneuropathy: KORA F4/FF4 Study. Diabetes Care. 2017; 40:569–76. https://doi.org/10.2337/dc16-2259

51. Milne RL, Fletcher AS, MacInnis RJ, Hodge AM, Hopkins AH, Bassett JK, Bruinsma FJ, Lynch BM, Dugué PA, Jayasekara H, Brinkman MT, Popowski LV, Baglietto L, et al. Cohort Profile: The Melbourne Collaborative Cohort Study (Health 2020). Int J Epidemiol. 2017; 46:1757–1757i. https://doi.org/10.1093/ije/dyx085

52. Bell B, Rose CL, Damon A. The Veterans Administration longitudinal study of healthy aging. Gerontologist. 1966; 6:179–84. https://doi.org/10.1093/geront/6.4.179

53. Burns F, Carney GM, Cruise S, Devine P, Devlin A, Donnelly M, French D, Kee F, Montgomery L, O'Reilly D, Scott A, Tully MA. (2017). Early key findings from a study of older people in Northern Ireland. The NICOLA Study. (Belfast: Queen's University, Belfast).

54. Ikram MA, Brusselle GG, Murad SD, van Duijn CM, Franco OH, Goedegebure A, Klaver CC, Nijsten TE, Peeters RP, Stricker BH, Tiemeier H, Uitterlinden AG, Vernooij MW, Hofman A. The Rotterdam Study: 2018 update on objectives, design and main results. Eur J Epidemiol. 2017; 32:807–50. https://doi.org/10.1007/s10654-017-0321-4

55. Ackermann-Liebrich U, Kuna-Dibbert B, Probst-Hensch NM, Schindler C, Felber Dietrich D, Stutz EZ, Bayer-Oglesby L, Baum F, Brändli O, Brutsche M, Downs SH, Keidel D, Gerbase MW, et al, and SAPALDIA Team. Follow-up of the Swiss Cohort Study on Air Pollution and Lung Diseases in Adults (SAPALDIA 2) 1991-2003: methods and characterization of participants. Soz Praventivmed. 2005; 50:245–63. https://doi.org/10.1007/s00038-005-4075-5

56. Pruijm M, Ponte B, Ackermann D, Paccaud F, Guessous I, Ehret G, Pechère-Bertschi A, Vogt B, Mohaupt MG, Martin PY, Youhanna SC, Nägele N, Vollenweider P, et al. Associations of urinary uromodulin with clinical characteristics and markers of tubular function in the general population. Clin J Am Soc Nephrol. 2016; 11:70–80. https://doi.org/10.2215/CJN.04230415

57. Elbaz A, Clavel J, Rathouz PJ, Moisan F, Galanaud JP, Delemotte B, Alpérovitch A, Tzourio C. Professional exposure to pesticides and Parkinson disease. Ann Neurol. 2009; 66:494–504. https://doi.org/10.1002/ana.21717

58. Kearney PM, Cronin H, O'Regan C, Kamiya Y, Savva GM, Whelan B, Kenny R. Cohort profile: the Irish Longitudinal Study on Ageing. Int J Epidemiol. 2011; 40:877–84. https://doi.org/10.1093/ije/dyr116

59. Raitakari OT, Juonala M, Rönnemaa T, Keltikangas-Järvinen L, Räsänen L, Pietikäinen M, Hutri-Kähönen N, Taittonen L, Jokinen E, Marniemi J, Jula A, Telama R, Kähönen M, et al. Cohort profile: the cardiovascular risk in Young Finns Study. Int J Epidemiol. 2008; 37:1220–26. https://doi.org/10.1093/ije/dym225

SUPPLEMENTARY MATERIAL

Please browse the links in Full Text version of this manuscript to see R Scripts.

SUPPLEMENTARY METHODS

Cohort description and variable definition

EPIC - Study participants were drawn from the Italian component of the European Prospective Investigation into Cancer and Nutrition (EPIC) cohort, a large general population cohort consisting of ~520,000 individuals, with standardized lifestyle and personal history questionnaires, measured anthropometric data and blood samples collected for DNA extraction [1]. Socio-economic, dietary and lifestyle-related variables were collected at study enrolment through the use of a validated questionnaires. The highest educational attainment was categorized as follow: 'low' = primary school or lower; 'medium' = secondary school, 'high' = university degree or higher. Smoking was categorized as 'never', 'former' and 'current' smokers based on self-reported information. Alcohol was categorized as 'abstainer', 'occasional' (less than 28 g/day) and 'habitual' drinkers (more than 28 g/day). Physical activity was assessed using the Cambridge Physical Activity Index which combines self-reported occupational activity with time participating in cycling and sports. Participants were divided into 3 categories: 'low' (sedentary job and no recreational activity), 'medium' (at least one of physical job and less than one hour of recreational activity per day), and 'high' (sedentary job with >1 hour of recreational activity per day, standing or physical job with some recreational activity, or a heavy manual job). Height and weight were measured at enrolment with a standardized protocol, and body mass index (BMI) was calculated as the ratio between weight in kg and squared height in meters, treated as categorical variable: normal weight = BMI ≤ 25; overweight = 25 < BMI ≤ 30; obese = BMI ≥ 30

Airwave - The Airwave Health Monitoring Study is an occupational cohort of employees of 28 police forces from across Great Britain. Full details of the cohort and methods are available in Elliott et al [2]. The study started recruitment in 2006 and now contains 53,280 participants. The study received ethical approval from the National Health Service Multi-Site Research Ethics Committee (MREC/13/NW/0588). At the baseline health screening, participants underwent health examination, self-completed a computer questionnaire and blood samples in EDTA tubes for DNA extraction. Blood samples were spun at the health clinic and the biological samples were stored in a Thermoporter (LaminarMedica) and frozen at -80 °C long term storage. Covariates in the analysis were categorised from self-report or clinical data as follows: Education was defined as low (completed GCSEs or equivalent only), medium (completed 'A' levels or equivalent only) or high (completed university or higher degree). Alcohol use was classed as non-drinker, occasional drinker (≤ 14 alcohol units/week for women and ≤ 21 alcohol units/week for men) or habitual drinker (> 14 alcohol units/week for women and >21 alcohol units/week for men). Physical activity was defined as low, moderate or high based on the scoring protocol of the International Physical Activity Questionnaire [3]. Smoking was categorized as 'never', 'former' and 'current' smokers based on self-reported information. Height and weight were measured at enrolment with a standardized protocol, and body mass index (BMI) was calculated as the ratio between weight in kg and squared height in meters, treated as categorical variable: normal weight = BMI ≤ 25; overweight = 25 < BMI ≤ 30; obese = BMI ≥ 30

The *ESTHER* study is an ongoing population-based cohort study conducted in the federal state of Saarland, Germany [4]. In brief, 9,949 older adults (50-75 years) were recruited by their general practitioners (GPs) during routine health check-ups (offered every two years to people older than 35 years in the German healthcare system) between 2000 and 2002, and followed up thereafter. During the baseline enrolment, epidemiological data (including socio-demographic characteristics, lifestyle factors, and history of major diseases) were collected via a standardized self-administered questionnaire completed by participants and via additional reports from participants' GPs, and biological samples (blood, stool, urine) were obtained and stored at −80 °C. Educational levels were defined as low [≤9 years], medium [10-11 years], and high [≥= 12 years]. Smoking behaviours were based on self-reported information and classified according to commonly used criteria. An ever-smoker was defined as a subject who had ever smoked ≥100 cigarettes during his or her lifetime, thus excluding rare occasional smoking. An ever-smoker was classified as a former smoker if he or she had stopped smoking for ≥1 year prior to the study. Body mass index (BMI) were categorized as underweight (<18.5 kg/m^2), normal weight (18.5 to <25.0 kg/m^2), overweight (25.0 to <30.0 kg/m^2), or obese (≥30.0 kg/m^2)]. Physical activity were categorized as inactive (<1 hour/week of physical activity), medium/high (≥2 hour/week of vigorous physical activity or ≥2 hour/week of light physical activity), or low (all others)]. Alcohol was categorized as abstainer (0 gram/day), occasional drinker (≤28 gram/day), and habitual drinkers (>28 gram/day). Two subsets of ESTHER participants were selected for DNA methylation assessment in the baseline blood samples: Subset I consists of 1,000 participants

consecutively enrolled during the first 3 months of recruitment; Subset II consists of 864 participants selected for a case-cohort design for mortality analysis [5]. The study was approved by the ethics committees of the University of Heidelberg and of the Medical Association of Saarland. All participants provided written informed consent.

KORA - This study is based on data from participants of four independent cross-sectional surveys (S1–S4) of the KORA (Cooperative Health Research in the Region of Augsburg) project between 1984 and 2001 [6], as well as from participants from KORA T2DM Family Study (T2DMFAM19 [7]), which was performed in 2001 / 2002. All probands were from the city or region of Augsburg. All participants were living in Germany and all were of European origin. DNA methylation was performed using the Illumina 450K BeadChip array. The highest educational attainment was categorized as follow: 'low' = primary school or lower; 'medium' = secondary school, 'high' = university degree or higher. Smoking was categorized as 'never', 'former' and 'current' smokers based on self-reported information. Alcohol was categorized as 'abstainer', 'occasional' (less than 28 g/day) and 'habitual' drinkers (more than 28 g/day). Height and weight were measured at enrolment with a standardized protocol, and body mass index (BMI) was calculated as the ratio between weight in kg and squared height in meters, treated as categorical variable: normal weight = BMI ≤ 25; overweight = 25 < BMI ≤ 30; obese = BMI ≥ 30.

MCCS - We used data from studies nested within the Melbourne Collaborative Cohort Study (Melbourne, Victoria, Australia), a prospective cohort study of 41,513 healthy adult volunteers (24,469 women) aged 27–76 years (99.3% were aged 40–69 years) at baseline between 1990 and 1994 (*MILNE, INT J EPIDEMIOL, 2017*). DNA samples used for the present analysis were extracted from peripheral blood drawn at the time of recruitment (1990–1994). For the majority (70%) of participants, the DNA source was dried blood spots collected onto Guthrie Card Diagnostic Cellulose filter paper (Whatman plc, Kent, United Kingdom) and stored in airtight containers at room temperature. The other sources of DNA were peripheral blood mononuclear cells and buffy coats stored at −80°C for 28% and 2% participants, respectively.

The study sample comprised Melbourne Collaborative Cohort Study participants selected as controls in nested case-control studies of breast, colorectal, kidney, lung, prostate, or urothelial cancer or mature B-cell malignancies [8-11]. Controls had been individually matched to cases on age (they had to be free of cancer at an age within 1 year of the age at diagnosis of the corresponding case), sex, country of birth, and blood DNA source (dried blood spot, peripheral blood mononuclear cells, or buffy coat). For all but the colorectal cancer study, controls were matched to cases on year of birth. For the lung cancer study, controls were matched on smoking status at the time of blood collection.

Socio-economic, dietary and lifestyle-related variables were collected at study enrolment through the use of a validated questionnaires. The highest educational attainment was categorized as follows: 'low' = primary school or lower; 'medium' = secondary school, 'high' = university degree or higher. Smoking was categorized as 'never', 'former' and 'current' smokers based on self-reported information. Alcohol was categorized as 'abstainer', 'occasional' (less than 28 g/day) and 'habitual' drinkers (more than 28 g/day). Height and weight were measured at enrolment with a standardized protocol, and body mass index (BMI) was calculated as the ratio between weight in kg and squared height in meters, treated as categorical variable: normal weight = BMI ≤ 25; overweight = 25 < BMI ≤ 30; obese = BMI ≥ 30. Physical activity was 1 to 4 and reflecting metabolic equivalents, as described in previous. Physical activity was defined by a summary score aimed to reflect the total energy expenditure as described in MacInnis et al., and was based on questions relating to frequency of walking, vigorous exercise (exercise 'making you sweat or feel out of breath, and includes such activities as swimming, tennis, netball, athletics, and running') and less vigorous exercise (exercise 'which did not make you sweat or feel out of breath and includes such activities as bike riding, dancing, etc.') over the last 6 months [12].

NAS - The Normative Aging Study (NAS) is an ongoing longitudinal male cohort established in 1963. Men were free of known chronic clinical conditions at enrolment and were subsequently invited to clinical examinations every 3 to 5 years [13]. At each visit, participants provided information on medical history, lifestyle, and demographic factors, and underwent a physical examination and laboratory tests. The NAS study was approved by the Institutional Review Boards (IRBs) of the participating institutions. Participants have provided written informed consent at each visit.

DNA samples were collected from 1999 to 2007 from the 675 active participants and used for DNA methylation analysis. We excluded participants who were not of European descent or had missing information on race, other covariates or with leukaemia or any blood cancer, leaving a total of 624 individuals for the analysis.

At each in-person examination visit, participants provide demographic information and completed a questionnaire enquiring about their smoking status, education, alcohol

consumption and life-style, including a measure of physical activity (metabolic equivalent of task: MET). Anthropometric measurements (height and weight) were also performed with participants in undershorts and socks. All variables were harmonized and categorized as in the EPIC study, except for smoking (ever/never), alcohol consumption (≤2 or >2 drinks/day), and physical activity (≤10 MET hours/week, 10<MET hours/week or >25 MET hours/week).

NICOLA – The Northern Ireland Cohort for the Longitudinal Study of Ageing is a longitudinal cohort representative of the non-institutionalized population of Northern Ireland over the age of 50 years (n=8,500) [14]. The study which was established in 2013 has three main components: a computer aided personal interview (CAPI), a self-completion questionnaire and health assessment. Dietary intake was also assessed by a food frequency questionnaire. The CAPI was extensive in scope and included assessment of demographic, social and health-related factors. Measures of cardiovascular, physical, cognitive and visual function were determined and a biobank of biological samples collected. Educational attainment was categorized as follows: 'low' = primary school or lower; 'medium' = secondary school, 'high' = higher education. Smoking was categorized as 'never', 'former' or 'current' based on self-reported information. Alcohol was categorized as 'abstainer', 'occasional' (on average less than one alcoholic beverage/day) and 'habitual' drinkers (on average one or more alcoholic beverages/day). Physical activity was defined as low, moderate or high based on the scoring protocol of the International Physical Activity Questionnaire (3). Physical activity was categorized as 'high' = highest tertile of metabolic equivalent (MET) computed based on self-reported frequency and duration of physical activity; 'medium' = middle tertile of MET; 'low' = lowest tertile of MET. Height and weight were measured at the health assessment with a standardized protocol, and body mass index (BMI) was calculated as the ratio between weight in kg and height in meters squared, treated as categorical variable: normal weight = BMI ≤ 25; overweight = 25 < BMI ≤ 30; obese = BMI ≥ 30.

Rotterdam Study - The Rotterdam Study is a prospective population based-study started in 1989. It is composed of residents of the neighborhood of Ommoord, Rotterdam, the Netherlands, aged 45 years and over. Data on socio-economic status, diet and lifestyle factors were assessed by standardized questionnaires, measured anthropometric data and blood samples collected for DNA extraction. Education was categorized into three groups: i) 'low' = primary school or lower; ii) 'medium' = secondary school; iii) 'high' = university degree or higher. Smoking was categorized as 'never', 'former' and 'current' smokers based on self-reported information. Physical activity level was measured with a self-administrated LASA Physical Activity questionnaire (LAPAQ). Later, the intensity of the reported activities was quantified using the metabolic equivalent of task (MET) hours per week, and then categorized in three groups: i) sedentary: <10 MET hours/week; ii) moderately active: 10–40 MET hours/week and iii) active: > 40 MET hours/week. Alcohol consumption was categorized into i) abstainer: consumption of 0 grams/day of alcohol, ii) moderate drinker: consumption > 0 grams/day and ≥ 28 grams/day, ii) habitual drinker: consumption > 28 grams/day. Height and weight were measured with a standardized protocol. Body mass index (BMI) was calculated as the ratio between weight in kg and squared height in meters, and then categorized in three groups: i) normal weight: BMI ≤ 25; ii) overweight: 25 < BMI ≤ 30; iii) obese = BMI ≥ 30

SAPALDIA - Swiss Study on Air Pollution and Lung and Heart Disease in Adults is a population cohort in Switzerland initiated in 1991 recruiting 9651 random samples from eight cities covering geographical, meteorological, and cultural diversity of the population. The SAPALDIA cohort has been described in detail previously [15]. Blood samples collected at the second follow-up in 2010-11 and stored at -80 °C. DNA methylation was analyzed in the framework of EXPOsOMICS for a total of 402 samples selected based on asthma status and the availability of archived blood samples and covariate information. Self-reported education level was categorized as 'high' = technical college or university; 'medium' = secondary school, middle school or apprenticeship; 'low' = primary school. Smoking status was categorized as 'never' or 'former' smokers based on self-reported information. Current smokers were excluded. Alcohol consumption was categorized as 'abstainer' = never; 'occasional' = rarely, 1-2 times per week, or several times per week; 'habitual' = once per day, twice per day, 3 times or more per day. Physical activity was categorized as 'high' = highest tertile of metabolic equivalent (MET) computed based on self-reported frequency and duration of physical activity; 'medium' = middle tertile of MET; 'low' = lowest tertile of MET. Vigorous physical activity was given 6 MET while moderate activity 3 MET. Weight and height were measured during the health examination. BMI was computed as weight in kilogram divided by squared height in meter and categorized as 'normal' = BMI < 25; 'overweight' = 25 ≤ BMI < 30; 'obese' = BMI ≥ 30.

SKIPOGH - The Swiss Kidney Project on Genes in Hypertension (SKIPOGH) study is a multicenter family-based population study initiated in 2009 to explore the genetic and environmental determinants of BP [16]. Study participants were recruited in the cantons of Bern and Geneva and the city of Lausanne. Recruitment began in

December 2009 and ended in April 2013. Inclusion criteria were: (i) written informed consent, (ii) minimum age of 18 years, (iii) Caucasian origin, and (iv) at least one, and preferably 3, first-degree family members also willing to participate. At the end of the recruitment period, the study population included 1,128 participants from 271 distinct family pedigrees. Of the individuals asked to participate in Bern, Geneva, and Lausanne, 21%, 22%, and 20% agreed, respectively. The SKIPOGH study was approved by the ethical committees of Lausanne University Hospital, Geneva University Hospital, and the University Hospital of Bern. Data were from the first follow-up which started in 2013, but highest attained education that was collected at baseline. Covariates in the analysis were categorised from self-report or clinical data as follows: Education was defined as low (no diploma or mandatory school or secondary vocational training), medium (secondary vocational training of superior level or superior non-university training) or high (university degree). Alcohol use was classed as non-drinker, occasional drinker (≤ 14 alcohol units/week for women and ≤ 21 alcohol units/week for men) or habitual drinker (> 14 alcohol units/week for women and >21 alcohol units/week for men). Physical activity was defined as low, moderate or high based on a question about overall physical activity: "Please indicate on a scale from 1-10 the physical efforts that you are doing on a daily basis, including those at work, during sports and your free time activities" [3]. Smoking was categorized as 'never', 'former' and 'current' smokers based on self-reported information. Height and weight were measured with a standardized protocol, and body mass index (BMI) was calculated as the ratio between weight in kg and squared height in meters, treated as categorical variable: normal weight = BMI ≤ 25; overweight = 25 < BMI ≤ 30; obese = BMI ≥ 30. An EDTA whole blood collection vessel was used (BD, Franklin Lakes, New Jersey). DNA was extracted using standard methods on a bead-based KingFisher Duo robot extraction system (ThermoFisher, Waltham, Massachusetts). DNA quality assessement and quantification was performed using a Nanodrop system (ThermoFisher, Waltham, Massachusetts). For bisulfite conversion, the protocol started with ~1.2ug of DNA extracted. For the PCR step: alternative incubation conditions was performed when using the Illumina Infinium® Methylation Assay (Appendix page 6 of bisulfite conversion protocol pdf). The final elution was done with 8ul of M-Elution Buffer. Processing pipeline of the beta methylation values was CPACOR [17].

TERRE: We used data from population-based controls of a case-control of Parkinson's disease in French farmers [18]. We included Parkinson's disease patients enrolled in the French health insurance system for farmers (MSA) from 62 metropolitan districts (1998-1999). We randomly selected eligible controls from among all MSA members who requested reimbursement for health expenses (participation rate = 83%). Controls were matched to cases on age, sex, and district of residency, and did not report cardinal signs of Parkinson's disease. DNA was extracted from peripheral blood leukocytes. Education was defined as low (no education or primary school), medium (certificate level) or high (secondary school to university degree). Alcohol use was classed as non-drinker, occasional drinker (e.g., events, family celebrations) or habitual drinker (regularly or daily drinker). Physical activity was not assessed. Smoking was categorized as 'never', 'former' and 'current' smokers based on self-reported information. Height and weight were self-reported, and body mass index (BMI) was calculated as the ratio between weight in kg and squared height in meters, treated as categorical variable: normal weight = BMI ≤ 25; overweight = 25 < BMI ≤ 30; obese = BMI ≥ 30.

The Irish Longitudinal Study on Ageing (TILDA) is a large prospective cohort study examining the social, economic and health circumstances of 8,175 community-dwelling older adults aged 50 years and over resident in the Republic of Ireland. The sample was generated using a 3-stage selection process and the Irish Geodirectory as the sampling frame. The Irish Geodirectory is a comprehensive listing of all addresses in the Republic of Ireland, which is compiled by the national post service and ordnance survey Ireland. Subdivisions of district electoral divisions pre-stratified by socio-economic status, age, and geographical location, served as the primary sampling units. The second stage involved the selection of a random sample of 40 addresses from within each PSU resulting in an initial sample of 25,600 addresses. The third stage involved the recruitment of all members of the household aged 50 years and over. Consequently, the response rate was defined as the proportion of households including an eligible participant from whom an interview was successfully obtained. A response rate of 62% was achieved at the household level. There were three components to the survey. Respondents completed a computer-assisted personal interview and a separate self-completion paper and pencil module which collected information that was considered sensitive. All participants were invited to undergo an independent health assessment at one of two national centers using trained nursing staff. Blood samples were taken during the clinical assessment with the consent of participants. A more detailed exposition of study design, sample selection and protocol is available elsewhere [19]. The present study sample included 500 healthy individuals: 125 for each of the four SES classes: stable professional, any downward mobility, any upward mobility, and stable unskilled (see socioeconomic position assessment). Buffy coat or peripheral blood mononuclear cells (PBMC) samples were available for all the individuals. Overall, after DNA

methylation data quality controls and sample filtering, 490 subjects were analyzed in this study.

Effects size comparison between SEMs and epigenetic clocks

For the three epigenetic clocks, the estimated differences presented in Table 2 (βs) represent the change in biological age (in years) compared with the reference group. In order to make the effect sizes of the logSEM variable comparable with those of the three epigenetic clocks (i.e. expressed as years of increasing biological age), we re-scaled both the effect sizes and the standard deviations by a factor $\sigma = \sigma_{EC} / \sigma_{SEMs}$, where σ_{EC} is the average standard deviation of the three epigenetic clocks and σ_{SEMs} is the standard deviation of the logSEM variable. In this way, the re-scaled effect size of logSEM can be interpreted as years of increasing biological age as is the case for the three epigenetic clocks.

SUPPLEMENTARY REFERENCES

1. Riboli E, Kaaks R. The EPIC Project: rationale and study design. European Prospective Investigation into Cancer and Nutrition. Int J Epidemiol. 1997 (Suppl 1); 26:S6–14. https://doi.org/10.1093/ije/26.suppl_1.S6

2. Elliott P, Vergnaud AC, Singh D, Neasham D, Spear J, Heard A. The Airwave Health Monitoring Study of police officers and staff in Great Britain: rationale, design and methods. Environ Res. 2014; 134:280–85. https://doi.org/10.1016/j.envres.2014.07.025

3. The IPAQ group. International Physical Activity Questionnaire. 2016. https://sites.google.com/site/theipaq/questionnaire_links

4. Raum E, Rothenbacher D, Löw M, Stegmaier C, Ziegler H, Brenner H. Changes of cardiovascular risk factors and their implications in subsequent birth cohorts of older adults in Germany: a life course approach. Eur J Cardiovasc Prev Rehabil. 2007; 14:809–14. https://doi.org/10.1097/HJR.0b013e3282eeb308

5. Zhang Y, Wilson R, Heiss J, Breitling LP, Saum KU, Schöttker B, Holleczek B, Waldenberger M, Peters A, Brenner H. DNA methylation signatures in peripheral blood strongly predict all-cause mortality. Nat Commun. 2017; 8:14617. https://doi.org/10.1038/ncomms14617

6. Wichmann HE, Gieger C, Illig T, and MONICA/KORA Study Group. KORA-gen--resource for population genetics, controls and a broad spectrum of disease phenotypes. Gesundheitswesen. 2005 (Suppl 1); 67:S26–30. https://doi.org/10.1055/s-2005-858226

7. Huth C, Illig T, Herder C, Gieger C, Grallert H, Vollmert C, Rathmann W, Hamid YH, Pedersen O, Hansen T, Thorand B, Meisinger C, Doring A, et al. Joint analysis of individual participants' data from 17 studies on the association of the IL6 variant -174G>C with circulating glucose levels, interleukin-6 levels, and body mass index. Ann Med. 2009; 41:128–38. https://doi.org/10.1080/07853890802337037

8. Severi G, Southey MC, English DR, Jung CH, Lonie A, McLean C, Tsimiklis H, Hopper JL, Giles GG, Baglietto L. Epigenome-wide methylation in DNA from peripheral blood as a marker of risk for breast cancer. Breast Cancer Res Treat. 2014; 148:665–73. https://doi.org/10.1007/s10549-014-3209-y

9. Dugué PA, Brinkman MT, Milne RL, Wong EM, FitzGerald LM, Bassett JK, Joo JE, Jung CH, Makalic E, Schmidt DF, Park DJ, Chung J, Ta AD, et al. Genome-wide measures of DNA methylation in peripheral blood and the risk of urothelial cell carcinoma: a prospective nested case-control study. Br J Cancer. 2016; 115:664–73. https://doi.org/10.1038/bjc.2016.237

10. Wong Doo N, Makalic E, Joo JE, Vajdic CM, Schmidt DF, Wong EM, Jung CH, Severi G, Park DJ, Chung J, Baglietto L, Prince HM, Seymour JF, et al. Global measures of peripheral blood-derived DNA methylation as a risk factor in the development of mature B-cell neoplasms. Epigenomics. 2016; 8:55–66. https://doi.org/10.2217/epi.15.97

11. Baglietto L, Ponzi E, Haycock P, Hodge A, Bianca Assumma M, Jung CH, Chung J, Fasanelli F, Guida F, Campanella G, Chadeau-Hyam M, Grankvist K, Johansson M, et al. DNA methylation changes measured in pre-diagnostic peripheral blood samples are associated with smoking and lung cancer risk. Int J Cancer. 2017; 140:50–61. https://doi.org/10.1002/ijc.30431

12. MacInnis RJ, English DR, Hopper JL, Haydon AM, Gertig DM, Giles GG. Body size and composition and colon cancer risk in men. Cancer Epidemiol Biomarkers Prev. 2004; 13:553–59.

13. Wilker E, Korrick S, Nie LH, Sparrow D, Vokonas P, Coull B, Wright RO, Schwartz J, Hu H. Longitudinal changes in bone lead levels: the VA Normative Aging Study. J Occup Environ Med. 2011; 53:850–55. https://doi.org/10.1097/JOM.0b013e31822589a9

14. Burns F, Carney GM, Cruise S, Devine P, Devlin A, Donnelly M, French D, Kee F, Montgomery L, O'Reilly D, Scott A, Tully MA. (2017). Early key findings from a study of older people in Northern Ireland. The NICOLA Study. (Belfast: Queen's University, Belfast).

15. Ackermann-Liebrich U, Kuna-Dibbert B, Probst-Hensch NM, Schindler C, Felber Dietrich D, Stutz EZ, Bayer-Oglesby L, Baum F, Brändli O, Brutsche M, Downs SH, Keidel D, Gerbase MW, et al, and SAPALDIA Team. Follow-up of the Swiss Cohort Study on Air Pollution and Lung Diseases in Adults (SAPALDIA 2) 1991-2003: methods and characterization of participants. Soz Praventivmed. 2005; 50:245–63. https://doi.org/10.1007/s00038-005-4075-5

16. Pruijm M, Ponte B, Ackermann D, Paccaud F, Guessous I, Ehret G, Pechère-Bertschi A, Vogt B, Mohaupt MG, Martin PY, Youhanna SC, Nägele N, Vollenweider P, et al. Associations of urinary uromodulin with clinical characteristics and markers of tubular function in the general population. Clin J Am Soc Nephrol. 2016; 11:70–80. https://doi.org/10.2215/CJN.04230415

17. 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

18. Elbaz A, Clavel J, Rathouz PJ, Moisan F, Galanaud JP, Delemotte B, Alpérovitch A, Tzourio C. Professional exposure to pesticides and Parkinson disease. Ann Neurol. 2009; 66:494–504. https://doi.org/10.1002/ana.21717

19. Whelan BJ, Savva GM. Design and methodology of the Irish Longitudinal Study on Ageing. J Am Geriatr Soc. 2013 (Suppl 2); 61:S265–68. https://doi.org/10.1111/jgs.12199

SUPPLEMENTARY TABLES

Table S1. Results of linear regressions using epigenetic aging biomarkers (WBC adjusted) as outcomes and lifestyle related risk factors as predictors.

		SEMs		HorvathEAA	
		Model 1	**Model 2**	**Model 1**	**Model 2**
Education (ref: High)	Medium	**0.28 (0.06; 0.49)***	**0.23 (-0.02; 0.48)**$^{+}$	0.1 (-0.08; 0.28)	0.09 (-0.1; 0.29)
	High	**0.32 (0.09; 0.54)****	**0.27 (0.04; 0.5)***	**0.19 (-0.01; 0.38)**$^{+}$	0.14 (-0.06; 0.34)
Smoking (ref: Never)	Former	**0.24 (0.08; 0.4)****	**0.28 (0.12; 0.43)*****	**0.18 (0.02; 0.35)***	**0.17 (0.01; 0.32)***
	Current	**0.54 (0.32; 0.76)*****	**0.54 (0.32; 0.77)*****	0.12 (-0.07; 0.31)	0.1 (-0.1; 0.3)
Obesity (ref: BMI < 25)	BMI < 30	0.02 (-0.17; 0.2)	0 (-0.18; 0.17)	**0.37 (0.22; 0.52)*****	**0.35 (0.19; 0.5)*****
	BMI ≥ 30	-0.11 (-0.33; 0.1)	-0.13 (-0.34; 0.08)	**0.45 (0.27; 0.63)*****	**0.44 (0.25; 0.62)*****
Alcohol (ref: Abstainer)	Occasional	**-0.16 (-0.35; 0.03)**$^{+}$	-0.14 (-0.33; 0.04)	-0.01 (-0.19; 0.16)	0.02 (-0.16; 0.19)
	Habitual	0.18 (-0.06; 0.42)	0.13 (-0.12; 0.38)	**0.2 (-0.03; 0.44)**$^{+}$	**0.26 (0.03; 0.49)***
Physical activity (ref: High)	Medium	-0.04 (-0.26; 0.18)	-0.06 (-0.24; 0.13)	0.08 (-0.08; 0.24)	0.07 (-0.09; 0.24)
	Low	0.03 (-0.27; 0.33)	-0.02 (-0.3; 0.26)	**0.22 (0.05; 0.4)***	**0.19 (0.01; 0.37)***
		HannumEAA		**LevineEAA**	
		Model 1	**Model 2**	**Model 1**	**Model 2**
Education (ref: High)	Medium	**0.3 (0.14; 0.46)*****	**0.24 (0.06; 0.42)****	**0.36 (0.05; 0.68)***	0.16 (-0.23; 0.55)
	High	**0.32 (0.16; 0.48)*****	**0.28 (0.12; 0.43)*****	**0.85 (0.53; 1.17)*****	**0.57 (0.25; 0.89)*****
Smoking (ref: Never)	Former	**0.15 (0.03; 0.26)***	**0.1 (-0.01; 0.22)**$^{+}$	**0.68 (0.49; 0.87)*****	**0.58 (0.38; 0.78)*****
	Current	**0.46 (0.26; 0.65)*****	**0.4 (0.21; 0.6)*****	**1.62 (1.35; 1.89)*****	**1.48 (1.19; 1.77)*****
Obesity (ref: BMI < 25)	BMI < 30	**0.2 (0.08; 0.32)****	**0.19 (0.06; 0.32)****	**0.59 (0.3; 0.88)*****	**0.5 (0.26; 0.74)*****
	BMI ≥ 30	**0.25 (0.11; 0.39)*****	**0.24 (0.1; 0.39)*****	**1.25 (0.92; 1.58)*****	**1.18 (0.87; 1.48)*****
Alcohol (ref: Abstainer)	Occasional	-0.02 (-0.15; 0.11)	0.05 (-0.09; 0.18)	-0.07 (-0.33; 0.19)	0.09 (-0.14; 0.32)
	Habitual	**0.24 (0.06; 0.42)****	**0.3 (0.1; 0.49)****	**0.91 (0.61; 1.2)*****	**0.99 (0.66; 1.31)*****
Physical activity (ref: High)	Medium	0.09 (-0.07; 0.24)	0.07 (-0.08; 0.22)	0.17 (-0.17; 0.52)	0.12 (-0.13; 0.37)
	Low	0.13 (-0.11; 0.37)	0.07 (-0.19; 0.33)	0.49 (-0.01; 0.98)$^{+}$	0.27 (-0.13; 0.67)

***** p < 0.001; ** p < 0.01; * p < 0.05; + p < 0.10**

Model 1 includes age, sex, and cohort specific covariates; Model 2 includes additional adjustment for education, smoking, BMI, alcohol and physical activity.

Table S2. Interaction with age and sex.

		SEMs		HorvathEAA	
		age	**sex**	**age**	**sex**
Education (ref: High)	Medium	0.02 (-0.01; 0.05)	**-0.08 (-0.16; 0)**$^{+}$	0 (-0.03; 0.03)	0.02 (-0.39; 0.43)
	High	0.03 (0; 0.05)$^{+}$	-0.05 (-0.13; 0.02)	-0.01 (-0.04; 0.02)	0.19 (-0.22; 0.59)
Smoking (ref: Never)	Former	0.01 (0; 0.03)	-0.07 (-0.15; 0)$^{+}$	**0.02 (0; 0.04)***	0 (-0.4; 0.4)
	Current	**0.04 (0.01; 0.06)****	-0.06 (-0.14; 0.02)	**0.03 (0.01; 0.05)***	0.21 (-0.19; 0.61)
Obesity (ref: BMI < 25)	BMI < 30	**0.03 (0.01; 0.06)***	-0.06 (-0.14; 0.02)	0 (-0.03; 0.03)	0.03 (-0.37; 0.43)
	BMI ≥ 30	**0.02 (0; 0.04)**$^{+}$	-0.05 (-0.13; 0.02)	0 (-0.02; 0.02)	0.23 (-0.17; 0.63)
Alcohol (ref: Abstainer)	Occasional	-0.01 (-0.04; 0.01)	-0.08 (-0.15; 0)$^{+}$	-0.01 (-0.03; 0.02)	0.01 (-0.39; 0.41)
	Habitual	0.01 (-0.01; 0.04)	-0.06 (-0.14; 0.02)	0 (-0.02; 0.03)	0.19 (-0.21; 0.59)
Physical activity (ref: High)	Medium	0 (-0.03; 0.02)	-0.08 (-0.16; 0)$^{+}$	-0.02 (-0.04; 0)	0 (-0.4; 0.4)
	Low	0 (-0.02; 0.03)	-0.06 (-0.14; 0.02)	-0.01 (-0.03; 0.02)	0.16 (-0.24; 0.57)
		HannumEAA		**LevineEAA**	
		Age	**Sex**	**Age**	**sex**
Education (ref: High)	Medium	0.01 (-0.01; 0.02)	0.03 (-0.32; 0.38)	0 (-0.04; 0.03)	0.13 (-0.44; 0.7)
	High	**-0.02 (-0.04; 0)***	0.1 (-0.22; 0.43)	-0.01 (-0.05; 0.02)	-0.08 (-0.74; 0.59)
Smoking (ref: Never)	Former	**0.01 (0; 0.03)***	-0.01 (-0.37; 0.35)	**0.02 (0; 0.04)**$^{+}$	0.09 (-0.47; 0.65)
	Current	**0.04 (0.02; 0.06)*****	0.11 (-0.21; 0.43)	**0.03 (0; 0.07)**$^{+}$	-0.06 (-0.72; 0.6)
Obesity (ref: BMI < 25)	BMI < 30	-0.01 (-0.03; 0.02)	0.05 (-0.29; 0.4)	0 (-0.03; 0.04)	0.11 (-0.45; 0.67)
	BMI ≥ 30	-0.01 (-0.02; 0.01)	0.13 (-0.19; 0.46)	0 (-0.03; 0.03)	-0.04 (-0.7; 0.63)
Alcohol (ref: Abstainer)	Occasional	0 (-0.02; 0.02)	0.03 (-0.31; 0.37)	0.01 (-0.02; 0.04)	0.08 (-0.48; 0.64)
	Habitual	0 (-0.02; 0.02)	0.1 (-0.22; 0.43)	0.01 (-0.02; 0.05)	-0.08 (-0.75; 0.58)
Physical activity (ref: High)	Medium	0 (-0.02; 0.01)	0 (-0.35; 0.35)	**-0.03 (-0.06; 0)**$^{+}$	0.12 (-0.44; 0.68)
	Low	0 (-0.02; 0.02)	0.07 (-0.25; 0.4)	0 (-0.04; 0.03)	-0.07 (-0.73; 0.59)

***** p < 0.001; ** p < 0.01; * p < 0.05; + p < 0.10**

Table S3. Results of the enrichment analyses using the classification of the UCSC Genome Browser for the relationship with CpG islands, open chromatin state and DNase hypersensitivity.

Relation to CpG island according to UCSC Genome Browser	Permutation based p-value for enrichment
Shores	0.02
Open chromatin evidence	0.03
CpG Island	0.27
Shelves	1
Non CpG Island	1
DNase hypersensitivity evidence	1

P-values were computed according to the algorithm implemented in the *regioneR* R package.

Table S4. Results of the enrichment analyses using the ENCODE classification for chromatin states in embryonic stem cell (H1-hESC).

Chromatin state according to ENCODE ChIP-Seq	Permutation based p-value for enrichment
Heterochromatin / Low signal / CNV	< 0.0001
Inactive / Poised promoter	< 0.0001
Polycomb Repressed	0.001
Transcriptional elongation / Transition	1
Weak Transcribed	1
Active promoter	1
Weak promoter	1
Strong enhancer	1
Weak / Poised enhancer	1
Insulator	1
Non regulatory elements	1

P-values were computed according to the algorithm implemented in the *regioneR* R package.

Table S5. Results of the enrichment analyses using the ENCODE classification for 58 protein TFBS in non-treated embryonic stem cell (H1-hESC).

Transcription Factor	Lab	Permutation based p-value for enrichment
EZH2	Broad	< 0.0001
SUZ12	USC	< 0.0001
CtBP2	USC	0.15
CHD1	Broad	0.27
NRSF	HudsonAlpha	0.88
BCL11A	HudsonAlpha	1
c-Myc	UT-A	1
POU5F1	HudsonAlpha	1
MafK	Stanford	1
RXRA	HudsonAlpha	1
NANOG	HudsonAlpha	1
CHD1	Stanford	1
c-Jun	Stanford	1
RFX5	Stanford	1
HDAC2	HudsonAlpha	1
FOSL1	HudsonAlpha	1
TCF12	Stanford	1
CEBPB	HudsonAlpha	1
TEAD4	USC	1
Max	UT-A	1
CTCF	Stanford	1
Rad21	Stanford	1
Rad21	HudsonAlpha	1
BRCA1	Stanford	1
CTCF	Broad	1
CTCF	Stanford	1
USF2	Stanford	1
SP2	HudsonAlpha	1
JARID1A	Broad	1
SIX5	HudsonAlpha	1
SRF	HudsonAlpha	1
USF-1	HudsonAlpha	1
ATF2	HudsonAlpha	1
JunD	Stanford	1
ATF3	HudsonAlpha	1
Bach1	Stanford	1
Nrf1	Stanford	1
c-Myc	Stanford	1

JunD	HudsonAlpha	1
GABP	HudsonAlpha	1
GTF2F1	Stanford	1
p300	HudsonAlpha	1
Egr-1	HudsonAlpha	1
Mxi1	Stanford	1
CHD2	Stanford	1
Znf143	Stanford	1
Sin3Ak-20	HudsonAlpha	1
RBBP5	Broad	1
YY1	HudsonAlpha	1
SP1	HudsonAlpha	1
SP4	HudsonAlpha	1
TAF7	HudsonAlpha	1
TBP	Stanford	1
Pol2	HudsonAlpha	1
SIN3A	Stanford	1
Pol2	UT-A	1
TAF1	HudsonAlpha	1
Pol2	HudsonAlpha	1

P-values were computed according to the algorithm implemented in the *regioneR* R package.