Research Paper

Blood DNA methylation sites predict death risk in a longitudinal study of 12,300 individuals

Elena Colicino[1,*], Riccardo Marioni[2,*], Cavin Ward-Caviness[3,33,*], Rahul Gondalia[4,*], Weihua Guan[5,*], Brian Chen[6,*], Pei-Chien Tsai[7,*], Tianxiao Huan[8,*], Gao Xu[9,*], Agha Golareh[9], Joel Schwartz[10], Pantel Vokonas[11], Allan Just[1], John M. Starr[12], Allan F. McRae[13], Naomi R. Wray[13], Peter M. Visscher[13], Jan Bressler[14], Wen Zhang[15], Toshiko Tanaka[6], Ann Zenobia Moore[6], Luke C. Pilling[16], Guosheng Zhang[17], James D. Stewart[4], Yun Li[4], Lifang Hou[18], Juan Castillo-Fernandez[7], Tim Spector[7], Douglas P. Kiel[19], Joanne M. Murabito[20], Chunyu Liu[21], Mike Mendelson[22], Tim Assimes[23], Devin Absher[24], Phil S. Tsaho[25], Ake T. Lu[25], Luigi Ferrucci[26], Rory Wilson[27,33], Melanie Waldenberger[27,33], Holger Prokisch[28], Stefania Bandinelli[29,#], Jordana T. Bell[7,#], Daniel Levy[30,#], Ian J. Deary[31,#], Steve Horvath[25,#], Jim Pankow[32,#], Annette Peters[33,#], Eric A. Whitsel[4,#], Andrea Baccarelli[9,#]

[1]Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA
[2]Centre for Genomic and Experimental Medicine, Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh EH4 2XU, UK
[3]US Environmental Protection Agency, Chapel Hill, NC 27514, USA,
[4]Gillings School of Global Public Health, University of North Carolina, Chapel Hill, NC 27514, USA
[5]Division of Biostatistics, School of Public Health, University of Minnesota, Minneapolis, MN 55455, USA
[6]Longitudinal Study Section, Translational Gerontology Branch, National Institute of Aging, Bethesda, MD 20892, USA
[7]Department of Twin Research and Genetic Epidemiology, King's College London, London SE1 7EH, UK
[8]National Heart, Lung, and Blood Institute, Bethesda, MD 20892, USA
[9]Columbia University Mailman School of Public Health, New York, NY 10032, USA
[10]Harvard T.H. Chan School of Public Health, Boston, MA 02115, USA
[11]VA Boston Healthcare System and Boston University Schools of Public Health and Medicine, Boston, MA 02215, USA
[12]Alzheimer Scotland Dementia Research Centre, University of Edinburgh, Edinburgh EH8 9JZ, UK
[13]Institute for Molecular Bioscience, University of Queensland, Brisbane, QLD, Australia
[14]University of Texas Health Science Center at Houston, Houston, TX 77030, USA
[15]Department of Biostatistics and Data Science, School of Public Health, University of Texas Health Science Center at Houston, Houston, TX 77030, USA
[16]Epidemiology and Public Health Group, University of Exeter Medical School, Exeter, UK
[17]Department of Genetics, University of North Carolina, Chapel Hill, NC 27514, USA
[18]Feinberg School of Medicine, Northwestern University, Chicago, IL 60611, USA
[19]Hebrew SeniorLife Institute for Aging Research and Department of Medicine, Beth Israel Deaconess Medical Center and Harvard Medical School, Boston, MA 02215, USA
[20]Section General Internal Medicine, Department of Medicine, Boston University School of Medicine, Boston, MA 02215, USA
[21]Boston University School of Public Health, Boston, MA 02215, USA
[22]Boston University School of Medicine, Boston, MA 02215, USA
[23]Stanford University School of Medicine, Stanford, CA 94305, USA

[24]Hudson Alpha Institute for Biotechnology, Huntsville, AL 35806, USA
[25]Department of Human Genetics, David Geffen School of Medicine, University of California Los Angeles, Los Angeles, CA 90095, USA
[26]National Institute of Aging, Bethesda, MD 20892, USA
[27]Research Unit of Molecular Epidemiology, Helmholtz Zentrum Munich, German Research Center for Environmental Health, Neuherberg D-85764, Germany
[28]Institute of Human Genetics, Helmholtz Zentrum Munich, German Research Center for Environmental Health, Neuherberg S-85764, Germany
[29]Geriatric Unit, Azienda Sanitaria Firenze, Florence, Italy
[30]Framingham Heart Study, Framingham, MA 01702, USA
[31]Department of Psychology, University of Edinburgh, Edinburgh EH8 9JZ, UK
[32]Division of Epidemiology and Community Health, School of Public Health, University of Minnesota, Minneapolis, MN 55455, USA
[33]Institute for Epidemiology II, Helmholtz Zentrum Munich, German Research Center for Environmental Health, Neuherberg D-85764, Germany
*First authors
[#]Last authors

Correspondence to: Elena Colicino; **email:** elena.colicino@mssm.edu
Keywords: DNA methylation, 450K, all-cause mortality, epigenome-wide association studies, aging
Received: February 10, 2020 **Accepted:** May 25, 2020 **Published:** July 22, 2020

ABSTRACT

DNA methylation has fundamental roles in gene programming and aging that may help predict mortality. However, no large-scale study has investigated whether site-specific DNA methylation predicts all-cause mortality. We used the Illumina-HumanMethylation450-BeadChip to identify blood DNA methylation sites associated with all-cause mortality for 12, 300 participants in 12 Cohorts of the Heart and Aging Research in Genetic Epidemiology (CHARGE) Consortium. Over an average 10-year follow-up, there were 2,561 deaths across the cohorts. Nine sites mapping to three intergenic and six gene-specific regions were associated with mortality ($P < 9.3x10^{-7}$) independently of age and other mortality predictors. Six sites (cg14866069, cg23666362, cg20045320, cg07839457, cg07677157, cg09615688)—mapping respectively to *BMPR1B, MIR1973, IFITM3, NLRC5*, and two intergenic regions—were associated with reduced mortality risk. The remaining three sites (cg17086398, cg12619262, cg18424841)—mapping respectively to *SERINC2, CHST12*, and an intergenic region—were associated with increased mortality risk. DNA methylation at each site predicted 5%–15% of all deaths. We also assessed the causal association of those sites to age-related chronic diseases by using Mendelian randomization, identifying weak causal relationship between cg18424841 and cg09615688 with coronary heart disease. Of the nine sites, three (cg20045320, cg07839457, cg07677157) were associated with lower incidence of heart disease risk and two (cg20045320, cg07839457) with smoking and inflammation in prior CHARGE analyses. Methylation of cg20045320, cg07839457, and cg17086398 was associated with decreased expression of nearby genes (*IFITM3, IRF, NLRC5, MT1, MT2, MARCKSL1*) linked to immune responses and cardiometabolic diseases. These sites may serve as useful clinical tools for mortality risk assessment and preventative care.

INTRODUCTION

The human epigenome contains DNA methylation marks that progressively change as we age. DNA methylation can influence gene expression and manifests in response to both environmental and hereditary factors [1, 2]. Biological age estimations, constructed from DNA methylation marks and referred to as "epigenetic aging clocks", have been associated with environmental exposures, morbidities, and mortality [9–13]. As these clocks were designed to track chronological age, not to predict mortality, further study is necessary to fully elucidate indicators of all-cause mortality. To date, no large-scale analysis has been conducted to identify variations in DNA methylation at individual 5'-cytosine-phosphate-guanosine-3' (CpG) sites associated with future mortality risk. Here, we present an epigenome-wide methylation analysis of 12,300 participants and 2, 561 (21%) deaths from 12 American and European cohorts to determine whether site-specific DNA methylation predicts all-cause mortality, independent of age, lifestyle factors, and clinical predictors of mortality including comorbidities. We also assessed the causal relationship of identified sites with age-related chronic diseases using Mendelian randomization approaches, and we related the sites to epigenetic aging clocks and a mortality risk score, an epigenetic indicator of mortality previously created and validated with DNA methylation arrays in two European cohorts.

RESULTS

Cohorts

Across studies in the Cohorts of the Heart and Aging Research in Genetic Epidemiology (CHARGE) Consortium, mortality rates ranged from 3%–70% of all participants, and the average time to death or censoring ranged from 4.4–16.6 years (Supplementary Table 1). Each study conducted epigenome-wide mortality analyses, adjusting for two sets of harmonized risk factors and confounders, and shared results for meta-analysis (Figure 1).

Meta-analysis

Inverse variance-weighted fixed-effects meta-analysis of 426, 724 CpGs identified 51 Bonferroni-significant and

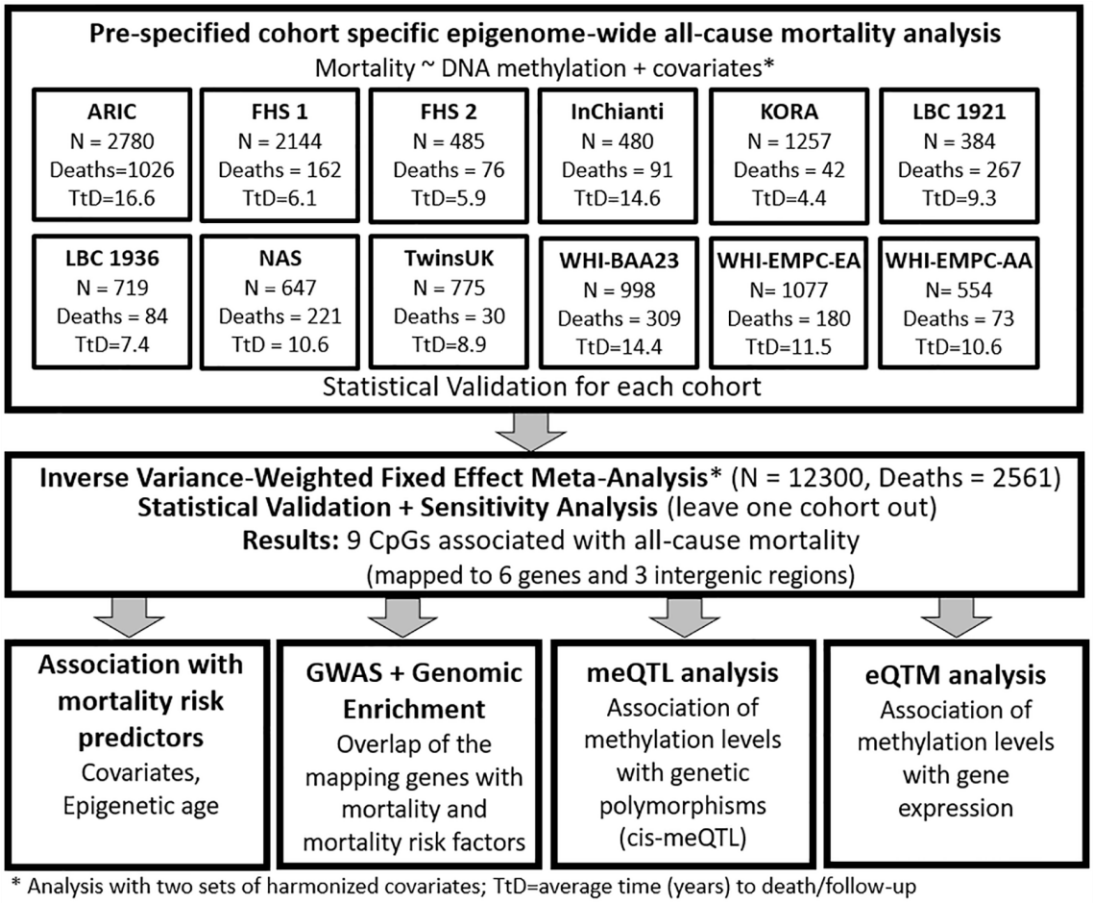


Figure 1. Workflow of the study.

257 FDR-significant ($P < 3.03x10^{-5}$) CpGs in a basic model adjusting for age, sex, technical covariates, and white blood cell proportions (Figure 2A and Supplementary Table 2). We also identified three Bonferroni-significant and nine FDR-significant ($P < 9.3x10^{-7}$) CpGs in a fully-adjusted model also adjusting for education, smoking status, pack-years smoked, body mass index, recreational physical activity, alcohol consumption, hypertension, diabetes, and history of cancer and coronary heart disease (Figures 2B, 3A and Supplementary Table 3). For 188 (73%) basic-adjusted FDR-significant CpGs and six (67%) fully-adjusted CpGs, higher blood DNA methylation was associated with lower all-cause mortality (Figure 2 and Supplementary Tables 2, 3).

All nine fully-adjusted FDR-significant CpGs had similar magnitude of associations with mortality in the basic model, although only five were also FDR-significant in the basic model (Figure 3B). Hazard ratios (HRs) of the nine fully-adjusted FDR-significant CpGs ranged 0.53–1.26 per 10% increase in DNA methylation levels, where 1 represents 100% methylation (Supplementary Table 4). Six sites (cg14866069, cg23666362, cg20045320, cg07839457, cg07677157, cg09615688) were associated with reduced mortality risk, while the remaining three sites (cg17086398, cg12619262, cg18424841) were associated with increased mortality risk (Figure 3A and Supplementary Tables 3, 4). Three fully-adjusted CpGs (cg07677157, cg09615688, cg18424841) were in intergenic regions; the remaining six (cg17086398, cg14866069, cg23666362, cg12619262, cg20045320, cg07839457) were within 10,000 bp of a gene, with two CpGs (cg07839457, cg23666362) mapped respectively to nucleotide-binding oligomerization domain-like receptor caspase recruitment domain containing 5 *(NLRC5)* and microRNA 1973 (*MIR1973)* within 1,500 bp of transcription start sites, and one (cg17086398) in the serine incorporator 2 (*SERINC2*) gene body (Supplementary Table 3).

Meta-analysis results did not appear to suffer from systematic bias due to unmeasured confounding, as assessed by genomic inflation (basic model: λ = 1.12,

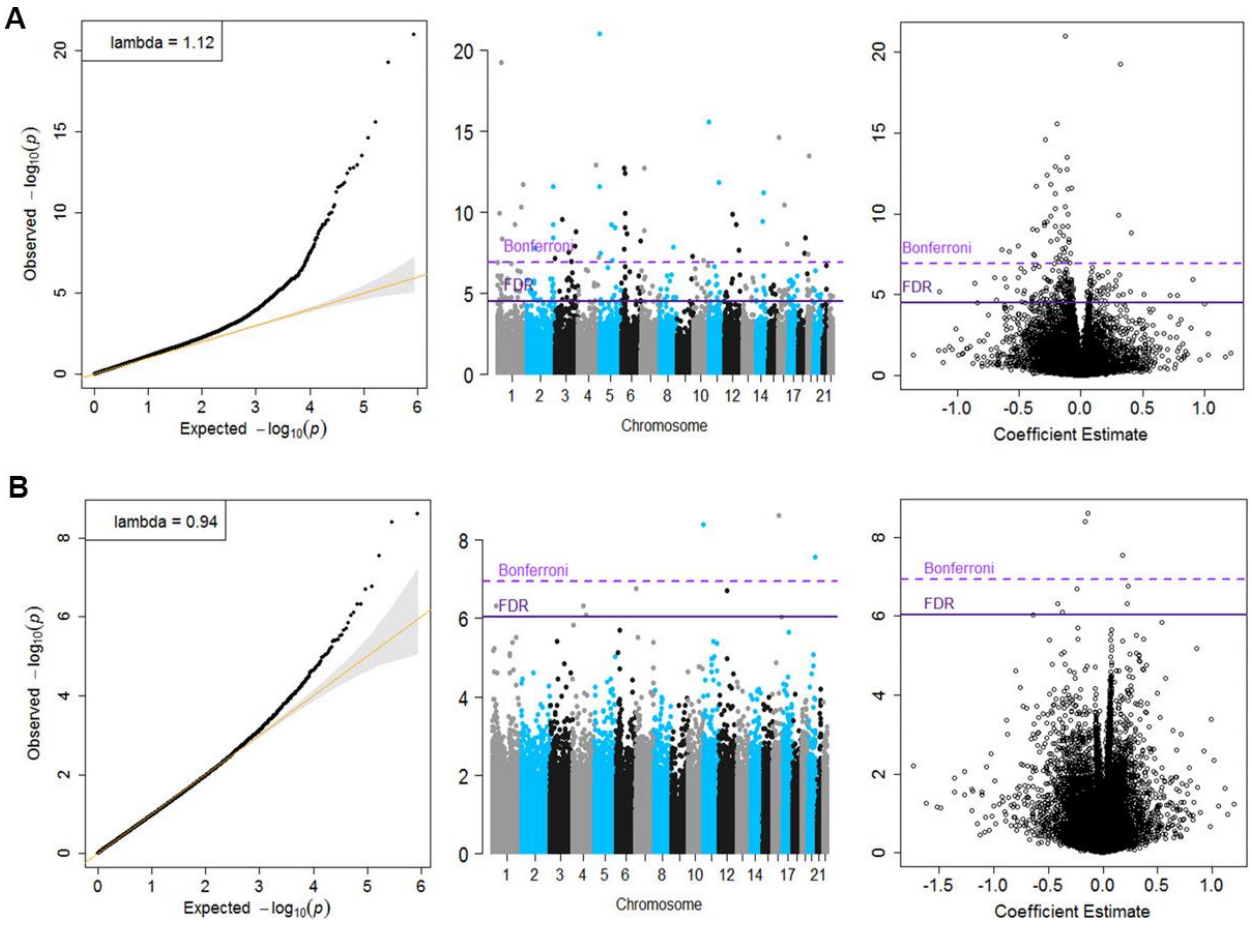


Figure 2. Quantile-Quantile plots, Manhattan and Volcano for the basic model (Panel A) and for the fully adjusted model (Panel B).

fully adjusted model λ = 0.94, Figure 2 and Supplementary Table 5). Cohort-specific inflation was also minimal, with lambdas close to one for most cohorts. Volcano plots showed symmetry in the direction of the associations with all-cause mortality (Figure 2). All nine fully-adjusted FDR-significant CpGs showed low/medium heterogeneity (Supplementary Tables 7) and consistent magnitude of the estimated HRs across studies (Figure 3A). We further validated our results by excluding cohorts with high proportion of deaths (30%) and inflation (λ > 1.5). In these sensitivity analyses, HRs for the nine FDR-significant CpGs were consistent with main results in terms of direction, magnitude, and statistical significance (Supplementary Figure 1 and Supplementary Tables 8, 9).

Three of the nine fully-adjusted FDR-significant CpGs (cg20045320, cg07677157, cg07839457) were associated with lower incidence of coronary heart disease rates (P < Bonferroni threshold of 0.005) (Figure 4 and Supplementary Table 10).

Miettinen's population attributable factor, epigenetic aging clocks, and mortality risk score

To assess the extent that methylation levels of each CpG predict all-cause mortality, we calculated Miettinen's population attributable fraction on data from the Normative Aging Study (NAS) and the Women Health Initiative-Epigenetic Mechanisms of Particulate Matter-Mediated Cardiovascular Disease (WHI-EMPC) for European and African American ancestries. DNA methylation levels above the average at each CpG predicted, individually and independently of other factors, 5%–15% of all deaths (Figure 3C and Supplementary Table 11). In the same datasets, all nine CpGs were associated with age, cumulative smoking, body mass index, and physical activity (P < 0.05). Seven out of nine CpGs (cg17086398, cg14866069, cg23666362, cg20045320, cg7677157, cg07839457, cg09615688) had negative relationships with age (Supplementary Table 12). Seven CpGs were strongly associated with epigenetic aging clocks and mortality risk scores; all significant associations had the same direction and similar magnitude across the four epigenetic aging clocks (Supplementary Table 13), even if none of those sites was included in any of the clocks. Those CpGs had consistent and independent association with all-cause mortality when adjusted for epigenetic aging clocks and mortality risk scores (Supplementary Tables 14, 15). In overall meta-analysis, we identified 57 out of 58 CpGs of the risk score, and those sites had low to moderate association with DNA methylation levels at our FDR-significant CpGs with a balance between

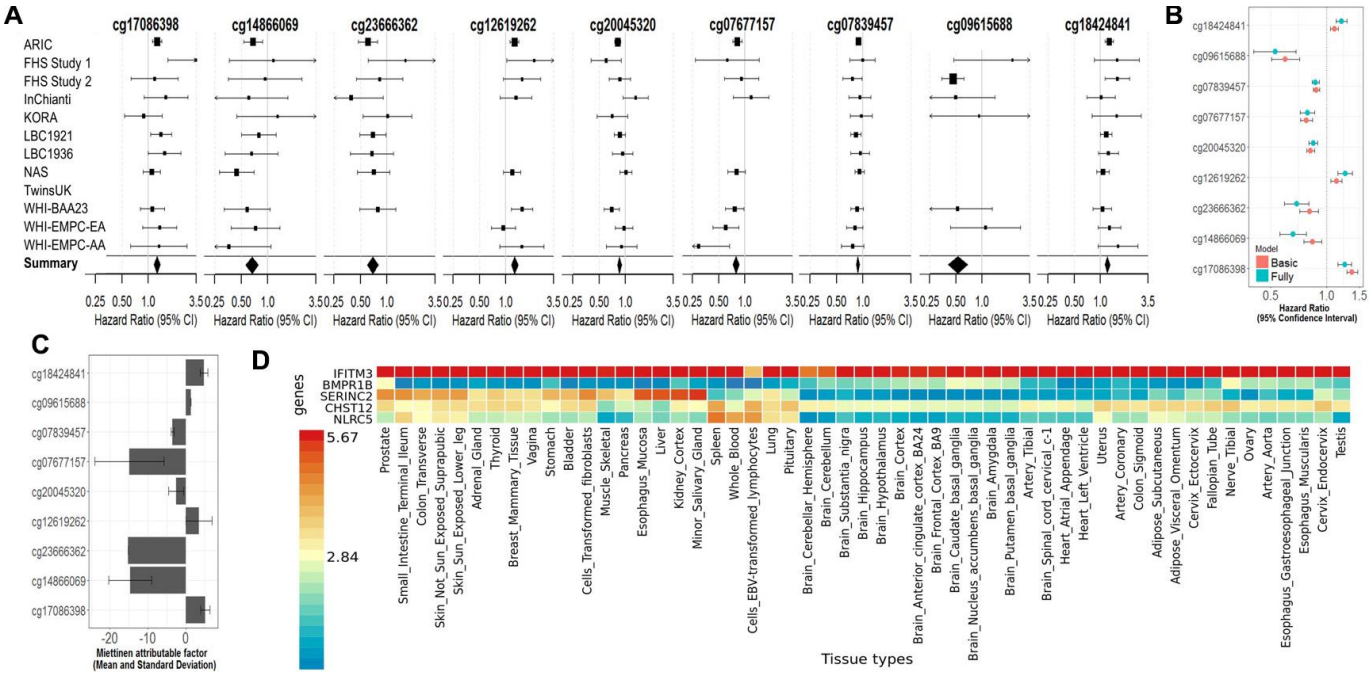


Figure 3. (**A**) Forest Plots for the association of methylation levels of the FDR-significant fully-adjusted CpGs with risk of all-cause mortality in the CHARGE consortium. (**B**) Sensitivity analysis. Comparison of the hazard ratio of the basic-adjusted and the fully-adjusted fixed effect meta-analysis. (**C**) Attributable factor. Predicted Contribution (%) of increased methylation levels (above the mean) of each CpG to the all-cause mortality associations in NAS, WHI-EMPC (EA) and WHI-EMPC (AA). (**D**) Functional Mapping and Annotation results in order to examine tissue specificity of the genes mapped to the FDR-significant fully-adjusted CpGs.

positive and negative correlations (Supplementary Figure 2A, 2B). In overall meta-analysis, the association between all-cause mortality and DNA methylation levels at the majority (34 out of 58) of mortality risk score CpGs had consistent direction with previous results. Among those CpGs, only two (cg25193885 and cg19859270) showed nominally significant association with mortality (Supplementary Figure 2A, 2B).

Pathways analyses and DNA methylation integration with quantitative trait loci analysis (meQTL) and with gene expression (eQTM)

Extended genome-wide enrichment analysis showed that two of the CpGs (cg07839457 and cg17086398) mapped to genes (*NLRC5* and *SERINC2*, respectively) previously associated with high-density lipoprotein cholesterol levels (FDR P = 0.02) and alcohol dependence (FDR P = 0.004) in genome-wide association studies (GWAS) analyses (Supplementary Table 16) [14]. We confirmed these results using Database for Annotation, Visualization and Integrated Discovery (DAVID) and KEGG, identifying and testing for enriched underlying biological processes in publicly available gene ontology databases (Supplementary Tables 17, 18).

To characterize the functional relevance of FDR-significant CpGs, we performed covariate-adjusted methylation quantitative trait locus (meQTL) analyses using available unique single-nucleotide-polymorphism (SNP)–CpG combinations from 713 participants in the Cooperative Health Research in the Region Augsburg (KORA) study [15]. We identified nine Bonferroni-significant unique cis-regulatory polymorphisms associated with two 1000 bp-distant CpGs (cg09615688, cg18424841) (Supplementary Figure 3A and Supplementary Table 19). None of the nine identified polymorphisms overlapped with previous genetic results from the National Human Genome Research Institute-EBI GWAS Catalog (Supplementary Table 16).

We also evaluated expression quantitative trait methylation (eQTM) associations using 998 KORA participants. We identified three CpGs with FDR-significant associations with decreased leukocyte expression levels of nearby genes, among the 13, 351 unique associations between gene-expression and DNA methylation levels at FDR-significant fully-adjusted CpGs. Namely, DNA methylation levels of cg07839457 (in *NLRC5*) were associated with *NLRC5* expression as well as with that of a ~300 Mb-distant set of *metallothionein* (*MT*) *1* and *2* genes, which are linked to oxidative stress and immune responses [16, 17]. DNA methylation of cg17086398 in *SERINC2* was inversely associated with myristoylated alanine-rich C-kinase substrate like 1 (*MARCKSL1*) expression, which is involved in migration of cancer cells [18]. DNA methylation at cg20045320 in *IFITM3* was associated with lower expression of *IFITM3* and *IRF*, which have a critical role in immune responses (Supplementary Figure 3B and Supplementary Table 20) [6, 19].

We finally used functional mapping and annotation to examine tissue-specific expression. Genes identified in the fully-adjusted model showed universal expression at varying levels across tissues. *IFITM3* was highly expressed in all tissues; *BMPR1B* showed low expression across all tissues, except for moderate expression in the prostate and tibial nerve. Remaining genes had moderate or low expression in a wide range of tissues, except for *SERINC2*, which showed high expression in the liver, kidney, salivary gland, and esophagus. *MIR1973* was not represented in the dataset (Figure 3D).

Mendelian randomization

To evaluate the causal relationship of FDR-significant CpGs to mortality-related risk factors and diseases, we included two sets of Mendelian randomization analysis using methQTL data from KORA and publicly available ARIES data. Only two FDR-significant CpGs

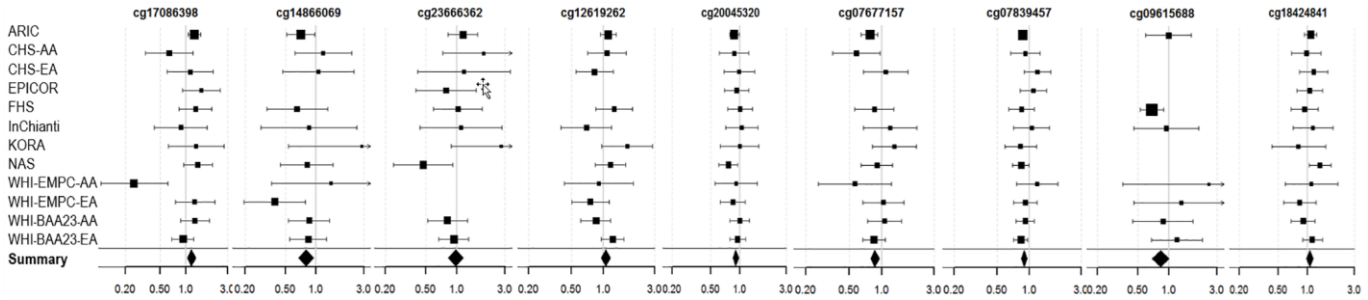


Figure 4. Forest Plots for the association of methylation levels of the FDR-significant fully-adjusted CpGs with risk of future incident coronary heart disease in the CHARGE consortium.

(cg18424841 and cg09615688) overlapped with methQTLs in either KORA or ARIES and with SNPs associated with coronary heart disease (CHD) or kidney function. A GWAS assessing longevity and age-related chronic diseases (CHD and kidney function) [34–38] showed no overlap with KORA and ARIES methQTLs even when using a moderate threshold for proxy variants (proxy $r^2 > 0.75$). In KORA, cg09615688 showed evidence of a positive causal effect on CHD (OR = 1.51; 95% CI = 1.02, 2.23; Wald ratio method), directionally consistent with the association of overall meta-analysis on mortality. However, this causal estimate at this site was not represented in ARIES methQTL data. Cg18424841 had multiple variants in KORA methQTL data and a single variant in ARIES methQTL data. We did not observe consistent evidence of a causal effect of cg18424841 on CHD. Indeed, weak evidence for a causal effect of cg18424841 on CHD was observed in ARIES using the Wald ratio method but not in KORA using pleiotropy-robust, multi-variant, or Wald ratio methods. We did not find evidence for a causal effect of cg18424841 on kidney function in either KORA or ARIES (Supplementary Table 21).

Cell-type fractions and all-cause mortality

Cell-type fractions, mostly neutrophil–lymphocyte ratio (NLR), have been often associated with comorbidities and mortality and have been recognized to influence DNA methylation levels [20–22]. We identified that NLR was significantly associated with all-cause mortality only when data were not adjusted for Houseman cell proportions using NAS data (Supplementary Table 22). Interestingly, NLR had no significant association with all-cause mortality when we adjusted for DNA methylation levels at cg07839457, mapped to immune-related gene *NLRC5*. However, the contribution of NLR on mortality at that specific site may be minimized due to adjustment of prior history of cancer and comorbidities in all models.

DISCUSSION

This study is the largest to date investigating site-specific DNA methylation and all-cause mortality. We identified new whole blood DNA methylation marks that predict all-cause mortality risk, independent from chronological age, lifestyle habits, and morbidity. These newly identified sites may be useful in developing clinical tools for risk assessment and mortality preventive intervention strategies.

All nine FDR-significant CpGs demonstrated novel association with all-cause mortality and were not part of epigenetic aging clocks or mortality risk scores [9, 11–13]. Further, the CpGs were associated with mortality independent from epigenetic aging and mortality signatures. All-cause mortality was associated with a mortality risk score in a model including seven FDR-significant CpGs, although those associations may be driven by the inclusion of CpGs related to our FDR-significant sites. This suggests that whole blood DNA methylation levels at FDR-significant CpGs may be sentinels for epigenetic disruptions leading to aging acceleration and contributing to mortality. In addition, the association between DNA methylation levels at FDR-significant CpGs with chronological aging may suggest that those CpGs are stronger independent biomarkers of aging than other epigenetic aging signatures.

In previous CHARGE meta-analyses [3, 4], DNA methylation of two of the newly-identified CpGs, cg20045320 and cg07839457 (mapping to interferon induced transmembrane protein 3 [*IFITM3*] and *NLRC5*) were respectively associated with smoking and cardiovascular-related chronic inflammation, both factors of mortality. Cardiovascular disease, especially CHD, is a major contributor to mortality [23]. The direction of association with incident heart disease was consistent with that of all-cause mortality. Thus, DNA methylation at these CpGs may contribute to development and progression of CHD and, consequently, to risk of death. To validate this idea, we used a Mendelian randomization approach and identified one site, cg09615688, with a causal effect on CHD in KORA data and weak evidence for the causal effect of cg18424841 on CHD in ARIES data.

Expression of several genes mapped to the fully-adjusted FDR-significant CpGs has been associated with mortality predictors and mortality. Elevated and persistent gene expression levels of *NLRC5*, a master regulator of the immune response [16], has demonstrated an inverse correlation with familial longevity and mortality predictors, such as elevated blood pressure, arterial stiffness, chronic levels of inflammatory cytokines, metabolic dysfunction, and oxidative stress [5]. In addition, expression of *IFITM3* provides an essential barrier to influenza A virus infection *in vivo* and *in vitro*. Absence of *IFITM3* leads to uncontrolled viral replication and a predisposition to morbidity and subsequent mortality [6]. Further, expression of *BMPR1B* enhances cancer cell migration, and approaches targeting *BMPR1B* inhibit metastatic activity in breast cancer [7]. Finally, expression of *MIR1973*, part of a family of microRNAs, increases resistant lung adenocarcinoma cells, with subsequent low apoptosis intensity [8]. This body of evidence may suggest an active role of DNA methylation levels in regulating relevant gene expression and reducing all-cause mortality risk.

The overall meta-analyses included 12 cohorts with varying biological age and mortality. There was a balance between six studies with long (≥10 years) and six cohorts with short (<10 years) average time to follow-up or death. All cohorts showed consistency in magnitudes and directionality for the association with mortality of four CpGs (cg12619262, cg20045320, cg07839457, cg18424841). Two studies (FHS study 1 and KORA) showed non-significant opposing directionality when compared with the rest of the cohorts for several CpGs (FHS-Study 1: cg14866069, cg23666362, cg09615688; KORA: cg17086398, cg14866069, cg23666362). However, both cohorts had among the shortest average time-to-death (FHS-Study 1: 6.1 years; KORA: 4.4 years) and youngest average population age (FHS-Study 1: 65 years; KORA: 61 years). Both cohorts also had limited contribution in our meta-analysis due to reduced number of deaths (FHS-Study 1: 62; KORA: 42). Our results may indicate that DNA methylation levels at these select CpGs were relevant for mortality risk prediction of longer time-to-death in both adults and older-age adults.

Cell-type fractions, including NLR, as related to cancer and systemic inflammation have been related to mortality in different populations [20–22]. When we excluded Houseman cell proportions, NLR was strongly associated with mortality at all CpGs except cg07839457, which is mapped to the immune-related gene *NLRC5*. This may suggest that the contribution of NLR on mortality is minimized when controlled for prior history of cancer and related comorbidities.

In summary, we identified nine CpGs with a novel association with all-cause mortality, responsive to several external stimuli including alcohol consumption and smoking, and more than 10 years before death. These sites thus may be considered sentinels for epigenetic disruptions leading to age-related disease, such as cardiovascular disease, and contributing to mortality. Further studies have to confirm these associations in other tissues and in different populations.

MATERIALS AND METHODS

Participating cohort studies

Our meta-analysis included 12,300 participants from 12 population-based cohorts of the Heart and Aging Research in Genetic Epidemiology Consortium (CHARGE; Supplementary material): Atherosclerosis Risk In Communities (ARIC), two studies from the Framingham Heart Study (FHS), Invecchiare in Chianti (InChianti), Kooperative Gesundheitsforschung in der Region Augsburg (KORA), Lothian Birth Cohort 1921 (LBC1921) and 1936 (LBC1936), Normative Aging Study (NAS), UK Adult Twin Registry (TwinsUK), and three studies from the Women's Health Initiative (WHI), including Broad Agency Announcement 23 (WHI-BAA23) and Epigenetic Mechanisms of PM-Mediated CVD Risk (WHI-EMPC), both European (WHI-EMPC-EA) and African American ancestries (WHI-EMPC-AA). For each participant, we derived years of follow-up using time between the blood draw used for DNA methylation analysis and death or last follow-up. Each cohort excluded participants with diagnosed leukemia (ICD-9: 203–208) or undergoing chemotherapy treatment, which both modify blood-derived data [24, 25]. All participating cohorts shared cohort descriptive statistics and results files from pre-specified in-house mortality analyses (Figure 1). Further information about death ascertainment, covariates measurement and harmonization, protocols, and methods of each cohort are included in the Supplemental Materials. The institutional review committees of each cohort approved this study, and all participants provided written informed consent. Data and analytical codes that support our findings are available from the corresponding author upon request.

Blood DNA methylation measurements and quality control

Each cohort independently conducted laboratory DNA methylation measurements and internal quality control. All samples underwent bisulfite conversion via the EZ-96 DNA Methylation kit (Zymo Research) and were processed with the Illumina Infinium HumanMethylation450 (450K) BeadChip (Illumina) at Illumina or in cohort-specific laboratories. Quality control of samples included exclusion on the basis of Illumina's detection *P*-value, low sample DNA concentration, sample call rate, CpG specific percentage of missing values, bisulfite conversion efficiency, gender verification with multidimensional scaling plots, and other quality control metrics specific to cohorts. Each cohort used validated statistical methods for normalizing methylation data on untransformed methylation beta values (ranging 0–1). Some cohorts also made independent probe exclusions. Further details are provided in the Supplemental Material. For meta-analysis, additional probe exclusions were made across all cohorts. In detail, we also excluded control probes, non-CpG sites, probes that mapped to allosomal chromosomes, cross-reactive CpGs, probes with underlying SNPs within 10 bp of the CpG sequence, non-varying CpGs defined by interquartile range of <0.1%, CpGs with ≥10% of missing information, and CpGs with non-converging results [26–28]. We included only CpGs that were available in more than three cohorts. A total of 426, 724 CpGs were included in the meta-analysis (Supplementary Table 5).

The official gene name of each CpG site was noted via Illumina's genome coordinate. We used the name provided by Illumina with the UCSC Genome Browser and annotation data in Bioconductor. All annotations use the human February 2009 (GRCh37/hg19) assembly.

Cohort-specific statistical analyses

Each cohort independently ran a common pre-specified statistical analysis in R.version 3.5.1. We estimated the association between locus-by-locus blood DNA methylation levels and all-cause mortality in each cohort using a Cox-regression model. Proportional hazard assumptions were confirmed for each model in all cohorts. Familial relationship was also accounted for, when appropriate, in the model; FHS analyses included cluster for family structure, and TwinsUK analyses used random intercepts for zygosity and family structure. To avoid non-convergent results, cohorts with low deaths (KORA and TwinsUK) used a two-step analysis, in which covariates were first linearly regressed on each probe, and then residuals were used to perform a Cox mortality analysis.

Each cohort adjusted for harmonized covariates in the basic model: age (categories for decades), sex, and technical covariates (plate, chip, row, and column). A second set of fully-adjusted analyses adjusted for this initial list of covariates in addition to education level, self-reported recreational physical activity, smoking status, cumulative smoking (pack-years), body mass index, alcohol intake, hypertension, diabetes, and any personal history of cancer. Cohorts independently estimated cell type proportions using the reference-based Houseman method, which was subsequently extended by Horvath. Cell type correction was applied by including estimated cell type proportions (CD4T, NK cells, monocytes, granulocytes, plasma B cells, CD8T naïve, and memory and effector T cells) as covariates in cohort-specific statistical models. Each cohort underwent statistical validation of Cox-proportional hazard assumptions before being included in the meta-analysis.

Meta-analysis

We performed inverse variance-weighted fixed-effects meta-analysis. Due to the variability of available CpG sites across cohorts after quality-control steps, we included only CpG sites that were available in three or more cohorts. We accounted for multiple testing by controlling at 5% both the Bonferroni correction and false discovery rate (FDR) using the Benjamini-Hochberg procedure.

For FDR-significant CpGs, we confirmed robustness of the models and results in additional analyses using the leave-one-out cohort validation method, by excluding one cohort at a time and then comparing model estimates for each CpG. We compared effect hazard ratio (HR) and 95% confidence interval (95% CI) for the model to estimates for our models to evaluate the consistency of our findings. For each CpG, we evaluated goodness of the meta-analysis model using the I^2 statistic measure of inter-study variability from random-effect meta-analyses.

Enrichment analysis

We enriched our results using a publicly available catalog of all published GWAS relating genetic variants with human diseases (National Human Genome Research Institute-EBI GWAS Catalog) to elucidate potential associations [14]. Enrichment analysis was performed in R using one-sided Fisher exact test. We controlled for false positives with the FDR procedure.

We evaluated whether CpG sites associated with mortality were enriched with genomic features provided in the Illumina annotation file (version 1.2; http://support.illumina.com/array/array_kits/infinium_humanmethylation450_beadchip_kit/downloads.html) to identify CpG location relative to the gene (i.e., body, first exon, 3'-UTR, 5'-UTR, within 200 bp of transcriptional start site [TSS200]), and within 1500 bp of transcriptional start site [TSS1500]) and relation of the CpG site to a CpG island, northern shelf, northern shore, southern shelf, and southern shore.

We also tested each gene mapped to the newly identified CpGs for tissue-specific expression using data from the Genotype Tissue Expression (GTEx) project as integrated by the Functional Mapping and Annotation (FUMA) tool [29], which allowed us to extract and interpret relevant biological information from publicly available repositories and provide interactive figures for prioritized genes. As a result, we obtained a heatmap of genes with normalized gene expression values (reads per kilo base per million). To obtain differentially expressed gene sets for each of 53 tissue types in the database, we used two-sided Student's *t*-tests on normalized expression per gene per tissue against all other tissues. We controlled for multiple comparison with Bonferroni correction. Finally, we distinguished between genes upregulated and downregulated in a specific tissue compared to other tissues by accounting for sign of the *t*-score [29].

Pathway analyses

To functionally interpret the genomic information identified from FDR-significant CpGs, we used the Kyoto Encyclopedia of Genes and Genomes (KEGG)

pathway database, which links genomic information with higher-order functional information. Genomic information stored in the GENES database is a collection of gene catalogs for all completely sequenced genomes and some partial genomes with up-to-date annotation of gene functions. Higher-order functional information stored in the PATHWAY database contains graphical representations of cellular processes, such as metabolism, membrane transport, signal transduction, and cell cycle [30]. We controlled our results for multiple comparisons with the FDR approach. We finally confirmed our results with the Database for Annotation, Visualization and Integrated Discovery (DAVID). We tested for enrichment in gene ontology biological processes and applied the Benjamini-Hochberg procedure to control for false positivity. We mapped each CpG significantly associated with mortality to genes on the basis of the 450K BeadChip annotation file. We excluded CpGs lacking annotated genes within 10 Mb (n = 3). Using topGO in R, we tested for gene enrichment over the background array (16, 119 unique annotated Entrez Gene IDs) by using Fisher's exact tests with a minimum of two genes per node.

Integrating DNA methylation with quantitative trait loci analysis (meQTL)

A subset of 713 KORA samples was genotyped on an Affymetrix Axiom array. We removed variants with a call rate of <0.98, Hardy-Weinberg equilibrium $P < 5x10^{-6}$, and minor allele frequency < 0.01. We considered only variants with an information score > 3. Imputation was performed using the 1000 Genomes Project phase I version 3 reference panel with IMPUTE 2.3.0. Phasing of data was performed using SHAPEIT v2. We retained approximately 10,000,000 variants for analyses. In each model, we used DNA methylation beta values as independent variables and SNPs as dependent variables. We adjusted each model for age, sex, body mass index, and white blood cell proportions. We used OmicABEL [31] for the analyses and genotype probabilities for each variant. Due to large size of the output, we retained only variants with $P < 1x10^{-4}$. We considered genome-wide significant results at $P < 1x10^{-14}$. We reported only associations with CpGs significant in the epigenome-wide association study.

Integrating DNA methylation with gene expression (eQTM)

In KORA, 998 individuals had both valid methylation and blood gene expression data, which we used to assess whether DNA methylation was correlated with gene expression. Gene expression data (Illumina HumanHT-12 v3 Expression BeadChip) was quality controlled with GenomeStudio, and samples with <6,000 detected genes were excluded from analysis. All samples were log2-transformed and quantile-normalized using the Bioconductor package lumi [32]. A total of 48,803 expression probes passed quality control. We used R (version 3.3.1) to run a linear mixed effects model adjusting for covariates (age, sex, blood cell proportions, and technical variables of RNA integrity number, sample storage time, and RNA amplification batch) and a random intercept for RNA amplification batch. Models were run for each of the nine newly-identified CpGs associated with mortality. We filtered results to report only CpG-expression probe pairs located on the same chromosome. Start and end sites for each gene were determined according to the Illumina HT annotation file. A cutoff of 500,000 bp was used to differentiate cis- vs. trans-eQTMs.

Miettinen's population attributable factor and mendelian randomization analysis

To assess the contribution of methylation levels of each CpG to all-cause mortality, we calculated Miettinen's population attributable fraction on data from the in-house Normative Aging Study (NAS) and Women Health Initiative-Epigenetic Mechanisms of Particulate Matter-Mediated Cardiovascular Disease (WHI-EMPC) for European and African American ancestries. Population attributable fraction takes into account strength of association between the risk factor (DNA methylation higher than the mean in specific CpG sites) and outcome (mortality) as well as prevalence of the risk factor in the population [33]. This metric provides estimates of the public health importance of risk factors, ascertaining what proportion of the outcome is due to exposure to the risk factor, and distinguishes between etiologic fraction attributable to or related to the given risk factor depending on whether all or just some confounding by extraneous factors was under control [33]. To support information about the population attributable factor, we also included two Mendelian randomization approaches.

We identified the causal effect on all-cause mortality of FDR-significant CpGs by using two sample Mendelian randomization analyses and summary statistics from published GWAS for chronic diseases and longevity [34] and chronic diseases, including CHD [35], kidney function (serum creatinine), [36] blood pressure, [37] and type 2 diabetes [38]. We extracted GWAS information with MR-base [14]. We also extracted SNP-methylation association summary statistics from both KORA and publicly available ARIES [39] methQTL data; for ARIES, we used MR-base [40]. To account for multiple variants and pleiotropy, we used multiple Mendelian randomization methods—when only one variant was present, we used the Wald Ratio method [41]; when we

had multiple variants, we used MR Egger [42], weighted median [43], and weighted mode [44], as these three methods use different assumptions to provide consistent causal effect estimates even with invalid instruments arising from horizontal pleiotropy, a primary source of bias in multi-variant Mendelian randomization analyses.

FDR-significant CpGs, DNA methylation-related aging measures, and mortality risk score

PhenoAge, a composite measure of CpG sites representing phenotypic age, captures differences between lifespan and health span. The Horvath clock is a linear combination of sites identifying the cumulative effect of an epigenetic maintenance system [1, 45]. Among the 513 CpGs comprising PhenoAge, 41 are shared with the Horvath clock. While both aging measures correlate strongly with age in every tissue and cell type tested, and both captured risks for mortality across multiple tissues and cells, PhenoAge is highly predictive of nearly every morbidity [1, 10]. Blood PhenoAge outperformed the Horvath clock with regard to predictions for a variety of aging outcomes, including all-cause mortality. The mortality risk score instead was based on results using discovery cohort ESTHER (61 years old on average) and both ESTHER and KORA for validation [11].

To investigate whether the association of FDR-significant CpGs with mortality was independent of DNA methylation aging measures and risk score, we included acceleration of PhenoAge and Horvath clock, defined respectively as discrepancies between age with PhenoAge and Horvath clock age and the risk score. We also identified the correlation between each CpG included in the risk score and our FDR-significant CpGs, and we compared our pooled meta-analysis results with previous findings.

Cell-type fractions and all-cause mortality

Cell-type fractions, mostly NLR, influence DNA methylation levels and have been associated with comorbidities and mortality [20–22]. To elucidate which cell proportions were associated with mortality when adjusting for DNA methylation at FDR-significant CpGs, we included NLR, which has been associated with lung cancer risk and mortality [21] as well was cardiovascular disease and mortality in prospective studies [22]. NLR computation was performed using DNA methylation data via Koestler et al. [46]

CONFLICTS OF INTEREST

The authors declare that they have no conflicts of interest.

FUNDING

ARIC has been funded in whole or in part with federal funds from the U.S. National Heart, Lung, and Blood Institute, National Institutes of Health (NIH), Department of Health and Human Services (HHSN 268201700001I, HHSN268201700002I, HHSN268201700003I, HHSN268201700004I, HSN268201700005I). The authors thank the staff and participants of the ARIC study for their important contributions. Funding was also supported by 5RC2HL102419 and R01NS087541. **FHS** is funded by the U.S. National Institutes of Health (N01-HC-25195 and HHSN268201500001I). The laboratory work for this investigation was funded by the Division of Intramural Research, National Heart, Lung, and Blood Institute. The analytical component of this project was funded by the Division of Intramural Research, National Heart, Lung, and Blood Institute. Dr. Kiel's time was funded by grants from the U.S. National Institute of Arthritis Musculoskeletal and Skin Diseases (R01 AR041398) and the U.S. National Institute on Aging (U34 AG051418). Dr. Murabito's time was funded by the U.S. National Institute of Aging (R56AG029451, U34 AG051418). The **InCHIANTI** was supported as a "targeted project" (ICS110.1/ RF97.71) by the Italian Ministry of Health and in part by the U.S. National Institute on Aging (263 MD 9164, 263 MD 821336); InCHIANTI Follow-up 1 (2001–2003) was funded by the U.S. National Institute on Aging (N.1-AG-1-1, N.1-AG-1-2111); InCHIANTI data are available on request at http://inchiantistudy.net/wp/inchianti-dataset. **KORA** was initiated and financed by the Helmholtz Zentrum München-Germany Research Center for Environmental Health, which is funded by the German Federal Ministry of Education and Research (BMBF) and by the State of Bavaria. Further, KORA research was supported within the Munich Center of Health Sciences, Ludwig Maximilians-Universität, as part of LMUinnovativ. This work was supported in part by BMBF within the framework of the e:Med research and funding concept (e:AtheroSysMed; 01ZX1313A-2014). The views in this manuscript do not necessarily reflect the views and policies of the US Environmental Protection Agency. **LBC** 1921 was supported by the UK's Biotechnology and Biological Sciences Research Council (BBSRC; 15/SAG09977), The Royal Society, and The Chief Scientist Office of the Scottish Government (CZB/4/505 and ETM/55). Phenotype collection in the Lothian Birth Cohort 1936 was supported by Age UK (The Disconnected Mind project). Methylation typing was supported by the Centre for Cognitive Ageing and Cognitive Epidemiology (Pilot Fund award), Age UK, The Wellcome Trust Institutional Strategic Support Fund, The University of Edinburgh, and The University of Queensland. REM, JMS, and IJD are members of the

University of Edinburgh Centre for Cognitive Ageing and Cognitive Epidemiology, which is supported by funding from the BBSRC, the Medical Research Council, and the University of Edinburgh as part of the cross-council Lifelong Health and Wellbeing initiative (MR/K026992/1). **NAS** is supported by grants from the U.S. National Institute of Environmental Health Sciences (NIEHS) (R01ES021733, R01ES 025225, R01ES027747). Other support comes from NIEHS (ES015172, ES014663, ES020010) and the Environmental Protection Agency and NIEHS (RD832416, P30ES009089). NAS is supported by the Cooperative Studies Program/ERIC and U.S. Department of Veterans Affairs and is a research component of the Massachusetts Veterans Epidemiology Research and Information Center. The views expressed in this paper are those of the authors and do not necessarily represent the views of the U.S. Department of Veterans Affairs. Additional support was provided by the U.S. Department of Agriculture, Agricultural Research Service (53-K06-510). The **TwinsUK** epigenetic study was supported by the Economic and Social Research Council (ES/N000404/1), the European Research Council (ERC 250157), and in part from the TwinsUK resource, which is funded by the Wellcome Trust, the European Community's Seventh Framework Programme (FP7/2007–2013), and the National Institute for Health Research BioResource, Clinical Research Facility, and Biomedical Research Centre, based at Guy's and St Thomas's NHS Foundation Trust and King's College London. **WHI-EMPC** is funded by the NIEHS (R01-ES020836; Whitsel, Baccarelli, Hou). The **WHI-BAA23** program is funded by the U.S. National Heart, Lung, and Blood Institute (HHSN268201100046C, HHSN 268201100001C, HHSN268201100002C, HHSN26820 1100003C, HHSN268201100004C, HHSN2712011000 04C). LH was funded by AHA (14SFRN20790000). The authors thank WHI investigators and staff for their dedication and study participants for making the program possible. A listing of WHI-BAA23 investigators can be found at http://www.whi.org/researchers/Documents%20%20Write%20a%20Paper/WHI%20Investigator%20Short%20List.pdf.

REFERENCES

1. Horvath S. DNA methylation age of human tissues and cell types. Genome Biol. 2013; 14:R115. https://doi.org/10.1186/gb-2013-14-10-r115 PMID:24138928

2. Argentieri MA, Nagarajan S, Seddighzadeh B, Baccarelli AA, Shields AE. Epigenetic pathways in human disease: the impact of DNA methylation on stress-related pathogenesis and current challenges in biomarker development. EBioMedicine. 2017; 18:327–50. https://doi.org/10.1016/j.ebiom.2017.03.044 PMID:28434943

3. Joehanes R, Just AC, Marioni RE, Pilling LC, Reynolds LM, Mandaviya PR, Guan W, Xu T, Elks CE, Aslibekyan S, Moreno-Macias H, Smith JA, Brody JA, et al. Epigenetic signatures of cigarette smoking. Circ Cardiovasc Genet. 2016; 9:436–47. https://doi.org/10.1161/CIRCGENETICS.116.001506 PMID:27651444

4. Aslibekyan S, Agha G, Colicino E, Do AN, Lahti J, Ligthart S, Marioni RE, Marzi C, Mendelson MM, Tanaka T, Wielscher M, Absher DM, Ferrucci L, et al. Association of methylation signals with incident coronary heart disease in an epigenome-wide assessment of circulating tumor necrosis factor α. JAMA Cardiol. 2018; 3:463–72. https://doi.org/10.1001/jamacardio.2018.0510 PMID:29617535

5. Furman D, Chang J, Lartigue L, Bolen CR, Haddad F, Gaudilliere B, Ganio EA, Fragiadakis GK, Spitzer MH, Douchet I, Daburon S, Moreau JF, Nolan GP, et al. Expression of specific inflammasome gene modules stratifies older individuals into two extreme clinical and immunological states. Nat Med. 2017; 23:174–84. https://doi.org/10.1038/nm.4267 PMID:28092664

6. Everitt AR, Clare S, Pertel T, John SP, Wash RS, Smith SE, Chin CR, Feeley EM, Sims JS, Adams DJ, Wise HM, Kane L, Goulding D, et al. GenISIS Investigators, and MOSAIC Investigators. IFITM3 restricts the morbidity and mortality associated with influenza. Nature. 2012; 484:519–23. https://doi.org/10.1038/nature10921 PMID:22446628

7. Allison SE, Chen Y, Petrovic N, Zimmermann S, Moosmann B, Jansch M, Cui PH, Dunstan CR, Mackenzie PI, Murray M. Activation of the pro-migratory bone morphogenetic protein receptor 1B gene in human MDA-MB-468 triple-negative breast cancer cells that over-express CYP2J2. Int J Biochem Cell Biol. 2016; 80:173–78. https://doi.org/10.1016/j.biocel.2016.10.004 PMID:27720933

8. Fomicheva KA, Knyazev EN, Mal'tseva DV. hsa-miR-1973 MicroRNA is significantly and differentially expressed in MDA-MB-231 cells of breast adenocarcinoma and xenografts derived from the tumor. Bull Exp Biol Med. 2017; 163:660–62. https://doi.org/10.1007/s10517-017-3873-0 PMID:28948560

9. 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 PMID:27690265

10. Marioni RE, Shah S, McRae AF, Chen BH, Colicino E, Harris SE, Gibson J, Henders AK, Redmond P, Cox SR, Pattie A, Corley J, Murphy L, et al. DNA methylation age of blood predicts all-cause mortality in later life. Genome Biol. 2015; 16:25. https://doi.org/10.1186/s13059-015-0584-6 PMID:25633388

11. 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 PMID:28303888

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 PMID:23177740

13. Weidner CI, Lin Q, Koch CM, Eisele L, Beier F, Ziegler P, Bauerschlag DO, Jöckel KH, Erbel R, Mühleisen TW, Zenke M, Brümmendorf TH, Wagner W. Aging of blood can be tracked by DNA methylation changes at just three CpG sites. Genome Biol. 2014; 15:R24. https://doi.org/10.1186/gb-2014-15-2-r24 PMID:24490752

14. MacArthur J, Bowler E, Cerezo M, Gil L, Hall P, Hastings E, Junkins H, McMahon A, Milano A, Morales J, Pendlington ZM, Welter D, Burdett T, et al. The new NHGRI-EBI catalog of published genome-wide association studies (GWAS catalog). Nucleic Acids Res. 2017; 45:D896–901. https://doi.org/10.1093/nar/gkw1133 PMID:27899670

15. Ligthart S, Marzi C, Aslibekyan S, Mendelson MM, Conneely KN, Tanaka T, Colicino E, Waite LL, Joehanes R, Guan W, Brody JA, Elks C, Marioni R, et al, and WHI-EMPC Investigators, and CHARGE epigenetics of Coronary Heart Disease. DNA methylation signatures of chronic low-grade inflammation are associated with complex diseases. Genome Biol. 2016; 17:255. https://doi.org/10.1186/s13059-016-1119-5 PMID:27955697

16. Meissner TB, Li A, Biswas A, Lee KH, Liu YJ, Bayir E, Iliopoulos D, van den Elsen PJ, Kobayashi KS. NLR family member NLRC5 is a transcriptional regulator of MHC class I genes. Proc Natl Acad Sci USA. 2010; 107:13794–99. https://doi.org/10.1073/pnas.1008684107 PMID:20639463

17. Ruttkay-Nedecky B, Nejdl L, Gumulec J, Zitka O, Masarik M, Eckschlager T, Stiborova M, Adam V, Kizek R. The role of metallothionein in oxidative stress. Int J Mol Sci. 2013; 14:6044–66. https://doi.org/10.3390/ijms14036044 PMID:23502468

18. Björkblom B, Padzik A, Mohammad H, Westerlund N, Komulainen E, Hollos P, Parviainen L, Papageorgiou AC, Iljin K, Kallioniemi O, Kallajoki M, Courtney MJ, Mågård M, et al. C-jun n-terminal kinase phosphorylation of MARCKSL1 determines actin stability and migration in neurons and in cancer cells. Mol Cell Biol. 2012; 32:3513–26. https://doi.org/10.1128/MCB.00713-12 PMID:22751924

19. Taniguchi T, Takaoka A. The interferon-Alpha/beta system in antiviral responses: a multimodal machinery of gene regulation by the IRF family of transcription factors. Curr Opin Immunol. 2002; 14:111–16. https://doi.org/10.1016/s0952-7915(01)00305-3 PMID:11790540

20. Palmerini T, Mehran R, Dangas G, Nikolsky E, Witzenbichler B, Guagliumi G, Dudek D, Genereux P, Caixeta A, Rabbani L, Weisz G, Parise H, Fahy M, et al. Impact of leukocyte count on mortality and bleeding in patients with myocardial infarction undergoing primary percutaneous coronary interventions: analysis from the harmonizing outcome with revascularization and stent in acute myocardial infarction trial. Circulation. 2011; 123:2829–37. https://doi.org/10.1161/CIRCULATIONAHA.110.985564 PMID:21632496

21. Grieshober L, Graw S, Barnett MJ, Thornquist MD, Goodman GE, Chen C, Koestler DC, Marsit CJ, Doherty JA. Methylation-derived neutrophil-to-lymphocyte ratio and lung cancer risk in heavy smokers. Cancer Prev Res (Phila). 2018; 11:727–34. https://doi.org/10.1158/1940-6207.CAPR-18-0111 PMID:30254071

22. Kim S, Eliot M, Koestler DC, Wu WC, Kelsey KT. Association of neutrophil-to-lymphocyte ratio with mortality and cardiovascular disease in the jackson heart study and modification by the duffy antigen variant. JAMA Cardiol. 2018; 3:455–62. https://doi.org/10.1001/jamacardio.2018.1042 PMID:29801037

23. He FJ, MacGregor GA. How far should salt intake be reduced? Hypertension. 2003; 42:1093–99. https://doi.org/10.1161/01.HYP.0000102864.05174.E8 PMID:14610100

24. Flanagan JM, Wilson A, Koo C, Masrour N, Gallon J, Loomis E, Flower K, Wilhelm-Benartzi C, Hergovich A, Cunnea P, Gabra H, Braicu EI, Sehouli J, et al. Platinum-based chemotherapy induces methylation changes in blood DNA associated with overall survival in patients with ovarian cancer. Clin Cancer Res. 2017; 23:2213–22. https://doi.org/10.1158/1078-0432.CCR-16-1754 PMID:27663594

25. Figueroa ME, Lugthart S, Li Y, Erpelinck-Verschueren C, Deng X, Christos PJ, Schifano E, Booth J, van Putten W, Skrabanek L, Campagne F, Mazumdar M, Greally JM, et al. DNA methylation signatures identify biologically distinct subtypes in acute myeloid leukemia. Cancer Cell. 2010; 17:13–27. https://doi.org/10.1016/j.ccr.2009.11.020 PMID:20060365

26. Chen YA, Lemire M, Choufani S, Butcher DT, Grafodatskaya D, Zanke BW, Gallinger S, Hudson TJ, Weksberg R. Discovery of cross-reactive probes and polymorphic CpGs in the illumina infinium HumanMethylation450 microarray. Epigenetics. 2013; 8:203–09. https://doi.org/10.4161/epi.23470 PMID:23314698

27. Teschendorff AE, Marabita F, Lechner M, Bartlett T, Tegner J, Gomez-Cabrero D, Beck S. A beta-mixture quantile normalization method for correcting probe design bias in illumina infinium 450 K DNA methylation data. Bioinformatics. 2013; 29:189–96. https://doi.org/10.1093/bioinformatics/bts680 PMID:23175756

28. Logue MW, Smith AK, Wolf EJ, Maniates H, Stone A, Schichman SA, McGlinchey RE, Milberg W, Miller MW. The correlation of methylation levels measured using illumina 450K and EPIC BeadChips in blood samples. Epigenomics. 2017; 9:1363–71. https://doi.org/10.2217/epi-2017-0078 PMID:28809127

29. Watanabe K, Taskesen E, van Bochoven A, Posthuma D. Functional mapping and annotation of genetic associations with FUMA. Nat Commun. 2017; 8:1826. https://doi.org/10.1038/s41467-017-01261-5 PMID:29184056

30. Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res. 2000; 28:27–30. https://doi.org/10.1093/nar/28.1.27 PMID:10592173

31. Fabregat-Traver D, Sharapov SZ, Hayward C, Rudan I, Campbell H, Aulchenko Y, Bientinesi P. High-Performance Mixed Models Based Genome-Wide Association Analysis with omicABEL software. F1000Res. 2014; 3:200. https://doi.org/10.12688/f1000research.4867.1 PMID:25717363

32. Schurmann C, Heim K, Schillert A, Blankenberg S, Carstensen M, Dörr M, Endlich K, Felix SB, Gieger C, Grallert H, Herder C, Hoffmann W, Homuth G, et al. Analyzing illumina gene expression microarray data from different tissues: methodological aspects of data analysis in the metaxpress consortium. PLoS One. 2012; 7:e50938. https://doi.org/10.1371/journal.pone.0050938 PMID:23236413

33. Laaksonen MA, Härkänen T, Knekt P, Virtala E, Oja H. Estimation of population attributable fraction (PAF) for disease occurrence in a cohort study design. Stat Med. 2010; 29:860–74. https://doi.org/10.1002/sim.3792 PMID:20213711

34. Pilling LC, Kuo CL, Sicinski K, Tamosauskaite J, Kuchel GA, Harries LW, Herd P, Wallace R, Ferrucci L, Melzer D. Human longevity: 25 genetic loci associated in 389,166 UK biobank participants. Aging (Albany NY). 2017; 9:2504–20. https://doi.org/10.18632/aging.101334 PMID:29227965

35. Nikpay M, Goel A, Won HH, Hall LM, Willenborg C, Kanoni S, Saleheen D, Kyriakou T, Nelson CP, Hopewell JC, Webb TR, Zeng L, Dehghan A, et al. A comprehensive 1,000 genomes-based genome-wide association meta-analysis of coronary artery disease. Nat Genet. 2015; 47:1121–30. https://doi.org/10.1038/ng.3396 PMID:26343387

36. Pattaro C, Teumer A, Gorski M, Chu AY, Li M, Mijatovic V, Garnaas M, Tin A, Sorice R, Li Y, Taliun D, Olden M, Foster M, et al. ICBP Consortium, AGEN Consortium, CARDIOGRAM, CHARGe-Heart Failure Group, and ECHOGen Consortium. Genetic associations at 53 loci highlight cell types and biological pathways relevant for kidney function. Nat Commun. 2016; 7:10023. https://doi.org/10.1038/ncomms10023 PMID:26831199

37. Warren HR, Evangelou E, Cabrera CP, Gao H, Ren M, Mifsud B, Ntalla I, Surendran P, Liu C, Cook JP, Kraja AT, Drenos F, Loh M, et al, and International Consortium of Blood Pressure (ICBP) 1000G Analyses, and BIOS Consortium, and Lifelines Cohort Study, and Understanding Society Scientific group, and CHD Exome Consortium, and ExomeBP Consortium, and T2D-GENES Consortium, and GoT2DGenes Consortium, Cohorts for Heart and Ageing Research in Genome Epidemiology (CHARGE) BP Exome Consortium, and International Genomics of Blood Pressure (iGEN-BP) Consortium, and UK Biobank CardioMetabolic

Consortium BP working group. Genome-wide association analysis identifies novel blood pressure loci and offers biological insights into cardiovascular risk. Nat Genet. 2017; 49:403–15.
https://doi.org/10.1038/ng.3768
PMID:28135244

38. Gaulton KJ, Ferreira T, Lee Y, Raimondo A, Mägi R, Reschen ME, Mahajan A, Locke A, Rayner NW, Robertson N, Scott RA, Prokopenko I, Scott LJ, et al, and DIAbetes Genetics Replication And Meta-analysis (DIAGRAM) Consortium. Genetic fine mapping and genomic annotation defines causal mechanisms at type 2 diabetes susceptibility loci. Nat Genet. 2015; 47:1415–25.
https://doi.org/10.1038/ng.3437
PMID:26551672

39. Gaunt TR, Shihab HA, Hemani G, Min JL, Woodward G, Lyttleton O, Zheng J, Duggirala A, McArdle WL, Ho K, Ring SM, Evans DM, Davey Smith G, Relton CL. Systematic identification of genetic influences on methylation across the human life course. Genome Biol. 2016; 17:61.
https://doi.org/10.1186/s13059-016-0926-z
PMID:27036880

40. Hemani G, Zheng J, Elsworth B, Wade KH, Haberland V, Baird D, Laurin C, Burgess S, Bowden J, Langdon R, Tan VY, Yarmolinsky J, Shihab HA, et al. The MR-base platform supports systematic causal inference across the human phenome. Elife. 2018; 7:e34408.
https://doi.org/10.7554/eLife.34408
PMID:29846171

41. Didelez V, Sheehan N. Mendelian randomization as an instrumental variable approach to causal inference. Stat Methods Med Res. 2007; 16:309–30.
https://doi.org/10.1177/0962280206077743
PMID:17715159

42. Bowden J, Davey Smith G, Burgess S. Mendelian randomization with invalid instruments: effect estimation and bias detection through egger regression. Int J Epidemiol. 2015; 44:512–25.
https://doi.org/10.1093/ije/dyv080
PMID:26050253

43. Bowden J, Davey Smith G, Haycock PC, Burgess S. Consistent estimation in mendelian randomization with some invalid instruments using a weighted median estimator. Genet Epidemiol. 2016; 40:304–14.
https://doi.org/10.1002/gepi.21965
PMID:27061298

44. Hartwig FP, Davey Smith G, Bowden J. Robust inference in summary data mendelian randomization via the zero modal pleiotropy assumption. Int J Epidemiol. 2017; 46:1985–98.
https://doi.org/10.1093/ije/dyx102
PMID:29040600

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

46. Koestler DC, Usset J, Christensen BC, Marsit CJ, Karagas MR, Kelsey KT, Wiencke JK. DNA methylation-derived neutrophil-to-lymphocyte ratio: an epigenetic tool to explore cancer inflammation and outcomes. Cancer Epidemiol Biomarkers Prev. 2017; 26:328–38.
https://doi.org/10.1158/1055-9965.EPI-16-0461
PMID:27965295

SUPPLEMENTARY MATERIALS

Supplementary material - cohort description

The atherosclerosis risk in communities (ARIC) study

ARIC cohort description

The ARIC Study is a population-based prospective cohort study of cardiovascular disease risk in four US communities [1]. Between 1987 and 1989, 7,082 men and 8,710 women aged 45–64 years were enrolled in Forsyth County, North Carolina; Jackson, Mississippi (African Americans only); suburban Minneapolis, Minnesota; and Washington County, Maryland. The ARIC Study protocol was approved by the institutional review board of each participating university, and participants provided written informed consent. Participants underwent a baseline clinical examination (Visit 1) and four subsequent follow-up clinical exams (Visits 2–5). The present analysis is restricted to African Americans from Jackson and Forsyth County centers. Baseline for mortality follow-up is either Visit 2 (1990–1992) or Visit 3 (1993–1995), when the DNA used for methylation quantification was collected. Covariates were measured at the time of blood draw, unless otherwise specified. Data on education, smoking status, smoking pack-years, alcohol intake, and physical activity were obtained by self-report at Visit 1. Trained technicians took fasting blood samples and measured height and weight using standard protocols. Diabetes was defined as a fasting blood glucose level of ≥126 mg/dL, non-fasting blood glucose level of ≥200 mg/dL, self-reported physician diagnosis of diabetes, or use of antidiabetic medication in the past 2 weeks. Hypertension was defined as systolic blood pressure ≥140 mm Hg, diastolic blood pressure ≥90 mm Hg, or self-reported use of antihypertensive medication in the past 2 weeks. History of cancer was defined by self-report or incident cancer cases found between Visit 1 and time of blood draw found through cancer registry and hospital linkage. History of coronary heart disease (CHD) was defined as self-reported history at baseline or an adjudicated event (Myocardial infarction (MI), silent MI, coronary artery bypass surgery, or angioplasty) found between Visit 1 and time of blood draw.

ARIC death ascertainment

Deaths among cohort participants were identified through December 2012 via annual telephone calls and by surveillance of local death certificates and obituaries. If a participant was lost to telephone follow-up, a National Death Index search was conducted.

ARIC DNA methylation quantification

Genomic DNA was extracted from peripheral blood leukocyte samples using the Gentra Puregene Blood Kit (Qiagen; Valencia, CA, USA) according to the manufacturer's instructions (https://www.qiagen.com). Bisulfite conversion of 1 μg genomic DNA was performed using the EZ-96 DNA Methylation Kit (Deep Well Format) (Zymo Research; Irvine, CA, USA) according to the manufacturer's instructions (https://www.zymoresearch.com). Bisulfite conversion efficiency was determined by PCR amplification of converted DNA before proceeding with methylation analyses on the Illumina platform using Zymo Research's Universal Methylated Human DNA Standard and Control Primers. The Illumina Infinium HumanMethylation450K Beadchip array (HM450K) was used to measure DNA methylation (Illumina, Inc.; San Diego, CA, USA). Background subtraction was conducted with the GenomeStudio software using built-in negative control bead types on the array. Positive and negative controls and sample replicates were included on each 96-well plate assayed. After exclusion of controls, replicates, and samples with integrity issues or failed bisulfite conversion, a total of 2,841 study participants had HM450K data available for further quality control (QC) analyses. We removed poor-quality samples with pass rate of <99% (i.e., if the sample had at least 1% of CpG sites with detection *P*-value > 0.01 or missing), indicative of lower DNA quality or incomplete bisulfite conversion, and samples with a possible gender mismatch based on evaluation of selected CpG sites on the Y chromosome. Additional details have been published elsewhere [2 , 3].

Framingham heart study offspring cohort (FHS)

FHS study participants

The FHS Offspring Cohort began enrollment in 1971 and included 5,124 offspring of the FHS original cohort as well as spouses of the offspring. Participants were eligible for the current study if they attended the eighth examination cycle (2005–2008) and consented to have their DNA used for genetic research. All participants provided written informed consent at the time of each examination visit. The study protocol was approved by the Institutional Review Board at Boston University Medical Center (Boston, MA). FHS data are available in dbGaP (accession number: phs000724.v2.p9).

FHS death ascertainment

Deaths among FHS participants that occurred before January 1, 2013 were ascertained using multiple strategies, including routine contact with participants for health history updates, surveillance at the local

hospital, obituaries in the local newspaper, and queries to the National Death Index. Death certificates, hospital and nursing home records before death, and autopsy reports were requested. When cause of death was undeterminable, the next of kin were interviewed. The date and cause of death were reviewed by an endpoint panel of three investigators.

FHS DNA methylation quantification

Peripheral blood samples were collected at the 8th examination. Genomic DNA was extracted from buffy coats using the Gentra Puregene DNA extraction kit (Qiagen) and bisulfite converted using the EZ DNA Methylation kit (Zymo Research). DNA methylation quantification was conducted in two laboratory batches using the Illumina Infinium HumanMethylation450 array. Methylation beta values were generated using the Bioconductor *minfi* package with background correction. Sample exclusion criteria included poor SNP matching of control positions, missing rate >1%, outliers from multi-dimensional scaling, and sex mismatch. In addition, we excluded individuals with leukemia and those who received chemotherapy. Additional sample exclusions included those with mismatches in their reported sex and methylation-predicted sex as well as methylation-predicted tissues that were not blood. Lastly, samples with correlation with our reference population of $r < 0.80$ were excluded. Predicted sex, tissues, correlation with reference population, and DNA methylation-predicted ages were computed using our online age calculator (http://labs.genetics.ucla.edu/horvath/dnamage). Background subtraction was applied using the *preprocessIllumina* command in the *minfi* Bioconductor package [4]. In total, 2,635 samples and 443,304 CpG probes remained for analysis.

Invecchiare in chianti (InCHIANTI) study

InChianti study participants

The InCHIANTI Study is a population-based prospective cohort study of residents ≥20 years old from two areas in the Chianti region of Tuscany, Italy. Sampling and data collection procedures have been described elsewhere [5]. Briefly, 1,326 participants donated a blood sample at baseline (1998–2000), of which 784 also donated a blood sample at 9-year follow-up (2007–2009). DNA methylation was assayed using the Illumina Infinium HumanMethylation450 platform in DNA samples corresponding to participants with sufficient DNA at both baseline and Year 9 visits (n = 499). All participants provided written informed consent to participate in this study. The study complied with the Declaration of Helsinki. The Italian National Institute of Research and Care on Aging Institutional Review Board approved the study protocol.

InChianti death ascertainment

Vital status was ascertained using data from the Tuscany Regional Mortality General Registry. Deaths were assessed until December 1, 2014.

InChianti DNA methylation quantification

Genomic DNA was extracted from buffy coat samples using an AutoGen Flex and quantified on a Nanodrop1000 spectrophotometer before bisulfite conversion. Genomic DNA was bisulfite converted using the Zymo EZ-96 DNA Methylation Kit (Zymo Research) per the manufacturer's protocol. CpG methylation status of 485, 577 CpG sites was determined using the Illumina Infinium HumanMethylation450 BeadChip per the manufacturer's protocol, as previously described [6]. Initial data analysis was performed using GenomeStudio 2011.1 (Model M Version 1.9.0, Illumina Inc.). Threshold call rate for inclusion of samples was 95%. Quality control of sample handling included comparison of clinically reported sex versus sex of the same samples determined by analysis of methylation levels of CpG sites on the X chromosome [6]. Background subtraction was applied using the *preprocessIllumina* command in the *minfi* Bioconductor package [4].

Cooperative health research in the region of augsburg (KORA) F4 cohort

KORA cohort description

The KORA study is an independent population-based cohort from Augsburg, Southern Germany. Whole blood samples of the KORA F4 survey (examination 2006–2008), a seven-year follow-up study of the KORA S4 cohort, were used. Out of 4,621 participants for the KORA S4 baseline study, 3,080 participants participated in the KORA F4 follow-up study [7]. Participants provided written informed consent, and the study was approved by the local ethics committee (Bayerische Landesärztekammer). For 1,799 subjects, methylation data as well as information about death ascertainment was available. Before analyses, all individuals with a detection *P*-value > 0.05 for >1% of probes were removed (375 individuals). Sex checks performed during calculation of DNAmAge resulted in the removal of another 167 individuals, 137 of whom had an "unsure" gender. This left 1,257 individuals for analysis. At the KORA F4 follow-up examination, all individuals completed questionnaires and physical examinations conducted by trained staff covering demographics, lifestyle, and medical history since the KORA S4 examination. Collected information included age, sex, years of education, smoking status (current regular, current irregular, former, never), pack-years, alcohol consumption

(g/day), physical activity (active, inactive), diabetes status, hypertension status, self-reported cancer diagnosis, and body mass index (BMI), among other clinical variables [7].

KORA mortality ascertainment

The vital status of all F4 participants was ascertained through the population registries inside and outside the study area in 2011 (cut-off date: December 31, 2011). Record linkage was based on name, sex, date of birth, and address. If the person died, the time and location of death was assessed via population registries, and a copy of the death certificate was obtained from the Regional Health Department. If the person moved out of the study area, time of the move and information on the new address was typically available. Vital status could not be assessed for those who had moved to a foreign country or to an unknown location in the country. Causes of death were ICD-9 revision coded. There were a total of 42 deaths, including 16 from cardiovascular disease and 17 from cancer.

KORA DNA methylation measures

Whole blood was drawn into serum gel tubes. We bisulfite-converted 1 μg of genomic DNA using the EZ-96 DNA Methylation Kit (Zymo Research) according to the manufacturer's procedure, with the alternative incubation conditions recommended when using the Illumina Infinium Methylation Assay. Genome-wide DNA methylation was analyzed in 1,799 subjects using the Illumina Infinium HumanMethylation450 BeadChip Array. Raw methylation data were extracted using the Illumina Genome Studio (version 2011.1) with the methylation module (version 1.9.0). Preprocessing was performed with R (version 3.0.1). Probes with signals from less than three functional beads and probes with a detection *P*-value > 0.01 were defined as low-confidence probes. Probes that covered SNPs (MAF in Europeans > 5%) were excluded from the data set. A color bias adjustment was performed with the R package lumi (version 2.12.0) by smooth quantile normalization and background correction based on negative control probes present on the Infinium HumanMethylation BeadChip. This was performed separately for the two-color channels and chips. β-values corresponding to low-confidence probes were set to missing. A 95% call rate threshold was applied on samples and CpG sites. Beta-mixture quantile normalization (BMIQ) was applied by using the R package wateRmelon, version 1.0.3. Plate and batch effects were investigated by principle component analysis and eigenR2 analysis, because KORA F4 samples were processed on 20 96-well plates across nine different batches.

Probes with a detection $P > 0.05$ for > 1% of samples were removed as well as all "ch" and "rs" probes, leaving a total of 431, 217 probes for analysis. Although raw beta values were used in Dr. Horvath's online calculator to determine cell counts, normalized data was used for the final analyses.

To reduce non-biological variability between observations, data were normalized using quantile normalization on raw signal intensities. Precisely, quantile normalization was stratified to six probe categories based on probe type and color channel (i.e., Infinium I signals from beads targeting methylated CpG sites obtained through red and green color channels, Infinium I signals from beads targeting unmethylated CpG sites obtained through red and green color channels, and Infinium II signals obtained through red and green color channels [8]) using the R package *limma*, version 3.16.5 [9]. Further, to correct the shift in the distribution of methylation values observed for the two different assay designs (Infinium I and Infinium II) on the BeadChip, BMIQ was applied [10] using the R package wateRmelon, version 1.0.3 [11].

Lothian birth cohorts of 1921 and 1936 (LBC1921 and LBC1936)

LBC cohort description

LBC1921 and LBC1936 are two longitudinal studies of aging [12, 13] that derive from the Scottish Mental Surveys of 1932 and 1947, respectively, when nearly all 11-year-old children in Scotland completed a test of general cognitive ability [14]. Survivors living in the Lothian area of Scotland were recruited in late-life at a mean age of 79 years for LBC1921 (n = 550) and mean age of 70 years for LBC1936 (n = 1,091). Follow-up took place at ages 70, 73, and 76 years in LBC1936 and ages 79, 83, 87, and 90 years in LBC1921. Collected data include genetic information, longitudinal epigenetic information, longitudinal brain imaging (LBC1936), numerous blood biomarkers, and anthropomorphic and lifestyle measures. Post-QC, DNA methylation data were available for 920 LBC1936 participants at age 70 years and for 446 LBC1921 participants at age 79 years. At each in-person visit, participants completed questionnaires regarding demography, lifestyle, and medical history. They reported chronological age, years of education, smoking status (never, former, current), pack-years consumption (continuous), alcohol consumption (light, moderate, and heavy drinkers), self-reported type 2 diabetes, cancer, and hypertension. BMI was computed from anthropometric measures. Participants were asked to remove their shoes before a SECA stadiometer was used to assess height in centimeters. Weight (after removing shoes and outer clothing) was measured in

kilograms using a digital readout from electronic SECA scales.

LBC mortality ascertainment

For both LBC1921 and LBC1936, mortality status was obtained via data linkage from the National Health Service Central Register, provided by the General Register Office for Scotland (now National Records of Scotland). Participant deaths and cause of death are routinely flagged to the research team about every 12 weeks. The last update available for the current project was 26th November 2014.

LBC DNA methylation measures

Detailed information about collection and QC steps on LBC methylation data have been reported previously [12, 15]. Briefly, the Illumina Infinium HumanMethylation450 BeadChip was used to measure DNA methylation in whole blood of consenting participants. Background correction was performed, and QC was used to remove probes with a low detection rate, low quality (manual inspection), and low call rate as well as samples with a poor match between genotypes and SNP control probes or incorrect predicted sex. Additional QC was performed to remove samples and probes in which >1% of probes or samples, respectively, had a detection $P > 0.05$. The working set included 442,227 CpG probes.

Normative aging study (NAS)

NAS cohort description

The ongoing longitudinal US Department of Veterans Affairs NAS was established in 1963 and included men 21–80 years old and free of known chronic medical conditions at entry [16]. Participants were invited to medical examinations every three to five years. At each visit, men provided information on medical history, lifestyle, and demographic factors and underwent physical examinations and laboratory tests. DNA samples were collected from 675 active participants between 1999–2007 [16]. We excluded participants who were non-white or who reported leukemia at the time of DNA extraction, leaving a total of 646 individuals with a single observation each. Participants provided written informed consent at each visit. The NAS study was approved by the institutional review boards of participating institutions. At each in-person visit, participants completed questionnaires regarding demography, lifestyle, and medical history. They reported chronological age, years of education, smoking status (never, former, current), pack-years consumption (continuous), alcohol consumption (<2, ≥2 drinks/day), physical activity (<12, 12–30, ≥30 metabolic equivalent hours [MET-h] per week), type 2 diabetes (self-reported diagnosis and/or use of diabetes medications), diagnosis of CHD (validated on medical records, ECG, and physician exams), diagnosis of malignant cancer in the five years prior the visit (diagnosed with ICD-9 code). High blood pressure was defined as antihypertensive medication use, systolic blood pressure ≥140 mmHg, or diastolic blood pressure ≥90 mmHg at study visit. BMI was computed from anthropometric measures, performed with participants in undershorts and socks [17].

NAS mortality ascertainment

Official death certificates were obtained for decedents from the appropriate state health departments and were reviewed by a physician. An experienced research nurse coded the cause of death using ICD-9. Both participant deaths and causes of death were routinely updated by the research team, and the last update available was December 31, 2013 [12].

NAS DNA methylation measures

DNA was extracted from buffy coats using the QIAamp DNA Blood Kit (Qiagen). We used 500 ng of DNA for bisulfite conversion using the EZ-96 DNA Methylation Kit (Zymo Research). To reduce chip and plate effects, we used a two-stage age-stratified algorithm to randomize samples and ensure similar age distributions across chips and plates; 12 samples that were sampled across all age quartiles were randomized to each chip, and then chips were randomized to plates (8 chips/plate).

QC analysis was performed to remove samples and probes, where >1% of probes or samples, respectively, had a detection $P > 0.05$. Remaining samples were preprocessed using the Illumina-type background correction [18] and normalized with dye-bias [19] and BMIQ [20] adjustments, which were used to generate beta methylation values. The working set included 477, 928 CpG probes. DNA methylation age was computed using the Horvath calculator from background-corrected methylation data, and QC analysis was performed only on samples, leaving 485, 512 CpG and CpH probes in the working set.

TwinsUK

TwinsUK study participants

The TwinsUK cohort was established in 1992 and recruited both monozygotic and dizygotic same-sex twins in the United Kingdom. The majority of participants are female, Caucasian, and mostly disease-free at time of ascertainment. There are >13, 000 twin participants in the cohort, of which 805 were included in the current study. Whole blood samples were collected during participants' clinical visits, along with questionnaire data on phenotype and lifestyle factors. All subjects provided written informed consent [21].

Information on physical activity, smoking pack-years, plate number, and chip position number were not available for subjects in the TwinsUK dataset and therefore were not adjusted as covariates in all analyses.

TwinsUK death ascertainment

Mortality data were collected using two approaches: 1) during routine contact for standard clinical visits in TwinsUK, and 2) using queries to the National Death Register. Date and cause of death were recorded.

TwinsUK DNA methylation quantification

DNA samples were extracted from whole blood using the DNeasy kit (Qiagen). DNA was bisulfite converted using the EZ DNA methylation kit (Zymo Research). Methylation levels were profiled using the Illumina Infinium HumanMethylation450 array, and methylation betas were generated using the R package *minfi* with background correction. Raw beta levels were subjected to BMIQ dilation to correct for technical effects. Probe exclusion criteria included probes that mapped to multiple locations in the reference sequence and probes in which >1% of subjects had detection $P > 0.05$. Individuals with >5% missing probes, with mismatched sex, and with mismatched genotypes were also excluded. Methylation-predicted sex, methylation-predicted blood cell types, correlations with the reference population, and DNA methylation-predicted age were computed using the online epigenetic age calculator (http://labs.genetics.ucla.edu/horvath/dnamage).

Women's health initiative–broad agency announcement 23 (WHI-BAA23)

WHI-BAA23 cohort description

Subjects included a subsample of participants of the WHI study, a national study that began in 1993 and enrolled postmenopausal women 50–79 years of age into one of three randomized clinical trials. Women were selected from one of two WHI large sub-cohorts that had previously undergone genome-wide genotyping as well as profiling for 7 cardiovascular disease-related biomarkers, including total cholesterol, high-density lipoprotein, low-density lipoprotein, triglycerides, C-reactive protein (CRP), creatinine, insulin, and glucose through two core WHI ancillary studies [22]. The first cohort is the WHI SNP Health Association Resource (SHARe) cohort of minorities that includes >8,000 African American (AA) women and >3,500 Hispanic women. Women were genotyped through the WHI core study M5-SHARe (www.whi.org/researchers/data/WHIStudies/StudySites/M5) and underwent biomarker profiling through WHI Core study W54-SHARe (...data/WHIStudies/StudySites/W54). The second cohort consists of a combination of European Americans (EA) from two hormonal therapy trials selected for GWAS and biomarkers in core studies W58 (.../data /WHIStudies/StudySites/W58) and W63 (.../data/WHIStudies/StudySites/W63). From these two cohorts, two sample sets were formed. Sample Set 1 is a sample set of 637 CHD cases and 631 non-CHD cases as of Sept 30, 2010. Sample Set 2 is a non-overlapping sample of 432 cases of CHD and 472 non-CHD cases as of September 17, 2012. The ethnic groups differed in terms of age distribution, as Caucasian women tended to be older. We acknowledge a potential for selection bias using the above-described sampling scheme in WHI but suspect that if such bias is present, it is minimal. First, selection bias is introduced by restricting our methylation profiling at baseline to women with GWAS and biomarker data from baseline as well, given the requirement that these subjects must have signed the WHI supplemental consent for broad sharing of genetic data in 2005. However, we believe that selection bias at this stage is minimized by inclusion of subjects who died between time of start of the WHI study and time of supplemental consent in 2005, which excluded only ~6%–8% of all WHI participants. Subjects unable or unwilling to sign consent in 2005 may not represent a random subset of all participants who survived to 2005. Second, some selection bias may also occur if similar gross differences exist in the characteristics of participants who consented to be followed in the two WHI extension studies beginning in 2005 and 2010 compared to non-participants at each stage. We believe these selection biases, if present, have minimal effects on our effect estimates. Data are available from this page: https://www.whi.org/researchers/Stories/June%202015%20WHI%20Investigators'%20Datasets%20Released.aspx, as well as https://www.whi.org/researchers/data/Documents/WHI%20Data%20Preparation%20and%20Use.pdf

WHI-BAA23 death ascertainment

We used the variable "DEATHALL" from form 124/120 that incorporated any report of death (as of August 2015).

WHI-BAA23 DNA methylation quantification

In brief, bisulfite conversion using the Zymo EZ DNA Methylation Kit (Zymo Research) as well as subsequent hybridization of the Illumina HumanMethylation450k Bead Chip and scanning (iScan, Illumina) were performed according to the manufacturer's protocols by applying standard settings. DNA methylation levels (β values) were determined by calculating the ratio of intensities between methylated (signal A) and un-methylated (signal B) sites. Specifically, β value was calculated from the intensity of methylated (M corresponding to signal A) and un-methylated (U corresponding to signal B) sites, as the ratio of fluorescent signals $\beta = \text{Max}(M,0)/[\text{Max}(M,0)+\text{Max}$

(U,0)+100]. Thus, β values range from 0 (completely un-methylated) to 1 (completely methylated).

Women's health initiative—epigenetic mechanisms of PM-Mediated CVD risk (WHI-EMPC)

WHI-EMPC cohort description

WHI-EMPC is an ancillary study of epigenetic mechanisms underlying associations between ambient particulate matter (PM) air pollution and cardiovascular disease in the WHI clinical trials (CT) cohort. It is funded by the National Institute of Environmental Health Sciences (R01-ES020836).

The WHI-EMPC study population is a stratified, random sample of 2,200 WHI CT participants who were examined in 1993–2001; had available buffy coats, core analytes, electrocardiograms, and ambient concentrations of PM; and were not taking anti-arrhythmic medications at the time.

As such, WHI-EMPC is representative of the larger, multiethnic WHI CT population from which it was sampled: 68 ,132 participants aged 50–79 years who were randomized to hormone therapy, calcium/vitamin D supplementation, and/or dietary modification in 40 U.S. clinical centers at baseline exam (1993–1998) and re-examined in the fasting state one, three, six, and nine years later [23, 24]. During participant visits, data on age, race/ethnicity, education, smoking status (current, former, never), pack-years of smoking, alcohol consumption (drinks per week), recreational physical activity (MET-hours/week), weight/height/BMI, systolic and diastolic blood pressure, medication use, CHD, type 2 diabetes, and cancer diagnosis were obtained.

Hypertension status was based on systolic blood pressure ≥140 mmHg or diastolic blood pressure ≥90 mmHg or antihypertensive medication use (angiotensin converting enzyme inhibitors, angiotensin II receptor antagonists, beta blockers, calcium channel blockers, thiazides). CHD was defined by a history of myocardial infarction (acute, hospitalized, definite or probable events supported by cardiac pain, electrocardiogram, and biomarker data) or revascularization procedure (coronary artery bypass graft, percutaneous coronary angioplasty, stent) and was self-reported at baseline and confirmed by physician-review, classification, and local/central adjudication of medical records during follow-up. Type 2 diabetes was defined by a self-reported history of physician-treated diabetes, fasting glucose ≥126 mg/dL, non-fasting glucose ≥200 mg/dL, or anti-diabetic medication use. Cancer was defined by a diagnosis of any cancer, excluding leukemia and other hematologic malignancies (Hodgkin's lymphoma, non-Hodgkin's lymphoma, multiple myeloma).

Current analyses involve information collected at the first available visit with available DNA methylation data and stratification by race/ethnicity [European (WHI-EMPC-EA) and African American (WHI-EMPC-AA) ancestries].

WHI-EMPC mortality ascertainment

All-cause mortality and sub-classification of the underlying cause of death to cardiovascular or cancer mortality were based on WHI physician review of death certificates, medical records, and autopsy reports. Cardiovascular disease mortality was defined as death due to definite or possible CHD, cerebrovascular disease, or other or unknown cardiovascular disease. Cancer mortality was defined as death due to any cancer. Participants affected by leukemia or other hematologic malignancies (i.e., Hodgkin's lymphoma, non-Hodgkin's lymphoma, multiple myeloma) were excluded due to known effects on red cell, white cell, and platelet counts.

WHI-EMPC DNA methylation quantification

Genome-wide DNA methylation at CpG sites was measured using the Illumina 450K Infinium Methylation BeadChip, quantitatively represented by beta (percentage of methylated cytosines over the sum of methylated and unmethylated cytosines) and quality-controlled using the following filters: detection $P > 0.01$ in >10% of samples, detection $P > 0.01$ or missing in >1% of probes, and probes with a coefficient of variation <5%, yielding values of beta at 293,171 sites. DNA methylation data were normalized using BMIQ [25] and stage-adjusted using ComBat [10]. Modeled epigenome-wide associations also adjusted for cell subtype proportions (CD8-T, CD4-T, B cell, natural killer, monocyte, and granulocyte) [26] and for technical covariates, including plate, chip, row, and column.

REFERENCES

1. The atherosclerosis risk in communities (ARIC) study: design and objectives. The ARIC investigators. Am J Epidemiol. 1989; 129:687–702. PMID:2646917

2. Demerath EW, Guan W, Grove ML, Aslibekyan S, Mendelson M, Zhou YH, Hedman ÅK, Sandling JK, Li LA, Irvin MR, Zhi D, Deloukas P, Liang L, et al. Epigenome-wide association study (EWAS) of BMI, BMI change and waist circumference in african american adults identifies multiple replicated loci. Hum Mol Genet. 2015; 24:4464–79. https://doi.org/10.1093/hmg/ddv161 PMID:25935004

3. Bose M, Wu C, Pankow JS, Demerath EW, Bressler J, Fornage M, Grove ML, Mosley TH, Hicks C, North K, Kao WH, Zhang Y, Boerwinkle E, Guan W. Evaluation of microarray-based DNA methylation measurement using technical replicates: the atherosclerosis risk in communities (ARIC) study. BMC Bioinformatics. 2014; 15:312. https://doi.org/10.1186/1471-2105-15-312 PMID:25239148

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

5. Ferrucci L, Bandinelli S, Benvenuti E, Di Iorio A, Macchi C, Harris TB, Guralnik JM. Subsystems contributing to the decline in ability to walk: bridging the gap between epidemiology and geriatric practice in the InCHIANTI study. J Am Geriatr Soc. 2000; 48:1618–25. https://doi.org/10.1111/j.1532-5415.2000.tb03873.x PMID:11129752

6. Moore AZ, Hernandez DG, Tanaka T, Pilling LC, Nalls MA, Bandinelli S, Singleton AB, Ferrucci L. Change in epigenome-wide DNA methylation over 9 years and subsequent mortality: results from the InCHIANTI study. J Gerontol A Biol Sci Med Sci. 2016; 71:1029–35. https://doi.org/10.1093/gerona/glv118 PMID:26355017

7. Rückert IM, Heier M, Rathmann W, Baumeister SE, Döring A, Meisinger C. Association between markers of fatty liver disease and impaired glucose regulation in men and women from the general population: the KORA-F4-study. PLoS One. 2011; 6:e22932. https://doi.org/10.1371/journal.pone.0022932 PMID:21850244

8. Bibikova M, Barnes B, Tsan C, Ho V, Klotzle B, Le JM, Delano D, Zhang L, Schroth GP, Gunderson KL, Fan JB, Shen R. High density DNA methylation array with single CpG site resolution. Genomics. 2011; 98:288–95. https://doi.org/10.1016/j.ygeno.2011.07.007 PMID:21839163

9. Smyth GK. Limma: linear models for microarray data. Bioinformatics and computational biology solutions using R and Bioconductor: Springer. 2005; 397–420. https://doi.org/10.1007/0-387-29362-0_23

10. Teschendorff AE, Marabita F, Lechner M, Bartlett T, Tegner J, Gomez-Cabrero D, Beck S. A beta-mixture quantile normalization method for correcting probe design bias in illumina infinium 450 K DNA methylation data. Bioinformatics. 2013; 29:189–96. https://doi.org/10.1093/bioinformatics/bts680 PMID:23175756

11. Pidsley R, Y Wong CC, Volta M, Lunnon K, Mill J, Schalkwyk LC. A data-driven approach to preprocessing illumina 450K methylation array data. BMC Genomics. 2013; 14:293. https://doi.org/10.1186/1471-2164-14-293 PMID:23631413

12. Marioni RE, Shah S, McRae AF, Chen BH, Colicino E, Harris SE, Gibson J, Henders AK, Redmond P, Cox SR, Pattie A, Corley J, Murphy L, et al. DNA methylation age of blood predicts all-cause mortality in later life. Genome Biol. 2015; 16:25. https://doi.org/10.1186/s13059-015-0584-6 PMID:25633388

13. Taylor AM, Pattie A, Deary IJ. Cohort profile update: the lothian birth cohorts of 1921 and 1936. Int J Epidemiol. 2018; 47:1042–42r. https://doi.org/10.1093/ije/dyy022 PMID:29546429

14. Marioni RE, McRae AF, Bressler J, Colicino E, Hannon E, Li S, Prada D, Smith JA, Trevisi L, Tsai PC, Vojinovic D, Simino J, Levy D, et al. Meta-analysis of epigenome-wide association studies of cognitive abilities. Mol Psychiatry. 2018; 23:2133–44. https://doi.org/10.1038/s41380-017-0008-y PMID:29311653

15. Shah S, McRae AF, Marioni RE, Harris SE, Gibson J, Henders AK, Redmond P, Cox SR, Pattie A, Corley J, Murphy L, Martin NG, Montgomery GW, et al. Genetic and environmental exposures constrain epigenetic drift over the human life course. Genome Res. 2014; 24:1725–33. https://doi.org/10.1101/gr.176933.114 PMID:25249537

16. 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 PMID:5342911

17. Troisi RJ, Heinold JW, Vokonas PS, Weiss ST. Cigarette smoking, dietary intake, and physical activity: effects on body fat distribution—the normative aging study. Am J Clin Nutr. 1991; 53:1104–11. https://doi.org/10.1093/ajcn/53.5.1104 PMID:1850574

18. Triche TJ Jr, Weisenberger DJ, Van Den Berg D, Laird PW, Siegmund KD. Low-level processing of illumina infinium DNA methylation BeadArrays. Nucleic Acids Res. 2013; 41:e90. https://doi.org/10.1093/nar/gkt090 PMID:23476028

19. Davis S, Du P, Bilke S, Triche T, Jr and Bootwalla M.

methylumi: Handle Illumina methylation R package version 2140 2015.

20. Teschendorff AE, Jones A, Fiegl H, Sargent A, Zhuang JJ, Kitchener HC, Widschwendter M. Epigenetic variability in cells of normal cytology is associated with the risk of future morphological transformation. Genome Med. 2012; 4:24. https://doi.org/10.1186/gm323 PMID:22453031

21. Nica AC, Parts L, Glass D, Nisbet J, Barrett A, Sekowska M, Travers M, Potter S, Grundberg E, Small K, Hedman AK, Bataille V, Tzenova Bell J, et al, and MuTHER Consortium. The architecture of gene regulatory variation across multiple human tissues: the MuTHER study. PLoS Genet. 2011; 7:e1002003. https://doi.org/10.1371/journal.pgen.1002003 PMID:21304890

22. Curb JD, McTiernan A, Heckbert SR, Kooperberg C, Stanford J, Nevitt M, Johnson KC, Proulx-Burns L, Pastore L, Criqui M, Daugherty S, and WHI Morbidity and Mortality Committee. Outcomes ascertainment and adjudication methods in the women's health initiative. Ann Epidemiol. 2003; 13:S122–28. https://doi.org/10.1016/s1047-2797(03)00048-6 PMID:14575944

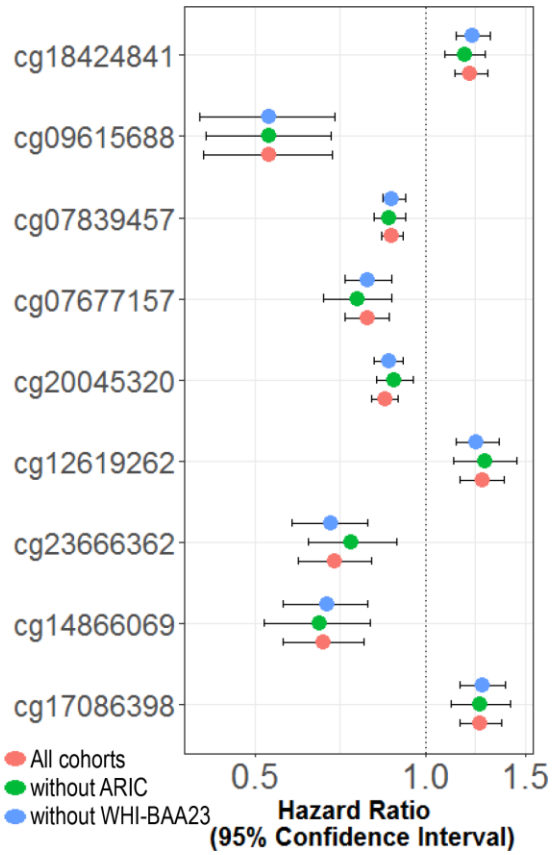
23. Design of the women's health initiative clinical trial and observational study. The women's health initiative study group. Control Clin Trials. 1998; 19:61–109. https://doi.org/10.1016/s0197-2456(97)00078-0 PMID:9492970

24. Anderson GL, Manson J, Wallace R, Lund B, Hall D, Davis S, Shumaker S, Wang CY, Stein E, Prentice RL. Implementation of the women's health initiative study design. Ann Epidemiol. 2003; 13:S5–17. https://doi.org/10.1016/s1047-2797(03)00043-7 PMID:14575938

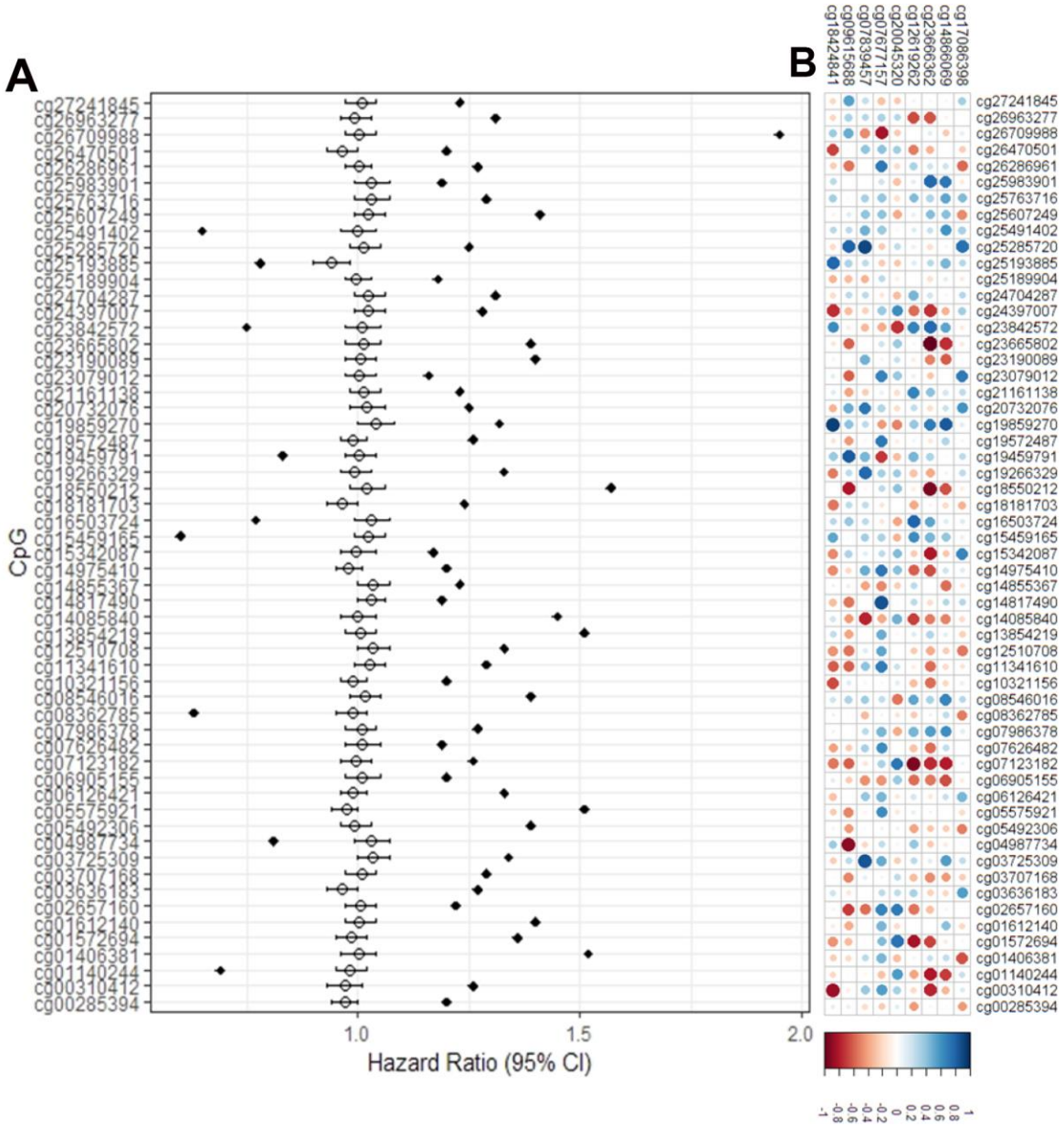
25. Leek JT, Johnson WE, Parker HS, Jaffe AE, Storey JD. The sva package for removing batch effects and other unwanted variation in high-throughput experiments. Bioinformatics. 2012; 28:882–83. https://doi.org/10.1093/bioinformatics/bts034 PMID:22257669

26. Houseman EA, Accomando WP, Koestler DC, Christensen BC, Marsit CJ, Nelson HH, Wiencke JK, Kelsey KT. DNA methylation arrays as surrogate measures of cell mixture distribution. BMC Bioinformatics. 2012; 13:86. https://doi.org/10.1186/1471-2105-13-86 PMID:22568884

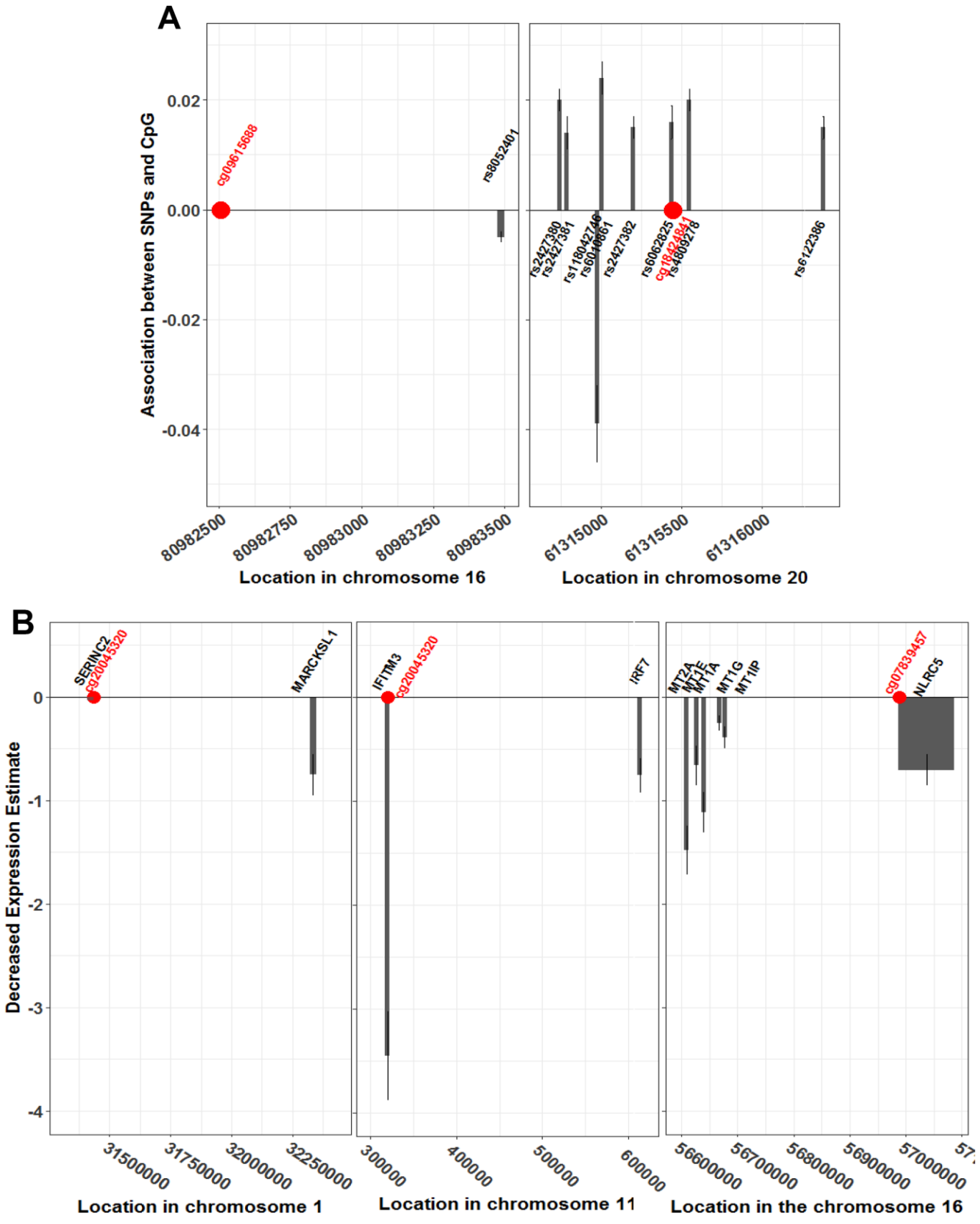
Supplementary Figures



Supplementary Figure 1. Sensitivity analysis comparing hazard ratios of the fully-adjusted meta-analysis, including all cohorts, all excluding ARIC, or all excluding WHI-BAA23.



Supplementary Figure 2. (**A**) All-cause mortality association of 57 out of 58 CpGs identified with mortality in Zhang et al. (black dots) and all-cause mortality association of the same CpGs in the pooled meta-analysis (white dots with 95% confidence intervals). (**B**) Association of methylation levels of 57 out of 58 CpGs identified with mortality in Zhang et al. and our FDR-significant CpGs in all cohorts.



Supplementary Figure 3. (**A**) Methylation quantitative trait loci (meQTL) analysis and (**B**) expression quantitative quantitative trait methylation (eQTM) analysis in KORA.

Supplementary Tables

Please browse Full Text version to see the data of Supplementary Tables 1, 2, 6, 10, 12, 17.

Supplementary Table 1. Covariates included in the analysis of each cohort (*variables that differ across cohorts due to type of cutoff used in data collection) SD = standard deviation; CHD = coronary heart disease.

Supplementary Table 2. CpG loci where blood DNA methylation was associated (FDR<0.05) with all-cause mortality in fixed-effect meta-analysis from the basic model (i.e., age, gender, technical covariates, and white cell subtypes).

Supplementary Table 3. CpG loci where blood DNA methylation was associated (FDR<0.05) with all-cause mortality in the fixed-effect meta-analysis from the fully adjusted model.

Probe name	CHR	Distance to nearest gene (bp)	Nearest gene (10 Mp)[a]	Gene group	Relation to CpG Island	HR[b]	95% CI	*p*	Mean methylation level	Bonferroni significance	FDR-significant in basic model	Methylation level (Mean±SD)
cg17086398	1	0	*SERINC2*	Body		1.25	(1.15;1.36)	4.86E-07	0.29		1	0.29 ± 0.06
cg14866069	4	579	*BMPR1B*			0.66	(0.56;0.78)	4.85E-07	0.85			0.85 ± 0.05
cg23666362	4	516	*MIR1973*	TSS1500		0.69	(0.59;0.8)	8.00E-07	0.82			0.81 ± 0.04
cg12619262	7	6276	*CHST12*			1.26	(1.16;1.37)	1.76E-07	0.75			0.75 ± 0.07
cg20045320	11	116	*IFITM3*		S_Shore	0.85	(0.8;0.9)	4.06E-09	0.54	1	1	0.54 ± 0.09
cg07677157	12		NA[a]			0.79	(0.72;0.86)	2.00E-07	0.16		1	0.18 ± 0.06
cg07839457	16	435	*NLRC5*	TSS1500	N_Shore	0.87	(0.84;0.91)	2.40E-09	0.46	1	1	0.45 ± 0.11
cg09615688	16		NA[a]			0.53	(0.41;0.68)	9.32E-07	0.91		1	0.90 ± 0.03
cg18424841	20		NA[a]		Island	1.2	(1.13;1.28)	2.80E-08	0.7	1		0.69 ± 0.09

[a]Nearest gene was far more than 10 Mp.
[b]Effect estimates represent hazard ratio per 10% increase in DNA methylation. CHR = chromosome; HR = hazard ratio; 95% CI = 95% confidence interval; p = p-value; SD = standard deviation.

Supplementary Table 4. Hazard ratios for FDR-significant fully-adjusted CpGs in basic and fully adjusted models CHR = chromosome; HR = hazard ratio; 95% CI = 95% confidence interval; p = p-value; SD = standard deviation.

Probe name	CHR	Fully adjusted model			Basic model		
		HR	95% CI	*p*	HR	95% CI	*p*
cg17086398	1	1.25	(1.15; 1.36)	4.86E-07	1.37	(1.28; 1.47)	5.32E-20
cg14866069	4	0.66	(0.56; 0.78)	4.85E-07	0.84	(0.75; 0.94)	2.21E-03
cg23666362	4	0.69	(0.59; 0.8)	8.04E-07	0.81	(0.72; 0.9)	2.02E-04
cg12619262	7	1.26	(1.15; 1.38)	1.76E-07	1.13	(1.05; 1.21)	6.77E-04
cg20045320	11	0.85	(0.80; 0.90)	4.06E-09	0.82	(0.78; 0.86)	2.61E-16
cg07677157	12	0.79	(0.72; 0.86)	2.00E-07	0.78	(0.72; 0.84)	1.31E-10
cg07839457	16	0.87	(0.84; 0.91)	2.40E-09	0.88	(0.85; 0.92)	3.40E-11
cg09615688	16	0.53	(0.41; 0.68)	9.32E-07	0.60	(0.51; 0.72)	8.96E-09
cg18424841	20	1.20	(1.13; 1.28)	2.80E-08	1.10	(1.05; 1.15)	2.00E-04

Supplementary Table 5. Summary of models.

Cohorts	Model used	Maximum # probes considered	Basic model		Fully adjusted model	
			Lambda	FDR-significant CpGs	Lambda	FDR-significant CpGs
ARIC	Cox Regression	406712	2.46	226	1.63	5
FHS Study 1	Cox Regression[b]	417934	0.94	3	1.04	0
FHS Study 2	Cox Regression[b]	407580	1.32	17	1.16	1
InChianti	Cox Regression	407179	1.41	0	1.44	0
KORA	Cox Regression[a]	330133	0.96	4	0.94	0
LBC 1921	Cox Regression	393529	1.15	0	1.22	0
LBC 1936	Cox Regression	385450	1.04	0	1	0
NAS	Cox Regression	395005	0.86	0	0.9	0
TwinsUK	Cox Regression [a,c]	426,120	1.02	0	0.98	0
WHI-BAA23	Cox Regression	419176	1.7	14	2.02	1
WHI-EMPC-EA	Cox Regression[d]	250537	1.23	9	0.93	0
WHI-EMPC-AA	Cox Regression[d]	218018	0.92	0	1.09	0
All cohorts	Fixed effect meta-analysis[e]	426724	1.12	257	0.94	9

[a]Cohort used as predictor of residuals from linear regression analysis between each probe and sets of covariates.
[b]Cohort included cluster for family structure.
[c]Cohort included random intercepts for zygosity and family structure.
[d]Cohort considered only CpGs with coefficient of variation >5%.
[e]Analysis of each CpG site included results of at least three cohorts.
Basic model: adjusted for age (categories), gender, technical variables, white blood cell count.
Fully adjusted model: adjusted for age (categories), gender, technical variables, white blood cell count, education level, physical activity, smoking status, smoking consumption (packyears), body mass index (categories), alcohol consumption, prior coronary heart disease (y/n), diabetes (y/n), hypertension (y/n), cancer (y/n).

Supplementary Table 6. I^2 measure of heterogeneity from random-effect meta-analysis in each FDR-significant basic-adjusted CpG.

Supplementary Table 7. I^2 measure of heterogeneity from random-effect meta-analysis in each FDR-significant fully-adjusted CpG.

Probe name	CHR	Distance to nearest gene (bp)	Nearest gene (10 Mp)[a]	Gene group	Relation to CpG island	I^2
cg17086398	1	0	*SERINC2*	Body		0.02
cg14866069	4	579	*BMPR1B*			0
cg23666362	4	516	*MIR1973*	TSS1500		0.01
cg12619262	7	6276	*CHST12*			0.01
cg20045320	11	116	*IFITM3*		S_Shore	53.53
cg07677157	12		NA[a]			0
cg07839457	16	435	*NLRC5*	TSS1500	N_Shore	0
cg09615688	16		NA[a]			29.9
cg18424841	20		NA[a]		Island	2.65

[a]Nearest gene was far more than 10 Mp.
[b]Effect estimates represent hazard ratio per 10% increase in DNA methylation.
CHR = chromosome; HR = hazard ratio; 95% CI = 95% confidence interval; p = p-value; SD = standard deviation.

Supplementary Table 8. FDR-significant fully-adjusted CpGs in fixed effect meta-analysis with exclusion of ARIC.

Probe name	CHR	Distance to nearest gene (bp)	Nearest gene (10 Mp)[a]	Gene group	Relation to CpG island	HR	95% CI	*p*	Mean DNA methylation level
cg17086398	1	0	*SERINC2*	Body		1.25	(1.11; 1.41)	3.80E-04	0.29
cg14866069	4	579	*BMPR1B*			0.65	(0.52; 0.8)	6.86E-05	0.84
cg23666362	4	516	*MIR1973*	TSS1500		0.74	(0.62; 0.89)	1.35E-03	0.8
cg12619262	7	6276	*CHST12*			1.27	(1.12; 1.45)	3.31E-04	0.74
cg20045320	11	116	*IFITM3*		S_Shore	0.88	(0.82; 0.95)	8.76E-04	0.54
cg07677157	12		NA[a]			0.76	(0.66; 0.87)	4.78E-05	0.19
cg07839457	16	435	*NLRC5*	TSS1500	N_Shore	0.86	(0.81; 0.92)	3.70E-06	0.45
cg09615688	16		NA[a]			0.53	(0.41; 0.68)	9.32E-07	0.89
cg18424841	20		NA[a]		Island	1.17	(1.08; 1.27)	1.90E-04	0.68

[a]Nearest gene was far more than 10 Mp.
[b]Effect estimates represent hazard ratio per 10% increase in DNA methylation.
CHR = chromosome; HR = hazard ratio; 95% CI = 95% confidence interval; p = p-value; SD = standard deviation.

Supplementary Table 9. FDR-significant fully-adjusted CpGs in fixed effect meta-analysis with exclusion of WHI-Study 1.

Probe name	CHR	Distance to nearest gene (bp)	Nearest gene (10 Mp)[a]	Gene group	Relation to CpG island	HR	95% CI	*p*	Mean DNA methylation level
cg17086398	1	0	*SERINC2*	Body		1.26	(1.15; 1.38)	4.30E-07	0.29
cg14866069	4	579	*BMPR1B*			0.67	(0.56; 0.79)	2.23E-06	0.85
cg23666362	4	516	*MIR1973*	TSS1500		0.68	(0.58; 0.79)	1.16E-06	0.81
cg12619262	7	6276	*CHST12*			1.23	(1.13; 1.35)	5.93E-06	0.74
cg20045320	11	116	*IFITM3*		S_Shore	0.86	(0.81; 0.91)	2.94E-07	0.54
cg07677157	12		NA[a]			0.79	(0.72; 0.87)	2.57E-06	0.17
cg07839457	16	435	*NLRC5*	TSS1500	N_Shore	0.87	(0.84; 0.92)	1.30E-08	0.45
cg09615688	16		NA[a]			0.53	(0.4; 0.69)	2.57E-06	0.9
cg18424841	20		NA[a]		Island	1.21	(1.13; 1.3)	1.53E-08	0.69

[a]Nearest gene was far more than 10 Mp.
[b]Effect estimates represent hazard ratio per 10% increase in DNA methylation.
CHR = chromosome; HR = hazard ratio; 95% CI = 95% confidence interval; p = p-value; SD = standard deviation.

Supplementary Table 10. Fixed effects meta-analysis results on incident coronary heart disease from the CHARGE Consortium.

Supplementary Table 11. Miettinen's population attributable factor for NAS, WHI-EMPC-EA, and WHI-EMPC-AA as well as weighted combination (average).

CpG	NAS	WHI-EMPC-EA	WHI-EMPC-AA	Mean	SD
cg17086398	6.47	5.55	3.57	5.20	1.21
cg14866069	-17.20	-6.77	-20.09	-14.69	5.72
cg23666362	-15.19	.	.	-15.19	0.00
cg12619262	7.33	-1.18	4.27	3.48	3.52
cg20045320	-5.34	-0.67	-1.42	-2.48	2.05
cg07677157	-7.29	-9.43	-27.86	-14.86	9.23
cg07839457	-4.08	-3.07	-3.43	-3.53	0.42
cg09615688	.	1.43	.	1.43	0.00
cg18424841	5.97	3.51	4.91	4.80	1.01

Supplementary Table 12. Standardized betas identifying the linear relationship between FDR-significant fully-adjusted CpGs and conventional risk factors.

Supplementary Table 13. Standardized betas identifying the linear relationship between FDR-significant fully-adjusted CpGs and epigenetic aging clock in NAS, after adjusting all conventional risk factors.

Epigenetic age	cg17086398	cg14866069	cg23666362	cg12619262	cg20045320	cg07677157	cg07839457	cg18424841
	Est (p)	Est (p)	Est (p)	Est (p)	Est (p)	Est (p)	Est (p)	Est (p)
Horvath epigenetic aging clock (years)	-0.09(0.1)	0.03(0.51)	0.01(0.78)	0.23(0)	-0.27(0)	-0.13(0.02)	-0.19(0)	0.11(0.03)
Hannum epigenetic aging clock (years)	-0.07(0.24)	-0.17(0)	-0.14(0.01)	0.36(0)	-0.08(0.23)	-0.06(0.34)	-0.28(0)	0.18(0)
Weidener epigenetic aging clock (years)	0.02(0.69)	-0.03(0.39)	-0.01(0.72)	0.08(0.1)	-0.07(0.11)	-0.04(0.46)	-0.04(0.41)	0.06(0.16)
PhenoAge (years)	-0.04 (0.5)	-0.11 (0.01)	-0.11 (0.02)	0.02 (0.76)	-0.27 (0)	0 (0.93)	-0.19 (0)	0.13 (0.01)
Mortality risk score	0.04 (0.44)	-0.22 (0)	-0.19 (0)	0.34 (0)	-0.19 (0)	-0.49 (0)	-0.03 (0.57)	-0.07 (0.22)

Est = estimate; p = p-value.

Supplementary Table 14. Association with all-cause mortality and DNA methylation levels at FDR-significant CpGs, adjusting for epigenetic acceleration ages in the Normative Aging Study (NAS).

Association with mortality	CpG alone	CpG + DNAmAge acceleration		CpG + PhenoAge acceleration	
		CpG	DNAmAge acceleration	CpG	PhenoAge acceleration
	HR (95% CI)	HR (95% CI)	HR (95% CI)	HR (95% CI)	HR (95% CI)
cg17086398	1.03 (0.83-1.27)	1.02 (0.82-1.27)	0.99 (0.97-1.03)	1.02 (0.82-1.26)	1.01 (0.98-1.03)
cg14866069	0.44 (0.30-0.70)	0.44 (0.30-0.69)	0.99 (0.97-1.02)	0.44 (0.28-0.70)	1.00 (0.98-1.03)
cg23666362	0.80 (0.52-1.23)	0.80 (0.52-1.23)	1.00 (0.97-1.03)	0.78 (0.50-1.20)	1.01 (0.99-1.03)
cg12619262	1.08 (0.88-1.33)	1.09 (0.89-1.34)	0.99 (0.97-1.02)	1.09 (0.88-1.33)	1.01 (0.98-1.03)
cg20045320	0.94 (0.81-1.09)	0.93 (0.81-1.08)	0.99 (0.97-1.02)	0.94 (0.82-1.09)	1.01 (0.98-1.03)
cg07677157	0.70 (0.52-0.94)	0.70 (0.52-0.94)	0.99 (0.97-1.02)	0.70 (0.52-0.94)	1.00 (0.98-1.03)
cg07839457	0.87 (0.78-0.97)	0.86 (0.77-0.96)	0.99 (0.97-1.02)	0.87 (0.78-0.97)	1.01 (0.98-1.03)
cg18424841	1.09 (0.94-1.26)	1.07 (0.92-1.25)	1.00 (0.97-1.02)	1.09 (0.94-1.26)	1.01 (0.98-1.03)

Supplementary Table 15. Association with all-cause mortality and DNA methylation levels at FDR-significant CpGs adjusting for mortality risk score in the Normative Aging Study (NAS).

Association with mortality	CpG alone	CpG + mortality risk score	
		CpG	Mortality risk score
	HR (95% CI)	HR (95% CI)	HR (95% CI)
cg17086398	1.03 (0.83-1.27)	1.12 (0.89-1.41)	1.68 (1.15-2.47)
cg14866069	0.44 (0.30-0.70)	0.45 (0.28-0.74)	1.39 (0.95-2.04)
cg23666362	0.80 (0.52-1.23)	0.87 (0.55-1.36)	1.55 (1.06-2.27)
cg12619262	1.08 (0.88-1.33)	1.01 (0.81 – 1.26)	1.60 (1.08-2.35)
cg20045320	0.94 (0.81-1.09)	0.96 (0.82-1.12)	1.57 (1.08-2.28)
cg07677157	0.70 (0.52-0.94)	0.74 (0.55-1.01)	1.46 (1.00-2.14)
cg07839457	0.87 (0.78-0.97)	0.88 (0.79-0.99)	1.51 (1.05-2.18)
cg18424841	1.09 (0.94-1.26)	1.07 (0.92-1.24)	1.59 (1.10-2.30)

Supplementary Table 16. Enrichment analysis for genes identified in GWAS of death-related factors.

Disease	Gene	Enrichment p-value	Enrichment_FDR
Alcohol dependence	*SERINC2*	0.002	0.004
HDL cholesterol	*NLRC5*	0.022	0.022

p = p-value; FDR = false discovery rate.

Supplementary Table 17. KEGG pathways for FDR-significant CpGs in the basic model.

Supplementary Table 18. Pathways analysis with DAVID.

Gene	Official gene name	Diseases	Disease class	*p*
NLRC5	NLR family CARD domain containing 5	Chronic renal failure, kidney failure, chronic coronary disease, erythrocyte count, type 2 diabetes	CARDIOVASCULAR, HEMATOLOGICAL, METABOLIC, RENAL	>0.05
BMPR1B	Bone morphogenetic protein receptor type 1B	Alcoholism, attention deficit disorder with hyperactivity, bone mineral density, cleft lip, cleft palate, hypertension, increased ovulation rate, juvenile polyposis, obesity, premature ovarian failure, polycystic ovarian syndrome, primary ovarian insufficiency, puberty (delayed), puberty (precocious), thrombophilia, tobacco use disorder	CARDIOVASCULAR, CHEMDEPENDENCY, DEVELOPMENTAL, METABOLIC, OTHER, PSYCH, REPRODUCTION	>0.05
CHST12	Carbohydrate sulfotransferase 12	Malaria, placenta diseases, pregnancy complications, parasitic	INFECTION	>0.05
IFITM3	Interferon induced transmembrane protein 3	Ulcerative colitis	IMMUNE	>0.05

Supplementary Table 19. Association between fully-adjusted FDR-significant CpGs and SNPs (meQTL analysis) in KORA.

Fully-adjusted FDR-significant CpGs			SNP			Est.	SE	*p*
Name	CHR	Position	Name	CHR	Position			
cg09615688	16	80982506	rs8052401	16	80983487	-0.005	0.001	3.81E-08
cg18424841	20	61315444	rs2427380	20	61314740	0.020	0.002	2.27E-16
cg18424841	20	61315444	rs2427381	20	61314785	0.014	0.003	2.50E-08
cg18424841	20	61315444	rs118042746	20	61314972	-0.039	0.007	2.47E-09
cg18424841	20	61315444	rs6010861	20	61315002	0.024	0.003	9.67E-17
cg18424841	20	61315444	rs2427382	20	61315199	0.015	0.002	6.91E-10
cg18424841	20	61315444	rs6062825	20	61315436	0.016	0.003	2.37E-06
cg18424841	20	61315444	rs4809278	20	61315545	0.020	0.002	6.81E-16
cg18424841	20	61315444	rs6122386	20	61316386	0.015	0.002	6.20E-10

CHR = chromosome; Est = estimate; SE = standard error; p = p-value.

Supplementary Table 20. Association between fully-adjusted FDR-significant CpGs and gene expression (eQTM analysis) in KORA.

Probe name	CHR	Distance to nearest gene (bp)	Nearest gene	Influenced gene name	Est	SE	*p*	FDR
cg17086398	1	0	*SERINC2*	*MARCKSL1*	-0.75	0.20	1.77E-04	5.92E-03
cg20045320	11	116	*IFITM3*	*IFITM3*	-3.45	0.43	3.19E-15	7.48E-13
cg20045320	11	116	*IFITM3*	*IRF7*	-0.76	0.17	8.78E-06	4.11E-04
cg07839457	16	435	*NLRC5*	*MT2A*	-1.48	0.24	7.92E-10	9.27E-08
cg07839457	16	435	*NLRC5*	*MT1E*	-0.66	0.19	5.17E-04	1.34E-02
cg07839457	16	435	*NLRC5*	*MT1A*	-1.11	0.20	1.98E-08	1.54E-06
cg07839457	16	435	*NLRC5*	*MT1G*	-0.25	0.07	6.70E-04	1.57E-02
cg07839457	16	435	*NLRC5*	*MT1IP*	-0.39	0.10	1.64E-04	5.92E-03
cg07839457	16	435	*NLRC5*	*NLRC5*	-0.70	0.15	3.67E-06	2.15E-04

CHR = chromosome; Est = estimate; SE = standard error; p = p-value; FDR = false discovery rate.

Supplementary Table 21. Causal association between coronary heart disease, kidney function (serum creatinine), and methylation at FDR-significant CpGs in KORA and ARIES.

Disease	Methylation locus	OR	95% LCI	95% UCI	P	methQTL cohort	N SNPs	MR method
	cg09615688	1.508	1.0199	2.2297	0.0395	KORA	1	Wald ratio
	cg18424841	1.0058	0.9994	1.0122	0.0743	ARIES	1	Wald ratio
	cg18424841	0.8506	0.7292	0.9922	0.0944	KORA	7	MR Egger
	cg18424841	1.01	0.9218	1.1068	0.8375	KORA	7	Weighted mode
	cg18424841	1.0007	0.9399	1.0656	0.9814	KORA	7	Weighted median
	cg09615688	0.8816	0.7449	1.0435	0.1429	KORA	1	Wald ratio
	cg18424841	1.0014	0.9989	1.0039	0.2771	ARIES	1	Wald ratio
	cg18424841	0.9279	0.8297	1.0377	0.4148	KORA	3	MR Egger
	cg18424841	0.9903	0.957	1.0246	0.6305	KORA	3	Weighted mode
	cg18424841	0.9939	0.9637	1.025	0.6969	KORA	3	Weighted median

Odds ratio (OR), lower 95% confidence interval (LCI), and upper 95% confidence interval (UCI) given per 10% higher methylation. Associations for coronary heart disease taken from (PMC4589895) and association for serum creatinine taken from (PMC4735748). Associations for ARIES methQTLs extracted from MR-base (PMC5976434) using middle age estimates for methQTLs from ARIES cohort (PMC4818469). MR = mendelian randomization; N SNPs = number of SNPs (instruments) used for the MR analyses; P = p-value.

Supplementary Table 22. Association of neutrophil–lymphocyte ratio (NLR) with all-cause mortality, with and without adjustment for cell type proportion in Normative Aging Study.

NLR association with mortality	Without adjusting for cell proportions		Adjusting for cell proportions	
	HR (95% CI)	*p*	HR (95% CI)	*p*
without any CpG inclusion	1.08 (1.00 – 1.17)	0.04	1.06 (0.92 – 1.21)	0.43
cg17086398	1.08 (1.00 – 1.17)	0.045	1.06 (0.93 – 1.21)	0.41
cg14866069	1.13 (1.05 – 1.22)	0.002	1.03 (0.90 – 1.18)	0.68
cg23666362	1.10 (1.02 – 1.19)	0.017	1.04 (0.90 – 1.20)	0.61
cg12619262	1.08 (1.00 – 1.17)	0.042	1.05 (0.92 – 1.21)	0.45
cg20045320	1.08 (1.00 – 1.17)	0.042	1.05 (0.92 – 1.20)	0.46
cg07677157	1.09 (1.01 – 1.18)	0.034	1.06 (0.93 – 1.21)	0.39
cg07839457	1.07 (0.99 – 1.15)	0.06	1.00 (0.88 – 1.15)	0.97
cg18424841	1.10 (1.02 – 1.19)	0.02	1.06 (0.93 – 1.21)	0.38