An integrative analysis of clinical and epigenetic biomarkers of mortality

Tianxiao Huan\textsuperscript{1,2,3+} (tianxiao.huan@umassmed.edu),
Steve Nguyen\textsuperscript{4†} (nguy2295@umn.edu),
Elena Colicino\textsuperscript{5} (elena.colicino@mssm.edu),
Carolina Ochoa-Rosales\textsuperscript{6} (c.ochoarosales@erasmusmc.nl),
W. David Hill\textsuperscript{7} (David.Hill@ed.ac.uk),
Jennifer A. Brody\textsuperscript{8,9} (jeco@uw.edu),
Mette Soerensen\textsuperscript{10,11,12} (msoerensen@health.sdu.dk),
Yan Zhang\textsuperscript{13} (y.zhang@Dkfz-Heidelberg.de),
Antoine Baldassari\textsuperscript{14} (baldassa@email.unc.edu),
Mohamed Ahmed Elhadad\textsuperscript{15,16,17} (mohamed.elhadad@helmholtz-muenchen.de),
Tanaka Toshiko\textsuperscript{18} (tanakato@mail.nih.gov),
Yinan Zheng\textsuperscript{19} (y-zheng@northwestern.edu),
Arce Domingo-Relloso\textsuperscript{20-22} (ad3531@cumc.columbia.edu),
Dong Heon Lee\textsuperscript{1,2} (dephs1042@gmail.com),
Jiantao Ma\textsuperscript{1,2,23} (jiantao.ma@tufts.edu),
Chenyu Liu\textsuperscript{24} (liuc@bu.edu),
Shih-Jen Hwang\textsuperscript{1,2} (hwangs2@nhlbi.nih.gov),
Roby Joehanes\textsuperscript{1,2} (roby.joehanes@nih.gov),
Myriam Fornage\textsuperscript{25} (Myriam.Fornage@uth.tmc.edu),
Jan Bressler\textsuperscript{26} (jan.bressler@uth.tmc.edu),
Joyce BJ van Meurs\textsuperscript{26} (j.vanmeurs@erasmusmc.nl),
Birgit Debrabant\textsuperscript{10} (bdebrabant@health.sdu.dk),
Jonas Mengel-From\textsuperscript{10,12} (jmengel-from@health.sdu.dk),
Jacob Hjelmborg\textsuperscript{10} (JHjelmborg@health.sdu.dk),
Kaare Christensen\textsuperscript{10,12} (KChristensen@health.sdu.dk),
Pantel Vokonas\textsuperscript{27-29} (pantel.vokonas@va.gov),
Joel Schwartz\textsuperscript{30} (joel@hsph.harvard.edu),

\textbf{NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.}
Sina A. Gahrib8,9 (sagharib@u.washington.edu),
Nona Sotoodehnia8,9 (nsotoo@uw.edu),
Colleen M. Sitlani8,9 (csitlani@uw.edu),
Sonja Kunze15,16 (sonja.kunze@helmholtz-muenchen.de),
Christian Gieger15,16,17 (christian.gieger@helmholtz-muenchen.de),
Annette Peters16,17,31,32 (peters@helmholtz-muenchen.de),
Melanie Waldenberger15,16,17 (waldenberger@helmholtz-muenchen.de),
Ian J. Deary7 (iand@exseed.ed.ac.uk),
Luigi Ferrucci18 (FerrucciLu@grc.nia.nih.gov),
Philip Greenland19 (p-greenland@northwestern.edu),
Donald M Lloyd-Jones19 (dlj@northwestern.edu),
Lifang Hou19 (l-hou@northwestern.edu),
Stefania Bandinelli33 (stefania1.bandinelli@uslcentro.toscana.it),
Trudy Voortman6 (trudy.voortman@erasmusmc.nl),
Brenner Hermann13,34 (h.brenner@Dkfz-Heidelberg.de),
Andrea Baccarelli35 (ab4303@cumc.columbia.edu),
Eric Whitsel14,36 (eric_whitsel@med.unc.edu),
James S. Pankow4* (panko001@umn.edu),
Daniel Levy1,2* (levyd@nhlbi.nih.gov)

1 The Framingham Heart Study, Framingham, MA, USA
2 The Population Sciences Branch, Division of Intramural Research, National Heart, Lung, and Blood Institute, National Institutes of Health, Bethesda, MD, USA
3 Department of Ophthalmology and Visual Sciences, University of Massachusetts Medical School, MA, USA
4 Division of Epidemiology & Community Health, School of Public Health, University of Minnesota, Minneapolis, MN, USA
5 Icahn School of Medicine at Mount Sinai, New York, NY, USA
6 Department of Epidemiology, Erasmus University Medical Center, Rotterdam, the Netherlands
7 Dept of Psychology, Univ of Edinburgh, Edinburgh, UK
8 Division of Pulmonary, Critical Care and Sleep Medicine, Center for Lung Biology, University of Washington, Seattle, WA, USA
9 Cardiovascular Health Research Unit, Seattle, WA, USA
10 Epidemiology, Biostatistics and Biodemography, Department of Public Health, University of Southern Denmark, J.B. Winsløws Vej 9B, 5000, Odense C, Denmark
11 Department of Clinical Biochemistry and Pharmacology, Center for Individualized Medicine in Arterial Diseases, Odense University Hospital, J.B. Winsløws Vej 4, 5000, Odense C, Denmark
Running title: Epigenetic prediction of mortality
Checklist:

1. total character count: **47,195**
2. word count of the Summary: **247**
3. the number of papers cited in the References: **52**
4. a listing of all Tables (Table 1, Table 2, etc.):

| Table 1: Clinical characteristics the 15,013 study participants. |
| Table 2: Trans-ethnic replicated all-cause mortality related CpGs. |
| Table 3: Performance robustness comparison of mortality predictors in FHS and ARIC cohorts. |

5. a listing of all Figures (Fig. 1, Fig. 2, etc.) including, (a) whether the Figure should be in colour, greyscale or black and white, (b) whether the Figure should appear in 1-column or 2-column format, (c) the size of the Figure at full scale (mm x mm), (d) the smallest font size used in the Figure at full scale.

<table>
<thead>
<tr>
<th>Figure No</th>
<th>Color</th>
<th>Greyscale</th>
<th>Black and white</th>
<th>Single</th>
<th>Double</th>
<th>Size</th>
<th>Smallest font size used in the figure at full scale</th>
</tr>
</thead>
<tbody>
<tr>
<td>Figure 1</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>80 mm</td>
<td>12</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Figure 2</td>
<td>Yes</td>
<td>Yes</td>
<td>80 mm</td>
<td>12</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Figure 3</td>
<td>Yes</td>
<td>Yes</td>
<td>80 mm</td>
<td>12</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Abstract

DNA methylation (DNAm) has been reported to be associated with many diseases and mortality. We hypothesized that the integration of DNAm with clinical risk factors would improve mortality prediction. We performed an epigenome-wide association study of whole blood DNAm in relation to mortality in 15 cohorts (n=15,013). During a mean follow-up of 10 years, there were 4314 deaths from all-causes including 1235 cardiovascular disease (CVD) deaths and 868 cancer deaths. Ancestry-stratified meta-analysis of all-cause mortality identified 163 CpGs in European ancestry (EA) and 17 in African ancestry (AA) participants at $P<1\times10^{-7}$, of which 41 (EA) and 16 (AA) were also associated with CVD death, and 15 (EA) and 9 (AA) with cancer death. We built DNAm-based prediction models for all-cause mortality that predicted mortality risk independent of clinical risk factors. The mortality prediction model trained by integrating DNAm with clinical risk factors showed a substantial improvement in prediction of cancer death with 11% and 5% increase in the C-index in internal and external replications, compared with the model trained by clinical risk factors alone. Mendelian randomization identified 15 CpGs in relation to longevity, CVD, or cancer risk. For example, cg06885782 (in KCNQ4) was positively associated with risk for prostate cancer (Beta=1.2, $P_{MR}=4.1\times10^{-4}$), and negatively associated with longevity (Beta=-1.9, $P_{MR}=0.02$). Pathway analysis revealed that genes associated with mortality-related CpGs are enriched for immune and cancer related pathways. We identified replicable DNAm signatures of mortality and demonstrated the potential utility of CpGs as informative biomarkers for prediction of mortality risk.

Key words: DNA methylation; machine learning; mortality; cardiovascular disease; cancer
Introduction

Despite substantial evidence of heritability of human longevity ($h^2 = 10$-$30\%$), genome-wide association studies (GWAS) have reported few loci associated with human longevity (Deelen et al., 2019; Pilling et al., 2017; Timmers et al., 2019; van den Berg, Beekman, Smith, Janssens, & Slagboom, 2017). DNA methylation (DNAm), the covalent binding of a methyl group to the 5’ carbon of cytosine-phosphate-guanine (CpG) dinucleotide sequences, reflects a wide range of environmental exposures and genetic influences at the molecular level and altered DNAm has been shown to regulate gene expression (Jones & Takai, 2001). Recent studies have reported DNAm patterns associated with age in humans (Hannum et al., 2013; Horvath, 2013; Levine et al., 2018; Lu et al., 2019). Estimates of biological age based on DNAm, referred to as "epigenetic age" or "DNAm age" have been validated in numerous studies, although the functions of these age-associated CpGs are largely unknown (Horvath et al., 2015; Lu et al., 2019; Marioni, Shah, McRae, Chen, et al., 2015; Marioni, Shah, McRae, Ritchie, et al., 2015). DNAm age also has been shown to be predictive of many age-related diseases and of all-cause mortality (Chen et al., 2016; Dugué et al., 2018; Levine et al., 2018; Lu et al., 2019; Marioni, Shah, McRae, Chen, et al., 2015).

Despite the association of DNAm age with a variety of age-associated outcomes, age-related CpGs are different from those that are most strongly associated with mortality. Relatively few DNAm studies have focused on mortality as the primary outcome (Colicino et al., 2020; Svane et al., 2018; Zhang et al., 2017). Moreover, due to sample size limitations, most DNAm mortality studies have not typically investigated cause-specific mortality such as death due to cardiovascular disease (CVD) and cancer. Additionally, little is known about the prediction performance of DNAm-based mortality models and whether or not such approaches improve mortality prediction above and beyond established clinical risk factors.

We hypothesized that inter-individual variation in DNAm is associated with all-cause mortality risk and with cause-specific mortality, and that we could build models incorporating CpGs that would improve mortality prediction beyond established clinical risk factors. In this study, we report the
results of a meta-analysis of epigenome-wide association studies (EWAS) of all-cause mortality and cause-specific mortality including death from CVD and cancer in up to 15,013 individuals from 15 prospective cohort studies in which DNAm was measured in whole blood. We built all-cause mortality risk prediction models using penalized regression and machine learning methods and integrated DNAm and established mortality clinical risk factors and validated the models’ performance. Additionally, using Mendelian randomization, we identified putatively causal CpGs for mortality. Last, we investigated the downstream gene expression and pathway changes of the mortality-related CpGs by testing associations between DNAm and gene expression. **Fig. S1** summarizes the multi-step study design.

### Results

### Study population

**Table 1** presents the major clinical characteristics of the 15,013 study participants including 11,684 European ancestry (EA, mean age 65, 55% women) and 3329 African ancestry (AA, mean age 59, 70% women) participants from 15 cohorts (**Table S1** summarizes additional clinical characteristics). Most studies had fewer than 15 years of mean follow-up (mean values ranged from 6.4 to 13.7 years), except ARIC (mean follow-up of 20.0 years in ARIC EA and 18.6 in ARIC AA participants, respectively). During follow-up of EA participants, 2907 died of any cause, 688 of CVD, and 546 of cancer; among AA participants, 1407 died of any cause, 547 of CVD, and 322 of cancer.

**Ancestry-stratified epigenome-wide meta-analysis of all-cause mortality**

At Bonferroni-corrected $P<1\times10^{-7}$ (~0.05/400,000), we identified 163 CpGs whose differential methylation in whole blood was associated with all-cause mortality in EA participants, and 17 CpGs in AA participants, after adjustment of age, sex, lifestyle factors, clinical risk factors, white blood cell types, and technical covariates. **Tables S2-S3** present the results for all CpGs at $P<1\times10^{-5}$. Overall genomic inflation in meta-analysis ($\lambda$) was estimated at 1.15 or less, indicating low inflation and low
risk of false-positive findings. Even though cohort-specific analysis showed slightly higher genomic
inflation in some cohorts (\( \lambda > 1.5 \) in two cohorts, Table S4), forest plots show that the results were not
driven by results from one or several cohorts (Fig. S2). Sensitivity analysis results including meta-
analysis after correcting for \( \lambda \) in each cohort, meta-analysis after excluding results from two cohorts
with \( \lambda > 1.5 \) and meta-analysis after excluding RS cohort are included in Table S5-S6. Results of the
sensitivity analysis remained consistent with the main results in terms of direction and effect estimates
with Pearson’s correlation \( r = 0.99 \) (in EA, corrected for \( \lambda \) in each cohorts), \( r = 1.00 \) (in EA, after
removing two cohorts with \( \lambda > 1.5 \)), \( r = 1.00 \) (in EA, after removing RS) and \( r = 1.00 \) (in AA, corrected
for \( \lambda \) in each cohorts).

Among the 177 all-cause mortality-related CpGs (union set of EA and AA results at \( P < 1 \times 10^{-7} \)), the
vast majority of significant CpGs (151, 85%) were inversely associated with mortality, with hazards
ratios (HRs) < 1 (range 0.72 to 0.89 per standard deviation [SD]). Methylation at the remaining 26
(15%) CpGs was positively associated with mortality, with HRs > 1 (range 1.13 to 1.32). The 177
CpGs are annotated to 121 genes and 43 intergenic regions.

**Transethnic replication and sensitivity analysis**

Of the 163 all-cause mortality related CpGs in EA participants, 18 (11%) had \( P < 0.0003 \) (0.05/163) in
AA participants; of the 17 CpGs in AA participants, 12 (71%) had \( P < 0.004 \) (0.05/17) in EA
participants. Table 2 displays the transethnic replicated CpGs including 27 unique CpGs. The top 3
transethnic replicated CpGs in EA participants remained the top 3 in AA participants, including
cg16743273 for MOBKL2A, cg18181703 for SOCS3, and cg21393163 at an intergenic region (Chr.1:
12217629).

Because ARIC had longer follow-up than the other cohorts, in sensitivity analysis, we truncated ARIC
follow up at 15 years. The HRs for the significant CpGs (at \( P < 1 \times 10^{-5} \)) remained consistent with the
main results in terms of direction and effect estimates with Pearson’s correlation \( r = 1.00 \) and \( r = 0.99 \)
in EA and AA participants, respectively (Table S2-S3 and Fig. S3).

**Associations of DNAm with CVD death and cancer death**
In comparison with results for all-cause mortality, fewer CpGs were associated with CVD death (at \(P<1\times10^{-7}\), n=4 in EA, and n=15 in AA) and cancer death (n=0 in EA, and n=1 in AA) Tables S7-S8 report the corresponding results at \(P<1\times10^{-5}\). Among the 163 all-cause mortality-related CpGs identified in EA participants at \(P<1\times10^{-7}\), 41 CpGs were associated with CVD death, 16 with cancer death, and 5 with both (at \(P<0.05/163\), Table S2). Among the 17 CpGs identified in AA participants at \(P<1\times10^{-7}\), 15 were associated with CVD death, 9 with cancer death, and 8 with both (at \(P<0.05/17\), Table S3). Fig. 1 shows the effect sizes and direction of effect for the top CpGs associated with all-cause mortality, and their consistency with the results of analyses of CVD death and cancer death. We found that if a CpG was positively correlated with all-cause mortality, then it also was positively correlated with CVD death and cancer death, and vice versa.

**Mortality prediction model**

To investigate if DNAm can be used to predict mortality risk, we constructed prediction models for all-cause mortality, and evaluated their prediction of all-cause mortality, CVD death, and cancer death. To ensure unbiased validation, we split the EA cohorts into separate discovery and replication sets (Fig. S1 shows the analysis flowchart). The discovery cohorts consisted of 8288 participants (including 2173 deaths from all-causes) from 10 cohorts, excluding FHS (n=2427) and ARIC (n=969), which were used as replication cohorts. The meta-analysis of the discovery set identified 74 CpGs at \(P<1\times10^{-7}\), 158 CpGs at \(P<1\times10^{-6}\), 357 CpGs at \(P<1\times10^{-5}\), 931 CpGs at \(P<1\times10^{-4}\), 2717 CpGs at \(P<1\times10^{-3}\), and 28,323 CpGs at \(P<0.05\). We evaluated three types of input features: a) clinical risk factors only (i.e., clinical risk factor models); b) CpGs identified in the meta-analysis of the discovery set (i.e., CpG models); and c) the input features including both CpGs and clinical risk factors (i.e., integrative models). We also compared four prediction methods including Elastic net - Cox proportional hazards (Elastic-coxph) (Friedman, Hastie, & Tibshirani, 2010), Random survival forest (RSF) (Ishwaran, Kogalur, Blackstone, & Lauer, 2008), Cox-nnet (Ching, Zhu, & Garmire, 2018), and DeepSurv (Katzman et al., 2018) (see Methods for details). In general, the four prediction methods did not show major differences in predicting mortality outcomes as assessed by multiple
evaluation metrics (Table S9 lists the evaluation metrics across all four methods). To simplify the
presentation of results, we focused on the Elastic-coxph method.

**Clinical risk factors strongly predict all-cause mortality and CVD death:** The C-index of the clinical
risk factor models (age, sex, and 12 clinical risk factors) was 0.80 for all-cause mortality, 0.81 for
CVD death, and 0.77 for cancer death in FHS (reflecting the average values of 10-fold cross-
validation). Among the 12 clinical risk factors, prevalent cancer status was the major contributor to
predicting cancer death. After excluding individuals with prevalent cancer at the time of blood draw
for DNAm measurements (i.e., the start of follow up), the C-index of the clinical risk factor model
was 0.57 for cancer death. Finally, two clinical risk models were built using the optimum parameters
selecting by cross-validation (see Methods). The first one was trained using all FHS participants and
included 10 risk factors selected by the Elastic-coxph method (to predict all-cause mortality and CVD
death, Table S10), and the second was trained using FHS participants excluding those with prevalent
cancer cases and including 10 risk factors (to predict cancer death, Table S11). The corresponding C-
index of the clinical risk factor model was 0.75 for all-cause mortality (HR=2.64 per SD in the risk
score, 95% CI [2.21, 3.15], \( P=4.4\times10^{-27} \)), 0.81 for CVD death (HR=3.51, 95% CI [2.58, 4.79], \( P=2.1 \times10^{-15} \)), and 0.71 for cancer death (excluding prevalent cancer samples, HR=2.35, 95% CI [1.74, 3.18],
\( P=2.3 \times10^{-8} \)) in ARIC EA participants with follow up truncated at 15 years (Table 3).

**DNAm predicts mortality independently of age and clinical risk factors:** The models using all-cause
mortality-related CpGs identified in the discovery cohorts as the sole input feature (the CpG model)
were predictive of all-cause mortality, CVD death, and cancer death in the replication set. As shown
in Fig. S4, when more discovery CpGs were added to the model, the prediction performance metrics
did not always improve. In FHS, the models with discovery CpGs at \( P<1\times10^{-3} \) showed the best
predictive performance for all-cause mortality (C-index =0.77) and CVD death (C-index =0.82), but
the model with discovery CpGs at \( P<1\times10^{-5} \) showed the best predictive performance for cancer death
(excluding prevalent cancer cases, [C-index =0.65]). The final CpG models that were trained using all
FHS participants are provided in Table S12 including 76 CpGs to predict all-cause mortality and
CVD death, and in Table S13 including 56 CpGs to predict cancer death (excluding prevalent cancer
cases). The C-index of the CpG models with the best predictive performance in ARIC were 0.72 for all-cause mortality (HR=2.21, 95% CI [1.86, 2.62], \( P=2.0 \times 10^{-20} \)), 0.77 for CVD death (HR=2.62, 95% CI [1.96, 3.51], \( P=9.9 \times 10^{-11} \)) and 0.73 for cancer death (HR=2.22, 95% CI [1.67, 2.95], \( P=3.2 \times 10^{-8} \), Table 3). The association of the mortality risk scores calculated by the CpG models with mortality outcomes remained significant after adjusting for age, sex, and clinical risk factors; for all-cause mortality (HR=1.68, 95% CI [1.37, 2.07], \( P=9.8 \times 10^{-7} \)), CVD death (HR=1.81, 95% CI [1.24, 2.64], \( P=0.002 \)), and cancer death (HR=2.04, 95% CI [1.46, 2.86], \( P=3.0 \times 10^{-5} \)).

The integrative model (trained by CpGs and clinical risk factors) moderately improved upon the clinical risk factor model for all-cause mortality and CVD death, and greatly improved the prediction of cancer death: As shown in Table 3, the integrative models demonstrated robustness for predicting mortality outcomes, with a good C-index, HR, and low brier error rate. The final integrative models trained using data from all FHS participants are provided in Table S14 including nine clinical risk factors and 36 CpGs to predict all-cause mortality and CVD death, and in Table S15 including seven clinical risk factors and 42 CpGs to predict cancer death (excluding prevalent cancer cases). The C-index values of the integrative models were 0.80 (FHS, reflecting the average values of 10-fold cross-validation) and 0.77 (ARIC) for all-cause mortality; 0.83 (FHS) and 0.80 (ARIC) for CVD death; and 0.69 (FHS) and 0.76 (ARIC) for cancer death. Kaplan-Meier survival curves for the mortality risk scores (split into high, middle, and low risk groups) in the ARIC EA cohort (computed by the integrative models using clinical risk factors and CpGs at discovery \( P<1 \times 10^{-6} \), Table S14-S15) illustrates the higher death rate for those with a higher mortality risk score (log-rank \( P<1 \times 10^{-6} \), Fig. 3). In comparison to the clinical risk factor models, the integrative models moderately improved prediction of all-cause mortality (0.7% increase in C-index with addition of CpGs in FHS and 2% increase in ARIC), and of CVD death (2% increase in C-index in FHS, but no increase in ARIC). We speculate that the reason for this minor increase is because the mortality-related CpGs capture the contributions of clinical risk factors for CVD death. For cancer death, the C-index of the integrative model revealed an 11% increase in FHS above and beyond the clinical risk factor model and a corresponding 5% increase in ARIC.
We also tested the mortality prediction models’ performance using the entire ARIC EA data (without truncation, Table S16). Due to the long follow-up time in this older cohort (mean age 59.8 at baseline, with 20 ± 5.5 years follow-up), the integrative model exhibits very similar performance features as the model using age and sex as the sole input features for predicting all-cause mortality and CVD death. The integrative model improved prediction of cancer death with 2% increase in the C-index versus the clinical risk factor model.

We further tested all-cause mortality prediction models in the CARDIA study (baseline age 45 ± 3 years). The CARDIA study has 12 years of follow-up, during which there were 27 deaths from all causes in 905 participants with DNA methylation. As shown in Table S17, the clinical risk factor model, the CpG model, and the integrative model each predicted all-cause mortality, and each outperformed the DNAm age models.

Comparing the mortality prediction model with DNAm age

We compared four DNAm age models (i.e., PhenoAge (Levine et al., 2018), Horvath Age (Horvath, 2013), Hannum Age (Hannum et al., 2013), and GrimAge (Lu et al., 2019)) with our mortality prediction models (CpG only models and integrative CpG plus 12 risk factor models) for all-cause mortality, CVD death, and cancer death in ARIC participants. The associations of mortality risk scores calculated by mortality prediction models with mortality outcomes were statistically significant, and the associations remained significant after adjusting for age and sex, and after additionally adjusting for the clinical risk factors. The four DNAm age models were significantly associated with mortality outcomes. After adjusting for age, sex and clinical risk factors, however, only GrimAge remained associated with all-cause mortality, CVD death, and cancer death. None of the other three DNAm age predictors was associated with mortality outcomes after additionally adjusting for clinical risk factors (Fig. 3). The mortality prediction models (both the CpG only model and the integrative model that included the clinical risk factors and CpGs) outperformed the GrimAge model in prediction of mortality outcomes in terms of HRs and P values. The associations of mortality risk scores with mortality outcomes remain significant after adjusting for the four DNAm age (Table S18).
Associations of DNA methylation (DNAm) with genetic variants and Mendelian randomization analysis

Among the 177 all-cause mortality-related CpGs (union of EA and AA results at $P<1 \times 10^{-7}$), 123 CpGs had significant associations with genetic variants (i.e. cis- or trans-meQTL variants). meQTL variants for 80 CpGs could be linked to 618 GWAS Catalog (Buniello et al., 2019) index SNPs associated with 432 complex traits or diseases (Table S18).

We further performed multiple instrumental variable (IV) MR analysis for the 17 CpGs having ≥3 independent cis-meQTL SNPs (pruned by LD $r^2<0.01$, as IVs, to model the causal relations of differential methylation at these CpGs (as the exposure) on the various outcomes, including longevity (Deelen et al., 2019), CVD, CVD risk factors, and cancer (Evangelou et al., 2018; Locke et al., 2015; Mikhailidou et al., 2017; Phelan et al., 2017; Schumacher et al., 2018; Scott et al., 2017; Wang et al., 2014; Willer et al., 2013). At $P_{MR}<0.05$, MR supported causal effects of 15 CpGs on one or more outcome (Table S19), and 4 CpGs were statistically significant at $P_{MR}<0.05/17$, including cg06885782 (within 1500 bases upstream of transcription start site [TSS1500] of KCNQ4) and cg04907244 (TSS1500 of SNORD93) in relation to prostate cancer (Schumacher et al., 2018) (Beta=1.2 and 2.1; and $P_{MR}=4.1 \times 10^{-4}$ and 0.003, respectively), cg07094298 (in the gene body of TNIP2) in relation to lung cancer (Wang et al., 2014) (Beta =2.2, and $P_{MR}=0.003$), and cg18241337 (in the gene body of SSR3) in relation to total cholesterol (Willer et al., 2013) (Beta=0.5, and $P_{MR}=0.003$). cg06885782 (KCNQ4) also was associated with longevity (Deelen et al., 2019) (Beta=-1.9, $P_{MR}=0.02$).

Associations of DNA methylation with gene expression, and pathway analysis

For the 177 all-cause mortality-related CpGs at $P<1 \times 10^{-7}$, we assessed associations of CpGs with nearby gene expression (i.e. cis gene expression; within +/- 1 Mb) and identified 15 cis-DNA methylation-mRNA associated pairs (13 CpGs and 15 mRNAs) at $P<3 \times 10^{-10}$. The genes located at these CpGs or cis-eQTM mRNAs were not enriched for any biological processes or pathways. For the 719 all-cause mortality-related CpGs at $P<1 \times 10^{-5}$, 495 genes located at CpG sites were enriched for positive regulation of transcription from RNA polymerase II promoter (Gene Ontology [GO]) (Ashburner et al.,
2000), fold change = 1.9, FDR=0.05), and pathways for cancer (Kyoto Encyclopedia of Genes and Genomes [KEGG] pathway (Kanehisa & Goto, 2000), fold change =2.2, FDR=0.13, Table S20).

There were 79 cis-DNAm-mRNA pairs (63 CpGs and 67 mRNAs, Table S21). The 67 cis-eQTM mRNAs were enriched for multiple immune functions including immune response (GO, fold change = 6.3, FDR=0.01).

Discussion

By performing EWAS using whole blood derived DNA from 15,013 individuals from 15 cohorts with the accrual of 4314 deaths during a mean follow up of more than 10 years, we identified robust DNAm signatures of all-cause and cause-specific mortality. We developed replicable mortality predictors by integrating mortality-related CpGs with traditional clinical risk factors. The integrative models that included clinical risk factors and CpGs showed modest improvement in prediction of all-cause mortality and CVD death, and a substantial improvement in prediction of cancer death compared to the traditional risk factor model in the FHS (internal cross-replication) and ARIC (external independent replication) cohorts.

Our study is one of the largest EWAS of mortality to date (Colicino et al., 2020; Svane et al., 2018; Zhang et al., 2017) and it revealed many replicable DNAm signatures for all-cause mortality. Our results are consistent with those from previous EWAS of all-cause mortality; the vast majority of CpGs (85% in our study, 84% in (Zhang et al., 2017), and 67 % in (Colicino et al., 2020)) were inversely associated with mortality suggesting a greater mortality risk with lower CpG methylation. Our study identified more CpGs in EA cohorts (n=163) than in AA cohorts (n=17). As shown in Table 2, the effect sizes (i.e., HR) of mortality-related CpGs in EA and AA participants were quite similar. We speculate that our study identified many more CpGs in EA participants than AA participants due the greater statistical power of the larger EA sample size. Using different DNAm data normalization methods (such as Noob (Triche Jr, Weisenberger, Van Den Berg, Laird, & Siegmund, 2013), SWAN (Maksimovic, Gordon, & Oshlack, 2012), BMIQ(Teschendorff et al., 2013), and
Dasen (Pidsley et al., 2013), see Additional File 1) in different cohorts may also affect the reproducibility of the results. Among the top mortality-associated CpGs, many were associated with common traits associated with diseases in prior EWAS including BMI (e.g., cg03725309 in SARS) (Mendelson et al., 2017), smoking (e.g., cg05575921 in AHRR) (Joehanes et al., 2016), blood pressure (e.g., cg03068497 and cg21429551 in GARS) (Richard et al., 2017), alcohol consumption (e.g., cg02583484 in HNRNPA1) (Liu et al., 2018), and diet (e.g., cg18181703 in SOCS3) (Ma et al., 2020).

Among the 177 all-cause mortality-related CpGs (union of EA and AA results at $P<1 \times 10^{-7}$), 123 CpGs had significant associations with genetic variants (i.e., cis- or trans-meQTL variants identified previously (Huan et al., 2019)). For the remaining 44 CpGs, however, this does not mean that their methylation levels have nothing to do with genetic variation. It is possible that the previous meQTL study lacked sufficient statistical power to identify meQTLs for those CpGs. The mortality-related CpGs are linked to hundreds of human complex diseases/traits via their cis-meQTL SNPs, which coincide with 618 GWAS Catalog (Buniello et al., 2019) index SNPs. This leads us to hypothesize that many disease/phenotype associated SNPs may contribute to disease processes via effects on mortality-related CpGs. In this way, the mortality-related CpGs may contribute causally to disease. To test this hypothesis, we conducted MR analyses that confirmed several putatively causal associations of mortality-related CpGs with longevity (Deelen et al., 2019), CVD (Nikpay et al., 2015), CVD risk factors, and several types of cancer (Evangelou et al., 2018; Locke et al., 2015; Michailidou et al., 2017; Phelan et al., 2017; Schumacher et al., 2018; Scott et al., 2017; Wang et al., 2014; Willer et al., 2013) (Table S19). Among the four CpGs passing a Bonferroni-corrected threshold in MR analyses, cg06885782 in KCNQ4 was reported to be associated with risk for prostate cancer (beta=1.2, $P_{MR}=4.1 \times 10^{-4}$), and negatively associated with longevity (beta=-1.9, $P_{MR}=0.02$). KCNQ4 (potassium voltage-gated channel subfamily Q member 4) was previously reported to be associated with age-related hearing impairment (Van Eyken et al., 2006), and it contains genetic variants associated with all-cause mortality and survival free of major diseases (Walter et al., 2011). cg07094298 in the gene body of TNIP2 was previously identified as causal for lung cancer. A recent study reported TNIP2-ALK fusion in lung adenocarcinoma (Feng et al., 2019). cg04907244 (in TSS1500 of SNORD93) was
identified as causal for prostate cancer by MR. SNORD93 and its methylation was reported to be associated with several cancer types including uveal melanoma (Gong et al., 2017), breast cancer (Patterson et al., 2017), and renal clear cell carcinoma (Zhao et al., 2020). Pathway analysis further supported a role of mortality-related CpGs in relation to cancer risk. The intragenic CpGs were enriched for genes in cancer pathways, possibly as a consequence of the expression of nearby genes (cis-eQTMs analysis, Table S21) related to immune function.

The 14 clinical risk factors for mortality were chosen based on prior knowledge. In contrast, there are far fewer established risk factors for cancer death other than age, sex, BMI, smoking, and alcohol consumption. It is not a surprise that the clinical risk factors themselves accurately predicted all-cause mortality (C-index = 0.80 in FHS, and 0.75 in ARIC) and CVD death (0.81 in FHS and 0.81 in ARIC), but not cancer death (0.57 in FHS and 0.71 in ARIC). Even though the clinical risk factors are important for stratifying CVD risk, clinical risk factors themselves are unable to reveal molecular mechanism and are thereby unable to highlight causal mechanisms or promising therapeutic targets.

After integrating clinical risk factors with DNAm in the all-cause mortality prediction model, the C-index only slightly increased (less than 2%) compared with the clinical risk factors model with regard to all-cause mortality and CVD death. As shown in Table S14, nine of the 14 clinical risk factors, including age, sex, physical activity, prevalent cancer, type II diabetes, hypertension, CHD, heart failure and stroke, as well as 36 CpGs that were selected as the representative features. Compared with clinical risk factors, the individual coefficients of the CpGs are much smaller. The small increase in the C-index and the small coefficients of the CpGs suggest that the contribution of CpGs to the prediction of death may overlap with these clinical risk factors. We also found that the mortality-related CpGs as the sole input features were still able to predict mortality outcomes after adjusting for clinical risk factors. This suggests that mortality-related CpGs may identify novel molecular mechanisms contributing to CVD mortality that cannot be captured by existing clinical risk factors.

In contrast to CVD and CVD mortality, for which established risk factors are highly predictive of risk, the prediction of cancer and cancer mortality has proved much more challenging. Owing to the lower prediction using clinical risk factors alone (0.57 in FHS and 0.71 in ARIC), the mortality-related
CpGs improved risk prediction of cancer death over and above the clinical risk factor model with an 11% increase in the C-index in FHS and a 5% increase in ARIC. We further tested whether the all-cause mortality prediction model can be used to predict mortality among all participants in the FHS with prevalent cancer (n=389). During a mean follow up of 9 years, there were 165 deaths in this group. The integrative mortality model predicted mortality risk among cancer cases (HR [95%CI]: 4.23 [2.63-6.80], \(P = 2.9 \times 10^{-9}\)). These results in conjunction with MR and pathway analysis, show strong evidence of potential causal relations between mortality-related CpGs and pathways in cancer. Based on these results, we hypothesize that mortality-related CpGs can shed light on the epigenetic regulation of molecular interactions and help to identify novel therapeutic targets to reduce mortality risk for both CVD and cancer death.

Recent studies have used DNAm of multiple CpG sites to predict chronological age (i.e. DNAm age), and showed that DNAm age was associated with all-cause mortality. We explored the prediction provided by these DNAm age models and show that PhenoAge (Levine et al., 2018), Horvath Age (Horvath, 2013), Hannum Age (Hannum et al., 2013), and GrimAge (Lu et al., 2019) were associated with mortality before accounting for risk factors. Only GrimAge, however, remained associated with mortality after adjusting for clinical risk factors. In contrast, the other three DNAm age models were no longer associated with mortality (Fig. 3). One possible explanation is that the three DNAm age predictors (i.e., PhenoAge, Horvath Age, and Hannum Age) identify CpGs associated with age, but are not specific for all-cause or cause-specific mortality risk. Of note, the CpGs that serve as DNAm mortality predictors and those that predict DNAm age in the three models do not overlap. Among the top CpGs (N=177) associated with all-cause mortality in our EWAS, only cg00687674 in TMEM84 is included in PhenoAge (Levine et al., 2018), and only cg19935065 in DNTT appears in Hannum Age (Hannum et al., 2013). GrimAge may have outperformed the other three DNAm age models in predicting mortality because the CpGs that it uses are associated with the levels of 80 CVD-related blood proteins, and with lifestyle and clinical risk factors (such as smoking), and mortality (Ho et al., 2018; Shah et al., 2019; Yao et al., 2018). However, because the CpGs in the GrimAge model are not disclosed (i.e. they are proprietary), we were unable to determine if any of the mortality-related CpGs
in our study overlap with CpGs in the GrimAge model. Of note, our mortality prediction models (both the CpG only model and the integrative model that included CpGs and the clinical risk factors) outperformed GrimAge in prediction of mortality outcomes.

We tested and compared four prediction methods including Elastic-coxph (Friedman et al., 2010), a regression based method, and three machine learning methods (Ching et al., 2018; Ishwaran et al., 2008; Katzman et al., 2018). The machine learning models did not outperform Elastic-coxph (Table S9 and Fig. S3). The clinical risk factor model trained by machine learning methods did not perform well in independent external replication. For example, the C-index of the clinical risk factor model for all-cause mortality was 0.67 using RSF17 versus 0.75 using Elastic-coxph in ARIC participants. Based on this metric, the machine learning methods did not outperform the regression-based methods when there were relatively few features as inputs.

The primary outcome of our study was all-cause mortality. We did not train prediction models for CVD death or cancer death, but we tested the prediction ability of the all-cause mortality predictor on CVD death and cancer death. The CpGs in the model were restricted to all-cause mortality related CpGs. As shown in Fig. 1, the top DNAm signatures for all-cause mortality showed the same direction of effect for CVD death and cancer death. It is possible that some CpGs show opposite directions in relation to CVD death and cancer death, but we did not train separate models for these outcomes. Therefore, developing separate prediction models for CVD death and cancer death with a very large sample size would be an important next step.

Conclusions

In conclusion, the ancestry-stratified epigenome-wide meta-analyses in 15 population-based cohorts identified replicable DNAm signatures of all-cause and cause-specific mortality. The top mortality-associated CpGs were linked with genes involved in immune and cancer related pathways, and were reported to be associated with human longevity, CVD risk factors, and several types of cancer. We
constructed and validated DNAm-based prediction models that predicted mortality risk independent of established clinical risk factors. The prediction model trained by integrating DNAm with clinical risk factors showed modest improvement in prediction of all-cause mortality and CVD death, and a substantial improvement in prediction of cancer death, compared with the model trained by clinical risk factors alone. The mortality-related CpG sites, and the DNAm-based prediction models may serve as useful clinical tools for assessing all-cause and cause-specific mortality risk and for developing new therapeutic strategies.

Methods

Study population

This study included 15,013 participants from 15 population-based cohorts. There were 11,684 European ancestry (EA) participants from 12 cohorts, including the Atherosclerosis Risk in Communities (ARIC) Study, the Cardiovascular Health Study (CHS), the Danish Twin Register sample (DTR), the Epidemiologische Studie zu Chancen der Verhütung, Früherkennung und optimierten Therapie chronischer Erkrankungen in der älteren Bevölkerung (ESTHER), the Framingham Heart Study (FHS), the Invecchiare in Chianti (InCHIANTI) Study, the Cooperative Health Research in the Region of Augsburg (KORA F4), the Lothian Birth Cohorts of 1921 (LBC1921) and 1936 (LBC1936), the Normative Aging Study (NAS), the Rotterdam Study (RS), and Women’s Health Initiative (WHI); and 3329 Africa ancestry (AA) participants from 3 cohorts, including ARIC, CHS, and WHI. For each participant, we calculated the follow-up time between the date of the blood draw for DNAm measurements and the date at death or last follow up. Mean follow up was less than 15 years (range 6.2 to 13.7) for most cohorts, except for ARIC (mean 20.0 for EA and 18.6 for AA). The protocol for each study was approved by the institutional review board of each cohort. Further details for each cohort were included in Additional file1.

Mortality ascertainment and clinical phenotypes
Outcomes including death from all causes, deaths from CVD, and deaths from cancer, were prospectively ascertained in each cohort. Survival status and details of death were ascertained using multiple strategies, including routine contact with participants for health history updates, surveillance at the local hospital, review of obituaries in the local newspaper, and National Death Index queries. Death certificates, hospital and nursing home records prior to death, and autopsy reports were requested and reviewed. Date and cause of death were determined separately for each cohort following review of all available medical records and/or were register-based.

The clinical and lifestyle risk factors (referred to as clinical risk factors for simplicity thereafter) used as covariates in this study included age, sex, body mass index (BMI), smoking, alcohol consumption, physical activity, educational attainment, and prevalent diseases including hypertension, coronary heart disease (CHD), heart failure, stroke, type-II diabetes, and cancer. Fourteen clinical risk factors were chosen based on prior knowledge; most of these are key CVD risk factors. The clinical risk factors were ascertained at the time of blood draw for DNAm measurements. BMI was calculated as weight (kg) divided by height squared (m²). Educational attainment (years of educational schooling), physical activity (frequency, intensity or the metabolic equivalent of task [MET] scores), smoking status (yes/no, or cigs/day), alcohol consumption (drinks per day) were self-reported or ascertained by an administered questionnaire at routine research clinic visits. Diabetes was defined as a measured fasting blood glucose level of >125 mg/dL or current use of glucose-lowering prescription medication. Hypertension was defined as a measured systolic blood pressure (BP) ≥140 mm Hg or diastolic BP ≥90 mm Hg or use of antihypertensive prescription medication. Cancer was defined as the occurrence of any type of cancer excluding non-melanoma skin cancer.

DNA methylation measurements and quality control

For each cohort, DNA was extracted from whole blood and bisulfite-converted using a Zymo EZ DNA methylation kit. DNAm was measured using the Illumina Infinium HumanMethylation450 (450K) BeadChip platform (Illumina Inc, San Diego, CA). Each cohort conducted independent
laboratory DNA methylation measurement, quality control (including sample-wise and probe-wise filtering, and probe intensity background correction; see Additional file1).

Cohort-specific epigenome-wide association analysis

The correction of methylation data for technical covariates was cohort specific. Each cohort performed an independent investigation to select an optimized set of technical covariates (e.g., batch, plate, chip, row and column), using measured or imputed blood cell type fractions, surrogate variables, and/or principal components. Most cohorts had previous publications using the same dataset for epigenome-wide association studies (EWAS) of different traits, such as EWAS of alcohol drinking and smoking (Mendelson et al., 2017; Michailidou et al., 2017). In this study, those cohorts used the same strategies as they did previously for correcting for technical variables including batch (see Additional file1). To avoid false positives driven by single CpG extreme values, in each cohort, we first performed rank-based inverse normal transformation (INT) of DNA methylation β-values (the ratio of methylated probe intensity divided by the sum of the methylation and unmethylated probe intensity). We then conducted time-to-event analyses using Cox proportional hazards models to test for associations between each CpG and mortality outcomes including all-cause mortality, CVD death, and cancer death using the `coxph()` function in the `survival` R library, adjusting for clinical risk factors (see Mortality ascertainment and clinical risk factors), technical confounders, and familial relatedness. Because ARIC cohorts had much longer follow-up than the other cohorts, ARIC follow up was truncated at 15 years and results were compared to those before truncation to determine if results were impacted by duration of follow up.

In this study, we performed INT of DNA methylation β-values to avoid false positives driven by extreme values of single CpGs. Using the FHS EWAS results as an example, Table S21 shows that the top CpGs associated with all-cause mortality (without INT) were no longer significant after performing INT. This finding suggests that if we directly use DNA methylation β-values, those extreme outlier values could lead to false positive results. Clearly, the distribution of DNA methylation β-values is non-normal and for this reason, we believe that the conservative INT approach we took protected against false positive results.

Meta-analysis
The meta-analysis was performed for all-cause mortality, CVD death, and cancer death in EA (n=11,684) and AA (n=3329) participants respectively, using inverse variance-weighted random-effects models implemented in metagen() function R packages (https://rdrr.io/cran/meta/man/metagen.html). We chose a random-effects model because of the heterogeneity in follow-up length and population demographics in the different cohorts (Table S1).

We excluded the EWAS results for a study with <20 deaths. We excluded probes mapping to multiple locations on the sex chromosomes or with an underlying SNP (MAF>5% in 1000 Genome Project data) at the CpG site or within 10bp of the single base extension. In addition, the meta-analysis was constrained to methylation probes passing filtering criteria in five or more cohorts (see Additional File1), which resulted in ~400,000 CpGs that were included in the final analyses. The statistical significance threshold was $P<0.05/400,000 \approx 1 \times 10^{-7}$.

Three types of sensitivity analyses were performed including 1) correcting for $\lambda$ values in each cohorts (Devlin, Roeder, & Wasserman, 2001), 2) excluding two cohorts with $\lambda > 1.5$ from the meta-analysis, and 3) excluding results of RS, because the cohort-specific analysis in RS having a strange distribution of top hits. There were 157 CpGs identified at $P<1e-7$ in the RS cohort-specific analysis. The number is much more than the number of all-cause mortality associated CpGs identified in the other cohorts.

**Mortality prediction models**

Mortality prediction models based on clinical risk factors and with the addition of DNAm were built and tested in EA cohorts. The analysis flowchart is shown in Fig. S1. To ensure unbiased validation, we split the EA cohorts into discovery and replication sets. The discovery cohorts consisted of 8288 participants from 10 cohorts, excluding FHS (n=2427) and ARIC (n=969), which were used as replication cohorts. To build and replicate a prediction model, the DNAm data were preprocessed utilizing the same strategy as in the EWAS analysis.

**Input features:** To evaluate the prediction performance of clinical risk factors and DNAm comprehensively, we tested 13 sets of features, Feature set 1 (F1) included age (years), sex (male as 1
and female as 2), and 12 other clinical risk factors including BMI (kg/m²), smoking (current smoker as 1, and former and never smoker as 0), alcohol consumption (grams/day), physical activity (MET scores), educational attainment (education years), and prevalent diseases (yes as 1 and no as 0) including hypertension, CHD, heart failure, stroke, type-2 diabetes, and cancer. F2-F7 were mortality-related CpGs selected by meta-analysis in the discovery cohorts by inverse-variance weighted random-effects models at a series of p value thresholds, including F2 CpGs at $P < 1 \times 10^{-7}$, F3 CpGs at $P < 1 \times 10^{-6}$, F4 CpGs at $P < 1 \times 10^{-5}$, F5 CpGs at $P < 1 \times 10^{-4}$, F6 CpGs at $P < 1 \times 10^{-3}$, and F7 CpGs at $P < 0.05$. F8-F13 are F1 (age, sex and 12 clinical phenotypes) plus F2-F7 respectively. In doing so, we were able to evaluate the prediction performance based on the clinical risk factors (F1) and the DNAm (F2-F7), and test if the combination of DNAm with clinical risk factors (F8-F13) could be able to improve the prediction performance by using clinical risk factors (F1) only and DNAm only (F2-F7).

Model building: We compared four methods of building prediction models, including 1) Elastic net - Cox proportional hazards method (Elastic-coxph, using glmnet, a R package) (Friedman et al., 2010); 2) Random survival forest (RSF, using randomForestSRC, a R package) (Ishwaran et al., 2008); 3) Cox-nnet (https://github.com/lanagarmire/cox-nnet, a Python package) (Ching et al., 2018), and 4) DeepSurv (https://github.com/jaredleekatzman/DeepSurv, a Python package) (Katzman et al., 2018). The first method is a penalized linear regression method, while the other three are non-linear machine learning methods.

Elastic-coxph is a Cox regression model regularized with elastic net penalty (Friedman et al., 2010). Performing this method requires to identify best values of two parameters, $\alpha$ and $\lambda$. We tuned each model by iterating over a number of $\alpha$ and $\lambda$ values under cross-validation. $\alpha$ indicated linearly combined penalties of the lasso ($\alpha=0$) and ridge ($\alpha=1$) regression. $\lambda$ is the shrinkage parameter, when $\lambda=0$ indicated no shrinkage, and as $\lambda$ increases, the coefficients are shrunk ever more strongly. Effectively this will shrink some coefficients close to 0 for optimizing a set of features. The $\alpha$ value was set to 0.5, and the $\lambda$ value was set to lambda.min when training models.
RSF is an ensemble tree model that is based on the random forest method for survival analysis (Ishwaran et al., 2008). The optimized values of parameters in RSF models, including the number of trees (nTrees=100) and nodeSize =15 were chosen by iterating over a number of values which maximized the accuracy of RSF models tested in the replication sets under cross-validation. RSF can compute feature importance scores for feature selection.

Cox-nnet is an artificial neural network based method for survival analysis (Ching et al., 2018). Cox-nnet includes two layer neural network: one hidden layer and one output layer. The output layer was used to perform Cox regression based on the activation levels of the hidden layer. Cox-nnet could also compute feature importance scores for feature selection. For each model training, the $L2$ regularization parameter is optimized using the $L2CVProfile$ Python function by iterating over a number of values under cross-validation.

DeepSurv is a deep learning-based survival prediction method (Katzman et al., 2018). DeepSurv uses a multi-layer feed forward neural network, of which the hidden layers consist of a fully-connected layer of nodes, followed by a dropout layer, and the output is a single node with a linear activation which estimated the log-risk function in the Cox model, parameterized by the weight of the network. The values of hyperparameters when using DeepSurv were $L2$ regularization = 0.8, dropout = 0.4, learning rate = 0.02, hidden layer size (4 layers with nodes 500, 200, 100 and 50), lr_decay = 0.001, momentum = 0.9 and the activation method (using Scaled Exponential Linear Units), which were optimized by iterating over a number of values each-by-each and under cross-validation. DeepSurv has not been used previously for selecting features.

**Cross-validation:** The 2427 FHS participants were randomly split into 5 equal sets (n=485 or 486 in each set), and each set included approximately equal numbers of deaths. We then used 3 of the 5 sets (60%) for model training and the remaining 2 sets (40%) for model testing. In doing so, we obtained 10 combinations. In each training / testing combination, we constructed a model using the training data, and then used the model to generate a mortality risk score based on the testing data. We assessed associations of the predicted mortality risk score (after inverse normal transformation) with all-cause
mortality, CVD death, and cancer death in the testing data using time-to-event proportional hazards models. This data partitioning and cross-validation strategy was only used to assess the robustness of prediction models when using different features and methods, and to select the optimized parameters for training models. The final models reported were built on all FHS participants using the optimized parameters. We also repeated the same analysis steps using FHS participants without cancer at baseline (n=2038; 238 deaths from all causes, 70 from CVD, and 42 from cancer).

Independent external validation: The prediction models built using all FHS participants were further tested in ARIC EA participants for the prediction of mortality outcomes. We performed tests on all-cause mortality and CVD death on all ARIC EA participants truncated at 15 years of follow up, and tests on cancer death after excluding prevalent cancer.

Evaluation of model performance: We used four evaluation metrics to assess model performance, including the concordance index (C-index) (Harrell Jr, Lee, & Mark, 1996), hazards ratio of predicted risk score (inversely-transformed) for prediction of mortality, the integrated brier score (IBS) (Brier, 1950), and Kaplan-Meier (KM) survival curves for high, medium and low risk groups (Kaplan & Meier, 1958). The C-index reflects the percentage of individuals whose predicted survival times are correctly ordered. A C-index of 0.50 reflects no improvement in prediction over chance. The brier score measures the mean of the difference between the observed and the estimated survival beyond a certain time. The brier score ranges between 0 and 1, and a larger score indicates higher inaccuracy. The integrated brier score is the brier score averaged over the entire time interval.

DNAm Age

We compared the prediction performance of DNAm age with our DNAm-based mortality prediction model in relation to all-cause mortality, CVD death, and cancer death in the ARIC EA cohort (truncating follow up at 15 years). Four measures of DNAm age were used in this study, including PhenoAge (Levine et al., 2018), Horvath age (Horvath, 2013), Hannum age (Hannum et al., 2013) and GrimAge (Lu et al., 2019). The Horvath Age is based on 353 CpGs, the Hannum age is based on 71 CpGs, and PhenoAge is based on 513 CpGs. DNAm age was calculated as the sum of the beta values.
multiplied by the reported effect size. Due to the GrimAge model was not publicly available, the
GrimAge was calculated by uploading the DNAm data to the website (http://dnamage.genetics.ucla.edu/). Proportional hazards regression models were used to test the association between inversely-rank transformed DNAm age (all 3 approaches) and mortality outcomes, adjusting for age, sex, and clinical covariates (see Mortality ascertainment and clinical phenotypes).

meQTLs

meQTLs (SNPs associated with DNA methylation) were identified from 4170 FHS participants as reported previously, including 4.7 million cis-meQTLs and 630K trans-meQTLs at $P < 2 \times 10^{-11}$ for cis and $P < 1.5 \times 10^{-14}$ for trans (Huan et al., 2019). The genotypes were measured using Affymetrix SNP 500K mapping and Affymetrix 50K gene-focused MIP arrays. Genotypes were imputed using the 1000 Genomes Project panel phase 3 using MACH / Minimac software. SNPs with MAF >0.01 and imputation quality ratio >0.3 were retained. cis-meQTLs were defined as SNPs residing within 1 Mb upstream or downstream of a CpG site. The FHS meQTL data resource includes 3.5 times more cis-, and 10 times more trans-meQTL SNPs than the other published studies to date (https://ftp.ncbi.nlm.nih.gov/eqtl/original_submissions/FHS_meQTLs/).

Mendelian randomization

Two-sample Mendelian randomization (MR) was used to identify putatively causal CpGs for human longevity, CVD and CVD risk factors, and cancer types using a multi-step strategy. Estimated associations and effect sizes between SNPs and traits were based on the latest published GWAS meta-analysis of human longevity (Deelen et al., 2019), coronary heart disease (CHD) (Nikpay et al., 2015); myocardial infarction (MI) (Nikpay et al., 2015); type-II diabetes (T2D) (Scott et al., 2017); body mass index (BMI) (Locke et al., 2015); lipids traits including high-density lipoprotein (HDL) cholesterol, low-density lipoprotein (LDL) cholesterol, total cholesterol (TC), and triglycerides (TG) (Willer et al., 2013); systolic blood pressure (SBP) and diastolic blood pressure (DBP) (Evangelou et al., 2018), and cancer types including breast cancer (Michailidou et al., 2017), prostate cancer
(Schumacher et al., 2018), lung cancer (Wang et al., 2014) and ovarian cancer (Phelan et al., 2017).

We were unable to include some other popular cancer types, because their GWAS data were not be accessible by us.

Instrumental variables (IVs) for each CpG site consisted of independent cis-meQTLs pruned at linkage disequilibrium (LD) $r^2<0.01$, retaining only one cis-meQTL variant with the lowest SNP-CpG $P$ value in each LD block. LD proxies were defined using 1000 genomes imputation in EA. Inverse variance weighted (IVW) MR tests were performed on CpGs with at least three independent cis-meQTL variants, which is the minimum number of IVs needed to perform multiple instruments MR. The multiple instruments improved the precision of IV estimates, and allowed the examination of underlying IV assumption (Palmer et al., 2012). Among 177 all-cause mortality related CpGs at $P<1\times10^{-7}$, MR tests were performed on 17 CpGs having $\geq 3$ independent cis-meQTL SNPs. To test the validity of IVW-MR results, we performed heterogeneity and MR-EGGER pleiotropy tests for all IVs. The statistical significance threshold for MR is $P_{MR}<0.05/17$, and both $P_{heter}$ and $P_{pleio}$ were required to be $>0.05$.

eQTM s

Association tests of DNAm and gene expression were performed in 3684 FHS participants with available DNAm and gene expression data. mRNA was extracted from whole blood (collected in PAXgene tubes) and profiled using the Affymetrix Human Exon 1.0 ST GeneChip platform. Raw gene expression data were first normalized using the RMA (robust multi-array average) from Affymetrix Power Tools (APT, thermofisher.com/us/en/home/life-science/microarray-analysis/affymetrix.html#1_2) with quantile normalization. Then, output expression values of 17,318 genes were extracted by APT based on NetAffx annotation version 31.

DNAm $\beta$ values were adjusted for age, sex, predicted blood cell fraction, the two top PCs of DNAm, and 25 surrogate variables (SVs), with DNAm as a fixed effect, and batch as a random effect by fitting LME models. Residuals (DNAm_resid) were retained. The gene expression values were adjusted for age, sex, predicted blood cell fraction, a set of technical covariates, the two top PCs and
25 SVs, with gene expression as a fixed effect, and batch as a random effect by LME, and residuals (mRNA_resid) were retained. Then, linear regression models were used to assess pair-wise associations between DNAm_resid and mRNA_resid. SVs were calculated using the SVA package in R. A cis-CpG-mRNA pair was defined as a CpG residing ±1Mb of the TSS of the corresponding gene encoding the mRNA (cis-eQTM). The annotations of CpGs and transcripts were obtained from annotation files of the HumanMethylation450K BeadChip and the Affymetrix exon array S1.0 platforms. We estimated that there were 1.6 x 10^8 potential cis-CpG-mRNA pairs. We only used cis-eQTMs in this study because trans-eQTMs were not replicated in independent external studies. The statistical significance threshold was \( P < 3 \times 10^{-10} \) (0.05/1.6x10^8).

**Gene ontology and pathway enrichment analysis**

Gene ontology and pathway enrichment analyses were performed on the genes annotated in relation to the 177 all-cause mortality related CpGs at \( P < 1 \times 10^{-7} \) or \( P < 1 \times 10^{-5} \) as well as the cis-eQTM genes associated with those CpGs. Hypergeometric tests were used to investigate over-representations of genes from multiple biological process and pathways. To improve focus in this study, we only used results of KEGG and Gene Ontology – biological process (GO-BP) terms. Enrichment tests used the online DAVID Bioinformatics Resources 6.8 (https://david.ncifcrf.gov/). The P-value was further corrected by the number of unique GO-BP terms and pathways. A Benjamini-Hochberg corrected FDR<0.2 was considered significant.

**Data Availability**

The DNA methylation data and phenotype data are available in dbGaP for some of the cohorts in this study (https://www.ncbi.nlm.nih.gov/gap/) including FHS (accession number phs000724.v5.p10) and WHI (accession number phs000200.v12.p3). For LBC, data are available in the European Genome-phenome Archive (https://www.ebi.ac.uk/ega/home), under accession number EGAS00001000910. For the other cohorts including ARIC, CHS, NAS, InCHIANTi, KORA, ESTHER, Danish, RS and
CARDIA, the data are available on request by contacting the principal investigators of each cohort.

Declarations

Ethics approval and consent to participate

This study included participants from 12 population-based cohorts studies, including the Atherosclerosis Risk in Communities (ARIC) Study, the Cardiovascular Health Study (CHS), the Danish Twin Register sample (DTR), the Epidemiologische Studie zu Chancen der Verhütung, Früherkennung und optimierten Therapie chronischer Erkrankungen in der älteren Bevölkerung (ESTHER), the Framingham Heart Study (FHS), the Invecchiare in Chianti (InCHIANTI) Study, the Cooperative Health Research in the Region of Augsburg (KORA F4), the Lothian Birth Cohorts of 1921 (LBC1921) and 1936 (LBC1936), the Normative Aging Study (NAS), the Rotterdam Study (RS), and Women’s Health Initiative (WHI). All of the 12 studies were approved by their institutional review committees (see details in Additional file 1). All study participants provided written informed consent.

Acknowledgements

The views expressed in this manuscript are those of the authors and do not necessarily represent the views of the National Heart, Lung, and Blood Institute; the National Institutes of Health; or the U.S. Department of Health and Human Services.

For a list of all the investigators who have contributed to WHI science, please visit: https://s3-us-west-2.amazonaws.com/www-whi-org/wp-content/uploads/WHI-Investigator-Long-List.pdf.

Funding

The Framingham Heart Study is funded by National Institutes of Health contract N01-HC-25195 and HHSN26820150001I. The laboratory work for this investigation was funded by the Division of Intramural Research, National Heart, Lung, and Blood Institute, National Institutes of Health. The
analytical component of this project was funded by the Division of Intramural Research, National Heart, Lung, and Blood Institute, and the Center for Information Technology, National Institutes of Health, Bethesda, MD.

The Cardiovascular Health Study is supported by NHLBI contracts HHSN268201200036C, HHSN268200800007C, HHSN268201800001C, N01HC55222, N01HC85079, N01HC85080, N01HC85081, N01HC85082, N01HC85083, N01HC85086; and NHLBI grants U01HL08295, U01HL130114, K08HL116640, R01HL087652, R01HL092111, R01HL103612, R01HL105756, R01HL103612, R01HL111089, R01HL116747 and R01HL120393 with additional contribution from the National Institute of Neurological Disorders and Stroke (NINDS). Additional support was provided through R01AG023629 from the National Institute on Aging (NIA), Merck Foundation / Society of Epidemiologic Research as well as Laughlin Family, Alpha Phi Foundation, and Locke Charitable Foundation. A full list of principal CHS investigators and institutions can be found at CHS-NHLBI.org. The provision of genotyping data was supported in part by the National Center for Advancing Translational Sciences, CTSI grant UL1TR000124, and the National Institute of Diabetes and Digestive and Kidney Disease Diabetes Research Center (DRC) grant DK063491 to the Southern California Diabetes Endocrinology Research Center. Infrastructure for the CHARGE Consortium is supported in part by the National Heart, Lung, and Blood Institute grant R01HL105756.

The DTR study was supported by The Danish Council for Independent Research—Medical Sciences (DFF-6110-00016), the European Union’s Seventh Framework Programme (FP7/2007–2011) under grant Agreement No. 259679 and The Danish National Program for Research Infrastructure 2007 (09-063256).

The WHI program is funded by the National Heart, Lung, and Blood Institute, National Institutes of Health, U.S. Department of Health and Human Services through contracts HHSN268201600018C, HHSN26820160002C, HHSN26820160003C, and HHSN26820160004C. Work in WHI was NIEHS-supported by R01-ES020836 (EAW; AB; LH).
Phenotype collection in the **Lothian Birth Cohort 1921** was supported by the UK’s Biotechnology and Biological Sciences Research Council (BBSRC), The Royal Society and The Chief Scientist Office of the Scottish Government. Phenotype collection in the **Lothian Birth Cohort 1936** was supported by Age UK (The Disconnected Mind project). Methylation typing was supported by 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.

IJD is a member of the University of Edinburgh Centre for Cognitive Ageing and Cognitive Epidemiology (CCACE), which is supported by funding from the BBSRC, the Medical Research Council (MRC), and the University of Edinburgh as part of the cross-council Lifelong Health and Wellbeing initiative (MR/K026992/1). W.D.H. is supported by a grant from Age UK (Disconnected Mind Project).

**Authors’ contributions**

T. H., D.L., and J.P. designed, directed, and supervised the project. T. H., and D.L. drafted the manuscript. T.H., S.N., E.C., C. R., D.H., J.B., M.S., Y.Z., A.B., E.M., and T.T. conducted the analyses. All authors participated in revising and editing the manuscripts. All authors have read and approved the final version of the manuscript.

**Competing Interests**

The authors declare no conflict of interest.
Table 1: Clinical characteristics the 15,013 study participants.

<table>
<thead>
<tr>
<th>Cohort</th>
<th>Total N</th>
<th>No. of all-cause death</th>
<th>No. of CVD death</th>
<th>No. of cancer death</th>
<th>Time to death / last follow-up years, mean (SD)</th>
<th>Age, mean (SD)</th>
<th>Sex (F, %)</th>
<th>BMI, mean (SD)</th>
<th>Type 2 Diabetes (n)</th>
<th>Coronary Heart Disease (n)</th>
<th>Heart Failure (n)</th>
<th>Stroke (n)</th>
<th>Hypertension (n)</th>
<th>Cancer (n)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>E</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ARIC</td>
<td>969</td>
<td>331</td>
<td>95</td>
<td>94</td>
<td>20.0 (5.2)</td>
<td>59.8 (5.5)</td>
<td>59</td>
<td>26.2 (4.5)</td>
<td>86</td>
<td>44</td>
<td>29</td>
<td>16</td>
<td>233</td>
<td>102</td>
</tr>
<tr>
<td>CHS</td>
<td>419</td>
<td>373</td>
<td>132</td>
<td></td>
<td>12.7 (6.1)</td>
<td>75.0 (4.9)</td>
<td>60</td>
<td>26.8 (4.9)</td>
<td>72</td>
<td>16</td>
<td>11</td>
<td>5</td>
<td>224</td>
<td>78</td>
</tr>
<tr>
<td>DTR</td>
<td>870</td>
<td>298</td>
<td>74</td>
<td>40</td>
<td>9.3 (3.4)</td>
<td>69.4 (7.9)</td>
<td>52</td>
<td>25.9 (3.9)</td>
<td>46*</td>
<td></td>
<td></td>
<td>37</td>
<td>269</td>
<td>129</td>
</tr>
<tr>
<td>ESTHER</td>
<td>1000</td>
<td>265</td>
<td>94</td>
<td>90</td>
<td>13.7 (3.5)</td>
<td>62.1 (6.5)</td>
<td>50</td>
<td>27.8 (4.3)</td>
<td>154</td>
<td>144</td>
<td>110</td>
<td>28</td>
<td>572</td>
<td>77</td>
</tr>
<tr>
<td>FHS</td>
<td>2427</td>
<td>403</td>
<td>91</td>
<td>155</td>
<td>9.1 (2.2)</td>
<td>66.3 (9.0)</td>
<td>55</td>
<td>28.3 (5.3)</td>
<td>279</td>
<td>226</td>
<td>53</td>
<td>116</td>
<td>107</td>
<td>389</td>
</tr>
<tr>
<td>InCHIANTI</td>
<td>488</td>
<td>104</td>
<td></td>
<td></td>
<td>10.0 (1.6)</td>
<td>62.4 (15.8)</td>
<td>52</td>
<td>27.0 (3.9)</td>
<td>42</td>
<td>31</td>
<td>9</td>
<td>10</td>
<td>232</td>
<td></td>
</tr>
<tr>
<td>KORA F-4</td>
<td>1727</td>
<td>89</td>
<td>31</td>
<td>35</td>
<td>6.4 (0.9)</td>
<td>61.0 (8.9)</td>
<td>51</td>
<td>28.1 (4.8)</td>
<td>158</td>
<td>105</td>
<td>41</td>
<td>47</td>
<td>789</td>
<td>154</td>
</tr>
<tr>
<td>LBC 1921</td>
<td>418</td>
<td>366</td>
<td></td>
<td></td>
<td>9.8 (4.7)</td>
<td>79.1 (0.6)</td>
<td>60</td>
<td>28.2 (4.0)</td>
<td>19</td>
<td>70</td>
<td></td>
<td>33</td>
<td>170</td>
<td></td>
</tr>
<tr>
<td>LBC 1936</td>
<td>900</td>
<td>192</td>
<td></td>
<td></td>
<td>10.2 (2.4)</td>
<td>69.6 (0.8)</td>
<td>50</td>
<td>27.7 (4.4)</td>
<td>72</td>
<td>221</td>
<td></td>
<td>46</td>
<td>364</td>
<td></td>
</tr>
<tr>
<td>NAS</td>
<td>640</td>
<td>221</td>
<td>123</td>
<td>72</td>
<td>10.5 (3.3)</td>
<td>72.8 (6.8)</td>
<td>0</td>
<td>28.1 (4.0)</td>
<td>117</td>
<td>181</td>
<td></td>
<td>42</td>
<td>447</td>
<td>316</td>
</tr>
<tr>
<td>RS</td>
<td>731</td>
<td>73</td>
<td></td>
<td></td>
<td>6.8 (1.5)</td>
<td>59.9 (8.2)</td>
<td>54</td>
<td>27.4 (4.5)</td>
<td>74</td>
<td>45</td>
<td></td>
<td>30</td>
<td>385</td>
<td>76</td>
</tr>
<tr>
<td>WHI</td>
<td>1095</td>
<td>192</td>
<td>48</td>
<td>60</td>
<td>11.5 (3.5)</td>
<td>62.9 (6.9)</td>
<td>100</td>
<td>28.8 (5.9)</td>
<td>60</td>
<td>20</td>
<td>5</td>
<td>11</td>
<td>469</td>
<td>14</td>
</tr>
<tr>
<td><strong>A</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ARIC</td>
<td>2446</td>
<td>1069</td>
<td>424</td>
<td>322</td>
<td>18.6 (6.6)</td>
<td>56.5 (5.8)</td>
<td>64</td>
<td>30.1 (6.2)</td>
<td>643</td>
<td>120</td>
<td>163</td>
<td>75</td>
<td>1373</td>
<td>87</td>
</tr>
<tr>
<td>CHS</td>
<td>325</td>
<td>264</td>
<td>96</td>
<td></td>
<td>12.9 (6.6)</td>
<td>73.1 (5.5)</td>
<td>62</td>
<td>28.6 (5.2)</td>
<td>68</td>
<td>2</td>
<td>0</td>
<td>2</td>
<td>235</td>
<td>36</td>
</tr>
<tr>
<td>WHI</td>
<td>558</td>
<td>74</td>
<td>27</td>
<td></td>
<td>10.6 (3.7)</td>
<td>61.6 (8.6)</td>
<td>100</td>
<td>31.5 (6.1)</td>
<td>76</td>
<td>18</td>
<td>11</td>
<td>12</td>
<td>369</td>
<td>2</td>
</tr>
</tbody>
</table>

*The diabetes cases in DTR included both type-I and type-II diabetes.*

The clinical risk factors were ascertained at the time of blood draw for DNAm measurements. BMI was calculated as weight (kg) divided by height squared (m2). Diabetes was defined as a measured fasting blood glucose level of >125 mg/dL or current use of glucose-lowering prescription medication. Hypertension was defined as a measured systolic blood pressure (BP) ≥140 mm Hg or diastolic BP ≥90 mm Hg or use of antihypertensive prescription medication. Cancer was defined as the occurrence of any type of cancer excluding non-melanoma skin cancer.
Table 2: Trans-ethnic replicated all-cause mortality related CpGs.

<table>
<thead>
<tr>
<th>CpG</th>
<th>Chr</th>
<th>Position</th>
<th>Gene</th>
<th>Meta-analysis EA cohorts</th>
<th>Meta-analysis AA cohorts</th>
<th>Transethnic replication</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>HR (95% CI) P-value</td>
<td>HR (95% CI) P-value</td>
<td>Bonferroni corrected P</td>
</tr>
<tr>
<td>cg16743273</td>
<td>19</td>
<td>2076833</td>
<td>MOBKL2</td>
<td>1.15 (1.1-1.21) 1.57E-09</td>
<td>1.24 (1.15-1.33) 1.28E-08</td>
<td>2.08E-06</td>
</tr>
<tr>
<td>cg18181703</td>
<td>17</td>
<td>76354621</td>
<td>SOCS3</td>
<td>0.83 (0.8-0.87) 6.15E-16</td>
<td>0.82 (0.77-0.88) 3.71E-08</td>
<td>6.05E-06</td>
</tr>
<tr>
<td>cg21393163</td>
<td>1</td>
<td>12217629</td>
<td></td>
<td>0.84 (0.8-0.88) 4.15E-12</td>
<td>0.84 (0.79-0.89) 7.48E-08</td>
<td>1.22E-05</td>
</tr>
<tr>
<td>cg15310871</td>
<td>8</td>
<td>2007936</td>
<td>ATP6V1B</td>
<td>1.18 (1.12-1.25) 1.42E-08</td>
<td>1.19 (1.11-1.26) 1.80E-07</td>
<td>2.94E-05</td>
</tr>
<tr>
<td>cg25933130</td>
<td>10</td>
<td>63753550</td>
<td>ARID5B</td>
<td>0.87 (0.83-0.91) 4.67E-10</td>
<td>0.86 (0.81-0.91) 3.71E-08</td>
<td>1.98E-04</td>
</tr>
<tr>
<td>cg05438378</td>
<td>15</td>
<td>67383736</td>
<td>SOCS3</td>
<td>0.88 (0.84-0.92) 1.52E-08</td>
<td>0.85 (0.79-0.91) 3.68E-06</td>
<td>6.00E-04</td>
</tr>
<tr>
<td>cg26470501</td>
<td>19</td>
<td>45252955</td>
<td>BCL3</td>
<td>0.84 (0.79-0.88) 8.38E-12</td>
<td>0.81 (0.74-0.89) 1.48E-05</td>
<td>2.42E-03</td>
</tr>
<tr>
<td>cg06126421</td>
<td>6</td>
<td>30720080</td>
<td></td>
<td>0.85 (0.79-0.87) 2.48E-10</td>
<td>0.84 (0.78-0.91) 1.69E-05</td>
<td>2.75E-03</td>
</tr>
<tr>
<td>cg02003183</td>
<td>14</td>
<td>103415882</td>
<td>CDC42BP</td>
<td>1.19 (1.13-1.26) 1.94E-11</td>
<td>1.16 (1.08-1.24) 2.00E-05</td>
<td>3.26E-03</td>
</tr>
<tr>
<td>cg10950251</td>
<td>1</td>
<td>204466432</td>
<td></td>
<td>0.86 (0.82-0.91) 4.05E-10</td>
<td>0.86 (0.80-0.92) 2.34E-05</td>
<td>3.81E-03</td>
</tr>
<tr>
<td>cg17501210</td>
<td>6</td>
<td>166970252</td>
<td>RPS6KA2</td>
<td>0.86 (0.81-0.91) 5.84E-09</td>
<td>0.87 (0.82-0.93) 2.71E-05</td>
<td>4.41E-03</td>
</tr>
<tr>
<td>cg23598089</td>
<td>12</td>
<td>233703120</td>
<td>GIGYF2</td>
<td>1.13 (1.08-1.18) 2.36E-08</td>
<td>1.14 (1.07-1.22) 4.19E-05</td>
<td>6.84E-03</td>
</tr>
<tr>
<td>cg21993290</td>
<td>2</td>
<td>233703120</td>
<td></td>
<td>0.88 (0.84-0.92) 6.13E-08</td>
<td>0.87 (0.81-0.93) 4.94E-05</td>
<td>8.06E-03</td>
</tr>
<tr>
<td>cg04987734</td>
<td>16</td>
<td>103415873</td>
<td>CDC42BP</td>
<td>1.23 (1.15-1.32) 2.53E-14</td>
<td>1.15 (1.07-1.23) 5.77E-05</td>
<td>9.41E-03</td>
</tr>
<tr>
<td>cg20813374</td>
<td>6</td>
<td>35696870</td>
<td>FKBP5</td>
<td>0.84 (0.78-0.89) 4.27E-08</td>
<td>0.84 (0.77-0.91) 7.19E-05</td>
<td>1.17E-02</td>
</tr>
<tr>
<td>cg11927233</td>
<td>5</td>
<td>170816542</td>
<td>NPM1</td>
<td>0.84 (0.80-0.94) 2.43E-09</td>
<td>0.89 (0.84-0.95) 2.41E-04</td>
<td>3.92E-02</td>
</tr>
<tr>
<td>cg24859433</td>
<td>11</td>
<td>315102</td>
<td>IFITM1</td>
<td>0.86 (0.80-0.93) 7.15E-10</td>
<td>0.88 (0.82-0.94) 2.70E-04</td>
<td>4.40E-02</td>
</tr>
<tr>
<td>cg01445100</td>
<td>16</td>
<td>88103339</td>
<td></td>
<td>1.23 (1.15-1.32) 2.53E-14</td>
<td>1.15 (1.07-1.23) 5.77E-05</td>
<td>9.41E-03</td>
</tr>
</tbody>
</table>

Abbreviation: HR, hazard ratio per standard deviation; CI, confidence interval; EA, European ancestry; AA, African ancestry.
Table 3: Performance robustness comparison of mortality predictors in FHS and ARIC cohorts.

<table>
<thead>
<tr>
<th>Model</th>
<th>FHS</th>
<th>ARIC</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>HR (95% CI)</td>
<td>C-index</td>
</tr>
<tr>
<td><strong>All-cause mortality</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Clinical Risk Factor Model</td>
<td>3.37 (2.21-3.15)</td>
<td>0.80</td>
</tr>
<tr>
<td>CpG Model</td>
<td>2.64 (2.21-3.15)</td>
<td>0.75</td>
</tr>
<tr>
<td>Integrative Model</td>
<td>2.95 (2.45-3.55)</td>
<td>0.77</td>
</tr>
<tr>
<td><strong>CVD Death</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Clinical Risk Factor Model</td>
<td>3.74 (2.57-4.79)</td>
<td>0.81</td>
</tr>
<tr>
<td>CpG Model</td>
<td>3.51 (2.57-4.79)</td>
<td>0.81</td>
</tr>
<tr>
<td>Integrative Model</td>
<td>3.90 (2.63-5.05)</td>
<td>0.80</td>
</tr>
<tr>
<td><strong>Cancer Death (excluding prevalent cancer cases)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Clinical Risk Factor Model</td>
<td>1.25 (1.74-3.18)</td>
<td>0.57</td>
</tr>
<tr>
<td>CpG Model</td>
<td>2.35 (1.74-3.18)</td>
<td>0.71</td>
</tr>
<tr>
<td>Integrative Model</td>
<td>1.78 (1.90-3.50)</td>
<td>0.68</td>
</tr>
</tbody>
</table>

Abbreviation: HR, hazard ratio per standard deviation; IBS: Integrated brier score;
* HR, C-index and IBS values in FHS reflect the average values of 10 times cross-validation.
†The results were obtained from ARIC European ancestry participants with follow up truncated at 15 years.
The clinical risk factor models were trained by using clinical risk factors as the sole input features. The CpG Models were trained by using CpGs selecting in the discovery meta-analysis. The integrative model was trained by using both clinical risk factors and CpGs selecting in the discovery meta-analysis.
The Clinical Risk Factor Model used to predict all-cause mortality and CVD death was shown in Table S10, and to predict cancer death (trained in samples excluding prevalent cancer cases) was shown in Table S11. The CpG model used to predict all-cause mortality and CVD death was shown in Table S12, and to predict cancer death (trained in samples excluding prevalent cancer cases) was shown in Table S13. The integrative model used to predict all-cause mortality and CVD death was shown in Table S14, and to predict cancer death (trained in samples excluding prevalent cancer cases) was shown in Table S15.
Figure Legends

**Figure 1:** The effect sizes (log hazards ratios) and 95% confidence intervals of CpGs related to mortality identified by meta-analysis, comparing the results for all-cause mortality, CVD death, and cancer death. A) Results of meta-analysis of European ancestry (EA); B) Results of meta-analysis of African ancestry (AA). These figures showed the CpGs associated with all-cause mortality identified by the meta-analysis, which were also associated with either CVD death or cancer death passing Bonferroni corrected threshold. Figure 1A shows 51 CpGs in EA, including 41 CpGs associated with CVD death, 16 with cancer death, and 5 with both. Figure 1B shows 16 CpGs in AA, including 15 CpGs associated with CVD death, 8 with cancer death, and 7 with both.

**Figure 2:** Kaplan–Meier estimates of mortality risk scores with respect to mortality outcomes in ARIC study. A) survival curves with respect to all-cause mortality; B) survival curves with respect to CVD death; C) survival curves with respect to cancer death. The results were obtained from ARIC European ancestry participants with follow up truncated at 15 years. For cancer death, we excluded samples who had any type of cancer before blood drawn for DNA methylation measurements. The mortality risk scores for A) and B) were computed by the model (Table S10), and for C) was computed by the model (Table S11).

**Figure 3:** Hazard ratios per standard deviation increment with 95% confidence intervals for mortality. A) with respect to all-cause mortality; B) with respect to CVD death; C) with respect to cancer death. The results were obtained from ARIC European ancestry participants with follow up truncated at 15 years. For cancer death, samples who had any type of cancer before blood drawn for DNA methylation measurements were excluded. Cox regression models were used to relate mortality outcomes to inversely-transformed mortality risk scores computed by Integrative models (Table S12-S13) and CpG models (Table S10-S11), and inversely-transformed DNAm age including GrimAge (Lu et al., 2019), PhenoAge(Levine et al., 2018), Horvath Age(Horvath, 2013), and Hannum Age(Hannum et al., 2013). *Adj age and sex* indicated the association further adjusted for age and sex.
Adj age, sex and risk factors indicated the association further adjusted for age, sex and the other clinical risk factors.

Figure 1

A

B
Figure 2

A

B

C
Figure 3

A

<table>
<thead>
<tr>
<th>Model</th>
<th>Hazard ratio</th>
<th>Pval</th>
</tr>
</thead>
<tbody>
<tr>
<td>Integrative Model</td>
<td>2.95</td>
<td>1.35e-30</td>
</tr>
<tr>
<td>Integrative Model (Adj age and sex)</td>
<td>2.4</td>
<td>5.1e-16</td>
</tr>
<tr>
<td>Integrative Model (Adj age, sex and risk factors)</td>
<td>1.85</td>
<td>1.14e-08</td>
</tr>
<tr>
<td>CpG Model</td>
<td>2.21</td>
<td>2e-20</td>
</tr>
<tr>
<td>CpG Model (Adj age and sex)</td>
<td>1.81</td>
<td>7.8e-10</td>
</tr>
<tr>
<td>CpG Model (Adj age, sex and risk factors)</td>
<td>1.68</td>
<td>9.76e-07</td>
</tr>
<tr>
<td>GrimAge (Adj age and sex)</td>
<td>2.24</td>
<td>5.24e-22</td>
</tr>
<tr>
<td>GrimAge (Adj age, sex and risk factors)</td>
<td>1.85</td>
<td>4.44e-09</td>
</tr>
<tr>
<td>PhenoAge (Adj age and sex)</td>
<td>1.59</td>
<td>0.00033</td>
</tr>
<tr>
<td>PhenoAge (Adj age, sex and risk factors)</td>
<td>1.22</td>
<td>0.03</td>
</tr>
<tr>
<td>HannumAge (Adj age and sex)</td>
<td>1.18</td>
<td>0.07</td>
</tr>
<tr>
<td>HannumAge (Adj age, sex and risk factors)</td>
<td>1.83</td>
<td>2.59e-12</td>
</tr>
<tr>
<td>HorvathAge (Adj age and sex)</td>
<td>1.21</td>
<td>0.06</td>
</tr>
<tr>
<td>HorvathAge (Adj age, sex and risk factors)</td>
<td>1.62</td>
<td>1.42e-08</td>
</tr>
<tr>
<td>HorvathAge (Adj age and sex)</td>
<td>1.1</td>
<td>0.37</td>
</tr>
</tbody>
</table>

B

<table>
<thead>
<tr>
<th>Model</th>
<th>Hazard ratio</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Integrative Model</td>
<td>3.65</td>
<td>6.25e-15</td>
</tr>
<tr>
<td>Integrative Model (Adj age and sex)</td>
<td>2.84</td>
<td>2.63e-08</td>
</tr>
<tr>
<td>Integrative Model (Adj age, sex and risk factors)</td>
<td>1.88</td>
<td>0.005</td>
</tr>
<tr>
<td>CpG Model</td>
<td>2.62</td>
<td>9.85e-11</td>
</tr>
<tr>
<td>CpG Model (Adj age and sex)</td>
<td>1.93</td>
<td>0.00016</td>
</tr>
<tr>
<td>CpG Model (Adj age, sex and risk factors)</td>
<td>1.81</td>
<td>0.002</td>
</tr>
<tr>
<td>GrimAge (Adj age and sex)</td>
<td>2.77</td>
<td>3.88e-12</td>
</tr>
<tr>
<td>GrimAge (Adj age, sex and risk factors)</td>
<td>2.12</td>
<td>0.000143</td>
</tr>
<tr>
<td>PhenoAge (Adj age and sex)</td>
<td>1.68</td>
<td>0.000392</td>
</tr>
<tr>
<td>PhenoAge (Adj age, sex and risk factors)</td>
<td>1.24</td>
<td>0.17</td>
</tr>
<tr>
<td>HannumAge (Adj age and sex)</td>
<td>1.83</td>
<td>0.22</td>
</tr>
<tr>
<td>HannumAge (Adj age, sex and risk factors)</td>
<td>2.56</td>
<td>3.26e-10</td>
</tr>
<tr>
<td>HorvathAge (Adj age and sex)</td>
<td>1.44</td>
<td>0.04</td>
</tr>
<tr>
<td>HorvathAge (Adj age, sex and risk factors)</td>
<td>1.41</td>
<td>0.07</td>
</tr>
<tr>
<td>HorvathAge (Adj age and sex)</td>
<td>2.19</td>
<td>1.54e-07</td>
</tr>
<tr>
<td>HorvathAge (Adj age, sex and risk factors)</td>
<td>1.32</td>
<td>0.13</td>
</tr>
</tbody>
</table>

C
References


Sc promotional text


