Epigenetic entropy, social disparity, and health and lifespan in the Women's Health Initiative

Khyobeni Mozhui[1,2,*], Athena Starlard-Davenport[2], Yangbo Sun[1], Aladdin H. Shadyab[3], Ramon Casanova[4], Fridtjof Thomas[1], Robert B Wallace[5], Jay H Fowke[1], Karen C Johnson[1]

[1]Department of Preventive Medicine, College of Medicine, University of Tennessee Health Science Center, Memphis, TN, USA.

[2]Department of Genetics, Genomics and Informatics, College of Medicine, University of Tennessee Health Science Center, Memphis, TN, USA.

[3]Herbert Wertheim School of Public Health and Human Longevity Science and Division of Geriatrics, Gerontology, and Palliative Care, Department of Medicine, University of California San Diego, La Jolla, CA, USA.

[4]Department of Biostatistics and Data Science, Wake Forest University School of Medicine, Winston-Salem, NC, USA.

[5]College of Public Health, University of Iowa, Iowa City, IA, USA.

[*]Corresponding author: kmozhui@uthsc.edu

ABSTRACT

The pace of aging varies between individuals and is marked by changes in DNA methylation (DNAm) including an increase in randomness or entropy. Here, we computed epigenetic scores of aging and entropy using DNAm datasets from the Women's Health Initiative (WHI). We investigated how different epigenetic aging metrics relate to demographic and health variables, and mortality risk. Income and education, two proxies of socioeconomics (SE), had consistent associations with epigenetic aging and entropy. Notably, stochastic increases in DNAm at sites targeted by the polycomb proteins were significantly related to both aging and SE. While higher income was associated with reduced age-related DNAm changes in White women, the protective effect of income was diminished in Black and Hispanic women, and on average, Black and Hispanic women had relatively more aged epigenomes. Faster pace of aging, as estimated by the DunedinPACE, predicted higher mortality risk, while the maintenance of methylation at enhancer regions was associated with improved survival. Our findings demonstrate close ties between social and economic factors and aspects of epigenetic aging, suggesting potential biological mechanisms through which societal disparities may contribute to differences in health outcomes and lifespan across demographic groups.

35 **INTRODUCTION**

36 Epigenetic clocks such as the Horvath, Hannum, PhenoAge , and DunedinPACE are machine-
37 learning (ML) based predictive models that estimate the biological age, or the rate-of-aging, of
38 an individual.[1-4] The underlying biomolecular data is DNA methylation (DNAm), the epigenetic
39 process that entails a chemical modification to the DNA—specifically, whether a cytosine
40 residue that is adjacent to a guanine (a CpG site) has a methyl- tag or is unmethylated. These
41 epigenetic markers can be measured from readily accessible biofluids such as blood or saliva
42 and are informative of whether a person is aging faster or slower relative to their chronological
43 age. The first-generation epigenetic clocks (e.g., Horvath, Hannum) are accurate predictors of
44 chronological age because the algorithms were mainly trained on the age variable.[1,2] The
45 second-generation clocks (e.g., PhenoAge, GrimAge) were trained on a broader panel of age-
46 related health parameters and perform remarkably well at assessing overall health, disease
47 states, and life expectancy.[3,5,6] The more recent DunedinPACE model was developed by tracking
48 longitudinal changes in health traits from age-matched individuals over decades, and it directly
49 measures the pace-of-aging.[4]

50 These existing models estimate biological age using preselected sets of CpGs that were picked
51 by the respective ML models and assigned specific weights, and the final value of aging is a
52 weighted combination of methylation levels at these distinct CpGs sites. The lists of preselected
53 CpGs are typically unique for each model, and it is unclear as to why the "mind of the
54 algorithm" selected those specific sites and were assigned those weights.[7-9] While there are
55 now several such ML-based biomarkers that are widely utilized in epidemiological research, the
56 mechanistic pathways and the extent to which these represent stochastic versus programmatic

57 processes of aging remain largely undefined.[10,11] However, several of the DNAm clocks do share
58 some common themes that hint at the underlying biology. For instance, relative to the genome-
59 wide background, bivalent chromatin states and DNA sequences bound by the polycomb
60 repressive complex (particularly PRC2) are consistently overrepresented among both the first-
61 and second-generation clocks.[3,6,12] The bivalent and PRC2 bound sequences are evolutionarily
62 conserved and play critical roles during embryonic development and cell fate determination.[13]
63 A common feature of aging is for CpGs in these chromatin states to drift from a low methylation
64 to a high methylation state, and the gain in methylation with aging has been implicated in
65 cancers and other health risks.[14] This is closely related to the concept of epigenetic entropy, a
66 term borrowed from Shannon's Information theory, that refers to the level of randomness that
67 can be computed from DNAm data either at the global genome-wide scale, or at regional
68 scale.[14-16] These stereotypical shifts in the epigenetic landscape and the accumulation of
69 stochastic variability likely contribute to some of the signal captured by the DNAm aging
70 biomarkers.[9,14,16,17]

71 The Women's Health Initiative (WHI) is a long-term prospective study of postmenopausal
72 women that began recruitment in 1993 and has been collecting extensive health, medical, and
73 lifespan data since that time.[18-20] Previous epigenetic studies of DNAm clocks in the WHI have
74 shown that a higher rate of epigenetic aging is predictive of lifespan. Prior studies also report
75 associations with lung cancer,[21] insomnia and immune aging,[22] cognitive impairment and
76 dementia,[23] diabetes related traits, and cardiovascular health.[24-27] Notably, the second-
77 generation PhenoAge clock has strong associations with social disparities, and WHI participants
78 with lower education exhibited more advance epigenetic aging. Liu et al. also found that an

79 accelerated PhenoAge partly explain the disparity in life expectancy between racial and ethnic
80 groups.[27] There is growing evidence that these epigenetic biomarkers not only tell us about the
81 intrinsic aging of cells but are strongly influenced by the larger social and environmental
82 context. In the United States, recent studies in young as well as older adults have shown that
83 the DunedinPACE detects more rapid aging among Black participants and among participants
84 with low education and income levels.[28,29] This sensitivity to social factors should not be
85 entirely surprising since variables such as education, income inequity, and experiences of
86 societal biases have profound and long-lasting impact on health, stress, and mental and
87 emotional well-being, and are likely linked to the differential rates of biological aging. In fact,
88 these close ties between the epigenetic models of aging and social variables have been
89 interpreted in light of the "weathering hypothesis", which proposes that the chronic exposure
90 to socioeconomic disadvantage, and to race-based and other stressors lead to accelerated
91 aging and higher disease burden among African Americans and other marginalized groups.[27-30]

92 However, a caveat to keep in mind is that these ML-based epigenetic biomarkers were initially
93 trained in cohorts that were predominantly, or in the case of the DunedinPACE, almost
94 exclusively of European ancestry.[4] The training conditions likely introduce some algorithmic
95 biases and may not generalize in full to other populations. In the present work, we attempt to
96 unlink the epigenetic readouts from potential training-based effects by computing measures of
97 epigenetic entropy and variability in CpG methylation that do not rely on training algorithms.
98 For the present study, the only pre-selection or sub-setting of CpGs we performed was based
99 on biologically informed chromatin states. The questions we ask are: (1) *how do these readouts*
100 *of epigenetic entropy and stochasticity, and gross chromatin states relate to the training-based*

101 *models of aging*? And, *are these non-ML readouts as informative of health and socioeconomic*
102 *variables as the ML-based models, and predictive of life expectancy*? Notably, we present novel
103 evidence linking measures of epigenetic stochasticity with social and racial disparities.
104 Furthermore, we uncover intriguing connections between methylation maintenance at active
105 enhancer regions and lifespan.

106 RESULTS

107 Datasets and participant characteristics

108 The present work is a secondary analysis of DNAm data generated by three WHI ancillary
109 studies: (1) WHI-EMPC/AS315, [26,31] (2) BA23, [21,32] and (3) AS311.[33] All three datasets measured
110 blood DNA methylation using the Illumina Infinium Methylation 450K BeadChip (Illumina Inc.).
111 Brief summaries of baseline characteristics are provided in **Table 1**, and sample
112 inclusion/exclusion chart is in **Fig S1**. BA23 and AS311 used blood collected at the time of
113 eligibility screening visit (SV) and prior to randomization to a study arm in the clinical trials (CT)
114 or enrollment in the observational study (OS). In AS315, DNA methylation was measured either
115 at SV (64%), annual visit 3 (29%), or annual visit 6 (7%) in the CT only. We used the larger and
116 more diverse AS315 dataset to describe the large-scale topology of the methylome, and all
117 analyses were initially performed in AS315 (including a sensitivity analysis using only blood
118 collected at the SV). The main findings were then tested for replication in BA23 and AS311 after
119 excluding participants who were also part of AS315 (**Fig S1**).

120 Epigenetic stochasticity and how it relates to chromatin states and aging

121 The methylation beta-values show the expected bimodal distribution with most CpGs at beta-
122 values close to 0 (i.e., most cells are unmethylated at that CpG) or 1 (most cell are methylated
123 at that CpG) (**Fig 1a**). With aging, cells begin to drift from their initial methylation states,[34] and
124 this results in an increase in entropy and a shift towards a beta = 0.5 (a "hemi-methylated"
125 state). We computed the methylome-wide entropy using all CpGs that had complete data in all
126 participants. To illustrate the "landscape erosion" towards a hemi-methylated and presumably
127 more random state (i.e., beta ~ 0.5), **Fig 1a** displays 5 participants with Shannon entropy > 0.85,
128 and 5 with entropy <0.4. This global measure of epigenetic discordance showed a wide
129 variation in the AS315 cohort (**Fig 1b**) but had only a weak positive correlation with age (r=0.08,
130 p = 0.0002; **Fig 1c**).

131 As methylation levels are highly dependent on the genomic context, we annotated each CpG
132 for the predicted chromatin state based on the Roadmap Epigenomics Consortium's
133 ChromHMM 15-states model.[35-38] This overlays the DNAm data with histone-based predicted
134 epigenetic states and regulatory factors (**Table 2**).[35] For each participant, we computed the
135 average beta-values at each of the 15 states thereby reducing the ~450,000 features to 15
136 values. The mean methylation at the chromatin states showed varying levels of correlation with
137 chronological age (**Fig 1d**; **Table 2**). The highest positive correlate of age was the repressed
138 bivalent transcription start site (TssBiv; r = 0.16, **Fig S2a**), and the highest negative correlate was
139 for enhancer regions (Enh; r = -0.19, **Fig 1f**).

140 Whether a CpG positively or negatively correlates with chronological age is partly dependent on
141 the average methylation of its chromatin state, and this is depicted in **Fig 1e**. For instance, CpGs

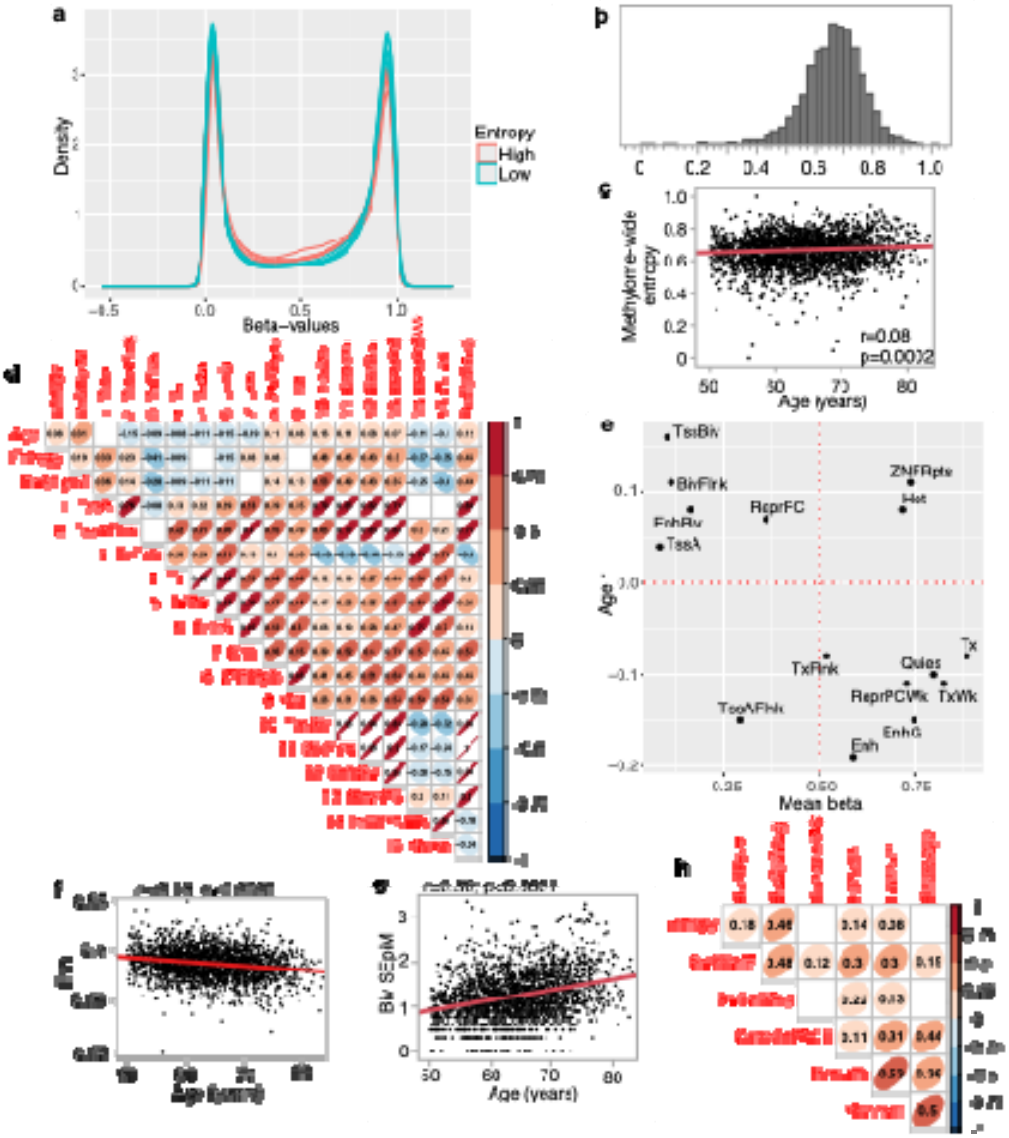


Fig 1. (a) Density plots of methylome-wide beta-values for 10 AS315 participants (all identified as non-Hispanic White). 5 have entropy ≤0.4 (cyan), and 5 have entropy ≥0.85 (soft red). **(b)** Histogram of methylome-wide entropy shows wide variability in AS315. **(c)** Weak but positive correlation between chronological age and methylome entropy in AS315. **(d)** Pair-wise correlations between age and the different non-training based epigenetic readouts including mean methylation by chromatin states (due to the large number of comparisons, the graph only displays correlations with $p < 0.001$). **(e)** The x-axis is the mean beta-values for the 15 chromatin states; y-axis is the Pearson r between these states and chronological age. Mean methylation at enhance CpGs decrease **(f)**, while levels of stochastic epimutations at the three bivalent sites (TssBiv, BivFlnk, EnhBiv) increase with chronological age **(g)**. **(h)** Pair-wise correlation for the non-training based measures (overall entropy, stochasticity at bivalent sites, and average beta-values for the bivalent CpGs) and four train-based measures of aging (only displays correlations with $p < 0.001$). All these are residual values adjusted for chronological age.

at the bivalent sites—TssBiv, BivFlnk, and EnhBiv— are known to have low mean methylation and gain methylation with aging, and this pattern is also seen in the WHI (plots to the top left

145 quadrant of **Fig 1e**).[39] These bivalents states also had the highest positive r with entropy (e.g.,
146 **Fig S2c**). For this reason, we explored the idea of "stochastic epimutations" (SEpiM) at these
147 sites. We implemented an outlier detection approach,[40,41] and counted for positive outliers at
148 TssBiv, BivFlnk, and EnhBiv. As these CpGs gain methylation with aging, a positive outlier with
149 higher beta-value is presumed to represent a more aged methylome. The outlier counts (what
150 we refer to as BivSEpiM) showed strong positive correlation with chronological age (**Fig 1g**). In
151 contrast to the bivalent sites, CpGs that flank actively transcribed genes (TxFlnk) had mean
152 methylation of beta-value ~0.5 and had the strongest negative correlation with entropy **(Fig**
153 **S2d**) and had modest inverse correlation with age (bottom right quadrant in **Fig 1e**). Weakly
154 repressed polycomb (ReprPCWk) and Quiescent (Quies) were also regions with higher mean
155 methylation (beta-values >0.70) and negatively correlated with both entropy and age.
156 Collectively, the results indicate that CpGs in low methylation sites (e.g., TssA, TssBiv, etc.) gain
157 methylation with aging, and contribute to higher entropy. In contrast, CpGs in sites with higher
158 mean methylation (TxFlnk, ReprPCWk and Quies) tend to lose methylation with aging, and
159 methylation maintenance at these sites may contribute to lower entropy.

160 **Relating epigenetic stochasticity to training-based DNAm biomarkers**

161 To evaluate how the non-training based DNAm readouts relate to the training-based models of
162 biological aging, we compared pace or rates of biological aging (i.e., age-deviation or age-
163 acceleration) estimated by the original Horvath and Hannum clocks, and the newer PhenoAge
164 and DunedinPACE.[1-4] After adjusting for chronological age, higher BivSEpiM was consistently
165 associated with a higher pace of aging as defined by the prior models (**Fig 1h**). The methylome-

166 wide entropy was positively correlated with higher age-acceleration as estimated by the
167 Horvath or Hannum clocks (**Fig 1h**). This indicates that the overall buildup of discordance in the
168 methylome partly, but not fully, contributes to these training-based biomarkers of aging.

169 **Epigenetic stochasticity and social and health variables at baseline**

170 We treated the age-regressed residuals of the DNAm readouts (i.e., age-deviation) as outcome
171 variables, and explored associations with baseline health and social variables. For rigor, we
172 performed the analysis with multiple regression models in the AS315 that included a sensitivity
173 analysis; for nominally significant associations (uncorrected $p \leq 0.05$), we tested for replication
174 in BA23 and AS311. Regarding the interpretation of age-deviation, as all the DNAm readouts
175 listed in **Table 3** increase in value with aging, a positive age-deviation indicates higher rate of
176 aging, and negative indicates slower rate of aging. Model 1 was a multivariable regression with
177 race/ethnicity, socioeconomic index (SE index, a combination of self-reported income and
178 education), smoking status, body weight index (BMI), and total weekly recreation energy
179 expenditure (TEXPWK; MET-hours/week) as predictors. To make the magnitudes of effects
180 comparable, **Fig 2a** displays the standardized regression estimates. For the three non-training-
181 based readouts (i.e., entropy, BivSEpiM, and BivCpGAvg), the most consistent association was
182 with SE index (**Table 3**). SE index also had significant negative associations with the
183 DunedinPACE and the PhenoAge clock (**Fig 2a**; unstandardized regression estimates in **Data**
184 **S1**.). A unit increase in SE index resulted in -0.004 lower entropy, -0.02 lower BivSEpiM, and -
185 0.005 slower DunedinPACE of aging (**Table 3**). The epigenetic clocks, which are in units of years
186 provide a more intuitive interpretation. For the PhenoAge, one unit increase in SE index

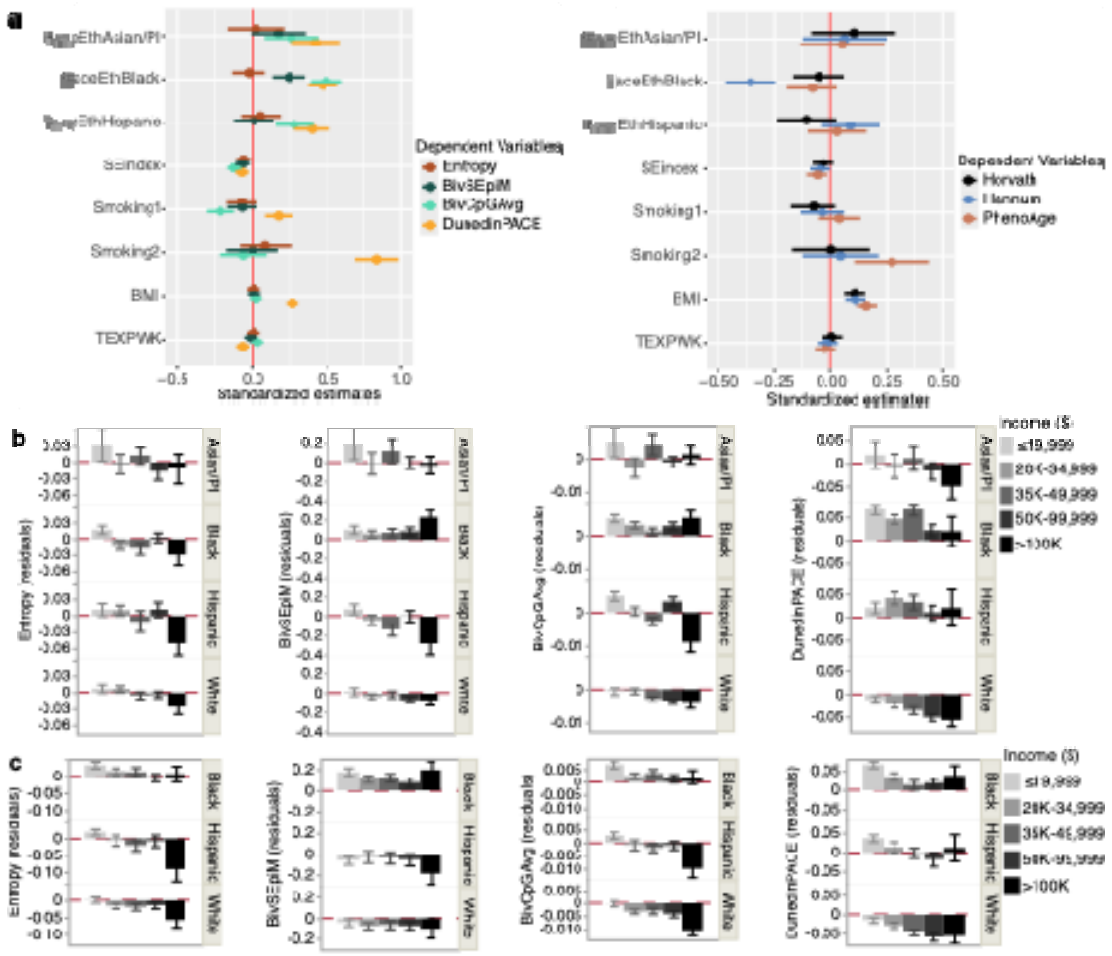


Fig 2. (a) Forest plots of standardized regression coefficients (95% confidence intervals) for the baseline predictor variables used in Model 1 for the AS315 group. The outcome variables are the age-adjusted DNAm readouts. Smoking1 and Smoking2 are past and present smokers, respectively, and regressions are related to never smokers. Estimates for race/ethnicity are relative to non-Hispanic White. **(b)** Bar graphs of age-adjusted entropy, BivSEpiM, BivAvg, and DunedinPACE plotted separately by self-reported race/ethnicity, and overlaid by income groups in AS315, and **(c)** in BA23 (bars shades denote income levels). Positive values indicate higher biological aging relative to chronological age. Error bars are standard errors.

resulted in roughly 0.25 years lower biological age relative to chronological age. We note that the bivalent readouts and DunedinPACE showed higher aging rates in all other race/ethnicity groups relative to White participants. Consistent with previous reports, the Hannum clock detected a significantly slower rate of aging among the Black participants.[42]

We computed these same readouts in BA23, and in the much smaller AS311. After excluding participants who were also part of the AS315, the negative associations between SE index and

194 bivalent readouts and entropy remained consistent in both data sets **(Figs S3** and **Data S2, S3**).
195 For example, in BA23, a unit increase in SE index resulted in 0.005 lower entropy, and 0.001
196 lower beta BivCpGAvg, a 0.009 slower DunedinPACE per chronological age, and a 0.46 years
197 younger biological age relative to chronological age for the PhenoAge (**Table 3**). To illustrate the
198 gradational effect of income on these measures of aging, we display bar plots of the non-
199 training epigenetic readouts and the DunedinPACE (after adjusting for chronological age) by
200 self-reported income for AS315 and BA23 (**Fig 2b, c**). The distribution and variance are shown as
201 violin and box plots in **Fig S4**. The potential age slowing effect of high income appears
202 particularly pronounced in the White population, whereas, for the Black participants, most of
203 the DNAm readouts had positive age-adjusted residual values even among the highest income
204 group (**Fig 2b**, **c**).

205 In Model 2, we included additional baseline health variables: history of hysterectomy, diabetes,
206 CVD, hypertension, and cancer; and in Model 3, we included all the variables in Model 2, plus
207 alcohol intake and the healthy eating index (2010 HEI) score to account for differences in
208 dietary intake (**Data S1–S3; Fig S5**). These models were implemented to verify that the negative
209 associations with SE index is robust. As sensitivity analysis, we repeated Model 2 in AS315 after
210 excluding all participants with DNA from follow-up annual visits (**Data S1**). All these showed
211 that the non-training based DNAm readouts, and the DunedinPACE and PhenoAge have
212 negative associations with SE index. The entropy and bivalent readouts were not significantly
213 associated with baseline health variables aside from slightly higher entropy for women who had
214 undergone hysterectomy in AS315 and BA23 (but not replicated in AS311; **Fig S5**). The training-
215 based biomarkers were more strongly related to baseline health and lifestyle variables. The

216 DunedinPACE and PhenoAge showed the expected age-accelerating effect of smoking, and the
217 DunedinPACE was slowed by higher total energy expenditure from recreational activity (MET-
218 hours/week) and higher HEI score. Higher BMI was positively associated with all the training-
219 based biomarkers, but not with entropy, BivSEpiM, or BivCpGAvg. In AS315, the BivCpGAvg
220 showed an unexpected significant negative regression estimate for prior smokers relative to
221 never smokers, but no difference between never and current smokers was observed; however,
222 this was not replicated in BA23 or AS311.

223 **Social disparity and methylation at the chromatin states**

224 To examine if the association with SE index was specific to the polycomb and bivalent CpGs, we
225 performed the Models 3 regressions with the age-adjusted mean methylation values for the
226 remaining 12 chromatin states in the AS315 and BA23 datasets. At a p threshold of 0.004
227 (Bonferroni corrected alpha = 0.05 for 12 tests), only the polycomb repressed state, ReprPC,
228 had significant negative association with SE index in AS315 and this was replicated in BA23 (**Fig**
229 **3a**; **Table S1**). Notably, like the bivalent states, ReprPC is a PRC2 targeted domain, and these
230 sites also have typically low methylation when young and accrue methylation with aging (all
231 plot to top left quadrant of **Fig 1e**). This pattern for the bivalent and polycomb sites suggests
232 that higher SE index is associated with a lower and presumably more "youthful" methylation
233 state. In both AS315 and BA23, higher SE index also had negative associations with methylation
234 levels at the active chromatin states, TssA and TssAFlnk (transcription start sites at or flanking
235 actively transcribed genes), which are also regions with low average methylation. However, at
236 the Bonferroni corrected alpha = 0.05, the negative association of TssA and TssAFlnk with SE

index was significant only in BA23 (**Fig 3a; Table S1**).

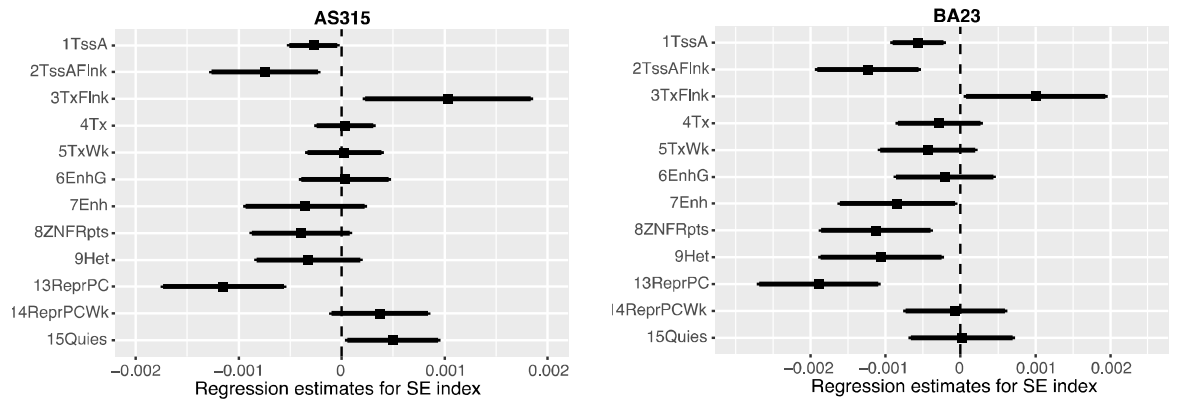


Fig 3. Forest plots of regression coefficients (95% confidence intervals) for SE index as predictor variable. Dependent variables are age-adjusted mean methylation at chromatin states (each fitted separately) after adjustment for other baseline variables (Model 3). One standard deviation higher SE index was associated with lower methylation at TssA, TssAFlnk, and ReprPC, and higher methylation at TxFlnk in both datasets.

Another chromatin state, TxFlnk (downstream of transcribed genes), had a positive association with SE index in both AS315 and BA23; however, the association was not significant after multiple test correction. Nonetheless, it is notable that in contrast to the bivalent states, CpGs in TxFlnk typically exist in hemi-methylated states (beta close to 0.5), loses methylation with aging, and contribute negatively to entropy (r=-0.41, **Fig 1d**). The positive association between TxFlnk and SE index suggests that higher SE index may be related to methylation maintenance at these CpGs.

DNAm readouts and all-cause mortality

Next, we tested whether these DNAm readouts predict life expectancy. Due to the low sample numbers, we performed this analysis only for the Black, White and Hispanic participants for AS315 and BA23. The PhenoAge clock is already known to be predictive of lifespan[3,27], and for

250 the present work, we examined only the non-training measures, and the DunedinPACE as a
251 training-based reference. We first performed a race/ethnicity stratified Cox regression for
252 censored survival time, and the DNAm readouts as primary predictors (entropy, BivSEpiM,
253 BivAvg, DunedinPACE; each fitted separately), and with SE index and baseline health variables
254 and the Clinical Trial (CT) dietary modification and hormone replacement arms as covariates. In
255 both AS315 and BA23, only the DunedinPACE was a significant predictor of lifespan, and one SD
256 increase in the pace of aging increased the risk of death by about 20% with hazard ratio (HR) of
257 1.23 (95% confidence interval CI = 1.13–1.33) in AS315 and HR = 1.16 (1.07–1.25) in BA23 (**Fig**
258 **4a, 4b**; **Table S2**). For SE index, one SD increase was associated with a modest decreased risk of
259 mortality of HR of 0.95 (CI = 0.87–1) in AS315, and HR 0.89 (0.83—0.96) in BA23. We further
260 examined if SE index and DunedinPACE were associated with lifespan when race/ethnicity are
261 analyzed separately (**Fig S6**; **Table S2**). In both AS315 and BA23, higher DunedinPACE increased
262 risk of mortality in all groups, and the added lifespan privilege of higher income was seen
263 mainly among the White participants.

264 We applied the same Cox regression to examine whether the average methylation levels of the
265 12 remaining chromatin states are predictive of all-cause mortality. Generally, higher
266 methylation at chromatin states that were negatively correlated with chronological age were
267 associated with reduced mortality risk (**Table S3**). This was nominally significant in both AS315
268 and BA23 for enhancer (Enh) CpGs (**Fig 4c**; **Table S3**). Higher mean methylation of Enh states
269 predicted lower mortality risk with HR of 0.90 (0.84—0.98) in AS315, and HR of 0.93 (0.88—
270 0.99) in BA23. Notably, the mean methylation of Enh regions also had the strongest negative
271 correlation with chronological age **(Fig 1f**). While the association with mortality risk is not

272 significant after Bonferroni correction, the overall pattern suggests that maintaining higher

273 methylation at these regulatory CpGs that typically lose methylation with aging could be related

274 to longer life expectancy.

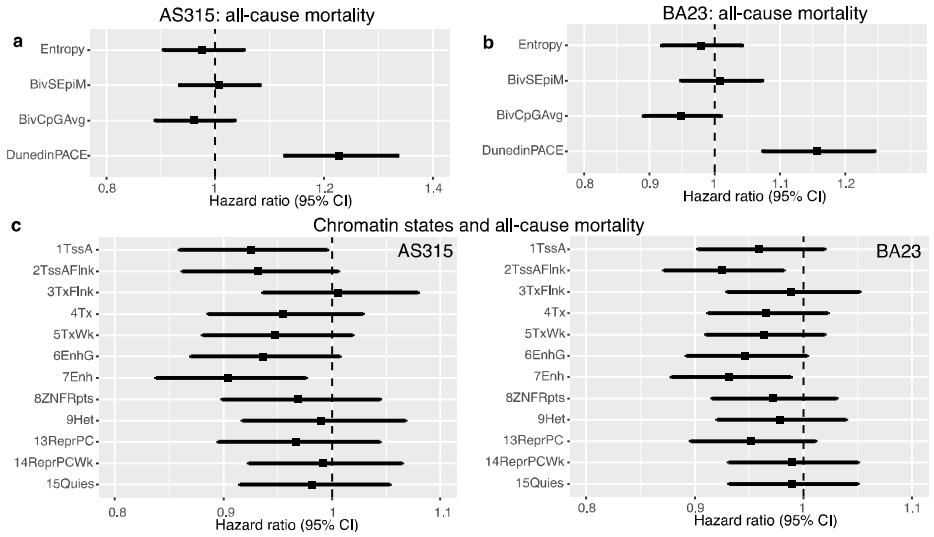


Fig 4. Effects of DNAm readouts on all-cause mortality risk in **(a)** AS315 and **(b)** BA23 based on race/ethnicity stratified multivariable Cox regression. The predictor variables (entropy, BivSEpiM, BivCpGAvg, and DunedinPACE) were fitted separately and the forest plots displays the hazard ratios associated with each. Only the DunedinPACE is predictive of survival time in both datasets. **(c)** HR associated with mean methylation at chromatin states. Higher methylation at Enh (enhancers) has a nominally significant association with reduced risk of death in both AS315 and BA23.

275

276 **DISCUSSION**

277 To summarize: **(1)** We found that nearly all the large-scale methylation patterns (i.e., entropy,

278 overall methylation averages of chromatin states, and outlier counts at the bivalent sites)

279 showed some degree of change with chronological age among postmenopausal women and the

280 typical pattern was for the low methylation CpGs to gain, and high methylation CpGs to lose

281 methylation with aging. **(2)** Age-dependent changes in the bivalent and polycomb targeted sites

282 contributed to the increase in epigenetic entropy. **(3)** The bivalent and polycomb CpGs were

283 consistently associated with SE index and varied between self-reported race/ethnicity. **(4)**

284 However, the untrained readouts were not significantly associated with baseline health
285 variables or predictive of mortality risk. **(5)** In contrast, biomarkers trained on health-related
286 parameters (i.e., the DunedinPACE and PhenoAge) were significantly associated with baseline
287 variables such as BMI, smoking, and energy expenditure; furthermore, a higher pace of aging
288 estimated by the DunedinPACE was associated with a higher risk for all-cause mortality. **(6)**
289 Among the chromatin states, we found evidence that maintaining methylation levels at
290 enhancer CpGs could be associated with a lower risk of death.

291 Of the DNAm readouts, the most "global" measurement we computed was the methylome-
292 wide entropy that was defined using nearly the full set of CpGs. This value estimates the overall
293 randomness and is related to the phenomenon of "epigenetic drift", the increase in
294 stochasticity and variance in DNAm with aging.[14] In cohorts with a wide age range, epigenetic
295 entropy has a strong positive correlation with chronological age.[1] Among the WHI participants,
296 the positive correlation between entropy and age was modest, albeit significant. This is possibly
297 because the WHI represents an older cohort, and the relationship between epigenetic aging
298 and chronological age is not perfectly linear and there is evidence that it begins to plateau
299 among older individuals.[9,43] Initially we had anticipated that health variables such as BMI and
300 history of cancer would show some association with epigenetic entropy. However, this was not
301 the case. Instead, the combination of self-reported education and income (what we refer to as
302 SE index) was the only variable at baseline that had a consistent inverse association with global
303 entropy in all three WHI datasets. Scoring high on the SE index was linked to lower age-adjusted
304 entropy. Since the increase in DNAm entropy will partly reflect the increase in cellular
305 heterogeneity with aging, we also attempted to account for blood cell heterogeneity by

306 including the DNAm-based estimated proportions of blood cell types, and in doing so, the link
307 between epigenetic entropy and SE index became stronger (see **Data S2–S3**). The lowest age-
308 adjusted entropy was seen among the White and Hispanic women at the highest income grade,
309 while the highest age-adjusted entropy was seen among the low-income minority groups. This
310 suggests that while higher income and the numerous psychosocial, environmental, and health
311 privileges of wealth could be protective against the buildup of epigenetic discordance with
312 aging, the impact of income is not the same across all racial groups. In fact, the effect is blunted
313 especially among Black women. This is interesting, yet not surprising as Black women who
314 achieve high education and socioeconomic status continue to experience chronic stressors over
315 their lifetime due systemic racism, discrimination, microaggressions and cultural
316 expectations.[44,45]

317 SE index also had a similar inverse relationship with methylation levels at bivalent CpGs, which
318 were the strongest positive correlates of methylome-wide entropy. Having a low SE index was
319 characterized by higher levels of stochastic epimutations and age-dependent gains at these
320 bivalent sites. Here, stochastic epimutation simply refers to the counts of outlier CpGs and this
321 is another way of quantifying the levels of epigenetic discordance.[40,41] Our finding is consistent
322 with previous work by Fiorito et al. [40] Theirs was a multi-cohort study that included both males
323 and females, and participants from a wide age-range. They also used education level as a proxy
324 of socioeconomic position, and found that methylome-wide stochastic epimutation was higher
325 among those with lower education.[40]

326 We must note that both education and income are readily recordable variables that are part of
327 an extremely complex social construct of ones status in society and referred to as
328 socioeconomic status.[46] The SE index therefore serves only as a global proxy that tags along
329 with a multitude of unmeasured variables such as differences in access to health care and
330 education, food and nutrition options, neighborhood and housing conditions, environmental
331 pollutants and toxins, chronic stress, etc. These exposures—the social and environmental
332 determinants of health—exert a strong influence on interrelated biological processes such as
333 the stress and glucocorticoid pathway, composition of circulating immune cells and
334 inflammatory state, metabolic health, etc. [47-49] Our results suggest that these large-scale non-
335 training based readouts from the epigenome are influenced by the larger socio-environmental
336 context, but lack the sensitivity to serve as predictors of specific health and lifestyle conditions
337 such as BMI, energy expenditure, smoking, and disease status.

338 In contrast to the global entropy and bivalent readouts, the widely used epigenetic clocks and
339 pace-of-aging biomarkers are derived from extensive training algorithms. For the initial training,
340 both the PhenoAge and DunedinPACE included blood biomarkers that are related to health and
341 correlated with aging (e.g., C-reactive protein, blood cell counts, albumin, cholesterol).[3,4,50] The
342 DunedinPACE also included waist-hip ratio, dental health, lung function, cardiorespiratory
343 fitness, etc. Notably, the DunedinPACE was trained on the New Zealand (NZ) based Dunedin
344 Longitudinal Cohort, which consists predominantly of European ancestry participants, with only
345 a very small minority of participants identifying as indigenous people of New Zealand.[4,50] It is
346 quite remarkable that a biomarker panel of aging developed on a NZ cohort can generalized to
347 the societal structure and inequities in the United States. The DunedinPACE also revealed a

348 potential “anti-aging” effect of higher income among the White women, but the effect of
349 income was not as pronounced among the minority women. Except for the small group of self-
350 identified Asians on the higher income scale, all other minority groups had positive pace of
351 aging according to the DunedinPACE. While the training population could partly contribute to
352 the differences between race/ethnic groups, there is evidence that DunedinPACE captures the
353 effects of adversity and negative experiences that is independent of genetic ancestry.[51] Having
354 a higher pace of aging likely has important health consequences, and in both AS315 and BA23,
355 higher pace of aging as measured by DunedinPACE predicted shorter lifespan independent of SE
356 index.

357 Segregating the CpGs by their predicted chromatin states and computing the average
358 methylation levels was a very broad stroke approach. However, we applied this data reduction
359 to determine the general patterns by which chromatin states contribute to the increase in
360 epigenetic entropy. The strongest positive correlates of entropy were the methylation levels at
361 the bivalent chromatin domains and regions bound by the PRC2 (TssBiv, BivFlnk, EnhBiv,
362 ReprPC). These sites mark repressed states with low average methylation, and have an age-
363 dependent drift towards higher methylation. The bivalent and polycomb target sites are highly
364 conserved regulatory regions, and are crucially involved in embryonic development and in
365 maintaining cellular identity and function.[13] Similar to the bivalent and polycomb domains,
366 CpGs at active promoters (TssA) also typically have low methylation levels, and in the WHI, the
367 mean methylation at TssA was positively correlated with entropy. Notably, the chromatin states
368 that had significant positive correlations with entropy (i.e., higher methylation related to higher
369 entropy) were all negatively associated with SE index. An interpretation we draw from this is

370 that higher SE index is associated with the age-dependent increase in discordance and drift at
371 these CpG sites. In contrast, the chromatin state with the strongest negative correlation with
372 entropy was TxFlnk, and this was the only chromatin domain that showed a significant and
373 replicated positive association with SE index (means higher SE index related to higher
374 methylation levels at TxFlnk). TxFlnk marks CpGs located at 3' and 5' ends of transcribed genes,
375 and the mean methylation of TxFlnk had a weak inverse correlation with age (i.e., methylation
376 loss with aging), and maintaining a higher methylation reduced the level of entropy. The
377 positive association with SE index suggests that high SE index could help maintain a higher
378 methylation level at these CpGs, and thereby contribute to lower entropy.

379 Intriguingly, when it came to predicting survival time, the general trend was for higher
380 methylation at active chromatin states (e.g., TssA, TssAFlnk, EnhG, and Enh) to be linked to
381 lower risk of all-cause mortality, and this reached statistical significance for Enh. Except for
382 TssA, which was uncorrelated with chronological age, all these chromatin states tended to have
383 a negative correlation with aging, and this was strongest for Enh. Enhancers are important
384 regulatory elements, and the ENCODE 15-states model divides them into three categories
385 based on histone marks: bivalent enhancers in repressed sites (EnhBiv) that have low
386 methylation, and enhancers in genic (EnhG) and non-genic (Enh) sites.[35] Both Enh and EnhG
387 mark actively transcribed regions, and other studies have also shown that these have higher
388 methylation and lose methylation with aging.[52-54] Taken together, our results suggest that being
389 able to maintain higher methylation levels at these enhancer CpGs may extend lifespan. To our
390 knowledge, this is the first time the global average methylation of Enh CpGs has been
391 associated with all-cause mortality in humans. There is however evidence from mouse studies

392 that interventions such as caloric restriction, which extends lifespan, suppress the age
393 associated methylation loss at enhancer CpGs.[52] Furthermore, in humans, aberrant methylation
394 at enhancer CpGs have been implicated in cancer progression and metastasis.[55] The work by
395 Cole et al.[55] also showed that higher methylation at the enhancer CpG of the metastasis gene
396 *KIT* increased survival time among cancer patients. This direction of association is therefore in
397 agreement with our observation that higher methylation at Enh is associated with a decreased
398 risk of all-cause mortality.

399 In the present work, we only examined time to all-cause mortality but did not perform a more
400 detailed analysis to test whether these DNAm readouts predict time to adjudicated diseases or
401 disease-related deaths. Given the links between enhancer methylation and cancer progression
402 and mortality, this will be particularly relevant in a follow-up analyses. We also acknowledge
403 that the results we present are all based on older postmenopausal women (age 50-79 at
404 baseline), and the significant associations between the bivalent and polycomb sites and SE
405 index may represent the effects of lifelong exposures to social inequities and weathering.
406 Further work is needed to examine whether the link between SE index and the bivalent CpGs is
407 also seen in younger populations. For the DunedinPACE, there is strong and growing evidence
408 that it serves as a sensitive biomarker of the age-accelerating effects of social inequities and
409 discrimination in a wide age range, and in both males and females.[28,29,56-58] Based on this corpus
410 of work, we also anticipate that the association between the social variables and the non-
411 training based DNAm readouts present in this work will also be generalizable, but that is yet to
412 be tested. The link between enhancers and lifespan also needs further replication. To facilitate

413 such follow-up studies, we have provided all the computational codes, and these can be easily
414 computed from existing datasets and can be adapted to the newer DNAm microarray datasets.

415 In conclusion, we have presented alternative ways to quantify aspects of epigenetic aging that
416 are based on epigenetic entropy and stochasticity, and have related the large-scale methylation
417 features to aging, as well as to extrinsic variables, and potentially to lifespan. The presented
418 work highlights the deep links between the larger social environment and the aging of the
419 epigenome and provides evidence that methylation at active chromatin states could be related
420 to life expectancy.

421 **METHODS**

422 **WHI DNA methylation datasets and sample inclusion**

423 The WHI is a multicenter long-term prospective study of postmenopausal women that began in
424 1993.[18-20] Women between the ages 50–79 were recruited in 1993–1998. Ancillary studies have
425 generated genomic data, and the present work uses the genome-wide DNA methylation
426 dataset by the three ancillary studies : (1) WHI-EMPC/AS315,[26,31] (2) BA23,[21,32] and (3) AS311.
427 These 3 datasets have been previously analyzed jointly as part of large meta-studies, and
428 detailed descriptions can be found in these publications.[3,59-61]

429 The EMPC study (Epigenetic Mechanisms of PM-mediated CVD Risk; AS315) randomly selected
430 2200 participants from within the Clinical Trial arm and measured DNA methylation at SV for
431 the majority or at annual visit 3 or 6, and with repeated measurements for a subset of the
432 samples [31]. We used only the first sampling for the present work, and after excluding eight

433 samples with low methylation calls, the primary analyses were done for 2192 AS315
434 participants. The statistical analysis for AS315 was repeated in the 1396 SV samples (**Fig S1**).
435 BA23 was designed as a case/control study of risk for future CHD. We received epigenetic data
436 for 2107 BA23 participants collected at SV, and after excluding samples that overlap with WHI-
437 AS315, we performed the analyses in 1989 participants. AS311 is a matched case-control study
438 of bladder cancer, and like BA23, used blood DNA collected at SV and prior to disease diagnosis.
439 We received 882 array data; after excluding an array with too many missing values (~262,970
440 CpGs had NA values) and samples that overlap with AS315, we performed analyses in 868
441 participants.

442 **Computing entropy, chromatin annotations, and estimates of biological aging**

443 All analyses were done using the processed data that were provided in NetCDF format. For
444 uniformity, we used the beta-mixture quantile (BMIQ) normalized data, which applied the same
445 data quality control and processing steps implemented by WHI-EMPC [31]. The R package "ncdf4"
446 was used to interface with the data matrices [62]. Before any downstream analyses, we first
447 counted how many CpGs had missing values per sample, and excluded one sample from AS311,
448 and eight samples from BA23 as those had unusually high missingness (>50%).

449 For methylome-wide discretized entropy calculation, we used the same method described in a
450 prior study.[16] The optimal number of bins was estimated to be 50 using the Freedman-Diaconis
451 rule [63]. The R codes used for each step are in supplemental file **DataS5**. We then used the R
452 package "entropy" [64,65] to compute the discretized entropy value for each participant and this
453 was scaled to a number between 0–1. Example density plots for the beta-values were

454 generated in R with ggplot2 [66] for ten participants of AS315; these were all non-Hispanic Whites
455 and five had low entropy, and five had high entropy.

456 To annotate for chromatin states, we downloaded the archived REMCChromHMM file from the
457 Zhou github page [38,67]. This contains the 15-states chromHMM consensus chromatin
458 annotations for each of the ~480,000 CpGs in the 450K platform [35]. These annotations were
459 matched by CpG probe IDs to the WHI BMIQ data matrices, and average methylation levels for
460 each chromatin state was computed for each participant. As an alternate measure of
461 randomness in the epigenetic data, we counted CpGs that were outliers in the three bivalent
462 states to compute epigenetic stochasticity: bivalent/poised transcription start site (TssBiv),
463 flanking bivalent TSS (BivFlnk), and bivalent enhancer (EnhBiv). While the entropy measures the
464 shift in the genome-wide methylome landscape for each participant and is not relative to the
465 study cohort, the outlier counts are estimates of how many CpGs are outliers for an individual
466 relative to the study cohort. We adapted the outlier detection method described in Gentilini
467 but counted only the positive outliers as these sites gain methylation with aging and a higher
468 beta-value is presumed to represent a more aged methylome, i.e., beta-values over 3-times the
469 interquartile range of the 3rd quartile [40,41]. The codes used to compute the bivalent stochastic
470 epimutations (BivSEpiM) are in **Data S4**.

471 For the first-generation epigenetic clocks (i.e., Horvath and Hannum), we used the pre-
472 computed clocks that were shared by the WHI. Specifically, we used the linear transformed
473 Horvath and Hannum DNAm age predictions [1,2]. We computed the PhenoAge (aka, Levine2018)
474 and the DunedinPACE using the dnaMethyAge R package [3,4,68,69].

475 **Statistical analyses**

476 Pearson correlations were used for bivariate comparisons of chronological age with the DNAm
477 readouts. The pair-wise correlation plots were generated using the corrplot R package [70], and
478 display significance threshold was set at p=0.001. For epigenetic clocks, which correlate
479 strongly with chronological age, the deviation of the predicted biological age from chronological
480 age (referred to as age-acceleration or age-deviation) is computed as the residuals of predicted
481 age regressed on chronological age [2]. As most of the DNAm readouts, and the DunedinPACE
482 also show varying levels of correlation with chronological age, we used the residuals to make
483 the values independent of chronological age.

484 We used the linear regression lm() function in R to test associations between the age-residuals
485 of the DNAm readouts and baseline variables. These baseline variables included BMI computed
486 from weight and height at baseline (kg/m^2), medical history at baseline, and personal habits
487 including smoking status (0 = never smoked; 1 = past smoker; and 2 = present smoker), and
488 recreational energy expenditure computed as measured by the TEXPWK variable (MET-
489 hours/week). TEXPWK was computed from questionnaire-based reports of recreational physical
490 activity that included walking, mild, moderate and strenuous physical activity in kcal/week/kg.[71]
491 Among the demographic variables, WHI collected self-reported education and income at time
492 of screening and are presented on an ascending scale from 1–11 for education, and 1–8 for
493 income (higher numbers represent higher education or income). We derived a single variable,
494 SEindex as the average of the two after excluding unreported or missing data.

495 We applied a simple model, Model 1: lm(y_i ~ race/ethnicity + SEindex + Smoking + BMI +
496 TEXPWK), and y_i is each of the DNAm readouts fitted independently. We limited the analysis to
497 participants who identified as non-Hispanic White (White), non-Hispanic Black (Black), Hispanic,
498 and Asian or Pacific Islander (Asian/PI) in AS315, and White, Black and Hispanic for BA23 and
499 AS311. Model 2 was the same as Model 1 and included additional medical history terms
500 collected at baseline: hysterectomy, diabetes, cardiovascular disease (CVD), hypertension, and
501 cancer ever. An expanded version of Model 1 included blood cell proportions (lymphocytes,
502 monocytes, granulocytes) that were indirectly estimated from the DNAm data. The DNAm data
503 for BA23 and AS311 were from baseline SV, and majority of the AS315 blood DNA was from the
504 SV prior to assignment of a study arm (n=1396). Model 2 was repeated for AS315 after
505 excluding the 633 blood DNA that were sampled at visit year 3, and 163 at visit year 4. Model 3
506 included all the variables in Model 2, plus the alcohol intake variable, and 2010 HEI total
507 score.[72,73] The lm() R codes and details on the variables are provided in **Data S4**.

508 We performed Cox regressions to test predictive value of the DNAm readouts using the 2022
509 versions of All Discovered Deaths, and Adjudicated Outcomes. This was done using the survival
510 R package [74,75]. For all-cause mortality, we used the variable DEATHALL (all discovered deaths
511 including from National Death Index) in conjunction with ENDFOLLOWALLDY (days enrollment
512 to end of follow up including uncensored death). Covariates for Cox regression were baseline
513 variables age, smoking status, BMI, diabetes, cancer, SE index, hormone therapy arm, dietary
514 modification arm, alcohol intake, 2010 HEI total score, and with race/ethnicity in the strata()
515 term. The coxph() R codes are provided in **DataS4**.

516 **Data availability:** All data are available upon request through the Women's Health
517 Initiative Study (https://www.whi.org).

518 **Acknowledgement:** The WHI program is funded by the National Heart, Lung, and Blood
519 Institute, National Institutes of Health, U.S. Department of Health and Human Services through
520 contracts HHSN268201600018C, HHSN268201600001C, HHSN268201600002C,
521 HHSN268201600003C, and HHSN268201600004C. WHI EMPC (AS315) was supported by NIEHS
522 grant R01-ES020836 (Whitsel, Eric A and other PIs). WHI AS311 was supported by American
523 Cancer Society award 125299-RSG-13–100-01-CCE (Parveen Bhatti). WHI-BAA23 was supported
524 by NHLBI Broad Agency Announcement contract HHSN268201300006C.

525 **Author contribution:** KM: study design, analyses and interpretation, initial
526 manuscript preparation, review and editing. AD, YS, AHS, RC, FJT, RBW, JHF: part
527 of WHI writing group and contributed to manuscript reviewing, editing, and
528 interpretation. KCJ: WHI sponsor and contributed to manuscript reviewing,
529 editing, and interpretation. All authors approved final manuscript.

530

531 **REFERENCES.**

532 1 Hannum, G. *et al.* Genome-wide methylation profiles reveal quantitative views of
533 human aging rates. *Mol Cell* **49**, 359-367, doi:10.1016/j.molcel.2012.10.016 (2013).
534 2 Horvath, S. DNA methylation age of human tissues and cell types. *Genome Biol* **14**, R115,
535 doi:10.1186/gb-2013-14-10-r115 (2013).
536 3 Levine, M. E. *et al.* An epigenetic biomarker of aging for lifespan and healthspan. *Aging*
537 *(Albany NY)* **10**, 573-591, doi:10.18632/aging.101414 (2018).
538 4 Belsky, D. W. *et al.* DunedinPACE, a DNA methylation biomarker of the pace of aging.
539 *Elife* **11**, doi:10.7554/eLife.73420 (2022).
540 5 Lu, A. T. *et al.* DNA methylation GrimAge strongly predicts lifespan and healthspan.
541 *Aging (Albany NY)* **11**, 303-327, doi:10.18632/aging.101684 (2019).
542 6 Horvath, S. & Raj, K. DNA methylation-based biomarkers and the epigenetic clock theory
543 of ageing. *Nat Rev Genet* **19**, 371-384, doi:10.1038/s41576-018-0004-3 (2018).
544 7 Liu, Z. *et al.* Underlying features of epigenetic aging clocks in vivo and in vitro. *Aging Cell*
545 **19**, e13229, doi:10.1111/acel.13229 (2020).
546 8 Dabrowski, J. K. *et al.* Probabilistic inference of epigenetic age acceleration from cellular
547 dynamics. *bioRxiv*, 2023.2003.2001.530570, doi:10.1101/2023.03.01.530570 (2023).
548 9 Meyer, D. H. & Schumacher, B. Aging clocks based on accumulating stochastic variation.
549 *Nat Aging* **4**, 871-885, doi:10.1038/s43587-024-00619-x (2024).
550 10 Ikram, M. A. The use and misuse of 'biological aging' in health research. *Nat Med* **30**,
551 3045, doi:10.1038/s41591-024-03297-9 (2024).
552 11 Gems, D., Virk, R. S. & de Magalhaes, J. P. Epigenetic clocks and programmatic aging.
553 *Ageing Res Rev* **101**, 102546, doi:10.1016/j.arr.2024.102546 (2024).
554 12 Moqri, M. *et al.* PRC2-AgeIndex as a universal biomarker of aging and rejuvenation. *Nat*
555 *Commun* **15**, 5956, doi:10.1038/s41467-024-50098-2 (2024).
556 13 Schuettengruber, B., Bourbon, H. M., Di Croce, L. & Cavalli, G. Genome Regulation by
557 Polycomb and Trithorax: 70 Years and Counting. *Cell* **171**, 34-57,
558 doi:10.1016/j.cell.2017.08.002 (2017).
559 14 Teschendorff, A. E. On epigenetic stochasticity, entropy and cancer risk. *Philos Trans R*
560 *Soc Lond B Biol Sci* **379**, 20230054, doi:10.1098/rstb.2023.0054 (2024).
561 15 Jenkinson, G., Pujadas, E., Goutsias, J. & Feinberg, A. P. Potential energy landscapes
562 identify the information-theoretic nature of the epigenome. *Nat Genet* **49**, 719-729,
563 doi:10.1038/ng.3811 (2017).
564 16 Mozhui, K. *et al.* Genetic loci and metabolic states associated with murine epigenetic
565 aging. *Elife* **11**, doi:10.7554/eLife.75244 (2022).
566 17 Sziraki, A., Tyshkovskiy, A. & Gladyshev, V. N. Global remodeling of the mouse DNA
567 methylome during aging and in response to calorie restriction. *Aging Cell* **17**, e12738,
568 doi:10.1111/acel.12738 (2018).
569 18 NIH. Design of the Women's Health Initiative clinical trial and observational study. The
570 Women's Health Initiative Study Group. *Control Clin Trials* **19**, 61-109,
571 doi:10.1016/s0197-2456(97)00078-0 (1998).

572 19 Anderson, G. L. *et al.* Implementation of the Women's Health Initiative study design.
573 *Ann Epidemiol* **13**, S5-17, doi:10.1016/s1047-2797(03)00043-7 (2003).
574 20 WHI. <www.whi.org>
575 21 Levine, M. E. *et al.* DNA methylation age of blood predicts future onset of lung cancer in
576 the women's health initiative. *Aging (Albany NY)* **7**, 690-700, doi:10.18632/aging.100809
577 (2015).
578 22 Carroll, J. E. *et al.* Epigenetic Aging and Immune Senescence in Women With Insomnia
579 Symptoms: Findings From the Women's Health Initiative Study. *Biol Psychiatry* **81**, 136-
580 144, doi:10.1016/j.biopsych.2016.07.008 (2017).
581 23 Shadyab, A. H. *et al.* Association of Epigenetic Age Acceleration With Incident Mild
582 Cognitive Impairment and Dementia Among Older Women. *J Gerontol A Biol Sci Med Sci*
583 **77**, 1239-1244, doi:10.1093/gerona/glab245 (2022).
584 24 Quach, A. *et al.* Epigenetic clock analysis of diet, exercise, education, and lifestyle
585 factors. *Aging (Albany NY)* **9**, 419-446, doi:10.18632/aging.101168 (2017).
586 25 Pottinger, T. D. *et al.* Association of cardiovascular health and epigenetic age
587 acceleration. *Clin Epigenetics* **13**, 42, doi:10.1186/s13148-021-01028-2 (2021).
588 26 Grant, C. D. *et al.* A longitudinal study of DNA methylation as a potential mediator of
589 age-related diabetes risk. *Geroscience* **39**, 475-489, doi:10.1007/s11357-017-0001-z
590 (2017).
591 27 Liu, Z. *et al.* The role of epigenetic aging in education and racial/ethnic mortality
592 disparities among older U.S. Women. *Psychoneuroendocrinology* **104**, 18-24,
593 doi:10.1016/j.psyneuen.2019.01.028 (2019).
594 28 Harris, K. M. *et al.* Sociodemographic and Lifestyle Factors and Epigenetic Aging in US
595 Young Adults: NIMHD Social Epigenomics Program. *JAMA Netw Open* **7**, e2427889,
596 doi:10.1001/jamanetworkopen.2024.27889 (2024).
597 29 Shen, B. *et al.* Association of Race and Poverty Status With DNA Methylation-Based Age.
598 *JAMA Netw Open* **6**, e236340, doi:10.1001/jamanetworkopen.2023.6340 (2023).
599 30 Geronimus, A. T. The weathering hypothesis and the health of African-American women
600 and infants: evidence and speculations. *Ethn Dis* **2**, 207-221 (1992).
601 31 Gondalia, R. *et al.* Methylome-wide association study provides evidence of particulate
602 matter air pollution-associated DNA methylation. *Environ Int* **132**, 104723,
603 doi:10.1016/j.envint.2019.03.071 (2019).
604 32 Levine, M. E. *et al.* Menopause accelerates biological aging. *Proc Natl Acad Sci U S A* **113**,
605 9327-9332, doi:10.1073/pnas.1604558113 (2016).
606 33 Jordahl, K. M. *et al.* Genome-Wide DNA Methylation in Prediagnostic Blood and Bladder
607 Cancer Risk in the Women's Health Initiative. *Cancer Epidemiol Biomarkers Prev* **27**, 689-
608 695, doi:10.1158/1055-9965.EPI-17-0951 (2018).
609 34 Teschendorff, A. E., West, J. & Beck, S. Age-associated epigenetic drift: implications, and
610 a case of epigenetic thrift? *Hum Mol Genet* **22**, R7-R15, doi:10.1093/hmg/ddt375 (2013).
611 35 Roadmap Epigenomics, C. *et al.* Integrative analysis of 111 reference human
612 epigenomes. *Nature* **518**, 317-330, doi:10.1038/nature14248 (2015).
613 36 Ernst, J. & Kellis, M. Chromatin-state discovery and genome annotation with
614 ChromHMM. *Nat Protoc* **12**, 2478-2492, doi:10.1038/nprot.2017.124 (2017).

615 37 Kaur, D. *et al.* Comprehensive Evaluation of The Infinium Human MethylationEPIC v2
616 BeadChip. *Epigenetics Commun* **3**, doi:10.1186/s43682-023-00021-5 (2023).
617 38 Zhou, W., Laird, P. W. & Shen, H. Comprehensive characterization, annotation and
618 innovative use of Infinium DNA methylation BeadChip probes. *Nucleic Acids Res* **45**, e22,
619 doi:10.1093/nar/gkw967 (2017).
620 39 Rakyan, V. K. *et al.* Human aging-associated DNA hypermethylation occurs preferentially
621 at bivalent chromatin domains. *Genome Res* **20**, 434-439, doi:10.1101/gr.103101.109
622 (2010).
623 40 Fiorito, G. *et al.* Socioeconomic position, lifestyle habits and biomarkers of epigenetic
624 aging: a multi-cohort analysis. *Aging (Albany NY)* **11**, 2045-2070,
625 doi:10.18632/aging.101900 (2019).
626 41 Gentilini, D. *et al.* Stochastic epigenetic mutations (DNA methylation) increase
627 exponentially in human aging and correlate with X chromosome inactivation skewing in
628 females. *Aging (Albany NY)* **7**, 568-578, doi:10.18632/aging.100792 (2015).
629 42 Horvath, S. *et al.* An epigenetic clock analysis of race/ethnicity, sex, and coronary heart
630 disease. *Genome Biol* **17**, 171, doi:10.1186/s13059-016-1030-0 (2016).
631 43 Snir, S., Farrell, C. & Pellegrini, M. Human epigenetic ageing is logarithmic with time
632 across the entire lifespan. *Epigenetics* **14**, 912-926,
633 doi:10.1080/15592294.2019.1623634 (2019).
634 44 Baker, T. A., Buchanan, N. T., Mingo, C. A., Roker, R. & Brown, C. S. Reconceptualizing
635 successful aging among black women and the relevance of the strong black woman
636 archetype. *Gerontologist* **55**, 51-57, doi:10.1093/geront/gnu105 (2015).
637 45 Henry, C. J. & Song, M. K. Use of the Strong Black Woman Construct in Research: An
638 Integrative Review. *ANS Adv Nurs Sci* **47**, E110-E120,
639 doi:10.1097/ANS.0000000000000501 (2024).
640 46 Antonoplis, S. Studying Socioeconomic Status: Conceptual Problems and an Alternative
641 Path Forward. *Perspect Psychol Sci* **18**, 275-292, doi:10.1177/17456916221093615
642 (2023).
643 47 Smarr, M. M. *et al.* Broadening the Environmental Lens to Include Social and Structural
644 Determinants of Women's Health Disparities. *Environ Health Perspect* **132**, 15002,
645 doi:10.1289/EHP12996 (2024).
646 48 Palma-Gudiel, H., Fananas, L., Horvath, S. & Zannas, A. S. Psychosocial stress and
647 epigenetic aging. *Int Rev Neurobiol* **150**, 107-128, doi:10.1016/bs.irn.2019.10.020
648 (2020).
649 49 Simons, R. L. *et al.* Discrimination, segregation, and chronic inflammation: Testing the
650 weathering explanation for the poor health of Black Americans. *Dev Psychol* **54**, 1993-
651 2006, doi:10.1037/dev0000511 (2018).
652 50 Belsky, D. W. *et al.* Quantification of the pace of biological aging in humans through a
653 blood test, the DunedinPoAm DNA methylation algorithm. *Elife* **9**,
654 doi:10.7554/eLife.54870 (2020).
655 51 Bourassa, K. J. *et al.* Demographic characteristics and epigenetic biological aging among
656 post-9/11 veterans: Associations of DunedinPACE with sex, race, and age. *Psychiatry Res*
657 **336**, 115908, doi:10.1016/j.psychres.2024.115908 (2024).

658 52 Cole, J. J. *et al.* Diverse interventions that extend mouse lifespan suppress shared age-
659 associated epigenetic changes at critical gene regulatory regions. *Genome Biol* **18**, 58,
660 doi:10.1186/s13059-017-1185-3 (2017).
661 53 Jain, N. *et al.* DNA methylation correlates of chronological age in diverse human tissue
662 types. *Epigenetics Chromatin* **17**, 25, doi:10.1186/s13072-024-00546-6 (2024).
663 54 Slieker, R. C., Relton, C. L., Gaunt, T. R., Slagboom, P. E. & Heijmans, B. T. Age-related
664 DNA methylation changes are tissue-specific with ELOVL2 promoter methylation as
665 exception. *Epigenetics Chromatin* **11**, 25, doi:10.1186/s13072-018-0191-3 (2018).
666 55 Bell, R. E. *et al.* Enhancer methylation dynamics contribute to cancer plasticity and
667 patient mortality. *Genome Res* **26**, 601-611, doi:10.1101/gr.197194.115 (2016).
668 56 Schmitz, L. L. *et al.* The Socioeconomic Gradient in Epigenetic Ageing Clocks: Evidence
669 from the Multi-Ethnic Study of Atherosclerosis and the Health and Retirement Study.
670 *Epigenetics* **17**, 589-611, doi:10.1080/15592294.2021.1939479 (2022).
671 57 Raffington, L. *et al.* Associations of socioeconomic disparities with buccal DNA-
672 methylation measures of biological aging. *Clin Epigenetics* **15**, 70, doi:10.1186/s13148-
673 023-01489-7 (2023).
674 58 Cuevas, A. G. *et al.* Multi-discrimination exposure and biological aging: Results from the
675 midlife in the United States study. *Brain Behav Immun Health* **39**, 100774,
676 doi:10.1016/j.bbih.2024.100774 (2024).
677 59 Do, W. L. *et al.* Associations between DNA methylation and BMI vary by metabolic
678 health status: a potential link to disparate cardiovascular outcomes. *Clin Epigenetics* **13**,
679 230, doi:10.1186/s13148-021-01194-3 (2021).
680 60 Agha, G. *et al.* Blood Leukocyte DNA Methylation Predicts Risk of Future Myocardial
681 Infarction and Coronary Heart Disease. *Circulation* **140**, 645-657,
682 doi:10.1161/CIRCULATIONAHA.118.039357 (2019).
683 61 Skinner, H. G. *et al.* Stressful life events, social support, and epigenetic aging in the
684 Women's Health Initiative. *J Am Geriatr Soc* **72**, 349-360, doi:10.1111/jgs.18726 (2024).
685 62 Pierce, D. *ncdf4*, <https://CRAN.R-project.org/package=ncdf4> (2023).
686 63 Freedman, D. & Diaconis, P. On the histogram as a density estimator:L2 theory.
687 *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete* **57**, 453-476,
688 doi:10.1007/BF01025868 (1981).
689 64 Hausser, J. & Strimmer, K. Entropy Inference and the James-Stein Estimator, with
690 Application to Nonlinear Gene Association Networks. *Journal of Machine Learning*
691 *Research* **10**, 1469-1484 (2009).
692 65 Hausser, J. & Strimmer, K. *entropy*, <https://CRAN.R-project.org/package=entropy >
693 (2021).
694 66 Wickham, H. *ggplot2: Elegant Graphics for Data Analysis*. (Springer-Verlag 2016).
695 67 Zhou, W. *zhou-lab github*, <https://github.com/zhou-
696 lab/ARCHIVED_KYCG_knowledgebases_hg38> (
697 68 Wang, Y., Grant, O. A., Zhai, X., McDonald-Maier, K. D. & Schalkwyk, L. C. Insights into
698 ageing rates comparison across tissues from recalibrating cerebellum DNA methylation
699 clock. *Geroscience* **46**, 39-56, doi:10.1007/s11357-023-00871-w (2024).
700 69 yiluyucheng. *dnaMethyAge*, <https://github.com/yiluyucheng/dnaMethyAge> (

70 Wei, T. & Simko, V. *R package 'corrplot': Visualization of a Correlation Matrix. (Version 0.92)*, <https://github.com/taiyun/corrplot> (2021).

71 Meyer, A. M., Evenson, K. R., Morimoto, L., Siscovick, D. & White, E. Test-retest reliability of the Women's Health Initiative physical activity questionnaire. *Med Sci Sports Exerc* **41**, 530-538, doi:10.1249/MSS.0b013e31818ace55 (2009).

72 Guenther, P. M. *et al.* Update of the Healthy Eating Index: HEI-2010. *J Acad Nutr Diet* **113**, 569-580, doi:10.1016/j.jand.2012.12.016 (2013).

73 Prentice, R. L. *et al.* Mortality Associated with Healthy Eating Index Components and an Empirical-Scores Healthy Eating Index in a Cohort of Postmenopausal Women. *J Nutr* **152**, 2493-2504, doi:10.1093/jn/nxac068 (2022).

74 Therneau, T. M. & Grambsch, P. M. *Modeling Survival Data: Extending the Cox Model*. (Springer, 2000).

75 Therneau, T. M. *A Package for Survival Analysis in R*, <https://CRAN.R-project.org/package=survival> (2024).

717 **Main Tables**

718 **Table 1. Datasets and baseline characteristics**

	EMPC/AS315	BA23[1]	AS311[1]
N	2192	1989	868
Age ± SD (years)	64 ± 7	65 ± 7	65.6 ± 7
BMI ± SD	29.4 ± 5.9	29.9 ± 6.1	28.0 ± 6.1
TEXPWK[2] (MET-hrs/wk)	9.7 ± 12.3	9.9 ± 12.6	11.5 ± 12.5
WHI study arm			
Clinical Trial (CT)	2191	1546	412
Observational Study (OS)		443	456
Race/Ethnicity			
Asian/Pacific Islander	133		15
Hispanic	318	402	25
Native American	51		1
Non-Hispanic Black	558	635	68
Non-Hispanic White	1097	952	749
Other	35		10
Education			
No data	18	19	5
Less Than HS	82	96	7
High School or GED	517	539	188
Vocation Degree	263	274	99
Some College or Graduate	793	654	341
Post College Graduate	519	407	228
Income ($)			
No data	124	119	55
≤19,999	460	589	148
20K–34,999	541	499	192
35K–49,999	414	366	174
50K–99,999	531	335	235
≥100K	122	81	64
SE index (range)[3]	5.4 ± 1.6 (1–11)	5.1 ± 1.6 (1–10)	5.7 ± 1.5 (1–11)
Smoking			
No data	29	31	17
Never	1159	1048	389
Past	821	713	389
Current	183	197	73
CVD ever			
No data	223	140	52
No	1687	1572	633
Yes	282	277	183
Hypertension ever			

No data	18	24	1
No	1362	1092	558
Yes	812	873	309
Cancer ever			
No data	24	14	
No	2088	1867	868
Yes	80	108	0
Diabetes ever			
No data		3	
No	2018	1734	814
Yes	174	252	54
Hysterectomy ever			
No	1201	1037	525
Yes	991	952	343

[1]Excludes participants sampled in AS315. [2]Total energy expenditure from recreational activity (TEXPWK or MET-hours/week). [3]Computed as the average of education level (WHI EDUC variable ranging from 1 = "*Didn't go to school*" to 11= "*Doctoral degree*") and income (WHI INCOME variable ranging from 1 = "*Less than $10,000*" to 8 = "*150,000 or more*").

Table 2. Predicted chromatin states for the >450,000 CpG

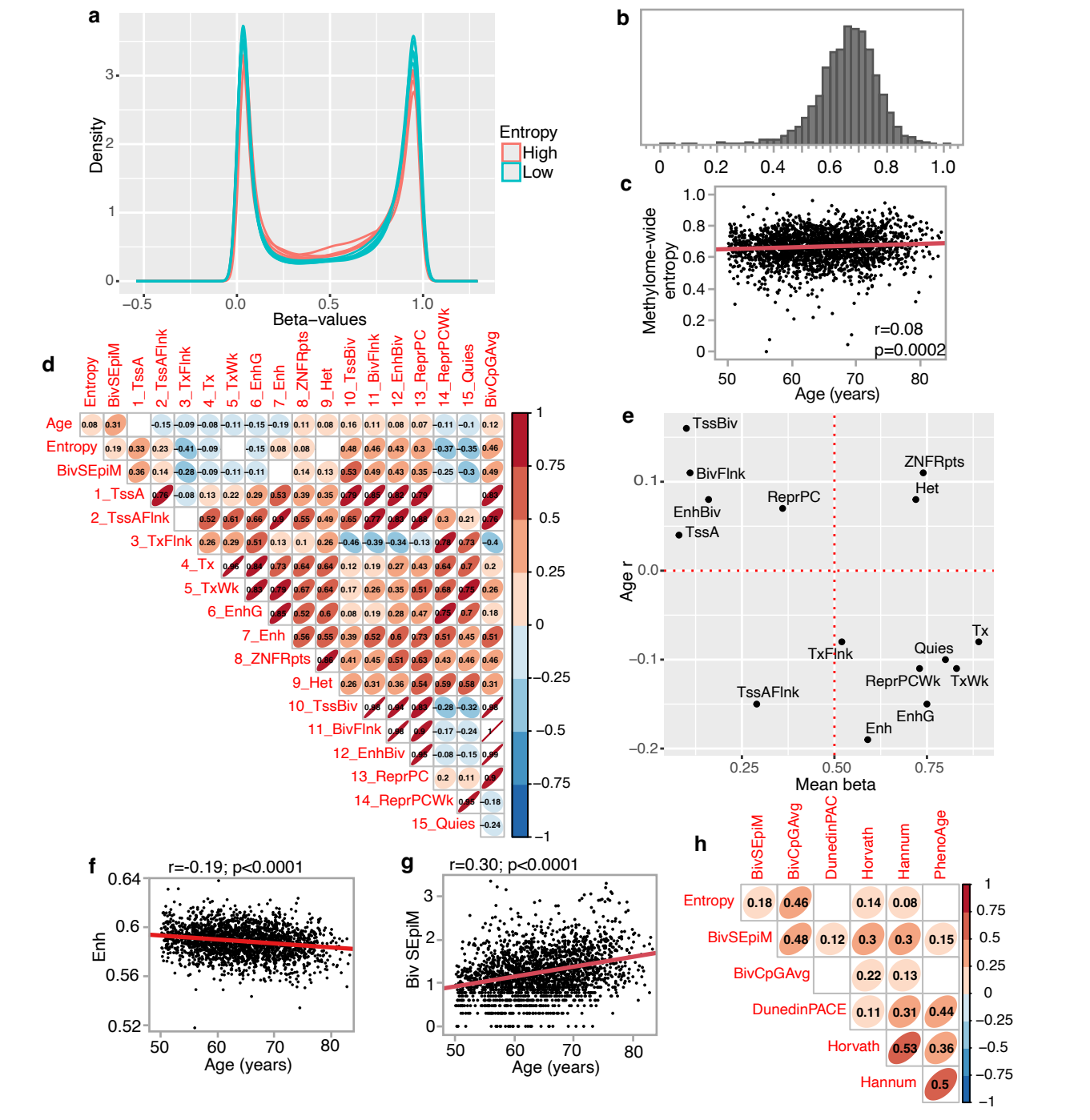
ChromHMM	Counts[1]	Repressed/ Active	Description	Non-coding conserved elements	Mean beta in AS315	r (age)[2]	p (age)[2]
1_TssA	124139	Active	Active TSS	Strongly enriched	0.08	0.04	0.0476
2_TssAFlnk	23774	Active	Flanking Active TSS	Strongly enriched	0.29	-0.15	<.0001
3_TxFlnk	1001	Active	Transcribed state at gene 5' and 3'	Moderately enriched	0.52	-0.08	<.0001
4_Tx	32162	Active	Strong transcription	No enrichment	0.89	-0.08	0.0001
5_TxWk	50018	Active	Weak transcription	No enrichment	0.83	-0.11	<.0001
6_EnhG	1545	Active	Genic enhancers	Moderately enriched	0.75	-0.15	<.0001
7_Enh	20980	Active	Enhancers	Strongly enriched	0.59	-0.19	<.0001
8_ZNF/Rpts	987	Active	ZNF genes & repeats	No enrichment	0.74	0.11	<.0001
9_Het	3468	Repressed	Heterochromatin	No enrichment	0.72	0.08	0.0002
10_TssBiv	6833	Repressed	Bivalent/Poised TSS	Strongly enriched	0.10	0.16	<.0001
11_BivFlnk	8290	Repressed	Flanking Bivalent TSS/Enh	Strongly enriched	0.11	0.11	<.0001
12_EnhBiv	4869	Repressed	Bivalent Enhancer	Strongly enriched	0.16	0.09	<.0001
13_ReprPC	32422	Repressed	Repressed PolyComb	Strongly enriched	0.36	0.07	0.0007
14_ReprPCWk	33551	Repressed	Weak Repressed PolyComb	Weakly enriched	0.73	-0.11	<.0001
15_Quies	136857	Repressed	Quiescent/Low	No enrichment	0.80	-0.10	<.0001
NA	1339						

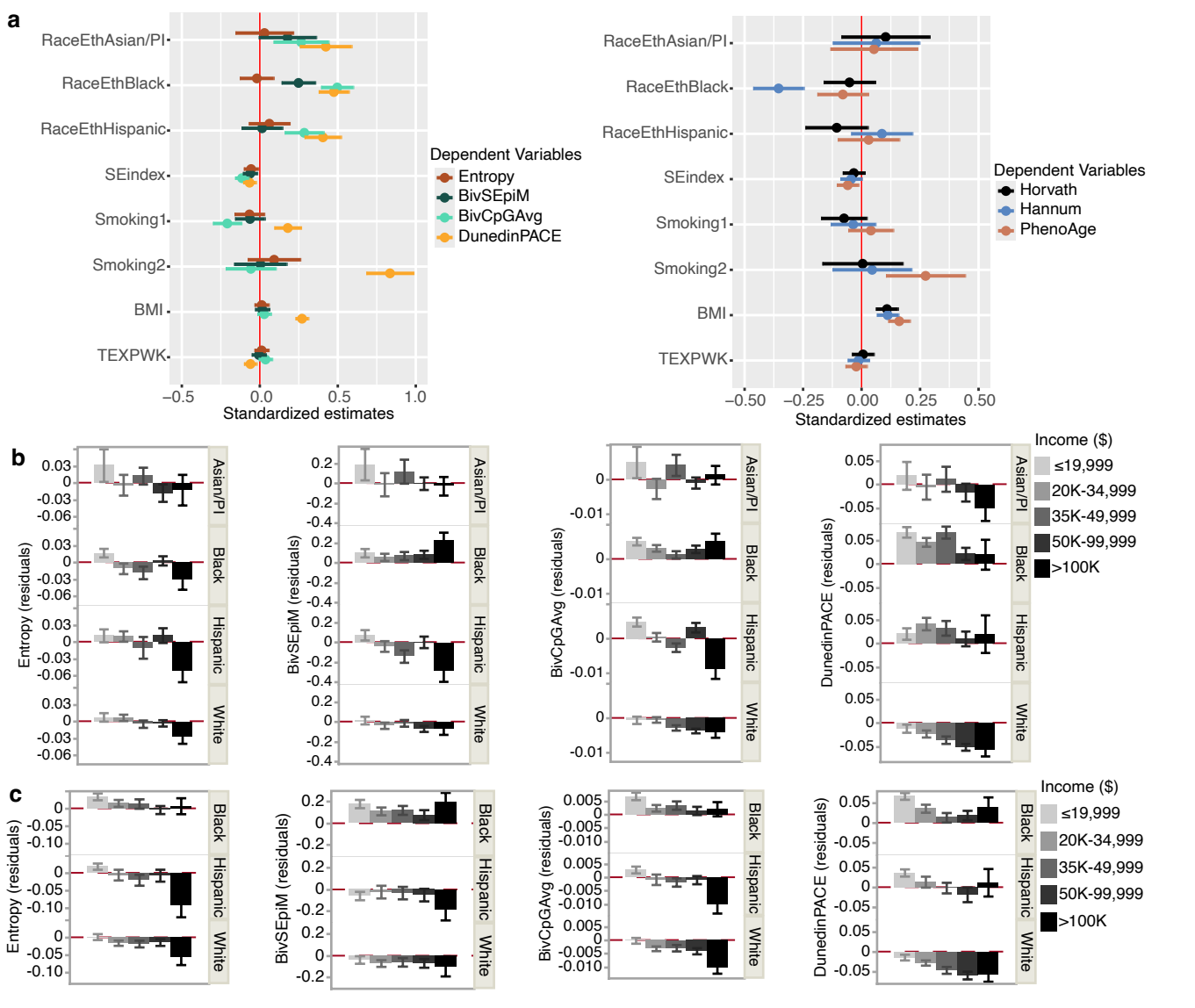
[1]Number of CpG probes on the Illumina 450K platform located in these predicted chromatin states. [2]Pearson correlation and p with chronological age in AS315

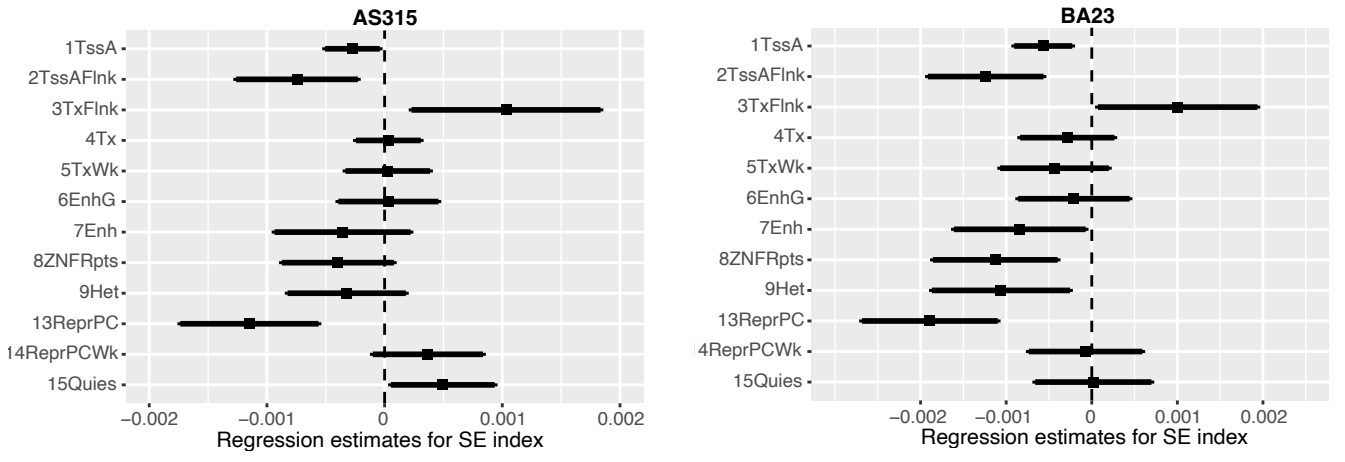
Table 3. Association between epigenetic readouts and socioeconomic index (SE index)

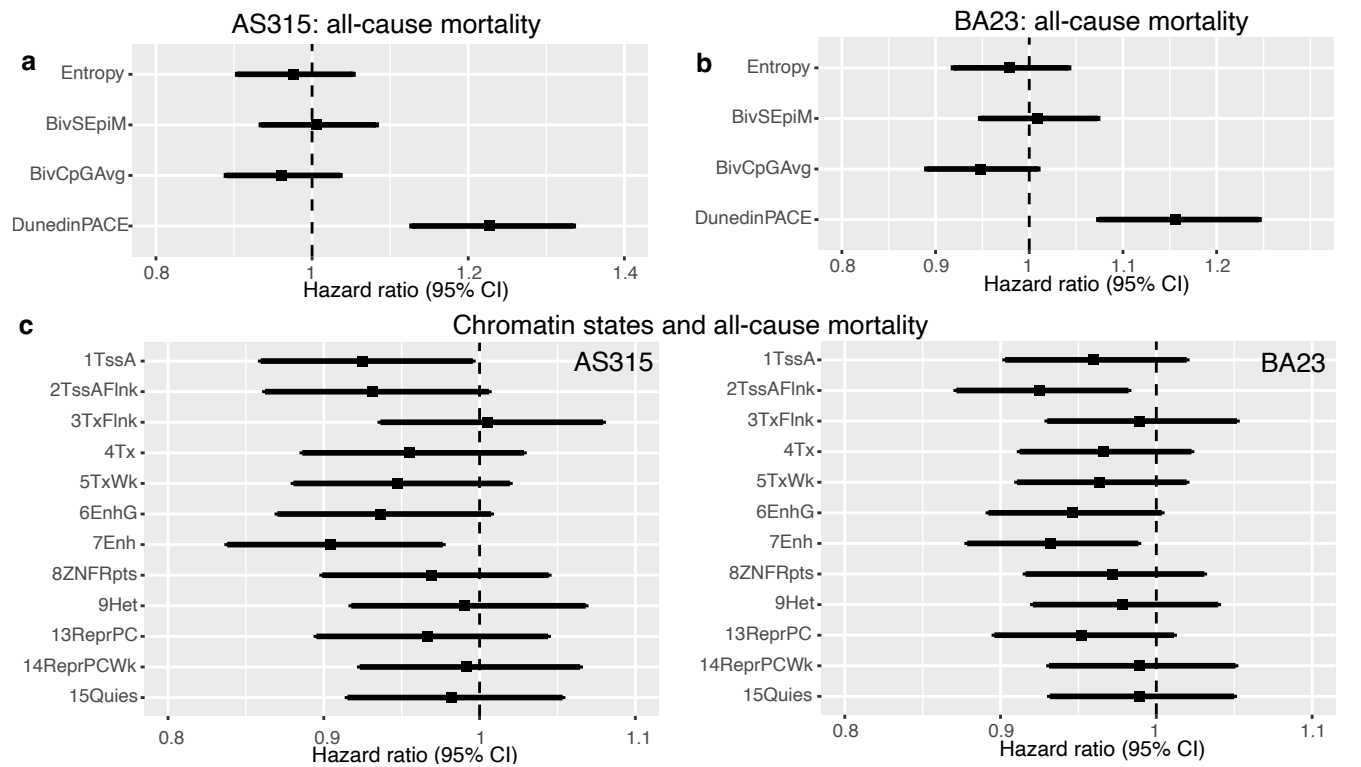
DNAm outcomes (age residuals)	AS315					BA23				
	Min	Max	SD	SE index coef[1]	p	Min	Max	SD	SE index coef[1]	p
DNAm Entropy	-0.655	0.344	0.10	-0.004	0.018	-0.548	0.452	0.13	-0.005	0.008
BivSEpiM (log10 outlier counts)	-1.437	2.312	0.50	-0.02	0.009	-1.453	2.373	0.51	-0.009	0.216
BivCpGAvg (beta)	-0.032	0.060	0.01	-0.001	4.20E-07	-0.059	0.098	0.02	-0.001	7.00E-07
DunedinPACE	-0.358	0.706	0.13	-0.005	0.002	-0.414	0.485	0.13	-0.009	4.91E-07
Horvath (years)	-27.71	93.40	6.22	-0.13	0.16	-29.16	54.18	6.64	-0.11	0.263
Hannum (years)	-21.07	48.44	5.35	-0.15	0.058	-23.86	31.59	5.83	-0.30	0.0007
PhenoAge (years)	-25.86	57.20	6.60	-0.25	0.012	-38.29	36.32	7.31	-0.46	3.30E-05

[1]Unstandardized regression estimates based on Model 1


a
Density
Beta−values
Entropy
High
Low
b
c
Methylome-wide entropy
Age (years)
r=0.08
p=0.0002
d
Entropy
BivSEpiM
1_TssA
2_TssAFlnk
3_TxFlnk
4_Tx
5_TxWk
6_EnhG
7_Enh
8_ZNFRpts
9_Het
10_TssBiv
11_BivFlnk
12_EnhBiv
13_ReprPC
14_ReprPCWk
15_Quies
BivCpGAvg
Age
e
Age r
Mean beta
TssBiv
BivFlnk
EnhBiv
TssA
ReprPC
ZNFRpts
Het
Tx
TxFlnk
Quies
ReprPCWk
TxWk
TssAFlnk
EnhG
Enh
f
r=-0.19; p<0.0001
Enh
Age (years)
g
r=0.30; p<0.0001
Biv SEpiM
Age (years)
h
BivSEpiM
BivCpGAvg
DunedinPAC
Horvath
Hannum
PhenoAge
Entropy
DunedinPACE


a
RaceEthAsian/PI
RaceEthBlack
RaceEthHispanic
SEindex
Smoking1
Smoking2
BMI
TEXPWK
−0.5
0.0
0.5
1.0
Standardized estimates
Dependent Variables
Entropy
BivSEpiM
BivCpGAvg
DunedinPACE
−0.50
−0.25
0.00
0.25
0.50
Horvath
Hannum
PhenoAge
b
Entropy (residuals)
BivSEpiM (residuals)
BivCpGAvg (residuals)
DunedinPACE (residuals)
Asian/PI
Black
Hispanic
White
Income ($)
≤19,999
20K-34,999
35K-49,999
50K-99,999
>100K
c


AS315
1TssA
2TssAFlnk
3TxFlnk
4Tx
5TxWk
6EnhG
7Enh
8ZNFRpts
9Het
13ReprPC
14ReprPCWk
15Quies
−0.002
−0.001
0
0.001
0.002
Regression estimates for SE index
BA23
1TssA
2TssAFlnk
3TxFlnk
4Tx
5TxWk
6EnhG
7Enh
8ZNFRpts
9Het
13ReprPC
4ReprPCWk
15Quies
−0.002
−0.001
0
0.001
0.002
Regression estimates for SE index


AS315: all-cause mortality
a
Entropy
BivSEpiM
BivCpGAvg
DunedinPACE
0.8
1
1.2
1.4
Hazard ratio (95% CI)
BA23: all-cause mortality
b
Entropy
BivSEpiM
BivCpGAvg
DunedinPACE
0.8
0.9
1
1.1
1.2
Hazard ratio (95% CI)
c
Chromatin states and all-cause mortality
AS315
1TssA
2TssAFlnk
3TxFlnk
4Tx
5TxWk
6EnhG
7Enh
8ZNFRpts
9Het
13ReprPC
14ReprPCWk
15Quies
0.8
0.9
1
1.1
Hazard ratio (95% CI)
BA23
1TssA
2TssAFlnk
3TxFlnk
4Tx
5TxWk
6EnhG
7Enh
8ZNFRpts
9Het
13ReprPC
14ReprPCWk
15Quies
0.8
0.9
1
1.1
Hazard ratio (95% CI)