PNAS NEXUS

PNAS Nexus, 2025, 4, pgaf121
https://doi.org/10.1093/pnasnexus/pgaf121
Advance access publication 16 April 2025
Research Report

Epigenetic biomarkers of socioeconomic status are associated with age-related chronic diseases and mortality in older adults

Lauren L. Schmitz[a,*], Lauren A. Opsasnick[b], Scott M. Ratliff[b], Jessica D. Faul[c], Wei Zhao[b,c], Timothy M. Hughes[d], Jingzhong Ding[d], Yongmei Liu[e] and Jennifer A. Smith[b,c,*]

[a]Robert M. La Follette School of Public Affairs, University of Wisconsin-Madison, Madison, WI 53706, USA
[b]Department of Epidemiology, School of Public Health, University of Michigan, Ann Arbor, MI 48109, USA
[c]Survey Research Center, Institute for Social Research, University of Michigan, Ann Arbor, MI 48104, USA
[d]Department of Gerontology and Geriatric Medicine, School of Medicine, Wake Forest University, Winston-Salem, NC 27157, USA
[e]Department of Medicine, Divisions of Cardiology and Neurology, Duke University Medical Center, Durham, NC 27710, USA
*To whom correspondence should be addressed: Email: llschmitz@wisc.edu (L.L.S.); Email: smjenn@umich.edu (J.A.S.)
Edited By Shibu Yooseph

Abstract

Later-life health is patterned by socioeconomic influences across the lifecourse. However, the pathways underlying the biological embedding of socioeconomic status (SES) and its consequences on downstream morbidity and mortality are not fully understood. Epigenetic markers like DNA methylation (DNAm) may be promising surrogates of underlying biological processes that can enhance our understanding of how SES shapes population health. Studies have shown that SES is associated with epigenetic aging measures, but few have examined relationships between early and later-life SES and DNAm sites across the epigenome. In this study, we trained and tested DNAm-based surrogates, or "biomarkers," of childhood and adult SES in two large, multiracial/ethnic samples of older adults—the Health and Retirement Study ($n = 3,527$) and the Multi-Ethnic Study of Atherosclerosis ($n = 1,182$). Both biomarkers were associated with downstream morbidity and mortality, and these associations persisted after controlling for measured SES, and in some cases, epigenetic aging clocks. Both childhood and adult SES biomarker CpG sites were enriched for genomic features that regulate gene expression (e.g. DNAse hypersensitivity sites and enhancers) and were implicated in prior epigenome-wide studies of inflammation, aging, and chronic disease. Distinct patterns also emerged between childhood CpGs and immune system dysregulation and adult CpGs and metabolic functioning, health behaviors, and cancer. Results suggest DNAm-based surrogate biomarkers of SES may be useful proxies for unmeasured social exposures that can augment our understanding of the biological mechanisms between social disadvantage and downstream health.

Keywords: socioeconomic status, DNA methylation, biomarkers, health, aging

Significance Statement

Information on DNA methylation (DNAm)—an epigenetic modification that plays a central role in regulating gene expression—is increasingly available in large epidemiological studies. Since DNAm is relatively stable but responsive to environmental influences, genome-wide signatures are promising surrogates or biomarkers of exposure that may both shed light on biological mechanisms between adverse environments and downstream health and/or act as proxies for unmeasured exposures. To better understand the biological embedding of social disadvantage, this study trained and tested DNAm-based surrogates of childhood and adult socioeconomic status (SES) in two US-based cohorts of older adults. Findings reveal distinct DNAm signatures of SES that connect social adversity across the lifecourse with dysregulated immune system responses, inflammatory pathways, poorer metabolic functioning, chronic diseases, and cancer.

Introduction

Health outcomes in later life are patterned in part by socioeconomic influences across the lifecourse (1, 2). However, the mechanistic pathways underlying the biological embedding of socioeconomic disadvantage and its consequences on morbidity and mortality are not fully understood. One potential pathway may be through modification of epigenetic mechanisms, such as DNA methylation (DNAm), that regulate gene expression. Alternatively, DNAm could be a marker of underlying biological response to environmental or contextual exposures. Studies

Competing Interest: The authors declare no conflicts of interests.
Received: May 21, 2024. **Accepted:** March 26, 2025

have shown that aspects of socioeconomic status (SES), including education, income, occupation, and neighborhood, are associated with both CpG-level differences in DNAm (3–7) and epigenetic aging measures (8–10). However, few studies have examined relationships between indicators of early and later-life SES and DNAm sites across the epigenome, particularly in diverse racial/ethnic samples (11).

DNAm is the covalent addition of a methyl group to the 5th carbon on a cytosine DNA base. In humans, DNAm occurs at sites on the genome where a cytosine is followed by a guanine, separated by a phosphodiester bond (CpG site). DNAm is the most well-studied type of epigenetic modification, and it is increasingly profiled in large epidemiological studies. DNAm patterns are tissue-specific, influenced by both genetics and environmental exposures, and are relatively stable compared to other cellular biomarkers such as gene expression (12, 13). However, DNAm patterns at some CpG sites are dynamic throughout life, responding to environmental stimuli, and have been proposed as a mechanism linking social exposures to adverse health outcomes. More recently, epigenome-wide differences in DNAm have been associated with current and/or past environmental exposures, suggesting global variation in DNAm could serve as a cumulative or time-specific biomarker of exposure, regardless of whether each specific CpG site is mechanistically involved in the disease process (14–16). Thus, epigenetic signatures may be useful proxies of exposure data in epidemiology studies, particularly when collection of these data are costly and/or subject to recall bias (15, 16).

DNAm-based surrogates or biomarkers are typically constructed by regressing the exposure or health outcome of interest on a set of CpG sites using a supervised machine-learning method (i.e. penalized regression) in a training sample. Currently, the most well-cited DNAm biomarkers are epigenetic aging measures or epigenetic "clocks" that are highly predictive of chronological age (17–24), biological age and mortality (25–27), and the rate of aging (28, 29). To date, over 30 epigenetic aging measures have been developed for adults and have been associated with numerous health, socioeconomic, and lifestyle factors (10, 30–36). These methods have been extended to calculate surrogate DNAm biomarkers for hundreds of health outcomes and disease-associated plasma proteins, including cardiovascular disease, diabetes, C-reactive protein, high-density lipoprotein cholesterol, and interleukin-6 (37–40). In the case of health behaviors like smoking and alcohol consumption, DNAm surrogates are typically more accurate than self-reports, thereby reducing misclassification and improving disease prediction and risk stratification (41). For past exposures where retrospective data collection is difficult or even impossible to collect, DNAm-based surrogates hold more promise as an exposure biomarker than other omics data because DNAm is more chemically stable than RNA or metabolites and does not degrade as easily with long-term storage (16, 42). For example, DNAm biomarkers have been successful in reflecting past exposures to prenatal smoking, prenatal alcohol, maternal diet, exposure to lead, pesticides, and other toxins (43–45). Additionally, for some exposures like smoking, DNAm surrogates have shown more promise in accurately estimating exposure timing and dosage than prior gold-standard smoking biomarkers like cotinine (16, 46).

Concerning SES, a DNAm-based surrogate may also be useful as a "discovery biomarker" that can indicate biological processes (BP) that arise from or are associated with social disadvantage and its consequences. Thus, as opposed to more clinical biomarkers where disease prediction or risk stratification is the central goal (47), DNAm-based surrogates of SES may also help elucidate "biological expressions of social inequality," or molecular pathways that link adverse social or environmental experiences with downstream health outcomes (48). To date, a large body of literature suggests that low SES is strongly associated with morbidity and mortality, including health problems such as cardiovascular disease, diabetes, hypertension, respiratory infections, and cancer (49, 50). Evidence suggests that the biological mechanisms linking SES with disease in industrialized populations operate through stress-related inflammatory pathways that can dysregulate immune system responses or metabolic functioning (51–53). In addition, mechanisms between socioeconomic adversity and health may vary depending on whether the exposure occurs in childhood, adulthood, or cumulatively over the lifespan. Empirical research indicates that while adult SES matters, early-life SES can have lasting effects, and better conditions in adulthood often cannot fully offset the impact of early adversity during sensitive periods of development (54–58). For example, the relationship between childhood adversity and later-life health may persist due to latent independent effects that disrupt critical or sensitive periods of development in early life (i.e. "biological embedding" or "programming" of adult disease) and/or via the role of early environment on subsequent life trajectories that have cumulative or "weathering' effects on health (52, 59–63).

Epigenome-wide association studies (EWAS) that independently test associations between individual CpG sites and SES-related outcomes generally corroborate these findings (11). However, more evidence is needed to understand the influence of timing and/or duration of socioeconomic disadvantage on DNAm in childhood and adulthood, as well as the degree of overlap in DNAm patterns across SES domains at both the individual and neighborhood levels (11). In addition, unlike penalized regression models that use shrinkage methods to construct DNAm biomarkers (e.g. lasso or elastic net), association weights from EWAS do not take the intercorrelation of CpG sites into account, which in turn may reduce the predictive accuracy of a biomarker or "methylation risk score" that uses EWAS weights to calculate a weighted sum of DNAm levels (64).

In this study, we trained DNAm-based surrogates or biomarkers of childhood and adult SES in the Health and Retirement Study (HRS)—a large, multiracial/ethnic sample of older adults ($N = 3{,}527$)—and then tested their performance in another diverse, US-based sample of older adults (the Multi-Ethnic Study of Atherosclerosis or MESA, $n = 1{,}182$). To capture the multidimensional nature of SES, DNAm biomarkers were trained on composite indices of childhood SES (cSES) and adult SES (aSES) that measure SES at both the individual and the sociological levels. We then examined whether SES biomarkers (cSES-BIO and aSES-BIO) were associated with downstream morbidity and mortality, both before and after adjusting for self-reported childhood and/or adult SES. Mediation analyses were also performed to assess the proportion of the SES-health relationship mediated by the biomarkers. Because epigenetic clocks may capture social-specific health signatures beyond general aging, we also tested if associations between SES and health persisted after controlling for GrimAge and DunedinPACE, two epigenetic clocks previously shown to strongly correlate with SES (9, 10). Finally, we explored potential biological mechanisms by functionally characterizing the individual CpG sites that comprise the SES biomarkers. Results indicate that SES biomarkers reveal additional information not captured in self-reported SES or next-generation epigenetic aging measures, suggesting they may be useful proxies for unmeasured social exposures that can augment current

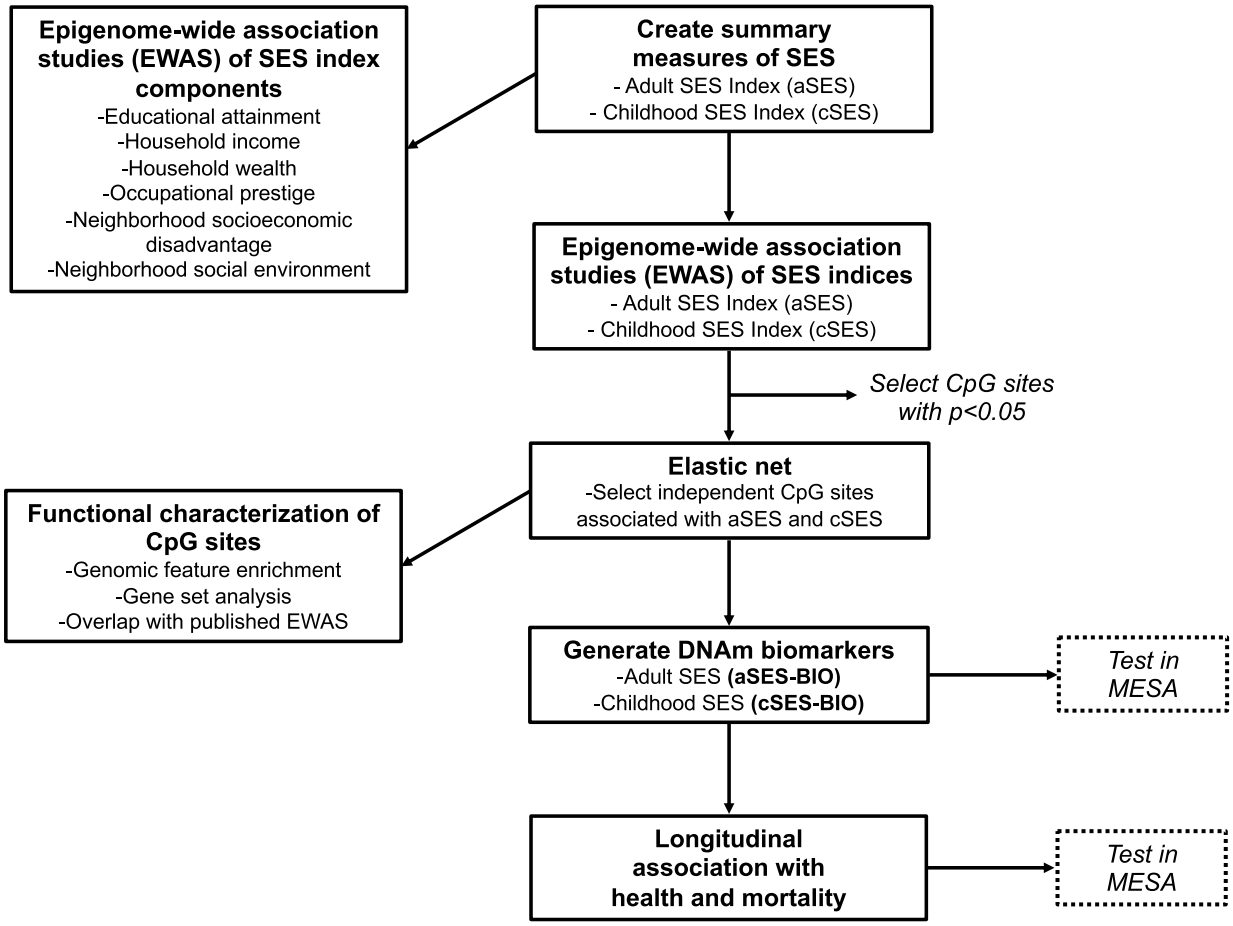

Fig. 1. Analytic pipeline in HRS and MESA.

knowledge about the biological impact of socioeconomic disadvantage on downstream health and aging.

Results

Sample statistics

The HRS, a nationally representative longitudinal study of older adults, was used for all initial analyses and biomarker training. DNAm was measured in 2016 from whole blood in a representative subset of participants using the Illumina Infinium Methylation EPIC BeadChip. In 2016, the HRS analytic sample ($n = 3{,}527$) had a mean age of 69.8 (SD = 9.7) and was 58.7% female (SI Appendix, Table S1). Approximately two-thirds (67.9%) were non-Hispanic white, with 15.6% non-Hispanic Black, 13.6% Hispanic, and 2.9% other. Most had a high school (HS) degree (81.5%), but only 25.6% had a college degree. Mean parental education was 10.9 years (SD = 3.9), with 50.5% of mothers and 44.4% of fathers completing at least 12 years of education or the equivalent of a HS degree. Mean childhood financial strain was 0.9 (SD = 1.2). In 2018, 3,120 participants with health information remained; sample statistics were comparable to 2016.

Data from MESA, a population-based longitudinal study of older adults focused on cardiovascular disease risk factors, were used for replication. DNAm was measured in 2010–2012 from monocytes using the Illumina Infinium Methylation450 BeadChip. The age distribution of the MESA replication sample ($n = 1{,}182$) was comparable to the HRS (69.6, SD = 9.3) but was slightly more diverse in terms of race (White = 48.5%; Black = 19.4%; Hispanic = 32.2%) and sex (female = 50.9%). Mean educational attainment was somewhat higher than HRS: 85.7% of the sample completed a HS degree, and 33.5% obtained a college degree. Parental education was lower but comparable, with 46.5% of mothers and 41.9% of fathers obtaining at least a HS degree. Information on childhood financial strain was not available in MESA.

SES index biomarkers

The analytic pipeline for biomarker creation and downstream analysis is shown in Fig. 1 and described in detail in Statistical methods. Due to the large number of CpG sites on the EPIC chip, we first conducted EWAS of the aSES and cSES indices separately in the training data (HRS) to reduce the number of CpGs evaluated by the elastic net procedure for biomarker construction. A total of 69,000 and 88,681 CpG sites on the EPIC chip were nominally associated ($P < 0.05$) with aSES and cSES in the EWAS, respectively. Manhattan plots for these EWAS can be found in SI Appendix, Figs. S1 and S2.

The elastic net procedure that minimized prediction error in the full sample selected a total of 25 CpG sites for aSES (lambda = 0.138, cross-validation [CV] $R^2 = 5.57\%$, Table 1) and 17 sites for cSES (lambda = 0.128, CV $R^2 = 3.35\%$, Table 2). For both aSES and cSES, the selected CpG sites tended to have relatively high ranks in the original EWAS (ranks range from 1 to 15,540 for aSES and 1 to 43 for cSES, all with EWAS P-values $< 4.6 \times 10^{-3}$). Four CpG sites at FDR $P < 0.10$ overlapped between the cSES and aSES index EWAS results (SI Appendix, aSES and cSES EWAS results). Two

Table 1. CpG sites selected for adult SES index biomarker (aSES-BIO) by elastic net in HRS.

CpG	E-net weight	Chr.	Position	N	EWAS P-value	EWAS rank	UCSC Ref gene name	UCSC Refgene group	Relation to UCSC CpG Island	Regulatory feature group	Phantom5 enhancer	DHS	eQTM	meQTL
cg18181703	−0.7348	17	76354621	3522	9.80E−22	1	*SOCS3*	Body	N_Shore	Promoter associated	0	0	1	0
cg11047325	−0.1265	17	76354934	3518	1.40E−19	2	*SOCS3*	Body	Island		0	1	0	0
cg21327712	0.6412	1	111744503	3522	6.90E−19	3	*DENND2D*	TSS1500; Body	N_Shore		0	1	0	0
cg01894508	0.0838	2	70189111	3517	5.30E−12	7	*ASPRV1*	5'UTR; 1stExon	S_Shore		0	0	0	1
cg18062721	0.1265	3	11643427	3518	1.30E−11	9	*VGLL4*	Body			0	1	0	0
cg03362418	−0.0698	22	50965563	3521	1.50E−11	11	*TYMP; SCO2*	Body; TSS1500	Island	Promoter associated	0	1	0	1
cg25008217	0.299	19	41882654	3499	2.30E−11	12	*TMEM91*	TSS1500; 1stExon; 5'UTR	Island		0	1	0	0
cg17461390	−0.1565	19	17428411	3524	1.60E−10	19	*DDA1*	Body		Unclassified cell type specific	0	1	0	1
cg02767093	−0.1932	13	99130655	3480	2.00E−10	21	*STK24*	Body	S_Shelf		1	1	1	1
cg25145728	−0.0126	16	89628775	3433	4.50E−10	28	*RPL13*	Body	S_Shore		0	1	0	1
cg02017926	0.0072	12	123754328	3517	4.90E−10	29	*CDK2AP1*	Body	Island	Promoter associated	0	1	1	0
cg10922280	0.0408	16	68034227	3517	7.80E−10	33	*DPEP2*	TSS1500			0	1	1	0
cg10206344	0.1281	16	31483277	3525	9.90E−10	39	*TGFB1I1*	TSS1500; TSS200	Island	Unclassified	0	1	0	1
cg11454468	−0.0387	3	142569221	3475	1.20E−09	44	*PCOLCE2*	Body			0	1	0	0
cg11183072	0.0696	17	37894397	3522	2.00E−09	48	*GRB7*	TSS200; 5'UTR			0	1	0	0
cg10842530	0.0079	14	69438358	3513	2.70E−09	52	*ACTN1*	Body			0	0	0	1
cg03031609	0.0759	10	7453871	3521	2.30E−08	82	*SFMBT2*	TSS1500	Island		0	1	0	1
cg24445316	0.0244	1	23889092	3525	3.40E−08	95			S_Shelf	Unclassified	1	1	1	0
cg06864083	0.2008	17	40832319	3520	9.00E−08	142	*CCR10*	Body	Island		0	1	0	1
cg03090734	0.0237	10	44525757	3519	3.00E−07	205				Unclassified	0	1	0	0
cg10946295	0.019	2	119913307	3505	1.30E−06	373			N_Shore		0	1	0	0
cg27637521	−0.331	17	76355202	3519	4.20E−06	576	*SOCS3*	5'UTR	Island		0	1	1	0
cg06021088	−0.0156	2	127822551	3515	6.70E−06	687	*BIN1*	Body		Unclassified	0	1	1	0
cg19574915	0.1381	15	89195555	3526	1.40E−04	2698	*ISG20*	Body		Promoter associated cell type specific	0	1	1	1
cg08852765	0.0296	6	33396407	3513	4.60E−03	15540	*SYNGAP1*	Body	S_Shore	Promoter associated	1	1	0	1

Chr., chromosome; EWAS, epigenome-wide association study; N_Shore, north shore; S_Shore, south shore, UCSC, University of California Santa Cruz Genome Browser; DHS, DNAse hypersensitive site; eQTM, expression quantitative trait methylation; meQTL, methylation quantitative trait locus.

Table 2. CpG sites selected for childhood SES index biomarker (cSES-BIO) by elastic net in HRS.

CpG	E-net weight	Chr.	Position	N	EWAS P-value	EWAS rank	UCSC ref gene name	UCSC ref gene group	Relation to UCSC CpG Island	Regulatory feature group	Phantom5 enhancer	DHS	eQTM	meQTL
cg04887278	0.6037	6	83903927	3505	2.40E−11	1	*PGM3*; *RWDD2A*	TSS1500; 5'UTR	S_Shore		0	1	0	0
cg03519157	0.4604	11	46367033	3526	4.90E−11	2	*DGKZ*	5'UTR; 1stExon; Body	Island		0	1	1	0
cg08469255	0.232	6	30851069	3451	6.70E−10	3	*DDR1*	TSS1500	N_Shore	Unclassified	0	1	1	1
cg17810176	0.1021	19	36035831	3523	1.70E−09	4	*GAPDHS*; *TMEM147*	Body; TSS1500	N_Shore		0	1	0	0
cg22505924	0.0791	2	64488157	3527	4.10E−09	6				Unclassified	1	1	0	0
cg27496526	0.4744	15	41805530	3526	6.90E−09	7	*LTK*	Body	Island		0	1	0	0
cg25563256	0.0349	17	7341641	3524	7.60E−09	8	*FGF11*	TSS1500	N_Shore	Unclassified	0	1	0	0
cg16329896	0.1303	7	47515060	3517	1.30E−08	10	*TNS3*	5'UTR		Unclassified cell type specific	1	1	0	1
cg10394832	0.0612	1	111889533	3525	2.50E−08	13	*C1orf88*	Body	Island	Promoter associated	0	0	0	1
cg17496659	0.0165	1	3568245	3520	3.60E−08	16	*TP73*	TSS1500	Island		0	1	0	0
cg21159993	0.019	5	139554405	3523	3.70E−08	17	*C5orf32*	TSS1500	Island	Promoter associated	0	1	1	0
cg19989295	0.003	14	24641077	3483	5.10E−08	19	*REC8*	TSS200	Island	Promoter associated	0	1	1	1
cg1382554	0.0131	9	140117257	3514	6.50E−08	21	*RNF208*	TSS1500	Island		0	1	0	0
cg05302489	0.0382	6	31760426	3499	1.00E−07	28	*VARS*	Body	N_Shelf		0	0	1	1
cg02170695	−0.0968	21	38909809	3525	1.40E−07	34					0	0	0	1
cg13067434	0.0325	15	63836865	3522	1.80E−07	37	*USP3*	Body			0	0	0	0
cg04382191	0.026	3	9437901	3525	2.20E−07	43	*SETD5*; *THUMPD3-AS1*	TSS1500; Body	N_Shore	Promoter associated cell type specific	0	1	0	0

Chr., chromosome; EWAS, epigenome-wide association study; N_Shore, north shore; S_Shore, south shore, UCSC, University of California Santa Cruz Genome Browser; DHS, DNAse hypersensitive site; eQTM, expression quantitative trait methylation; meQTL, methylation quantitative trait locus.

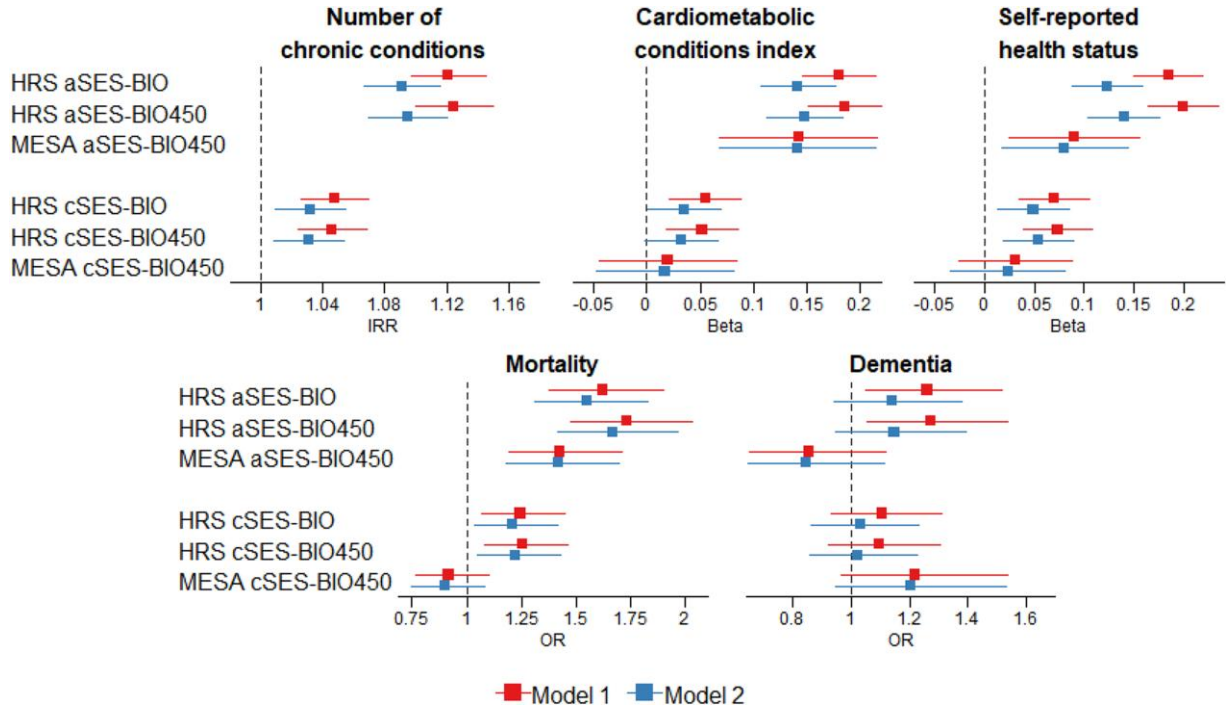

Fig. 2. Associations between SES biomarkers and health outcomes in HRS and MESA. Notes: The figure displays estimates from separate regressions of health or mortality in HRS and MESA on aSES-BIO and cSES-BIO (adult and child SES biomarkers constructed from EPIC chip methylation sites) and aSES-BIO450 and cSES-BIO450 (adult and child SES biomarkers constructed from 450 K chip methylation sites). All SES biomarkers were standardized for analysis. Model 1 controls for age, sex, race, current smoking, former smoking, and *APOE* e4 carrier status (dementia models only). Model 2 adds controls for measured aSES (in aSES-BIO or aSES-BIO450 regressions) or cSES (in cSES-BIO or cSES-BIO450 regressions). Comparable data on the number of chronic conditions was not available in MESA. For chronic conditions, a Poisson model was used; for the CMCI and SRHS, linear models were used; for mortality and dementia, logistic models were used. 95% CI. IRR, incidence rate ratio; OR, odds ratio. Sample sizes (HRS): number of chronic conditions = 3120; CMCI = 3120; SRHS = 3118; mortality = 3527; dementia = 3120. Sample sizes (MESA): CMCI = 827; SRHS = 848; mortality = 1180; dementia = 1180.

overlapping sites were selected by elastic net for cSES-BIO, and none were selected for aSES-BIO. EWAS of the individual components of the SES indices showed that education, household income, wealth, and neighborhood socioeconomic disadvantage had the strongest associations with many of the CpG sites that comprise aSES-BIO (SI Appendix, Table S2). Parental education had the strongest associations with many of the CpG sites that comprise cSES-BIO (SI Appendix, Table S3).

Since DNAm in MESA was profiled on the 450 K array, we also trained 450 K versions of the aSES and cSES biomarkers (aSES-BIO450 and cSES-BIO450) in HRS using the same parameters and model specifications for CpG sites common to both the 450 K and EPIC arrays. Elastic net selected 29 of 34,095 sites for aSES (lambda = 0.138, CV R^2 = 5.34%, SI Appendix, Table S4) and 15 of 46,548 sites for cSES (lambda = 0.128, CV R^2 = 3.28%, SI Appendix, Table S5). The aSES-BIO450 included 14 of the same CpG sites as aSES-BIO, and the two biomarkers were highly correlated (r = 0.89) (SI Appendix, Table S6). The cSES-BIO450 was also highly correlated with cSES-BIO (r = 0.99) and shared 13 CpG sites.

Correlations between SES indices, SES biomarkers, and epigenetic aging measures

In HRS, the correlation between the measured aSES and cSES indices (r = 0.43) was higher than the aSES-BIO and cSES-BIO correlation (r = 0.28) (SI Appendix, Table S6). The aSES index was correlated with aSES-BIO (r = 0.24) and the cSES index was correlated with cSES-BIO (r = 0.19). aSES-BIO was correlated with both aging clocks (r = 0.18 for GrimAge and r = 0.50 for DunedinPACE), and cSES-BIO was weakly correlated with GrimAge (r = 0.05) and DunedinPACE (r = 0.16).

Associations between SES biomarkers and downstream health outcomes

Figure 2 displays associations between aSES-BIO or cSES-BIO and downstream health outcomes measured ~2 years after DNAm measurement in the HRS. Both aSES-BIO and cSES-BIO were associated with the number of chronic conditions, the cardiometabolic conditions index (CMCI), self-reported health status (SRHS), and mortality after adjusting for age, sex, race, and smoking (all $P <$ 0.0001, model 1). Associations between aSES-BIO and these same health outcomes persisted in magnitude and statistical significance after adjusting for measured aSES (model 2). SI Appendix, Table S7 shows that these aSES-BIO associations were slightly attenuated but persisted after adjusting for epigenetic aging clocks (models 3a and 4a) and for both measured aSES and the clocks (models 4a and 4b). Conversely, associations with cSES-BIO were statistically insignificant after adjusting for both cSES and DunedinPACE (SI Appendix, Table S7, model 4b), and for CMCI, after adjustment for DunedinPACE (model 3b) or both cSES and GrimAge (model 4a). Associations between aSES-BIO and Langa-Weir dementia were fully attenuated after adjusting for measured aSES (models 2, 4a, and 4b), but persisted after adjusting for GrimAge and DunedinPACE (models 3a and 3b). There were no significant associations between cSES-BIO and Langa-Weir dementia. Associations between health outcomes and the aSES-BIO450 or cSES-BIO450 biomarkers were similar to EPIC chip biomarker associations (SI Appendix, Table S8).

To assess the performance of the SES biomarkers relative to measured SES, we also estimated a version of model 1 (model 1*)

Table 3. Mediation analysis of associations between measured SES and health outcomes mediated by SES biomarkers in the HRS.

Outcome	Adult SES				
	Estimate type	Total effect	Direct effect	Indirect effect	% Mediated
Number of chronic conditions	IRR	1.16*** (1.13, 1.19)	1.13*** (1.10, 1.16)	1.03*** (1.02, 1.04)	19.86*** (13.48, 26.23)
Cardiometabolic conditions index	Beta	0.21*** (0.17, 0.25)	0.16*** (0.12, 0.21)	0.05*** (0.03, 0.06)	21.82*** (14.74, 28.91)
Self-reported health status	Beta	0.30*** (0.26, 0.34)	0.26*** (0.22, 0.30)	0.04*** (0.03, 0.05)	13.20*** (8.79, 17.61)
Mortality	OR	1.48** (1.17, 1.78)	1.28** (1.01, 1.55)	1.16*** (1.09, 1.22)	41.82** (18.88, 64.75)
Langa-Weir dementia	OR	1.92** (1.44, 2.39)	1.84** (1.38, 2.30)	1.04 (0.98, 1.10)	8.10 (−4.28, 20.48)
	Childhood SES				
Outcome	Estimate Type	Total Effect	Direct Effect	Indirect Effect	% Mediated
Number of chronic conditions	IRR	1.09*** (1.07, 1.12)	1.08*** (1.06, 1.11)	1.01*** (1.00, 1.01)	8.38** (2.00, 14.76)
Cardiometabolic conditions index	Beta	0.12*** (0.08, 0.15)	0.11*** (0.07, 0.15)	0.01* (−0.00, 0.02)	6.90* (−0.34, 14.14)
Self-reported health status	Beta	0.12*** (0.08, 0.16)	0.11*** (0.07, 0.15)	0.01** (0.00, 0.02)	9.26** (1.89, 16.63)
Mortality	OR	1.23** (1.01, 1.45)	1.18* (0.97, 1.39)	1.04** (1.01, 1.08)	22.57* (−2.28, 47.41)
Langa-Weir dementia	OR	1.44** (1.15, 1.73)	1.43** (1.14, 1.72)	1.01 (0.97, 1.05)	1.91 (−11.3, 15.12)

Models control for age, sex, race, current smoking, and former smoking. Sample sizes: number of chronic conditions = 3120; CMCI = 3120; SRHS = 3118; mortality = 3527; dementia = 3120. *$P < 0.1$; **$P < 0.05$; ***$P < 0.0001$.

where aSES or cSES replaced the biomarkers (SI Appendix, Table S7). In general, SES biomarker associations were not significantly different from aSES or cSES associations in our training sample, except for SRHS and Langa-Weir dementia, where measured associations were significantly higher in magnitude. Model 1 biomarker associations were robust after adjusting for either measured cSES (in the case of aSES-BIO) or aSES (in the case of cSES-BIO) (SI Appendix, Table S7, model 5).

Biomarker mediation analysis

Finally, we tested the proportion of the total effect of measured aSES (or cSES) on health outcomes mediated by aSES-BIO (or cSES-BIO) (Table 3). For aSES, the proportion mediated by aSES-BIO was largest for mortality (41.8%, $P < 0.05$) with smaller effects for SRHS (13.2%), chronic conditions (20%), and CMCI (22%) (all $P < 0.0001$). For cSES, cSES-BIO mediated 8.4 to 9.3% of the effects on chronic conditions and SRHS, respectively ($P < 0.05$).

Replication in MESA

In the replication sample, the adult SES biomarker (aSES-BIO450) was correlated with aSES and explained ~4.4% of the variation in aSES (SI Appendix, Table S9). The variance explained was similar after accounting for age and sex (partial $R^2 = 3.88\%$, $P < 0.0001$). aSES-BIO450 correlations with GrimAge ($r = 0.36$) and Dunedin PACE ($r = 0.60$) were higher than correlations in HRS (SI Appendix, Table S9). The biomarker also had significant associations with downstream health outcomes that were similar in magnitude to HRS associations, including CMCI, SRHS, and mortality ($P < 0.05$, model 1, Fig. 2), and these associations persisted after controlling for aSES (model 2). Importantly, aSES-BIO450 appears to be a better predictor of downstream CMCI and mortality than aSES, whereas for SRHS, aSES outperforms the aSES biomarker (SI Appendix, Table S10, model 1* and model 1). Associations between aSES-BIO450 and SRHS were also fully attenuated after adjusting for GrimAge or DunedinPACE. For CMCI, associations persisted in models that controlled for GrimAge (models 3a and 4a) but were statistically insignificant after adjusting for DunedinPACE (models 3b and 4b). Mortality associations persisted across all models but were attenuated in magnitude and significance when adjusting for DunedinPACE. No associations were found between the biomarker and International Classification of Diseases (ICD)-based all-cause dementia.

For childhood SES, we were limited to self-reports of parental education and therefore could not test the correlation between cSES-BIO450 and the complete cSES index in MESA. The cSES-BIO450 was marginally correlated with parental education ($r = 0.07$). The cSES biomarker was not correlated with GrimAge or DunedinPACE (SI Appendix, Table S9). Associations between cSES-BIO450 and downstream health outcomes were insignificant at $P < 0.05$ (SI Appendix, Table S10 and Fig. S2).

Functional characterization of biomarker CpGs

CpG sites selected for the aSES-BIO were enriched in DNAse hypersensitivity sites (DHS) and enhancer regions, while CpGs included in the cSES-BIO were enriched in promoter regions (SI Appendix, Table S11; all $P < 0.05$). Further, compared to CpGs that were not included in the biomarker indices, CpGs selected for both EPIC and 450 K SES biomarkers were enriched for expression quantitative trait methylation (eQTM) expression ($P < 8.3E-04$) (SI Appendix, Table S11). Taken together, these results suggest that the selected CpGs may be important for the regulation of gene expression.

Based on Illumina annotation file mapping, CpGs selected for the aSES-BIO and cSES-BIO mapped to 21 and 18 unique genes, respectively. After FDR correction, gene-set enrichment of gene ontology (GO) BP terms, and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways were not observed for either set of CpGs when mapped to their proximal genes (FDR $q < 0.01$). When utilizing eQTM data from Keshawarz et al. (2023), CpGs selected for the aSES-BIO and cSES-BIO mapped to the expression of 230 and 66 genes, respectively, in *cis* or in *trans*. Additionally, CpGs selected for aSES-BIO were associated with the expression of genes enriched for eight GO BP terms related to cell death, T cell differentiation and activation, cell migration, and cell–cell signaling (SI Appendix, Table S12 and Fig. S3). CpGs selected for the cSES-BIO were associated with the expression of genes enriched for 13 GO BP terms and 1 KEGG pathway, including chemotaxis, T cell differentiation, immune system process, and cell adhesion molecules (SI Appendix, Tables S13 and S14; Fig. S4). Results from the GO enrichment analysis for the cSES-BIO450 were similar (SI Appendix, Table S15 and Fig. S5). There was no observed enrichment of GO or KEGG pathways for genes associated with CpGs selected for the aSES-BIO450. Finally, a relatively high percentage of CpGs selected for both the EPIC and 450 K SES biomarkers (35–72%) mapped to *cis*-methylation quantitative trait loci (meQTLs)

identified in the EPIGEN MeQTL and GoDMC databases, respectively (65, 66).

Biomarker CpG overlap with published EWAS

Several CpGs selected for both the aSES and cSES biomarkers have been previously associated with diseases and traits identified in large-scale EWAS, including age, chronic diseases, and inflammation. In general, CpG sites in cSES-BIO overlapped more with prior EWAS of immune system processes, whereas aSES-BIO CpGs overlapped with a broader range of traits or molecular processes related to metabolic functioning, health behaviors, or cancer (SI Appendix, Tables S16–S19).

Discussion

We developed DNAm biomarkers of child and adult SES in a large, population-representative sample of older adults (HRS) and validated their performance and association with health outcomes in one of the few comparable US samples with DNAm data (MESA). The aSES-BIO explained ∼4% of the variance in measured aSES in MESA. The explanatory power of cSES-BIO could not be fully ascertained in MESA because data on childhood financial strain were unavailable. While the explained variance is modest, these findings are significant given that (i) not all variation in SES across individuals will translate into discernible biological changes reflected in the epigenome, (ii) successful replication across different tissue types and platforms demonstrates robustness, and (iii) the aSES biomarker was a better predictor of downstream cardiometabolic conditions and mortality than measured aSES, suggesting potential clinical utility.

In HRS and MESA, aSES-BIO was associated with chronic disease, cardiometabolic conditions, and mortality 2 to 6 years after DNAm profiling, and these associations persisted after controlling for measured SES. Results were similar for cSES-BIO health associations in HRS, but were insignificant in MESA. Mediation results suggest SES DNAm signatures may mediate up to 42% of the SES-health relationship; however, due to challenges in meeting the causal assumptions of mediation analysis and the potential for reverse causation, we caution against overinterpreting these results. Taken together, these findings suggest that DNAm-based surrogates of SES may serve as valuable proxies for unmeasured social exposures and provide insights into the biological mechanisms linking SES to health beyond what is captured by self-reported SES.

Additionally, we examined whether epigenetic aging measures attenuated the association between SES-DNAm biomarkers and health. Epigenetic clocks may reflect social gradients in health by capturing the cumulative impact of social and environmental exposures on biological aging processes. These factors, including chronic stress, limited access to resources, and early-life adversity, may induce epigenetic changes that contribute to health disparities. In HRS and MESA, SES biomarkers were correlated with GrimAge ($r = 0.05$–0.36) and DunedinPACE ($r = 0.16$–0.60). Depending on the health outcome, aSES biomarker health associations were attenuated by 4–37% and 5–61% after controlling for GrimAge and DunedinPACE, respectively, and cSES-biomarker associations were attenuated by 1–40% and 3–60%, respectively. These results indicate that (i) the SES biomarkers capture additional SES-related health information beyond what is reflected in these aging measures and (ii) epigenetic aging measures themselves contain social-specific health signatures. Therefore, SES-DNAm biomarkers could be valuable in epigenetic aging studies to account for the potential confounding influence of social gradients on aging clock estimates.

Further analyses of the CpG sites that comprise the SES biomarkers revealed enrichment in enhancer and promoter regions, DNS, and eQTMs, suggesting they may be critical for the regulation of gene expression. In gene-set enrichment analysis, selected CpGs were associated with the expression of genes enriched in pathways related to chemotaxis, T cell activation, immune system response, cell migration, and cell death. CpG sites in both biomarkers were associated with prior EWAS of age, age-related chronic diseases, inflammation (e.g. chronic pain, chronic obstructive pulmonary disease, and C-reactive protein), and clear cell renal carcinoma (67). However, unique patterns for each biomarker also emerged. CpG sites in aSES-BIO were identified in prior EWAS of type 2 diabetes and health behaviors, including smoking, alcohol use, diet quality, and body mass index. Additionally, CpGs were associated with pancreatic cancer (68) and circulating tumor necrosis factor receptor 2 (sTNFR2) (69), a known biomarker for several types of cancer, including colorectal cancer, non-Hodgkin's lymphoma, and hepatocarcinoma (70), that impacts tumor activation and progression (71). Of note, the top CpG selected for both aSES-BIO and aSES-BIO-450 K (cg18181703) maps to the body of the suppressor of cytokine signaling 3 (*SOCS3*) gene, which has a role in regulating cytokine and hormone activity. Both inflammation and infection can stimulate the expression of *SOCS3* in different cell populations (72). cg18181703 was also identified in prior EWAS of educational attainment and social disadvantage (4, 73). Finally, although associations between aSES-BIO and dementia were weaker when compared to other health outcomes, one of the CpG sites selected for aSES-BIO (cg06021088) maps to the Bridging Integrator 1 (*BIN1*) gene, which is considered the second most significant genetic risk factor for late-onset Alzheimer's disease (74, 75).

Conversely, CpGs in cSES-BIO were found more in prior EWAS of immune and autoimmune diseases, including atopy, Sjorgren's syndrome, and rheumatoid arthritis, and a subset of CpGs were associated with the expression of genes enriched for immune system processes and T cell differentiation. These findings support prior research that has shown childhood disadvantage is a risk factor for immune dysregulation (76) and changes in T cell activity (77), which in turn can increase the risk of mental and physical health issues in adulthood, including depression, cardiovascular disease, type 2 diabetes, and cancer (78, 79). This supports the hypothesis of biological embedding of adult disease in early life, potentially through adverse childhood circumstances that may have cumulative or interactive effects on adult health.

A relatively high percentage of the CpGs in the biomarkers (35–72%) are meQTLs, meaning they are at least partially influenced by proximal genetic variants. This is unsurprising, given that a large percentage of CpGs are under genomic control, with 35 and 45% of all CpGs in EPIGEN and GoDMC, respectively, identified as *cis*-meQTLs. Although genetic variants influencing these CpGs may be associated with key demographics (e.g. race), we controlled for genetic principal components (PCs), which likely minimized the effects of ancestry on CpG selection.

This study has several limitations. First, although we include covariates in our models to mitigate any confounding from sex, race, smoking, genetic ancestry, and technical noise, we cannot draw causal conclusions regarding the relationship between socioeconomic position in childhood or adulthood and DNAm signatures. In addition, self-reported SES may be subject to recall bias or measurement error. Thus, the CpG sites chosen by the machine-learning algorithm likely detect a mix of features that are not directly causal, and some may be uninformative correlates

of SES. Second, although EWAS and elastic net analyses indicated limited overlap between cSES-BIO and aSES-BIO signatures, further investigation using RNA-seq data is needed to fully understand their biological independence and significance. Third, some features may be specific to the US context and may not be generalizable to other countries, particularly in high-income countries where different welfare states or social programs affect the depth and distribution of social disadvantage and/or in low- and middle-income countries where the biosocial determinants of health may vary due to vastly different socioeconomic and epidemiological contexts (80). Consequently, SES biomarkers developed in US data may exhibit limited generalizability in other contexts, and further research is necessary to determine whether country- or region-specific biomarkers can identify unique methylation signatures that reflect the distinct interplay between social disadvantage and health within diverse cultural settings. Similarly, since the biomarkers were trained in older adults, their generalizability in younger populations remains uncertain, particularly considering the significant impact of pregnancy on methylation patterns (81). Fourth, technical noise from batch effects or unreliable probes may limit generalizability, and SES biomarkers trained in the HRS EPIC array data could not be validated in the MESA 450 K array data. Finally, DNAm was profiled in different tissue types in HRS (whole blood) and MESA (monocytes). Although our results appear to replicate across these two tissue types, further replication in other cell/tissue samples is needed.

Despite these limitations, this study provides a foundation for the rigorous creation of epigenetic markers for social exposures. A major strength of this study is the use of large, multiracial/ethnic population-representative samples with rich socioeconomic data from childhood and adulthood. Within HRS and MESA, we were also able to test longitudinal associations between SES biomarkers and downstream health and mortality. Future analyses include evaluating SES biomarkers in each racial/ethnic group separately, investigating the developmental trajectory of SES-related methylation signatures across the lifespan, and examining associations between the identified CpG sites and forthcoming RNA-seq data in HRS.

Materials and methods

Data

Health and retirement study

The HRS is a nationally representative, longitudinal panel study of individuals over the age of 50 and their spouses that began in 1992 (82, 83). HRS participants are interviewed every 2 years about their income and wealth, health and use of health services, work and retirement, and family connections. Additional cohorts are introduced every 6 years to offset attrition and death and to maintain a sample size of ~20,000 per wave. DNAm data were profiled in 4,103 racially and socioeconomically diverse HRS participants, and 4,018 samples passed quality control (84). Of these, 491 were missing genomic data. A total of 3,527 participants with DNAm, genetic, and socioeconomic data were included in the analyses.

Demographic and socioeconomic data were taken from the RAND HRS Longitudinal File 2020 (V1) (85). Race and ethnicity were self-reported. Neighborhood social disadvantage data were taken from the 2006–2016 Psychosocial and Lifestyle Questionnaire and merged to respondents at the Census tract level using restricted data on the location of primary residence from the 2006–2016 waves (86, 87). Neighborhood socioeconomic disadvantage measures were constructed using the restricted HRS Contextual Data Resource, which contains data from the 2005–2018 American Community Survey linked to respondents at the Census tract level (88). Restricted three-digit Census occupation codes from the 1992–2016 (89) waves were used to merge occupational prestige measures from the General Social Survey (90).

DNA methylation

DNAm was measured from whole blood samples collected in the 2016 HRS Venous Blood Study (91). Methylation was measured using the Illumina Infinium Methylation EPIC BeadChip. The *minfi* package in R software was used for data preprocessing and quality control. Background correction with dye-bias normalization was performed using the normal-exponential out-of-band method (Noob) (92). Samples with mismatched sex or an average median intensity <8.5 were removed. A detection P-value < 0.01 was used to remove probes or samples with >5% missing data. Cross-reactive probes were removed before elastic net analyses (93). After quality control, a total of 836,660 CpGs were available for analysis. DNAm at each CpG site was quantified using beta values, which approximate the proportion of methylation. The Houseman method was used to estimate white blood cell (WBC) proportions (94). We also used two epigenetic aging measures, GrimAge (27) and Dunedin Pace of Aging (DunedinPACE) (29). GrimAge was calculated and released by HRS (84), and we calculated PACE using the *DunedinPACE* R package (29).

Genotypes

We used genotype data to estimate genetic PCs of ancestry and to obtain *APOE* ε4 carrier status. Genotype data for over 18,000 HRS participants was measured using Illumina HumanOmni2.5 BeadChips (HumanOmni2.5-4v1, HumanOmni2.5-8v1) (95). Individuals with call rates <98%, single nucleotide polymorphisms (SNPs) with call rates <98%, Hardy-Weinberg equilibrium P-value < 0.0001, chromosomal anomalies, and first-degree relatives in the HRS were removed. PCs were calculated using the *SNPRelate* package in R (96). *APOE* isoforms were inferred from SNP genotypes determined by TaqMan assays or imputed using the HumanOmni2.5 data (97). People with at least one copy of the ε4 allele were considered ε4 carriers.

Socioeconomic status

As in our previous work (10), to capture the multidimensional nature of SES, we used six measures to construct an aSES index: educational attainment, household income, wealth, occupational prestige, neighborhood socioeconomic disadvantage, and neighborhood social environment. Additional details on SES index construction are provided in SI Appendix, Section S1. Briefly, we took the unweighted average of the nonmissing quintiles for each indicator and reverse-coded the indicators so that a higher quintile corresponds to worse SES. Educational attainment was measured in years and was coded as 1 ≥ 16 years, 2 = 16 years, 3 = 13–15 years, 4 = 12 years, and 5 ≤ 12 years. Similarly, quintiles were constructed after averaging values across all HRS waves up to 2016 for household income (inflation-adjusted to 2016 dollars), wealth (total assets inflation-adjusted to 2016 dollars), and occupational disadvantage (98) (attained by matching occupational prestige codes from the General Social Survey to respondent three-digit occupation codes). Quintiles for the neighborhood socioeconomic disadvantage score (a summary of six Census tract-level metrics reflecting education, occupation, income/wealth, poverty, employment, and housing) and neighborhood social disadvantage score (a summary of conditional Bayes' estimates for aggregate-

level responses on esthetic quality, safety, and social cohesion) were averaged across 2012–2016 waves and 2006–2016 waves, respectively, to reflect cumulative neighborhood exposures (99, 100). Respondents were required to have nonmissing values for at least three of the six measures to be assigned an aSES index. As a robustness check, we also used PC analysis of these six variables to construct a PC-based SES measure, which was highly correlated with the SES index ($r = 0.97$). We opted to use the method above to construct the SES index because the PC method required all score elements to be nonmissing.

We created a cSES index by taking the unweighted average of two measures: parental education and childhood financial strain (101). Parental education was the average of maternal and paternal education, coded as 0 = 16 years or more, or 1 = 13–15 years, 2 = 12 years, 3 = 8–11 years, 4 ≤ 8 years. For missing values of maternal or paternal education, we first imputed continuous education years from demographic and SES variables using a multivariate, regression-based procedure in Imputation and Variance Estimation (IVEware) software (http://www.isr.umich.edu/src/smp/ive/) as described in Faul et al. (101). For HRS participants from the earliest wave of the study, parental education was collected as a dichotomous variable; for these participants, we used parental education imputations from Vable et al. (102) rounded to the nearest year. Childhood financial strain was calculated by taking the sum of nonmissing indicators for whether the respondent ever moved as a child, received financial help from relatives, had a period of paternal unemployment, or had poor/varied financial status.

Health outcomes

To assess health and aging in HRS, we used data from the 2018 wave, which was collected 2 years after DNAm measurement. Number of chronic conditions (range 0–8) was a count of the self-reported number of conditions that the respondent had ever experienced, including hypertension, diabetes, cancer, lung disease, heart problems, stroke, psychiatric problems, and arthritis. The CMCI was the mean standardized sum of four mean standardized doctor-diagnosed chronic diseases (diabetes, heart problems, stroke, and hypertension). SRHS (range 1–5) was the respondent rating of their overall health, ranging from 1 = excellent to 5 = poor. Mortality was an indicator (1 = yes) of whether the respondent died between the 2016 and 2018 data collection waves. Dementia was evaluated using the Langa-Weir definition (103, 104), which sets dementia classification (1 = yes) if an individual scores 0–6 points on a 27-point cognition scale.

Replication sample: MESA

We evaluated the performance of our SES biomarkers in MESA. DNAm was measured from monocytes at exam 5 (2010–2012) using the Illumina Infinium Methylation450 BeadChip. See Liu et al. (105) for a description of data processing. Briefly, the *neqc* function of the *limma* package was used for background correction and quantile normalization (106). Probes were excluded if they had "detected' methylation levels in <90% of MESA samples (detection P-value threshold = 0.05), overlapped with a nonunique genomic region, or were designed to assay polymorphic SNPs rather than methylation (107). Genomic PCs were estimated with the *EIGENSTRAT* program using genotype data from the Affymetrix 6.0 SNP array (108). *APOE* genotypes were estimated from imputed genotype data, and individuals with at least one copy of the $\varepsilon 4$ allele were considered $\varepsilon 4$ carriers. Race and ethnicity were self-reported.

The same six SES measures (educational attainment, household income, wealth, occupational disadvantage, neighborhood socioeconomic disadvantage, and neighborhood social environment) were used to construct the adult SES index in MESA. These were coded similarly to HRS except for income and wealth, both of which were not measured continuously (see SI Appendix, Section S1 for details). For income, respondents selected their household's income bracket from a series of unfolding brackets rather than reporting a continuous value. For wealth, a combination of home and land ownership, car ownership, and investments were used to construct a 5-level categorical variable to take the place of the quintiles in the aSES index (1 = home or land + car + investments; 2 = home or land + car or investments; 3 = home or land only; 4 = car or investments; 5 = none). In MESA, measures of childhood SES were limited to parental education. Here, we took the average of maternal and paternal education, which was coded as a categorical variable (4 = no schooling; 3 = some schooling but did not complete HS; 2 = HS degree, 1 = some college, 0 = college/graduate professional degree).

Health outcomes were constructed in the same manner as HRS, except for the number of chronic diseases, which were not available in MESA, and dementia, assessed using ICD-based all-cause dementia diagnoses through 2018 from the MESA dementia events file. Apart from dementia, all health outcomes were measured at exam 6 ~6 years after DNAm data collection (2016–2019).

Statistical methods

Epigenome-wide association studies

Due to the large number of CpG sites present on the EPIC chip, we used a prescreening procedure to reduce the number of CpGs evaluated when constructing the SES biomarkers by first conducting EWAS for aSES and cSES, separately, in the training data (HRS). EWAS CpGs with $P < 0.05$ were then carried forward into the biomarker construction stage (Fig. 1). Our EWAS model was as follows:

$$\mathrm{CpG}_{iprc} = \alpha + \beta \mathrm{SES}_i + \mathrm{X}_i'\delta + \mathrm{WBC}_i'\theta + \mathrm{PC}_i'\gamma + u_p + \mathrm{z}_r + n_c + \varepsilon_{iprc} \quad (1)$$

where SES_i is either the aSES or cSES index for individual i. The matrix X_i contains individual characteristics including age, sex, and race; due to the strong effects of smoking on the methylome and the well-established socioeconomic gradient in smoking, we also include dichotomous indicators for current and/or former smoking (omitted category = never smoker). Because DNAm was profiled in whole blood and methylation patterns differ by blood cell type, we accounted for differences in white blood cell (WBC_i) composition across samples, or the proportion of monocytes, natural killer cells, B-cells, CD4 T-cells, and CD8 T-cells (omitted category = proportion granulocytes). The first four genetic ancestry PCs were included to account for potential confounding due to ancestry (PC_i). Technical covariates for plate (p), plate row (r), and plate column (c) were included as random effects to account for potential batch effects, or u_p, z_r, and n_c, respectively. Manhattan plots for these EWAS can be found in SI Appendix, Figs. S1 and S2.

To examine which of the SES index components had the strongest associations with the CpG sites selected into the SES index biomarkers, we also conducted separate EWAS for each of the components of the aSES and cSES indices (Fig. 1). We used the same model, adding years of education as a covariate when neighborhood socioeconomic disadvantage or neighborhood social environment was the exposure.

SES index biomarker construction

To minimize overfitting and identify the CpG sites most predictive of the SES indices, we applied elastic net, a regression method that combines variable selection with regularization (shrinkage) (109). For aSES and cSES separately, we first selected all CpG sites with $P < 0.05$ from the corresponding EWAS and then adjusted these CpGs for all covariates in the EWAS model. Next, we used the residual CpGs as input for elastic net, using the following parameters in the *glmnet* package in R: $\alpha = 0.5$ (equally balancing ridge regression and lasso) and 5-fold CV to find the optimal value of lambda (penalty strength parameter) for minimizing prediction error (110). To construct the SES biomarker for adult SES (aSES-BIO) and childhood SES (cSES-BIO), we summed the methylation value at each CpG site selected by elastic net at the optimal value of lambda, weighted by its corresponding beta coefficient, and standardized the sums. We then calculated the correlation between the methylation-based biomarkers (aSES-BIO and cSES-BIO), measured markers of SES (aSES and cSES), and the epigenetic aging measures (GrimAge and DunedinPACE).

We recognize that many studies, including our replication sample (MESA), have DNAm measured using the Illumina 450 K chip rather than the EPIC chip. Therefore, we also constructed versions of the aSES and cSES biomarkers by limiting the analysis to EPIC CpG sites on the 450 K chip and then applying the same parameters and model specifications for elastic net. We assessed the level of CpG overlap and correlation between these 450 K SES biomarkers (aSES-BIO450 and cSES-BIO450) and the EPIC chip biomarkers (aSES-BIO and cSES-BIO).

Associations between SES biomarkers and downstream health outcomes

We used Poisson regression (number of chronic conditions), linear regression (CMCI, SRHS), or logistic regression (mortality, dementia) to assess the relationship between the SES biomarkers (aSES-BIO, cSES-BIO) and downstream health and mortality in 2018, or 2 years after DNAm was profiled in the HRS, using the following model (model 1):

$$\text{Health}_i = \alpha + \rho \text{SES}_{\text{BIO}_i} + X_i'\theta + \varepsilon_i, \quad (2)$$

where SES_{BIO} is the aSES-BIO or cSES-BIO biomarker for individual i and the matrix X_i includes controls for age, sex, race, and smoking status as in Equation 1. The aSES, cSES, aSES-BIO, and cSES-BIO measures were standardized for analysis. For dementia models, we also included an indicator for APOE e4 status. We then examined associations between the SES biomarkers and health outcomes after accounting for the measured SES index (aSES or cSES; model 2), GrimAge or DunedinPACE (models 3a and 3b), both measured SES and GrimAge or DunedinPACE (models 4a and 4b), and either measured cSES (for aSES-BIO) or aSES (for cSES-BIO) to examine SES-BIO associations after accounting for SES in childhood or adulthood (model 5). To assess the performance of the SES biomarkers relative to measured SES, we also estimated a version of model 1 (model 1*) that used the aSES or cSES indices instead of the SES biomarkers. Coefficient P-values < 0.05 were considered statistically significant.

Replication in MESA

For each MESA participant, aSES-BIO450 or cSES-BIO450 was constructed by multiplying the elastic net beta coefficients by the methylation beta values at the corresponding CpG sites (measured in exam 5) and taking the standardized sum. We used raw weights to construct the biomarkers without further adjustment for WBC contamination or technical effects in MESA since WBCs were already accounted for in the training data for all EWAS and elastic net analyses, and studies seeking to create these biomarkers in their own samples may choose to input unadjusted CpG data because confounding from batch effects varies across studies. In MESA, adjusted and unadjusted SES biomarkers were highly correlated ($r = 0.99$). We then tested the association of aSES-BIO450 or cSES-BIO450 with measured aSES or cSES and with health outcomes (CMCI, SRHS, mortality, and dementia) collected in exam 6. The number of chronic conditions was not analyzed in MESA because the data did not permit the creation of a variable comparable to that in HRS.

Functional characterization of biomarker CpGs

Genomic feature enrichment analysis

We performed genomic feature enrichment analysis on the sets of CpG sites selected for each SES biomarker separately. For each set of CpGs, we examined whether their genomic locations were enriched for features including gene promoters, enhancers, DHS, CpG islands (CGI), and CpG island flanking shores or shelves (<4 kb from CGI). We used annotation files from Illumina and the UCSC genome browser to identify target genes and genomic features associated with each CpG (111, 112). A CpG site was in the promoter region if it was located 0–1500 bases upstream of a transcription start site. To identify CpGs associated with gene expression, we used results from eQTM analysis performed using the 450 K (113) and EPIC (114) chips in the Framingham Heart Study (FHS). For each set of CpGs selected for a given biomarker, we compared whether the set of CpGs were more likely to be eQTMs than CpGs that were not selected by comparing the proportion of eQTMs in the two groups. All enrichment analyses were conducted using a two-sided Fisher's exact test with a significance threshold of P-value < 0.05.

Gene-set analysis

To better understand the functional pathways of the CpGs selected for each SES biomarker, we conducted GO and KEGG pathway analyses using two distinct methods. First, we performed a gene-set analysis using GOmeth from the *missMethyl* package in R (115). This method accounts for potential sources of bias, including the uneven distribution of probes across the epigenome and the lack of independence between CpGs and their associated genes. However, a major limitation of this method is that it uses the chromosomal position of the CpGs to map them to proximal genes. This may not be optimal because CpG sites do not always act on the nearest gene but instead may influence the expression of distal genes. As a second method, we performed GO and KEGG enrichment analysis using the *clusterProfiler* package in R to identify the enrichment of BP (116). This method allows us to utilize methylation-gene expression pairs (eQTMs) to define the gene list, which emphasizes the potential functional role of the CpG. For each SES biomarker, our gene signal list consisted of all genes from an FHS-identified eQTM that mapped to any of the CpGs included in the biomarker. The background gene list included all genes that were part of an eQTM that overlapped with CpGs that were present in our study. FDR $q < 0.05$ was considered significant for all GO and KEGG pathway analyses.

Genetic annotation of biomarker CpGs

To identify CpGs at least partially influenced by proximal genetic variants, we annotated the biomarker-selected CpGs with results from *cis*- meQTL analysis. We used the EPIGEN MeQTL Database

for the EPIC chip (2,478 samples across three cohorts, meQTL declared if $P < 2.21 \times 10^{-4}$) (66) and the GoDMC Database for the 450 K chip (27,750 samples across 36 cohorts, meQTL declared if $P < 1 \times 10^{-5}$ in at least one cohort) (65).

CpG associations in published EWAS

We examined whether each biomarker CpG had been previously associated with any diseases or traits using The EWAS Catalog (117). For each CpG, all traits with $P < 0.0001$ are reported.

This study was approved by the Institutional Review Board at the University of Wisconsin-Madison (2019-0924-CP003). Informed consent was obtained from all participants.

Acknowledgments

The authors thank the other investigators, staff, and participants of the HRS and MESA studies for their valuable contributions.

Supplementary Material

Supplementary material is available at *PNAS Nexus* online.

Funding

This study was supported by the National Institute on Aging (NIA, R00 AG056599, P30 AG017266 [L.L.S.]), the National Heart, Lung, and Blood Institute (NHLBI, R01 HL141292 (J.A.S.)), and the National Human Genome Research Institute (NHGRI, T32 HG000040 (L.A.O.)). The NIA supports the Health and Retirement Study (U01AG009740) and its genotyping (RC2 AG0336495, RC4 AG039029). Research in MESA is supported by contracts 75N92020D00001, HHSN268201500003I, N01-HC-95159, 75N92020D00005, N01-HC-95160, 75N92020D00002, N01-HC-95161, 75N92020D00003, N01-HC-95162, 75N92020D00006, N01-HC-95163, 75N92020D00004, N01-HC-95164, 75N92020D00007, N01-HC-95165, N01-HC-95166, N01-HC-95167, N01-HC-95168, and N01-HC-95169 from the NHLBI and by grants UL1-TR-000040, UL1-TR-001079, and UL1-TR-001420 from the National Center for Advancing Translational Sciences (NCATS). The provision of genotyping data was supported in part by the National Center for Advancing Translational Sciences, CTSI grant UL1TR001881, and a National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK) grant (DK063491) to the Southern California Diabetes Endocrinology Research Center. Funding for SHARe genotyping was provided by NHLBI Contract N02-HL-64278. A full list of participating MESA investigators and institutions can be found at http://www.mesa-nhlbi.org [mesa-nhlbi.org]. Funding for MESA SHARe genotyping was provided by NHLBI Contract N02-HL-64278. Genotyping was performed at Affymetrix (Santa Clara, California, USA) and the Broad Institute of Harvard and MIT (Boston, Massachusetts, USA) using the Affymetrix Genome-Wide Human SNP Array 6.0. The MESA Epigenomics & Transcriptomics Studies were funded by NHLBI, NIDDK, and NIA grants R01HL101250, R01HL119962, R01DK101921, R01HL135009, and 1RF1AG054474. This article has been reviewed and approved by the MESA Publications and Presentations Committee. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIA, NHLBI, NHGRI, or NCATS.

Author Contributions

L.L.S. and J.A.S. conceptualized the study and methodology with input from L.A.O., S.M.R., J.D.F., and W.Z. S.M.R., L.A.O., and W.Z. conducted all formal analyses. T.M.H., J.D., and Y.L. contributed data and content area expertise. L.L.S. and J.A.S. co-wrote the initial draft together with L.A.O. and S.M.R., who contributed to the write-up of the methods and results. All authors read and approved the final manuscript.

Preprints

This manuscript was posted as a preprint: https://doi.org/10.1101/2024.05.21.24307701.

Data Availability

This study used restricted individual-level information from the HRS, and our contractual agreement does not permit public dissemination of the data. Details on how to access restricted data for the HRS can be found at https://hrs.isr.umich.edu/data-products/restricted-data. Data used in this analysis from MESA can be obtained through the MESA Data Coordinating Center (https://www.mesanhlbi.org/) and the database of Genotypes and Phenotypes (dbGaP; phs000209). Analysis code is posted on github: https://github.com/laurenschmitz/epigenetic-SES-biomarker.

References

1 Danese A, McEwen BS. 2012. Adverse childhood experiences, allostasis, allostatic load, and age-related disease. *Physiol Behav*. 106:29–39.
2 Cole SW. 2013. Social regulation of human gene expression: mechanisms and implications for public health. *Am J Public Health*. 103 Suppl 1:S84–S92.
3 Karlsson Linnér R, *et al.* 2017. An epigenome-wide association study meta-analysis of educational attainment. *Mol Psychiatry*. 22:1680–1690.
4 van Dongen J, *et al.* 2018. DNA methylation signatures of educational attainment. *NPJ Sci Learn*. 3:7.
5 McDade TW, *et al.* 2019. Genome-wide analysis of DNA methylation in relation to socioeconomic status during development and early adulthood. *Am J Phys Anthropol*. 169:3–11.
6 Laubach ZM, *et al.* 2019. Socioeconomic status and DNA methylation from birth through mid-childhood: a prospective study in project viva. *Epigenomics*. 11:1413–1427.
7 Dunn EC, *et al.* 2019. Sensitive periods for the effect of childhood adversity on DNA methylation: results from a prospective, longitudinal study. *Biol Psychiatry*. 85:838–849.
8 Schmitz LL, *et al.* 2023. Associations of early-life adversity with later-life epigenetic aging profiles in the multi-ethnic study of atherosclerosis. *Am J Epidemiol*. 192:1991–2005.
9 Crimmins EM, Klopack ET, Kim JK. 2024. Generations of epigenetic clocks and their links to socioeconomic status in the health and retirement study. *Epigenomics*. 16:1031–1042.
10 Schmitz LL, *et al.* 2022. The socioeconomic gradient in epigenetic ageing clocks: evidence from the multi-ethnic study of atherosclerosis and the health and retirement study. *Epigenetics*. 17:589–611.
11 Cerutti J, Lussier AA, Zhu Y, Liu J, Dunn EC. 2021. Associations between indicators of socioeconomic position and DNA methylation: a scoping review. *Clin Epigenetics*. 13:221.
12 Byun HM, *et al.* 2012. Temporal stability of epigenetic markers: sequence characteristics and predictors of short-term DNA methylation variations. *PLoS One*. 7:e39220.
13 Shang L, *et al.* 2023. meQTL mapping in the GENOA study reveals genetic determinants of DNA methylation in African Americans. *Nat Commun*. 14:2711.
14 Ladd-Acosta C. 2015. Epigenetic signatures as biomarkers of exposure. *Curr Environ Health Rep*. 2:117–125.

15 Ladd-Acosta C, Fallin MD. 2016. The role of epigenetics in genetic and environmental epidemiology. *Epigenomics*. 8:271–283.
16 Ladd-Acosta C, Fallin MD. 2019. DNA methylation signatures as biomarkers of prior environmental exposures. *Curr Epidemiol Rep*. 6:1–13.
17 Horvath S. 2013. DNA methylation age of human tissues and cell types. *Genome Biol*. 14:R115.
18 Bocklandt S, *et al.* 2011. Epigenetic predictor of age. *PLoS One*. 6: e14821.
19 Horvath S, *et al.* 2018. Epigenetic clock for skin and blood cells applied to Hutchinson Gilford Progeria syndrome and ex vivo studies. *Aging (Albany NY)*. 10:1758–1775.
20 Garagnani P, *et al.* 2012. Methylation of ELOVL2 gene as a new epigenetic marker of age. *Aging Cell*. 11:1132–1134.
21 Hannum G, *et al.* 2013. Molecular cell resource genome-wide methylation profiles reveal quantitative views of human aging rates. *Mol Cell*. 49:359–367.
22 Florath I, Butterbach K, Müller H, Bewerunge-Hudler M, Brenner H. 2014. Cross-sectional and longitudinal changes in DNA methylation with age: an epigenome-wide analysis revealing over 60 novel age-associated CpG sites. *Hum Mol Genet*. 23: 1186–1201.
23 Horvath S, *et al.* 2012. Aging effects on DNA methylation modules in human brain and blood tissue. *Genome Biol*. 13:R97.
24 Weidner CI, *et al.* 2014. Aging of blood can be tracked by DNA methylation changes at just three CpG sites. *Genome Biol*. 15: R24.
25 Zhang Y, *et al.* 2017. DNA methylation signatures in peripheral blood strongly predict all-cause mortality. *Nat Commun*. 8:1–11.
26 Levine ME, *et al.* 2018. An epigenetic biomarker of aging for lifespan and healthspan. *Aging (Albany NY)*. 10:573–591.
27 Lu AT, *et al.* 2019. DNA methylation GrimAge strongly predicts lifespan and healthspan. *Aging (Albany NY)*. 11:303–327.
28 Belsky DW, *et al.* 2020. Quantification of the pace of biological aging in humans through a blood test, the DunedinPoAm DNA methylation algorithm. *Elife*. 9:1–56.
29 Belsky DW, *et al.* 2022. DunedinPACE, A DNA methylation biomarker of the pace of aging. *Elife*. 11:e73420.
30 Duan R, Fu Q, Sun Y, Li Q. 2022. Epigenetic clock: a promising biomarker and practical tool in aging. *Ageing Res Rev*. 81:101743.
31 Faul JD, *et al.* 2023. Epigenetic-based age acceleration in a representative sample of older Americans: associations with aging-related morbidity and mortality. *Proc Natl Acad Sci U S A*. 120:e2215840120.
32 Levine ME, Lu AT, Bennett DA, Horvath S. 2015. Epigenetic age of the pre-frontal cortex is associated with neuritic plaques, amyloid load, and Alzheimer's disease related cognitive functioning. *Aging (Albany NY)*. 7:1198–1211.
33 Perna L, *et al.* 2016. Epigenetic age acceleration predicts cancer, cardiovascular, and all-cause mortality in a German case cohort. *Clin Epigenetics*. 8:64.
34 Marioni RE, *et al.* 2015. DNA methylation age of blood predicts all-cause mortality in later life. *Genome Biol*. 16:25.
35 Crimmins EM, Thyagarajan B, Levine ME, Weir DR, Faul J. 2021. Associations of age, sex, race/ethnicity, and education with 13 epigenetic clocks in a nationally representative U.S. sample: the health and retirement study. *J Gerontol A Biol Sci Med Sci*. 76:1117–1123.
36 Raffington L, *et al.* 2021. Socioeconomic disadvantage and the pace of biological aging in children. *Pediatrics*. 147:e2020024406.
37 McCartney DL, *et al.* 2018. Epigenetic prediction of complex traits and death. *Genome Biol*. 19:136.
38 Hillary RF, Marioni RE. 2021. MethylDetectR: a software for methylation-based health profiling. *Wellcome Open Res*. 5:283.
39 Cappozzo A, *et al.* 2022. A blood DNA methylation biomarker for predicting short-term risk of cardiovascular events. *Clin Epigenetics*. 14:121.
40 Gadd DA, *et al.* 2022. Epigenetic scores for the circulating proteome as tools for disease prediction. *Elife*. 11:e71802.
41 Zhang Y, *et al.* 2016. Smoking-associated DNA methylation markers predict lung cancer incidence. *Clin Epigenetics*. 8:127.
42 Walker RM, *et al.* 2019. Assessment of dried blood spots for DNA methylation profiling. *Wellcome Open Res*. 4:44.
43 Joehanes R, *et al.* 2016. Epigenetic signatures of cigarette smoking. *Circ Cardiovasc Genet*. 9:436–447.
44 Richmond RC, Suderman M, Langdon R, Relton CL, Davey Smith G. 2018. DNA methylation as a marker for prenatal smoke exposure in adults. *Int J Epidemiol*. 47:1120–1130.
45 Colicino E, *et al.* 2021. Blood DNA methylation biomarkers of cumulative lead exposure in adults. *J Expo Sci Environ Epidemiol*. 31: 108–116.
46 Guida F, *et al.* 2015. Dynamics of smoking-induced genome-wide methylation changes with time since smoking cessation. *Hum Mol Genet*. 24:2349–2359.
47 Moqri M, *et al.* 2023. Leading edge biomarkers of aging for the identification and evaluation of longevity interventions. *Cell*. 186:3758–3775.
48 Krieger N. 2001. A glossary for social epidemiology. *J Epidemiol Community Health*. 55:693–700.
49 Crimmins EM. 2020. Social hallmarks of aging: suggestions for geroscience research. *Ageing Res Rev*. 63:101136.
50 Link BG, Phelan J. 1995. Social conditions as fundamental causes of disease. *J Health Soc Behav*. Spec No:80–94.
51 Muscatell KA, Brosso SN, Humphreys KL. 2018. Socioeconomic status and inflammation: a meta-analysis. *Mol Psychiatry*. 25: 2189–2199.
52 Lupien SJ, McEwen BS, Gunnar MR, Heim C. 2009. Effects of stress throughout the lifespan on the brain, behaviour and cognition. *Nat Rev Neurosci*. 10:434–445.
53 Needham BL, *et al.* 2015. Life course socioeconomic status and DNA methylation in genes related to stress reactivity and inflammation: the multi-ethnic study of atherosclerosis. *Epigenetics*. 10:958–969.
54 Cohen S, Janicki-Deverts D, Chen E, Matthews KA. 2010. Childhood socioeconomic status and adult health. *Ann N Y Acad Sci*. 1186:37–55.
55 Phelan JC, Link BG. 2015. Is racism a fundamental cause of inequalities in health? *Annu Rev Sociol*. 41:311–330.
56 Case A, Lubotsky D, Paxson C. 2002. Economic status and health in childhood: the origins of the gradient. *Am Econ Rev*. 92: 1308–1334.
57 Haas S. 2008. Trajectories of functional health: the 'long arm' of childhood health and socioeconomic factors. *Soc Sci Med*. 66: 849–861.
58 Haas SA. 2007. The long-term effects of poor childhood health: an assessment and application of retrospective reports. *Demography*. 44:113–135.
59 Power C, Hertzmant C. 1997. Social and biological pathways linking early life and adult disease. *Br Med Bull*. 53:210–221.
60 Galobardes B, Smith GD, Lynch JW. 2006. Systematic review of the influence of childhood socioeconomic circumstances on risk for cardiovascular disease in adulthood. *Ann Epidemiol*. 16: 91–104.

61 Gavrilov LA, Gavrilova NS. 2004. Early-life programming of aging and longevity: the idea of high initial damage load (the HIDL hypothesis). *Ann N Y Acad Sci*. 1019:496–501.
62 Barker DJP. 1990. The fetal and infant origins of adult disease. *BMJ*. 301:1111.
63 Schmitz LL, Duque V. 2022. In utero exposure to the great depression is reflected in late-life epigenetic aging signatures. *Proc Natl Acad Sci U S A*. 119:e2208530119.
64 Hüls A, Czamara D. 2020. Methodological challenges in constructing DNA methylation risk scores. *Epigenetics*. 15:1–11.
65 Min JL, *et al.* 2021. Genomic and phenotypic insights from an atlas of genetic effects on DNA methylation. *Nat Genet*. 53: 1311–1321.
66 Villicaña S, *et al.* 2023. Genetic impacts on DNA methylation help elucidate regulatory genomic processes. *Genome Biol*. 24: 176.
67 Wozniak MB, *et al.* 2013. Integrative genome-wide gene expression profiling of clear cell renal cell carcinoma in Czech Republic and in the United States. *PLoS One*. 8:e57886.
68 Nones K, *et al.* 2014. Genome-wide DNA methylation patterns in pancreatic ductal adenocarcinoma reveal epigenetic deregulation of SLIT-ROBO, ITGA2 and MET signaling. *Int J Cancer*. 135: 1110–1118.
69 Mendelson MM, *et al.* 2018. Epigenome-wide association study of soluble tumor necrosis factor receptor 2 levels in the Framingham Heart Study. *Front Pharmacol*. 9:207.
70 Kartikasari AER, *et al.* 2022. Elevation of circulating TNF receptor 2 in cancer: a systematic meta-analysis for its potential as a diagnostic cancer biomarker. *Front Immunol*. 13:918254.
71 Sheng Y, Li F, Qin Z. 2018. TNF receptor 2 makes tumor necrosis factor a friend of tumors. *Front Immunol*. 9:1170.
72 Carow B, Rottenberg ME. 2014. SOCS3, a major regulator of infection and inflammation. *Front Immunol*. 5:58.
73 Wang YZ, *et al.* 2022. DNA methylation mediates the association between individual and neighborhood social disadvantage and cardiovascular risk factors. *Front Cardiovasc Med*. 9:848768.
74 Sudwarts A, *et al.* 2022. BIN1 is a key regulator of proinflammatory and neurodegeneration-related activation in microglia. *Mol Neurodegener*. 17:33.
75 Saha O, *et al.* 2024. The Alzheimer's disease risk gene BIN1 regulates activity-dependent gene expression in human-induced glutamatergic neurons. *Mol Psychiatry*. 29:2634–2646.
76 Ravi S, Shanahan MJ, Levitt B, Harris KM, Cole SW. 2024. Socioeconomic inequalities in early adulthood disrupt the immune transcriptomic landscape via upstream regulators. *Sci Rep*. 14:1255.
77 Elwenspoek MMC, *et al.* 2017. Proinflammatory T cell status associated with early life adversity. *J Immunol*. 199:4046–4055.
78 Chen MA, *et al.* 2021. Immune and epigenetic pathways linking childhood adversity and health across the lifespan. *Front Psychol*. 12:788351.
79 Ziol-Guest KM, Duncan GJ, Kalil A, Boyce WT. 2012. Early childhood poverty, immune-mediated disease processes, and adult productivity. *Proc Natl Acad Sci U S A*. 109:17289–17293.
80 McDade TW. 2023. Three common assumptions about inflammation, aging, and health that are probably wrong. *Proc Natl Acad Sci U S A*. 120:e2317232120.
81 Ryan CP, *et al.* 2024. Pregnancy is linked to faster epigenetic aging in young women. *Proc Natl Acad Sci U S A*. 121: e2317290121.
82 Juster FT, Suzman R. 1995. An overview of the health and retirement study. *J Hum Resour*. 30:S7.
83 Sonnega A, *et al.* 2014. Cohort profile: the health and retirement study (HRS). *Int J Epidemiol*. 43:576–585.
84 Crimmins E, Kim JK, Fisher J, Faul J. 2020. *HRS epigenetic clocks-release 1*. Ann Arbor, MI: Survey Research Center, Institute for Social Research, University of Michigan.
85 [dataset] Health and Retirement Study. 2024. RAND HRS Longitudinal File 2020 (V2) public use dataset. Produced by the RAND Center for the Study of Aging, with funding from the National Institute on Aging and the Social Security Administration. Santa Monica, CA.
86 [dataset] Health and Retirement Study, (Psychosocial and Lifestyle Questionnaire) public use dataset. 2022. Produced and distributed by the University of Michigan with funding from the National Institute on Aging (grant number NIA U01AG009740). Ann Arbor, MI.
87 [dataset] Health and Retirement Study, (Detailed Cross-Wave Geographic Information) restricted dataset. 2020. Produced and distributed by the University of Michigan with funding from the National Institute on Aging (grant number NIA U01AG009740). Ann Arbor, MI.
88 [dataset] Ailshire J, Mawhorter S, Choi EY. 2020. Contextual Data Resource (CDR): United States Decennial Census and American Community Survey Data, 1990–2018, Version 2.0. Los Angeles, CA: USC/UCLA Center on Biodemography and Population Health.
89 [dataset] Health and Retirement Study, (Industry and Occupation Data) restricted dataset. 2022. Produced and distributed by the University of Michigan with funding from the National Institute on Aging (grant number NIA U01AG009740). Ann Arbor, MI.
90 [dataset] Smith TW, Son J. 2014. The General Social Survey, Occupational Prestige Ratings. NORC, University of Chicago.
91 Crimmins E, Faul J, Thyagarajan B, Weir D. 2017. Venous blood collection and assay protocol in the 2016 Health and Retirement Study: 2016 Venous Blood Study (VBS).
92 Triche TJ, Weisenberger DJ, Van Den Berg D, Laird PW, Siegmund KD. 2013. Low-level processing of Illumina Infinium DNA Methylation BeadArrays. *Nucleic Acids Res*. 41:e90.
93 Pidsley R, *et al.* 2016. Critical evaluation of the Illumina MethylationEPIC BeadChip microarray for whole-genome DNA methylation profiling. *Genome Biol*. 17:208.
94 Houseman EA, *et al.* 2012. DNA methylation arrays as surrogate measures of cell mixture distribution. *BMC Bioinformatics*. 13:86.
95 [dataset] Health and Retirement Study, (Genetic Data) sensitive dataset. 2021. Produced and distributed by the University of Michigan with funding from the National Institute on Aging (grant number NIA U01AG009740). Ann Arbor, MI.
96 Zheng X, *et al.* 2012. A high-performance computing toolset for relatedness and principal component analysis of SNP data. *Bioinformatics*. 28:3326–3328.
97 [dataset] Health and Retirement Study, (APOE and Serotonin Transporter Alleles, Early Release) sensitive dataset. 2021. Produced and distributed by the University of Michigan with funding from the National Institute on Aging (grant number NIA U01AG009740). Ann Arbor, MI.
98 Hout M, Smith TW, Marsden PV. 2016. *Prestige and socioeconomic scores for the 2010 census codes*. GSS Methodological Report 124.
99 Roux AVD, *et al.* 2001. Neighborhood of residence and incidence of coronary heart disease. *N Engl J Med*. 345:99–106.
100 Mujahid MS, Diez Roux AV, Morenoff JD, Raghunathan T. 2007. Assessing the measurement properties of neighborhood scales: from psychometrics to ecometrics. *Am J Epidemiol*. 165:858–867.

101 Faul JD, *et al.* 2022. Trans-ethnic meta-analysis of interactions between genetics and early-life socioeconomic context on memory performance and decline in older Americans. *J Gerontol A Biol Sci Med Sci*. 77:2248–2256.

102 Vable AM, Gilsanz P, Nguyen TT, Kawachi I, Glymour MM. 2017. Validation of a theoretically motivated approach to measuring childhood socioeconomic circumstances in the health and retirement study. *PLoS One*. 12:e0185898.

103 Crimmins EM, Kim JK, Langa KM, Weir DR. 2011. Assessment of cognition using surveys and neuropsychological assessment: the health and retirement study and the aging, demographics, and memory study. *J Gerontol B Psychol Sci Soc Sci*. 66B:i162–i171.

104 Langa KM, Weir DR, Kabeto M, Sonnega A. 2023. *Langa-Weir classification of cognitive function (1995-2020)*. Ann Arbor, MI: Survey Research Center, Institute for Social Research, University of Michigan.

105 Liu Y, *et al.* 2013. Methylomics of gene expression in human monocytes. *Hum Mol Genet*. 22:5065–5074.

106 Smyth GK, Michaud J, Scott HS. 2005. Use of within-array replicate spots for assessing differential expression in microarray experiments. *Bioinformatics*. 21:2067–2075.

107 Pidsley R, *et al.* 2013. A data-driven approach to preprocessing Illumina 450K methylation array data. *BMC Genomics*. 14:293.

108 Price AL, *et al.* 2006. Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet*. 38: 904–909.

109 Zou H, Hastie T. 2005. Regularization and variable selection via the elastic net. *J R Stat Soc Series B Stat Methodol*. 67:301–320.

110 Friedman J, Hastie T, Tibshirani R. 2010. Regularization paths for generalized linear models via coordinate descent. *J Stat Softw*. 33:1.

111 Bibikova M, *et al.* 2011. High density DNA methylation array with single CpG site resolution. *Genomics*. 98:288–295.

112 Kent WJ, *et al.* 2002. The human genome browser at UCSC. *Genome Res*. 12:996–1006.

113 Yao C, *et al.* 2021. Epigenome-wide association study of whole blood gene expression in Framingham Heart Study participants provides molecular insight into the potential role of CHRNA5 in cigarette smoking-related lung diseases. *Clin Epigenetics*. 13:60.

114 Keshawarz A, *et al.* 2023. Expression quantitative trait methylation analysis elucidates gene regulatory effects of DNA methylation: the Framingham Heart Study. *Sci Rep*. 13:12952.

115 Phipson B, Maksimovic J, Oshlack A. 2016. missMethyl: an R package for analyzing data from Illumina's HumanMethylation450 platform. *Bioinformatics*. 32:286–288.

116 Yu G, Wang LG, Han Y, He QY. 2012. ClusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS*. 16:284–287.

117 Battram T, *et al.* 2022. The EWAS catalog: a database of epigenome-wide association studies. *Wellcome Open Res*. 7:41.