OXFORD

Human Molecular Genetics, 2023, Vol. 32, 5, 709–719
https://doi.org/10.1093/hmg/ddac171
Advance access publication date 28 July 2022
Original Article

Socioeconomic changes predict genome-wide DNA methylation in childhood

Jiaxuan Liu[1,2,†], Janine Cerutti[1,†], Alexandre A. Lussier[1,3,4], Yiwen Zhu[1,2], Brooke J. Smith[1], Andrew D.A.C. Smith[5] and Erin C. Dunn[1,3,4,6,*]

[1]Center for Genomic Medicine, Massachusetts General Hospital, Boston, MA 02114, USA
[2]Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, MA 02115, USA
[3]Department of Psychiatry, Harvard Medical School, Boston, MA 02115, USA
[4]Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge, MA 02142, USA
[5]Mathematics and Statistics Research Group, University of the West of England, Bristol BS8 1QU, UK
[6]Harvard Center on the Developing Child, Harvard University, Cambridge, MA 02138, USA
*To whom correspondence should be addressed at: Psychiatric and Neurodevelopmental Genetics Unit, Center for Genomic Medicine, Massachusetts General Hospital, 185 Cambridge Street, Simches Research Building 6th Floor, Boston, MA 02114, USA. Tel: +1 6177269387; Fax: +1 6177260830; Email: edunn2@mgh.Harvard.edu
†Joint First Authors, who contributed equally to the manuscript.

Abstract

Childhood socioeconomic position (SEP) is a major determinant of health and well-being across the entire life course. To effectively prevent and reduce health risks related to SEP, it is critical to better understand when and under what circumstances socioeconomic adversity shapes biological processes. DNA methylation (DNAm) is one such mechanism for how early life adversity 'gets under the skin'. In this study, we evaluated the dynamic relationship between SEP and DNAm across childhood using data from 946 mother–child pairs in the Avon Longitudinal Study of Parents and Children. We assessed six SEP indicators spanning financial, occupational and residential domains during very early childhood (ages 0–2), early childhood (ages 3–5) and middle childhood (ages 6–7). Epigenome-wide DNAm was measured at 412 956 cytosine-guanines (CpGs) from peripheral blood at age 7. Using an innovative two-stage structured life-course modeling approach, we tested three life-course hypotheses for how SEP shapes DNAm profiles—accumulation, sensitive period and mobility. We showed that changes in the socioeconomic environment were associated with the greatest differences in DNAm, and that middle childhood may be a potential sensitive period when socioeconomic instability is especially important in shaping DNAm. Top SEP-related DNAm CpGs were overrepresented in genes involved in pathways important for neural development, immune function and metabolic processes. Our findings highlight the importance of socioeconomic stability during childhood and if replicated, may emphasize the need for public programs to help children and families experiencing socioeconomic instability and other forms of socioeconomic adversity.

Introduction

Socioeconomic position (SEP) is a fundamental determinant of health and disease across the lifespan (1). As defined by Krieger *et al.* (1997) (2), SEP is an 'aggregate concept' composed of diverse components of economic and social well-being across individual-, household- and neighborhood-level domains, including both resources (e.g. weekly income) and rank-based characteristics (e.g. occupational prestige). SEP therefore can be measured across time by various indicators, like job stability, ability to afford basic household needs and neighborhood quality, which are known to play related, yet distinct roles in health and life outcomes (3–5).

Dozens of observational and quasi-experimental studies examining these indicators have shown that children growing up in low SEP families have increased risk for both short- and long-term cognitive, socioemotional, behavioral and physical/mental health deficits compared to their high SEP counterparts (6–9). Some of these SEP-related disparities are evident very early in development, starting shortly after birth (10–13). Yet, the biological mechanisms that explain these well-established SEP and health relationships remain relatively unknown, limiting our ability to disentangle specific pathways of pathophysiology and design targeted interventions.

In the past two decades, epigenetic studies have exploded as a means of potentially unraveling the biological pathways through which SEP 'gets under the skin'. Most epigenetic studies have focused on DNA methylation (DNAm) (14), which occurs when methyl groups are added to cytosines in the DNA sequence, typically within cytosine-guanine (CpG) dinucleotides (15). These DNA modifications do not alter the sequence of the genome, but can influence how genes are expressed in ways that can have important short and long-term health consequences (16).

Recent reviews summarizing the effects of SEP on epigenetic patterns suggest that SEP is linked to DNAm differences in childhood and adulthood (17–19). In fact, over 30 studies have found a relationship between childhood SEP and DNAm. However, less than a quarter of these studies were longitudinal by design (i.e. including repeated measures of SEP exposure across time). Further, less than half were epigenome-wide association studies (EWAS) analyzing SEP-related DNAm variations. In one recent comprehensive review of the SEP-DNAm literature, the number of significant, SEP-associated CpGs reported across prior EWAS

Received: March 25, 2022. **Revised:** July 1, 2022. **Accepted:** July 22, 2022

studies ranged from 1 to 2546 (median = 10), yet relatively no consistent patterns in SEP-associated DNAm changes emerged between studies (see Cerutti *et al.* (19)). One possible explanation for these mixed results is that studies have conflated both the type of SEP indicator measured and the timing of SEP measurement (19). Indeed, few studies have investigated the effects of SEP type and/or timing on DNAm, even though it is well known that both features of SEP can influence the extent of its impact (20).

Prior studies that have analyzed the associations between multiple types of SEP indicators and DNAm have found little to no overlap in DNAm changes across SEP measures (21–23), suggesting that different SEP indicators may result in distinct biological signatures and subsequent cascading health risks. Yet, it remains relatively unknown whether exposure to distinct SEP indicators (e.g. low household income vs. neighborhood disadvantage) during childhood impacts later DNAm to a similar extent.

Even fewer studies have investigated the impact of SEP timing on DNAm, likely because it is difficult to collect multiple, repeated measures across time in large, epigenetic datasets. In some notable exceptions, studies comparing the time-dependent effects of childhood SEP (24–27) on DNAm have found timing differences with respect to SEP's impact, consistent with the idea that there may be sensitive periods of elevated plasticity during childhood when adversity-induced biological changes are most likely to occur. However, whether different aspects of the socioeconomic environment across developmental stages differentially influence DNAm remains largely unexplored.

The current study aimed to address this gap by utilizing a large, longitudinal birth cohort with multiple, repeated measures of socioeconomic-related hardships assessed prospectively across childhood before epigenome-wide DNAm collection at age 7. We specifically sought to assess how different indicators of the socioeconomic environment (e.g. neighborhood quality, job loss, low household income) measured repeatedly across the first seven years of life associated with child epigenetic alterations. Given that different socioeconomic domains may impact health via related, but distinct pathways (4, 28), we analyzed exposure to seven distinct socioeconomic-related hardships. Additionally, because socioeconomic adversity could have multiple time-varying effects on DNAm, we tested three commonly examined hypotheses from the life-course epidemiology literature (29) to evaluate the circumstances under which childhood socioeconomic adversity associates with DNAm changes at age 7: 1) accumulation hypothesis, where the impact of low SEP increases with the number of time periods exposed, regardless of when it occurs; 2) sensitive period hypothesis, where the impact of low SEP is larger in magnitude during a certain developmental period compared to any other; and 3) mobility hypothesis, where the impact of SEP on DNAm is driven by an upward or downward change in SEP between adjacent developmental time periods.

Uncovering the dynamic relationships between SEP and DNAm across childhood will not only highlight the biological mechanisms driving the effects of SEP on long-term health, but also will offer clearer insights to guide targeted interventions aimed at reducing the negative consequences of socioeconomic-related adversity in childhood.

Results

Sample characteristics and prevalence of socioeconomic adversity

We analyzed data from 946 mother–child pairs from the Avon Longitudinal Study of Parents and Children (ALSPAC), a longitudinal birth-cohort in the United Kingdom (UK). Children included in our analytic sample were mostly White (97.1%) and from both sexes (49.9% female) (Supplementary Material, Table S1). Among the six SEP indicators analyzed (i.e. job loss, income reduction, low family income, financial hardship, major financial problem and neighborhood disadvantage), job loss was the least reported socioeconomic adversity (11.5% ever-exposed), and income reduction was the most common (73.8% ever-exposed) (Table 1). The prevalence of all adversities decreased over time (Table 1, Supplementary Material, Fig. S1). The six SEP indicators were moderately correlated with each other during all three childhood periods (Supplementary Material, Fig. S2), suggesting they captured distinct aspects of the socioeconomic environment.

Childhood socioeconomic adversities were associated with differential DNAm at 62 CpGs

We next examined possible time-dependent associations between each of the SEP indicators and DNAm at individual CpGs using a two-stage structured life-course modeling approach (SLCMA) (30–32), which identified the life-course hypothesis most supported in the observed data and estimated the magnitude of associations. In this and the following three sections, we summarize 1) the top CpGs associated with socioeconomic adversity, 2) the most selected life-course hypotheses, 3) the robustness of findings evaluated through a variety of sensitivity analyses and 4) the biological relevance of findings.

We identified 62 CpGs where exposure to socioeconomic adversity explained more than 3% of the variance in DNAm ($R^2 > 3\%$, Supplementary Material, Table S2). Most of the 62 CpGs were linked to the two least commonly-reported adversities in this sample: neighborhood disadvantage (17 CpGs) and job loss (15 CpGs, Table 2). Only four of the 62 CpGs identified using the R^2 cutoff also passed a false discovery rate (FDR) < 0.05 significance threshold, all of which were associated with neighborhood disadvantage (Table 2).

Of note, 61 of these CpGs showed the same direction of effect as that reported in at least two prior EWASs examining SEP and DNAm. Furthermore, 17 out of 62 (27%) CpGs showed at least a nominal ($P < 0.05$) association in at least two prior EWASs. Of these 17 CpGs, two (cg23685969 and cg19260606) exceeded a statistical significance threshold of FDR < 0.05 in at least one prior EWAS (Supplementary Material, Table S3, Supplementary Material, Fig. S3).

Mobility and sensitive period hypotheses were most often selected

The SLCMA allowed us to determine which of the following three life-course hypotheses were most supported in the observed data: accumulation, sensitive period and mobility (Fig. 1). Of the life-course hypotheses we tested, mobility and sensitive period effects showed the strongest associations with DNAm (Fig. 2A).

We first focused on the four socioeconomic adversities for which we tested all three life-course hypotheses (low family income, financial hardship, major financial problem and neighborhood disadvantage, Supplementary Material, Table S4). Here, 44 CpGs ($R^2 > 3\%$) were identified, of which four passed an FDR < 0.05 threshold. The majority of CpGs reflected mobility (20 CpGs) or sensitive period (22 CpGs) relationships. The most selected life-course hypothesis varied by socioeconomic adversity. Sensitive period hypotheses were selected for all nine CpGs identified from financial hardship, with middle childhood selected for eight of them (Fig. 2A). By contrast, mobility (worsening SEP) explained more DNAm variability resulting from neighborhood disadvantage (11 of 17 CpGs) and major financial problem (4 of 5 CpGs). The time period when mobility had the greatest

Table 1. Prevalence of exposure to socioeconomic adversity by developmental period in the ARIES analytic sample

	Job loss ($N = 667$)	Income reduction ($N = 711$)	Low family income ($N = 619$)	Financial hardship ($N = 697$)	Major financial problem ($N = 710$)	Neighborhood disadvantage ($N = 687$)
Very early childhood (0–2 years)	42 (6.3%)	458 (64.4%)	95 (15.4%)	127 (18.2%)	138 (19.4%)	83 (12.1%)
Early childhood (3–5 years)	32 (4.8%)	220 (30.9%)	79 (12.8%)	46 (6.6%)	69 (9.7%)	36 (5.2%)
Middle childhood (6–7 years)	18 (2.7%)	134 (18.9%)	55 (8.9%)	29 (4.2%)	60 (8.5%)	29 (4.2%)
Ever-exposed[a]	77 (11.5%)	525 (73.8%)	130 (21.0%)	147 (21.1%)	184 (25.9%)	98 (14.3%)
Average correlation over time[b]	0.49	0.34	0.87	0.70	0.50	0.80

The first four rows present the number (%) of children who were exposed to the specific type of socioeconomic adversity at each developmental period or ever exposed throughout the three periods. [a]Children who were exposed during at least one period were defined as ever-exposed for the specific type of socioeconomic adversity. [b]Polychoric correlations are presented, characterizing the average correlation over time within the given type of exposure. The average within-SEP correlations were moderate to high, suggesting these measures were variable across time, which allowed for detecting differences across periods. Exposures with correlations in excess of 0.90 typically cannot be used in the SLCMA.

Table 2. Summary of the SLCMA results for the 62 CpGs with $R^2 > 3\%$

Adversity	Number of $R^2 > 3\%$ CpGs	Range of R^2	Range of (*P*-values	Number of FDR < 0.05 CpGs
Neighborhood disadvantage	17	3.0–4.2%	1.3×10^{-7}—7.1×10^{-6}	4[a]
Job loss	15	3.1–3.7%	5.8×10^{-7}—8.8×10^{-6}	-
Low family income	13	3.0–3.8%	1.7×10^{-6}– 2.5×10^{-5}	-
Financial hardship	9	3.0–3.7%	5.9×10^{-7}– 8.5×10^{-6}	-
Major financial problem	5	3.0–3.8%	2.6×10^{-7}– 4.7×10^{-6}	-
Income reduction	3	3.0–3.3%	1.5×10^{-6}– 4.5×10^{-6}	-

The R^2 values reflect the increase in the variance of DNA methylation explained by the first hypothesis chosen after accounting for covariates. *P*-values were calculated using selective inference, which assesses the significance of the increase in R^2 explained. See Supplementary Material, Table S2 for the full list of the 62 CpGs. SLCMA = structured life-course modeling approach. [a]Four CpGs for neighborhood disadvantage passed an FDR < 0.05 significance threshold: cg20102336, cg08638097, cg23405172 and cg14212190.

impact differed across SEP indicators, with very early to early childhood most often selected for neighborhood disadvantage, and early to middle childhood most selected for major financial problem (Fig. 2A). Accumulation was only selected for two CpGs, linked to low family income. Of note, mobility hypotheses were selected for all four FDR-significant CpGs, with a worsening hypothesis (meaning downward mobility) selected for three of them (Supplementary Material, Table S2). Fig. 2B shows at these three CpGs, children exposed to worsening SEP had the greatest shift in DNAm as compared to children with other types of SEP trajectories, including those who had persistently low SEP, worsening SEP, improved SEP or persistently high SEP.

For our instability indicators (job loss and income reduction), which innately capture the effects of socioeconomic mobility, we only tested accumulation and sensitive period hypotheses (Supplementary Material, Table S4). The strongest evidence was again for sensitive period effects, with middle childhood (age 3–5) most selected for job loss (9 of 15 CpGs) and very early childhood (age 0–2) most selected for income reduction (2 of 3 CpGs, Fig. 2A). Accumulation was only selected for one CpG linked to job loss.

Overall, exposure to socioeconomic changes (captured through instability indicators or mobility hypotheses) was associated with, on average, a 3.8% difference in DNAm levels, explaining 3.4% of the variance in DNAm across CpG sites after controlling for covariates (Supplementary Material, Table S2). The same patterns were found at the epigenome-wide level, with most CpGs showing most variability in response to adversity from mobility and sensitive periods, rather than the accumulation of exposure across development (Supplementary Material, Fig. S4).

SLCMA results were robust to sensitivity analyses

Additional covariate adjustment had minimal impact on results

To assess residual bias in the identified SEP-DNAm associations and further ensure the robustness of our findings, we additionally controlled for time-invariant SEP indicators, population substructure estimated from epigenetic data, cord blood DNAm, genetic variation and exposure to the other five time-varying SEP indicators. After additional covariate adjustments, the life-course hypothesis selected by Least Angle Regression (LARS) remained the same for all 62 CpGs with $R^2 > 3\%$ (Supplementary Material, Table S5, Supplementary Material, Table S6). Almost all CpGs remained significant at the nominal $P < 0.05$ threshold after adjusting for time-invariant SEP indicators (60 CpGs), population substructure (61 CpGs), cord blood DNAm (61 CpGs) and exposure to the other five SEP indicators (62 CpGs, Supplementary Material, Table S5). The associations between socioeconomic adversities and DNAm were also independent of genetic variation previously linked to significant CpGs (Supplementary Material, Table S6).

Mobility hypotheses improved our ability to identify CpGs related to SEP changes

SEP mobility during childhood had never been previously tested on childhood DNAm to our knowledge. Therefore, we assessed the insights gained from adding mobility hypotheses. We re-analyzed the CpGs with an $R^2 > 3\%$ for low family income, financial hardship, major financial problem and neighborhood disadvantage using only accumulation and sensitive period hypotheses. Considering only accumulation and sensitive period hypotheses, we were unable to fully detect shifts in DNAm patterns related to changes in socioeconomic environment. When mobility hypotheses were omitted from the SLCMA analyses, there were minimal changes to the main results showing effects of sensitive period on DNAm ($n = 22$ CpGs), as the same hypothesis was selected with similar effect estimates (Supplementary Material, Table S7). However, for CpGs originally linked to mobility ($n = 20$), there were substantial attenuations in the estimated SEP-DNAm associations: sensitive period hypotheses were selected instead, which in turn, showed smaller R^2 (ranging from 0.04% to 1.6%) and much larger *P*-values (ranging from 0.001 to 0.84, Supplementary Material, Table S7).

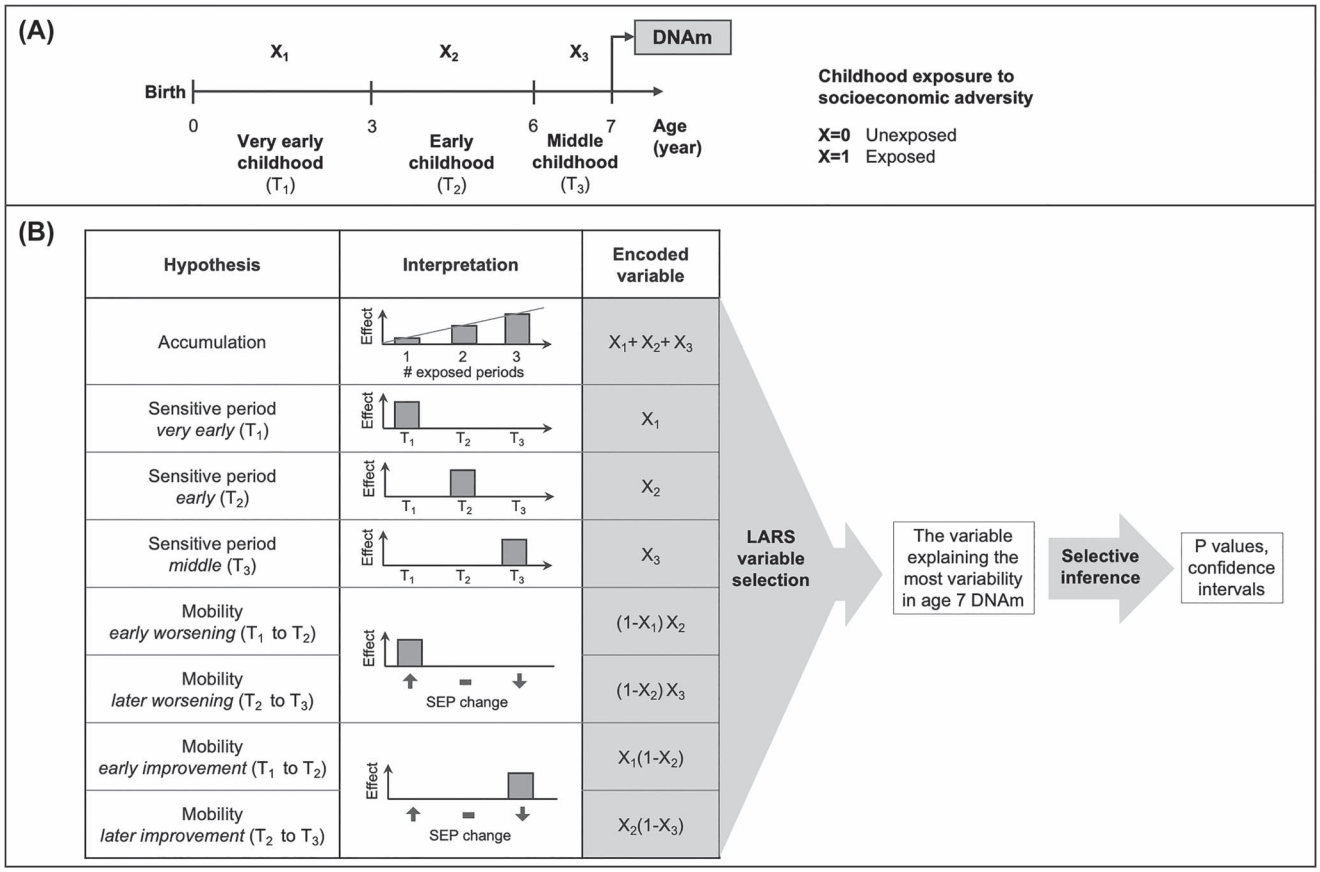

Figure 1. Study design and the conceptual life-course models used in the structured life-course modeling approach (SLCMA). (**A**) Measurement of childhood socioeconomic adversity (X) and DNA methylation (DNAm) over time (T). Exposure to socioeconomic adversities, or indicators of low socioeconomic position (SEP), was measured repeatedly across three childhood periods: very early (0–2 years, T_1), early (3–5 years, T_2) and middle childhood (6–7 years, T_3). DNAm was measured around age 7. (**B**) Illustration of the life-course hypotheses tested in the SLCMA, the least angle regression (LARS) variable selection procedure and selective inference test. Accumulation, sensitive period and mobility hypotheses were examined in this study. Accumulation assumes that the effect of low SEP increases with the number of exposed periods. Sensitive period assumes that low SEP is particularly impactful during one of the three time periods. Mobility assumes that changes in SEP across specific periods are particularly impactful. Early worsening and early improvement refer to adversity getting worse (↑SEP, i.e. increase in exposure) or better (↓SEP, i.e. decrease in exposure) from very early to early childhood, respectively; later worsening and later improvement refer to adversity getting worse or better from early to middle childhood, respectively. For each socioeconomic adversity, hypotheses were encoded into variables and then entered into the LARS variable selection procedure to identify the one explaining the most variability in DNAm at age 7 at each CpG site. We then performed post-selection inference to test the association between the selected variable and DNAm as well as estimate confidence intervals. See Supplemental Methods for more details about SLCMA.

These findings suggest that when the underlying association structure is misspecified, important DNAm signatures may not be identified.

EWAS of ever-exposed vs. never-exposed failed to identify time-dependent associations

To evaluate the loss (or gain) of information from the SLCMA compared to more conventional epigenetic approaches, we performed an epigenome-wide association study (EWAS) of any exposure to each type of SEP adversity before age 7 and DNAm, thus ignoring the timing or change of SEP over time. For 59 of the top 62 CpGs (including the 4 FDR-significant CpGs), the effect estimates from the SLCMA were larger in magnitude than those from EWAS (Supplementary Material, Fig. S5). In addition, no CpGs with an FDR < 0.05 were identified using EWAS of any exposure, meaning ever-exposed vs. never-exposed. These findings suggest the SLCMA was better able to identify developmentally sensitive effects of socioeconomic adversity on DNAm profiles, whereas EWAS might fail to detect signals if the true underlying hypothesis was time-dependent (24).

Biological significance of SLCMA findings

DNAm at top CpGs was weakly correlated across blood and brain

To examine the relevance of SEP-related DNAm pattern identified in peripheral blood tissues to brain health, we examined the correlation of DNAm at the top 62 CpGs in blood and brain samples, using data from the Blood Brain DNA Methylation Comparison Tool (http://epigenetics.essex.ac.uk/bloodbrain) (33). Overall, DNAm was weakly, but positively, correlated between blood and brain regions (Supplementary Material, Table S8) (prefrontal cortex: $r_{avg} = 0.06$; entorhinal cortex: $r_{avg} = 0.10$; superior temporal gyrus: $r_{avg} = 0.08$; cerebellum: $r_{avg} = 0.09$). Some top CpGs showed particularly strong correlations between blood and brain (e.g. cg24938210, $r = 0.78$ to 0.81 across brain regions).

Distinct biological pathways emerged across SEP indicators

The top 62 CpGs showed no significant differences in distributions of genomic features, CpG island locations or enhancers, as compared to all tested CpGs (Chi-squared tests $P > 0.05$, Supplementary Material, Fig. S6).

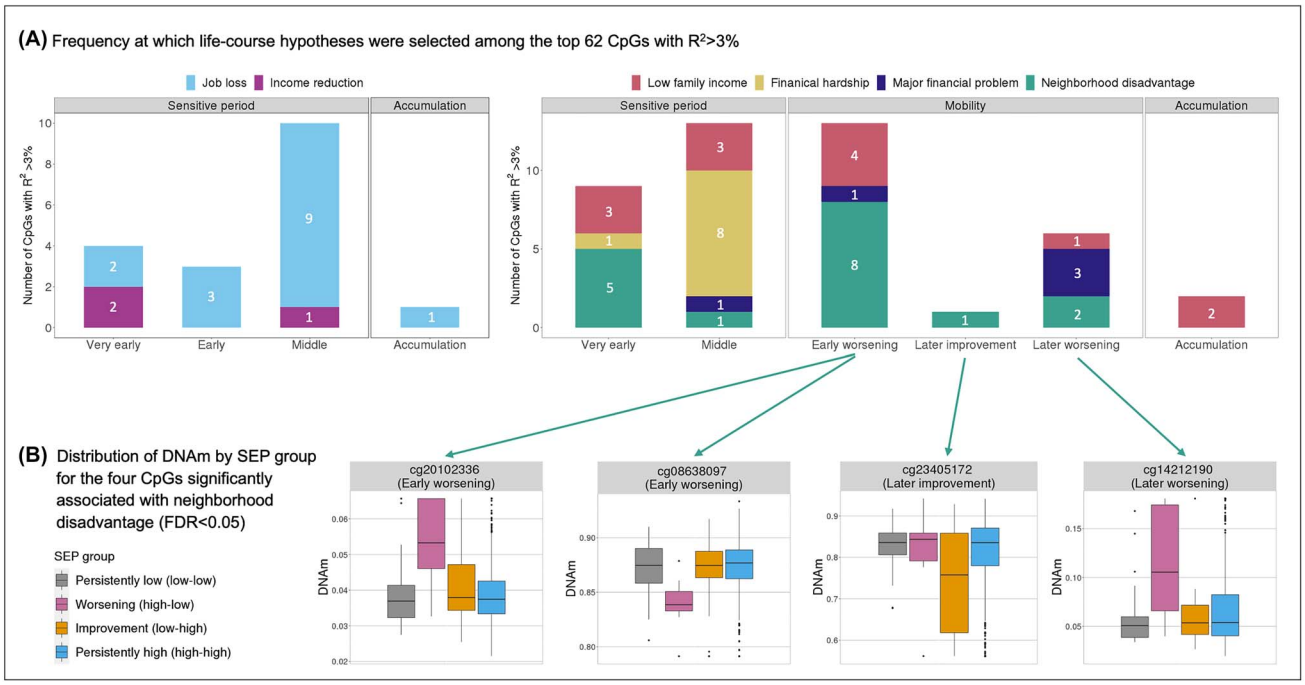

Figure 2. Mobility and sensitive period hypotheses were most often selected among the top 62 CpGs linked with socioeconomic adversity (or socioeconomic position, SEP) that explained > 3% variance in DNA methylation (DNAm). (**A**) Frequency at which each life-course hypothesis was selected among the 62 CpGs. For job loss and income reduction, we tested accumulation and sensitive period hypotheses, and middle childhood was the most selected hypothesis. For the other four socioeconomic adversities, we tested accumulation, sensitive period and mobility hypotheses. Mobility hypotheses, specifically worsening SEP, were most selected. Very early, Early and Middle refer to sensitive period hypotheses related to the three childhood periods: very early (0–2 years), early (3–5 years) and middle childhood (6–7 years). Early worsening/improvement refers to mobility hypotheses for changes between very early and early childhood, and later worsening/improvement refers to mobility hypotheses for changes between early and middle childhood. (**B**) For the four CpGs associated with neighborhood disadvantage at an FDR < 0.05, SEP mobility group implied by the selected mobility hypothesis showed the greatest shift in DNAm. The distribution of DNAm by SEP mobility group is shown in boxplots, where the center line indicates the median, box limits indicate the 25th and 75th percentiles, whiskers extend up to 1.5 inter-quartile range (IQR) from the box limits and individually plotted data points were values further than 1.5 IQR from the box limits. SEP mobility group was defined based on the exposure status at two consecutive childhood periods (very early and early, or early and middle) involved in the mobility hypothesis chosen for each CpG; persistently low was defined as being exposed to socioeconomic adversity during both periods; worsening SEP was defined as being unexposed during the former period but exposed during the later period; improving SEP was defined as being exposed during the former period but unexposed during the later period; persistently high was defined as being unexposed to socioeconomic adversity during both periods.

Gene set enrichment showed that SEP-related DNAm patterns were more likely to occur within or near genes involved in neural system regulation, developmental processes, immune functions, metabolic processes, substance localization and membrane transport (Supplementary Material, Fig. S7, Supplementary Material, Fig. S8). However, there was little overlap observed in the significant gene ontology (GO) terms across SEP indicators (Supplementary Material, Fig. S7), except for one GO term (morphogenesis of a branching epithelium), which emerged in the enrichment analysis for both financial hardship and major financial problem. These findings suggest different socioeconomic adversities may lead to shifts in distinct biological pathways.

Discussion

The main finding from this study was that changes in the socioeconomic environment may coincide with subsequent changes at a biological level as measured through DNAm signatures. Reports of a change in the socioeconomic environment, particularly worsening neighborhood quality (i.e. mobility) and parental job loss during middle childhood (i.e. sensitive period), were associated, on average, with a 3.8% difference in DNAm levels. These patterns were detected even after accounting for other dimensions of the socioeconomic environment, ancestry, DNAm levels at birth and genetic variation. To our knowledge, this study is the first to evaluate the role of socioeconomic changes in relation to epigenome-wide DNAm within childhood.

Our study extends prior literature on the effects of childhood SEP, providing new insights into the biological embedding of the socioeconomic environment. Only three studies to our knowledge have examined the relationship between socioeconomic mobility and DNAm (22, 34, 35). Each of these three studies included just two timepoints of SEP measures, one in childhood and another in adulthood, and only assessed DNAm in adulthood. Our results suggest that acute changes in children's socioeconomic environment, compared to exposure to more stable socioeconomic adversity, might play a role in shaping DNAm profiles in childhood as early as age 7. Although our study is the first to measure the impact of exposure to socioeconomic changes on DNAm levels in childhood, our results parallel previous findings on SEP-related outcomes in the child development literature. For example, non-epigenetic studies focused on other SEP-related outcomes in childhood have shown that an episode of parental job loss may have a larger impact on child health and behavior than stable employment in low-income jobs (36–38). Indeed, the developmental literature largely suggests that children benefit from stable, predictable environments (39–41) and that changes in the socioeconomic environment can impact cognitive development and other mechanisms implicated in future risk of health and behavioral problems (36–38, 42, 43). Future

studies are needed to replicate our findings and investigate how SEP-associated DNAm alterations may influence subsequent health and behavioral outcomes. Insights from such studies will be critical to discern whether SEP-related DNAm changes influence children's vulnerability to disease and other negative health/behavioral outcomes.

We found more evidence for the importance of the developmental timing of SEP on DNAm rather than its accumulation. These results parallel previous findings from the ALSPAC cohort (24) and elsewhere (44), suggesting that sensitive period effects can be detected in the epigenome. Our results also specifically point to the importance of middle childhood as a potential sensitive period when the socioeconomic environment might be particularly impactful. SEP plays an important role during school-age years (39, 45), corresponding to our middle childhood time period findings, when children in the cohort began school. Socioeconomic disruptions during school-age years may lead to changes in parent–child interactions, afterschool care center attendance or extracurricular activities.

Consistent with prior epigenome-wide studies (21, 22), we found little overlap between the top CpGs across SEP domains, suggesting that various aspects of the SEP construct may trigger distinct mechanisms that lead to different alterations in DNAm patterns (19, 46). Across our six SEP indicators, the greatest number of detected CpGs (17 of 62) were related to neighborhood disadvantage, with 4 being the only CpGs to pass an FDR < 0.05 significance threshold. These findings point to the important role that neighborhood-level indicators, including more ubiquitous social and physical exposures experienced daily by larger segments of a population, may play in shaping the epigenome during child development. For example, we found that the DNAm alterations linked to neighborhood disadvantage were more likely to occur in genes related to peroxisomes, which are a key component of the biological response to various environmental pollutants (47). By contrast, we found that experiences of financial hardship (e.g. difficulty in affording common household necessities like food, clothing, heat and rent) and income reduction were linked to biological pathways related to diet quality, such as nutrient transport and metabolic processes. Overall, different clusters of biological pathways emerged across distinct DNAm-associated SEP domains, suggesting that socioeconomic adversities may affect child health through multiple mechanisms.

Many of the genes in which our top CpGs were located on or near have been linked to human health and diseases. For example, *OAS3*, in which our most significant CpG (cg20102336) resides, encodes an enzyme that plays a critical role in innate antiviral response (48) and has been linked with the incidence and severity of illness caused by coronavirus disease 2019 (COVID-19) (49, 50). *TGFBR3*, the nearest gene to another significant CpG (cg08638097), encodes a key receptor in the transforming growth factor-β (TGF-β) superfamily signaling pathways and has been implicated in various human cancers including prostate cancer and bladder cancer (51–54). Furthermore, one of the top CpGs showing strong evidence of replication across studies (cg24121967; same direction of effect and $P < 0.05$ in 8 and 3 other studies, respectively) was located in a putative oncogene *MYEOV* whose overexpression has been documented in many cancers such as gastric cancer (55), myeloma (55) and pancreatic cancer (56). These findings suggest that early life socioeconomic adversities are associated with biological disruptions that may ultimately lead to a wide constellation of health risks later in life.

While the current study uncovered many insights into SEP and DNAm associations, a major unanswered question is whether these DNAm changes are adaptive or maladaptive, in both the short- and long-term. Teicher and others have noted that early neurobehavioral changes that occur in response to experiences of childhood adversity often enhance immediate survival at the cost of long-term functioning (57). Thus, are specific epigenomic fluctuations in the face of family socioeconomic adversity reflective of increased risk, resilience or both? Although we found DNAm differences when comparing children who were exposed vs. unexposed to socioeconomic adversity, we do not know if these SEP-induced shifts represent systemic alterations of biological functions across tissue types, which may cause key impairments that lead to behavioral changes and increase disease risks. With existing publicly available data, we could only compare the potential implications of our findings to DNAm levels in brain tissue. Additional research comparing DNAm levels between different tissues is warranted to better understand the systemic effects of socioeconomic hardship.

Should these DNAm markers of socioeconomic adversity be replicated and identified as harmful (rather than adaptive) to health, our findings suggest at least two paths forward for prevention and intervention. First, our results suggest that children and families, especially lower-income families who may lack a safety-net to draw from during times of parental job loss or other socioeconomic transitions (58), might benefit from extending policies and social programs aimed at minimizing socioeconomic instability, such as the Supplemental Nutrition Assistance Program (59) and the American Families Plan (60). Second, prevention programs aimed at promoting socioeconomic stability during childhood might benefit from adopting a multisystemic approach that considers the social determinants of health (61) at multiple levels (62). In fact, interventions at the household-level (e.g. parenting-based) and neighborhood-level (e.g. community-based) have revealed measurable biological impacts on children's DNAm profiles (63, 64) and on other biomarkers (65–67).

The current study should be interpreted in light of several limitations. First, like other epigenome-wide studies of this sample size, we identified few specific CpGs passing a stringent correction for multiple testing. However, following the recent movement to move beyond *P*-value thresholds alone (68, 69), we explored the patterns and implications of SEP-related DNAm profiles among top CpGs passing an effect-size-based threshold. The top CpGs passing this threshold were robust to various sensitivity analyses, and there was consistent evidence for the patterns of CpGs observed, with the majority showing effects in the same direction as previously published findings and two CpGs showing significance in other studies after correcting for multiple testing. Nevertheless, the results from individual CpGs should be interpreted with caution and validated in larger samples. Second, because this was a population-based sample, extreme cases of socioeconomic disadvantage were likely underrepresented in the ALSPAC cohort. Our results suggest that more severe forms of adversity may have more potent effects, as we identified most top DNAm CpGs (32 out of 62) from the two socioeconomic adversities that showed the lowest prevalence (job loss and neighborhood disadvantage). Future research in populations with more diverse SEP distributions capturing a wider gradient (i.e. extreme poverty) will help fully disentangle the impact of SEP on DNAm patterns. Third, the ALSPAC cohort is mostly White, which limits its generalizability of these findings to other individuals and populations of non-European descent. Prior studies (see review (70)) show ancestry-related variation in DNA methylation that may lead to differences in gene regulation across populations. Thus, future replication efforts are needed in more diverse and

representative populations. Finally, this study was observational and based on self-report measures of SEP, which could have been influenced by reporter bias, wherein participant responses may have been shaped by factors like social desirability or recall biases, leading to over- or underestimates of observed associations (71). Although self-reporting bias is common among survey/questionnaire data in observational studies, previous research has shown that individual-level SEP measures like education and income, compared to more objective measures assessed at the census tract-level, can more accurately capture the impact of SEP on a number of health outcomes, such as blood pressure and height (72). Future randomized experiments will help determine the causal effect of socioeconomic adversity on DNAm.

In summary, this study adds to a growing literature showing that early-life socioeconomic adversity can leave biological memories in the form of DNAm differences in childhood. Uniquely, our findings on socioeconomic mobility and instability suggest changes in the socioeconomic environment during childhood are especially impactful and associated with epigenetic disruptions related to various health outcomes. Ultimately, these findings will enable researchers to build toward better intervention and prevention efforts aimed at reducing socioeconomic disparities and promoting health across the life course.

Materials and Methods

Sample and procedures

Data came from the Accessible Resources for Integrated Epigenomics Studies (ARIES) (73), a subsample of 1018 mother–child pairs from the ALSPAC. ALSPAC is a prospective, longitudinal birth cohort in the UK designed to investigate genetic and environmental determinants of health across the lifespan (74–76). Women living in the county of Avon, UK, with estimated delivery dates between April 1991 and December 1992 were invited to participate. Mother–child pairs in the ARIES were randomly selected from ALSPAC based on availability of DNA samples across five waves of data collection (73). We analyzed data from 946 singletons in ARIES with blood-based DNAm profiles generated at age 7. Ethical approval for the study was obtained from the ALSPAC Ethics and Law Committee and the Local Research Ethics Committee. Note that the ALSPAC study website contains details of all the data that is available through a fully searchable data dictionary and variable search tool (http://www.bristol.ac.uk/alspac/researchers/our-data). See Supplemental Methods for full ALSPAC details.

Measures

Early-life socioeconomic position (SEP)

We analyzed six SEP indicators, spanning financial, occupational and residential domains: 1) job loss, 2) income reduction, 3) low family income, 4) financial hardship, 5) major financial problem and 6) neighborhood disadvantage. These were the only available, time-varying SEP indicators that were measured repeatedly via maternal report through mailed questionnaires during three developmental time periods (Fig. 1A): very early childhood (0–2 years), early childhood (3–5 years) and middle childhood (6–7 years).

For each SEP indicator, children were classified as exposed or unexposed at each period, using criteria described in Supplemental Methods (Supplementary Material). With these repeated, self-reported SEP indicators, we could identify changes occurring between time-periods for indicators capturing time-varying status of SEP. For job loss and income reduction, the measures inherently captured change within a certain developmental period, because they asked about socioeconomic mobility. To distinguish job loss and income reduction from other indicators, we refer to them throughout the manuscript as 'instability indicators'.

DNA methylation (DNAm)

DNAm was measured from peripheral blood at age 7 using the Illumina Infinium HumanMethylation450 BeadChip microarray (Illumina, San Diego, CA). DNAm wet laboratory procedures, preprocessing analyses and quality control are described in Supplemental Methods (Supplementary Material). A total of 412 956 CpGs on autosomal chromosomes passed quality control and were included in this analysis. For each CpG, DNAm level is expressed as a 'beta' value (β-value) ranging from 0 to 1, which represents the proportion of cells methylated at each interrogated CpG.

Covariates

To adjust for baseline demographic differences in ARIES and technical variation in DNAm assessment, we controlled for the following variables measured at birth in all analyses: child age in months at blood draw, child race/ethnicity, child sex, child birthweight, maternal age, number of previous pregnancies, sustained maternal smoking during pregnancy and cell type proportions estimated using the Houseman method (77). Details can be found in the Supplemental Methods (Supplementary Material).

Data analysis

All analysis code is available through our GitHub page: https://github.com/thedunnlab/sep-dnam.

Structured life-course modeling approach

We used the two-stage structured life-course modeling approach (SLCMA) (30–32) to evaluate the time-dependent effects of socioeconomic adversity on DNAm. SLCMA is a method that leverages repeated exposure data to simultaneously investigate the relationship between exposure and outcome under multiple a priori-defined life-course hypotheses. In our analyses, we tested three life-course hypotheses, described previously, which were parameterized as follows (Fig. 1B).

First, to test the accumulation hypothesis, we created a sum score (ranging from 0 to 3), which captured the number of time periods across the three developmental stages that children were exposed. Second, to test the sensitive period hypothesis, we created three binary variables, one for each of the three developmental periods, to classify children's exposure status (0 = unexposed during the period; 1 = exposed during that period). Third, to test the mobility hypothesis, we created a pair of indicator variables for change in SEP between very early and early childhood, and a pair of indicator variables for change in SEP between early and middle childhood. Each pair consisted of an indicator variable for worsening (1 = change from unexposed to exposed, 0 = other) and an indicator variable for improvement (1 = change from exposed to unexposed, 0 = other).

We tested all three hypotheses for low family income, financial hardship, major financial problem and neighborhood disadvantage. Only the accumulation and sensitive period hypotheses were tested for job loss and income reduction, as these two instability indicators inherently reflect SEP changes (Supplementary Material, Table S4).

We performed the SLCMA in two stages: 1) life-course hypothesis model selection followed by 2) post-selection inference (Fig. 1B,

Supplemental Methods). In the first stage, we tested the variables described above using a Least Angle Regression (LARS) variable selection procedure (78) to identify the life-course hypothesis most supported in the observed data (i.e. explaining the most variation in DNAm). In the second stage, we used selective inference (30, 79) to test the association between the selected variable and DNAm and estimate confidence intervals.

Defining CpGs of interest

We used two thresholds to identify associations between SEP and CpG CpGs for further investigation. Given recent recommendations discouraging the use of *P*-values alone for statistical inference (68, 69), we used an effect-size-based threshold of $R^2 > 3\%$, meaning that the SEP exposure explained more than 3% of the variance in DNAm. This cutoff was selected based on the effect sizes observed in previous epigenome-wide analyses of childhood adversity in ALSPAC (24, 26) and other well-established environmental exposures, including tobacco smoking (80). We also performed multiple-testing correction using the Benjamini-Hochberg method (81) at a 5% FDR to assess the significance of top CpGs.

Sensitivity analyses

We conducted three sensitivity analyses to evaluate the robustness of our SLCMA results. First, we additionally controlled for 1) time-invariant SEP indicators (e.g. maternal education at baseline), 2) population substructure estimated from epigenetic data, 3) cord blood DNAm (to account for differences in DNAm that might have been present at birth), 4) genetic variation (at methylation quantitative trait loci, or mQTL) or 5) exposure to the other five time-varying SEP indicators. Second, we reran the analyses of the CpGs with an $R^2 > 3\%$ for low family income, financial hardship, major financial problem and neighborhood disadvantage using only accumulation and sensitive period hypotheses and compared the results from analysis with and without mobility tested. Third, we performed an EWAS of any exposure to each type of SEP adversity before age 7 and DNAm and compared the findings with SLCMA results. See (Supplemental Methods, Supplementary Material) for details.

Secondary analyses

To interpret our findings and place them in the context of prior literature, we conducted two secondary analyses. First, we compared the effect estimates of $R^2 > 3\%$ CpGs to those reported in previous SEP-related EWAS studies (19) (Supplemental Methods, Supplementary Material). Second, we also evaluated the biological significance of our findings by examining the correlation between DNAm in blood and brain tissue for the $R^2 > 3\%$ CpGs and testing for the enrichment of genomic features, regulatory elements and Gene Ontology (GO) terms (Supplemental Methods, Supplementary Material).

Supplementary material

Supplementary Material is available at *HMG* online.

Acknowledgements

This publication is the work of the authors, each of whom serve as guarantors for the contents of this paper. We are extremely grateful to all the families who took part in the ALSPAC study, the midwives for their help in recruiting them and the whole ALSPAC team, which includes interviewers, computer and laboratory technicians, clerical workers, research scientists, volunteers, managers, receptionists and nurses. Supplementary funding to generate DNA methylation data which is included in ARIES has been obtained from the MRC, ESRC, NIH and other sources; ARIES is maintained under the auspices of the MRC Integrative Epidemiology Unit at the University of Bristol (MC_UU_12013/2 and MC_UU_12013/8); a comprehensive list of grants funding is available on the ALSPAC website (http://www.bristol.ac.uk/alspac/external/documents/grant-acknowledgements.pdf).

Conflict of Interest statement. None declared.

Funding

Russell Sage Foundation and the Ford Foundation (# 96-17-05 to E.C.D., awarded July 2017; any opinions expressed are those of the author(s) alone and should not be construed as representing the opinions of either foundation); National Institute of Mental Health at the National Institutes of Health (grant number R01MH113930 to E.C.D); UK Medical Research Council and the Wellcome Trust (Grant ref: 217065/Z/19/Z); University of Bristol (to ALSPAC); BBSRC (BBI025751/1 and BB/I025263/1 to ARIES).

Ethical standards

All ethical guidelines were followed per research involving use of human subjects. Ethical approval for the study was obtained from the ALSPAC Ethics and Law Committee and the Local Research Ethics Committee. This study was approved with oversight by the Mass General Brigham Institutional Review Boards (IRB) (Protocol ID 2017P001110).

References

1. Braveman, P. and Gottlieb, L. (2014) The social determinants of health: it's time to consider the causes of the causes. *Public Health Rep.*, **129**(Suppl 2), 19–31.
2. Krieger, N., Williams, D.R. and Moss, N.E. (1997) Measuring social class in US public health research: concepts, methodologies, and guidelines. *Annu. Rev. Public Health*, **18**, 341–378.
3. Vable, A.M., Gilsanz, P., Nguyen, T.T., Kawachi, I. and Glymour, M.M. (2017) Validation of a theoretically motivated approach to measuring childhood socioeconomic circumstances in the health and retirement study. *PLoS One*, **12**, e0185898.
4. Elgar, F.J., McKinnon, B., Torsheim, T., Schnohr, C.W., Mazur, J., Cavallo, F. and Currie, C. (2016) Patterns of socioeconomic inequality in adolescent health differ according to the measure of socioeconomic position. *Soc. Indic. Res.*, **127**, 1169–1180.
5. d'Errico, A., Ricceri, F., Stringhini, S., Carmeli, C., Kivimaki, M., Bartley, M., McCrory, C., Bochud, M., Vollenweider, P., Tumino, R. *et al.* (2017) Socioeconomic indicators in epidemiologic research: a practical example from the LIFEPATH study. *PLoS One*, **12**, e0178071.
6. Caspi, A., Taylor, A., Moffitt, T.E. and Plomin, R. (2000) Neighborhood deprivation affects children's mental health: environmental risks identified in a genetic design. *Psychol. Sci.*, **11**, 338–342.
7. Duncan, G.J., Magnuson, K. and Votruba-Drzal, E. (2017) Moving beyond correlations in assessing the consequences of poverty. *Annu. Rev. Psychol.*, **68**, 413–434.
8. Bradley, R.H. and Corwyn, R.F. (2002) Socioeconomic status and child development. *Annu. Rev. Psychol.*, **53**, 371–399.

9. Hackman, D.A., Farah, M.J. and Meaney, M.J. (2010) Socioeconomic status and the brain: mechanistic insights from human and animal research. *Nat. Rev. Neurosci.*, **11**, 651–659.
10. Noble, K.G., Houston, S.M., Brito, N.H., Bartsch, H., Kan, E., Kuperman, J.M., Akshoomoff, N., Amaral, D.G., Bloss, C.S., Libiger, O. *et al.* (2015) Family income, parental education and brain structure in children and adolescents. *Nat. Neurosci.*, **18**, 773–778.
11. Stringhini, S., Carmeli, C., Jokela, M., Avendaño, M., Muennig, P., Guida, F., Ricceri, F., d'Errico, A., Barros, H., Bochud, M. *et al.* (2017) Socioeconomic status and the 25 × 25 risk factors as determinants of premature mortality: a multicohort study and meta-analysis of 1.7 million men and women. *Lancet*, **389**, 1229–1237.
12. Chetty, R., Stepner, M., Abraham, S., Lin, S., Scuderi, B., Turner, N., Bergeron, A. and Cutler, D. (2016) The association between income and life expectancy in the United States, 2001-2014. *JAMA*, **315**, 1750–1766.
13. Marmot, M. (2020) Health equity in England: the Marmot review 10 years on. *BMJ*, **368**, m693.
14. Szyf, M., McGowan, P. and Meaney, M.J. (2008) The social environment and the epigenome. *Environ. Mol. Mutagen.*, **49**, 46–60.
15. Jin, B., Li, Y. and Robertson, K.D. (2011) DNA methylation: superior or subordinate in the epigenetic hierarchy? *Genes Cancer*, **2**, 607–617.
16. Portela, A. and Esteller, M. (2010) Epigenetic modifications and human disease. *Nat. Biotech.*, **28**, 1057–1068.
17. McCrory, C., McLoughlin, S. and O'Halloran, A.M. (2019) Socioeconomic position under the microscope: getting 'under the skin' and into the cells. *Curr. Epidemiol. Rep.*, **6**, 403–411.
18. Wood, N.M., Trebilco, T. and Cohen-Woods, S. (2020) Scars of childhood socioeconomic stress: a systematic review. *Neurosci. Biobehav. Rev.*, **118**, 397–410.
19. Cerutti, J.K., Lussier, A.A., Zhu, Y., Liu, J. and Dunn, E.C. (2021) Associations between indicators of socioeconomic position and DNA methylation: a scoping review. *Clin. Epigenetics*, **13**(1), 1–20.
20. Smith, K.E. and Pollak, S.D. (2021) Rethinking concepts and categories for understanding the neurodevelopmental effects of childhood adversity. *Perspect. Psychol. Sci.*, **16**, 67–93.
21. Bush, N.R., Edgar, R.D., Park, M., MacIsaac, J.L., McEwen, L.M., Adler, N.E., Essex, M.J., Kobor, M.S. and Boyce, W.T. (2018) The biological embedding of early-life socioeconomic status and family adversity in children's genome-wide DNA methylation. *Epigenomics*, **10**(11), 1445–1461.
22. McDade, T.W., Ryan, C.P., Jones, M.J., Hoke, M.K., Borja, J., Miller, G.E., Kuzawa, C.W. and Kobor, M.S. (2019) Genome-wide analysis of DNA methylation in relation to socioeconomic status during development and early adulthood. *Am. J. Phys. Anthropol.*, **169**, 3–11.
23. Alfano, R., Guida, F., Galobardes, B., Chadeau-Hyam, M., Delpierre, C., Ghantous, A., Henderson, J., Herceg, Z., Jain, P., Nawrot, T.S. *et al.* (2019) Socioeconomic position during pregnancy and DNA methylation signatures at three stages across early life: epigenome-wide association studies in the ALSPAC birth cohort. *Int. J. Epidemiol.*, **48**, 30–44.
24. Dunn, E.C., Soare, T.W., Zhu, Y., Simpkin, A.J., Suderman, M.J., Klengel, T., Smith, A., Ressler, K.J. and Relton, C.L. (2019) Sensitive periods for the effect of childhood adversity on DNA methylation: results from a prospective, longitudinal study. *Biol. Psychiatry*, **85**, 838–849.
25. Marini, S., Davis, K.A., Soare, T.W., Zhu, Y., Suderman, M.J., Simpkin, A.J., Smith, A.D.A.C., Wolf, E.J., Relton, C.L. and Dunn, E.C. (2020) Adversity exposure during sensitive periods predicts accelerated epigenetic aging in children. *Psychoneuroendocrino.*, **113**, 104484.
26. Lussier, A.A., Zhu, Y., Smith, B.J., Simpkin, A.J., Smith, A.D.A.C., Suderman, M.J., Walton, E., Ressler, K.J. and Dunn, E.C. (2022) Updates to data versions and analytic methods influence the reproducibility of results from epigenome-wide association studies. *Epigenetics*, 1–16.
27. Lussier, A.A., Zhu, Y., Smith, B.J., Simpkin, A.J., Smith, D.A.C., Suderman, M.J., Walton, E., Relton, C, Ressler, K.J., and Dunn, E.C. (in press) Sensitive periods for the effect of childhood adversity on DNA methylation: Updated results from a prospective, longitudinal study. *Biological Psychiatry Global Open Science*. https://doi.org/10.1016/j.bpsgos.2022.04.002.
28. Vable, A.M., Duarte, C.D., Cohen, A.K., Glymour, M.M., Ream, R.K. and Yen, I.H. (2020) Does the type and timing of educational attainment influence physical health? A novel application of sequence analysis. *Am. J. Epidemiol.*, **189**, 1389–1401.
29. Ben-Shlomo, Y. and Kuh, D. (2002) A life course approach to chronic disease epidemiology: conceptual models, empirical challenges, and interdisciplinary perspectives. *Int. J. Epidemiol.*, **31**, 285–293.
30. Zhu, Y., Simpkin, A.J., Suderman, M.J., Lussier, A.A., Walton, E., Dunn, E.C. and Smith, A.D.A.C. (2021) A structured approach to evaluating life course hypotheses: moving beyond analyses of exposed versus unexposed in the omics context. *Am. J. Epidemiol.*, **190**, 1101–1112.
31. Smith BJ, Smith ADAC, Dunn EC. 2021. Statistical Modeling of Sensitive Period Effects Using the Structured Life Course Modeling Approach (SLCMA). In: Andersen SL, editor. *Sensitive Periods of Brain Development and Preventive Interventions*. Berlin, Heidelberg: Springer. p 215–234.
32. Smith, A.D., Heron, J., Mishra, G., Gilthorpe, M.S., Ben-Shlomo, Y. and Tilling, K. (2015) Model selection of the effect of binary exposures over the life course. *Epidemiology*, **26**, 719–726.
33. Hannon, E., Lunnon, K., Schalkwyk, L. and Mill, J. (2015) Interindividual methylomic variation across blood, cortex, and cerebellum: implications for epigenetic studies of neurological and neuropsychiatric phenotypes. *Epigenetics*, **10**, 1024–1032.
34. Needham, B.L., Smith, J.A., Zhao, W., Wang, X., Mukherjee, B., Kardia, S.L., Shively, C.A., Seeman, T.E., Liu, Y. and Diez Roux, A.V. (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.
35. Stringhini, S., Polidoro, S., Sacerdote, C., Kelly, R.S., van Veldhoven, K., Agnoli, C., Grioni, S., Tumino, R., Giurdanella, M.C., Panico, S. *et al.* (2015) Life-course socioeconomic status and DNA methylation of genes regulating inflammation. *Int. J. Epidemiol.*, **44**, 1320–1330.
36. Johnson, R.C., Kalil, A. and Dunifon, R.E. (2012) Employment patterns of less-skilled workers: links to children's behavior and academic progress. *Demography*, **49**, 747–772.
37. Oreopoulos, P., Page, M. and Stevens, A.H. (2008) The intergenerational effects of worker displacement. *J. Labor Econ.*, **26**, 455–000.
38. Brand, J.E. (2015) The far-reaching impact of job loss and unemployment. *Ann. Rev. Soc.*, **41**, 359–375.
39. Sandstrom, H. and Huerta, S. (2013) *The Negative Effects of Instability on Child Development: A Research Synthesis*. Urban Institute, Washington, D.C., in press.
40. Ellis, B.J., Figueredo, A.J., Brumbach, B.H. and Schlomer, G.L. (2009) Fundamental dimensions of environmental risk : the impact of harsh versus unpredictable environments on the

evolution and development of life history strategies. *Hum. Nat.*, **20**, 204–268.
41. Belsky, J., Schlomer, G.L. and Ellis, B.J. (2012) Beyond cumulative risk: distinguishing harshness and unpredictability as determinants of parenting and early life history strategy. *Dev. Psychol.*, **48**, 662–673.
42. Levesque, A.R., MacDonald, S., Berg, S.A. and Reka, R. (2021) Assessing the impact of changes in household socioeconomic status on the health of children and adolescents: a systematic review. *Adol. Res. Rev.*, **6**(2), 91–123.
43. Shonkoff, J.P. (2010) Building a new biodevelopmental framework to guide the future of early childhood policy. *Child Dev.*, **81**, 357–367.
44. Essex, M.J., Boyce, W.T., Hertzman, C., Lam, L.L., Armstrong, J.M., Neumann, S.M. and Kobor, M.S. (2013) Epigenetic vestiges of early developmental adversity: childhood stress exposure and DNA methylation in adolescence. *Child Dev.*, **84**, 58–75.
45. Hosokawa, R. and Katsura, T. (2018) Effect of socioeconomic status on behavioral problems from preschool to early elementary school - a Japanese longitudinal study. *PLoS One*, **13**, e0197961.
46. Ellwood-Lowe, M.E., Humphreys, K.L., Ordaz, S.J., Camacho, M.C., Sacchet, M.D. and Gotlib, I.H. (2018) Time-varying effects of income on hippocampal volume trajectories in adolescent girls. *Dev. Cogn. Neurosci.*, **30**, 41–50.
47. Cajaraville, M.P., Cancio, I., Ibabe, A. and Orbea, A. (2003) Peroxisome proliferation as a biomarker in environmental pollution assessment. *Microsc. Res. Tech.*, **61**, 191–202.
48. Kristiansen, H., Gad, H.H., Eskildsen-Larsen, S., Despres, P. and Hartmann, R. (2011) The oligoadenylate synthetase family: an ancient protein family with multiple antiviral activities. *J. Interf. Cytokine Res.*, **31**, 41–47.
49. Pairo-Castineira, E., Clohisey, S., Klaric, L., Bretherick, A.D., Rawlik, K., Pasko, D., Walker, S., Parkinson, N., Fourman, M.H., Russell, C.D. *et al.* (2021) Genetic mechanisms of critical illness in COVID-19. *Nature*, **591**, 92–98.
50. Velavan, T.P., Pallerla, S.R., Rüter, J., Augustin, Y., Kremsner, P.G., Krishna, S. and Meyer, C.G. (2021) Host genetic factors determining COVID-19 susceptibility and severity. *eBioMedicine*, **72**, 103629.
51. Gatza, C.E., Oh, S.Y. and Blobe, G.C. (2010) Roles for the type III TGF-beta receptor in human cancer. *Cell. Signal.*, **22**, 1163–1174.
52. Sharifi, N., Hurt, E.M., Kawasaki, B.T. and Farrar, W.L. (2007) TGFBR3 loss and consequences in prostate cancer. *Prostate*, **67**, 301–311.
53. Liu, X.L., Xiao, K., Xue, B., Yang, D., Lei, Z., Shan, Y. and Zhang, H.T. (2013) Dual role of TGFBR3 in bladder cancer. *Oncol. Rep.*, **30**, 1301–1308.
54. Fang, W.Y., Kuo, Y.Z., Chang, J.Y., Hsiao, J.R., Kao, H.Y., Tsai, S.T. and Wu, L.W. (2020) The tumor suppressor TGFBR3 blocks lymph node metastasis in head and neck cancer. *Cancers (Basel)*, **12**(6), 1375.
55. Leyden, J., Murray, D., Moss, A., Arumuguma, M., Doyle, E., McEntee, G., O'Keane, C., Doran, P. and MacMathuna, P. (2006) Net1 and Myeov: computationally identified mediators of gastric cancer. *Br. J. Cancer*, **94**, 1204–1212.
56. Liang, E., Lu, Y., Shi, Y., Zhou, Q. and Zhi, F. (2020) MYEOV increases HES1 expression and promotes pancreatic cancer progression by enhancing SOX9 transactivity. *Oncogene*, **39**, 6437–6450.
57. Teicher, M.H., Samson, J.A., Anderson, C.M. and Ohashi, K. (2016) The effects of childhood maltreatment on brain structure, function and connectivity. *Nat. Rev. Neurosci.*, **17**, 652–666.
58. Hardy, B., Hill, H.D. and Romich, J. (2019) Strengthening social programs to promote economic stability during childhood. *Soc. Policy Rep.*, **32**, 1–36.
59. Nestle, M. (2019) The supplemental nutrition assistance program (SNAP): history, politics, and public health implications. *Am. J. Public Health*, **109**, 1631–1635.
60. Tankersley, J. and Goldstein, D. (2019) Biden Details $1.8 Trillion Plan for Workers, Students and Families. *New York Times*, published April 28, 2021. https://www.nytimes.com/2021/04/28/us/politics/biden-american-families-plan.html.
61. Nishimi, K., Howe, E. and Dunn, E.C. (2019) Social Determinants of Pyschiatric Illness. In Trinh, N.-H.T. and Chen, J.A. (eds.), *Sociocultural Issues in Psychiatry: A Casebook and Curriculum*. Oxford University Press, New York, NY.
62. Benzies, K. and Mychasiuk, R. (2009) Fostering family resiliency: a review of the key protective factors. *Child Fam. Soc. Work*, **14**, 103–114.
63. Hoye, J.R., Cheishvili, D., Yarger, H.A., Roth, T.L., Szyf, M. and Dozier, M. (2020) Preliminary indications that the attachment and Biobehavioral catch-up intervention alters DNA methylation in maltreated children. *Dev. Psychopathol.*, **32**, 1486–1494.
64. Purewal Boparai, S.K., Au, V., Koita, K., Oh, D.L., Briner, S., Burke Harris, N. and Bucci, M. (2018) Ameliorating the biological impacts of childhood adversity: a review of intervention programs. *Child Abuse Negl.*, **81**, 82–105.
65. Poehlmann-Tynan, J., Engbretson, A., Vigna, A.B., Weymouth, L.A., Burnson, C., Zahn-Waxler, C., Kapoor, A., Gerstein, E.D., Fanning, K.A. and Raison, C.L. (2020) Cognitively-based compassion training for parents reduces cortisol in infants and young children. *Infant Ment. Health J.*, **41**, 126–144.
66. Slopen, N., McLaughlin, K.A. and Shonkoff, J.P. (2014) Interventions to improve cortisol regulation in children: a systematic review. *Pediatrics*, **133**, 312–326.
67. Raby, K.L., Bernard, K., Gordon, M.K. and Dozier, M. (2020) Enhancing diurnal cortisol regulation among young children adopted internationally: a randomized controlled trial of a parenting-based intervention. *Dev. Psychopathol.*, **32**, 1657–1668.
68. Amrhein, V. and Greenland, S. (2018) Remove, rather than redefine, statistical significance. *Nat. Hum. Behav.*, **2**, 4–4.
69. McShane, B.B., Gal, D., Gelman, A., Robert, C. and Tackett, J.L. (2019) Abandon statistical significance. *Am. Stat.*, **73**, 235–245.
70. Gurdasani, D., Barroso, I., Zeggini, E. and Sandhu, M.S. (2019) Genomics of disease risk in globally diverse populations. *Nat. Rev. Genet.*, **20**, 520–535.
71. Althubaiti, A. (2016) Information bias in health research: definition, pitfalls, and adjustment methods. *J. Multidiscip. Healthc.*, **9**, 211–217.
72. Krieger, N. (1992) Overcoming the absence of socioeconomic data in medical records: validation and application of a census-based methodology. *Am. J. Public Health*, **82**, 703–710.
73. Relton, C.L., Gaunt, T., McArdle, W., Ho, K., Duggirala, A., Shihab, H., Woodward, G., Lyttleton, O., Evans, D.M., Reik, W. *et al.* (2015) Data resource profile: accessible resource for integrated Epigenomic studies (ARIES). *Int. J. Epidemiol.*, **44**, 1181–1190.
74. Boyd, A., Golding, J., Macleod, J., Lawlor, D.A., Fraser, A., Henderson, J., Molloy, L., Ness, A., Ring, S. and Davey Smith, G. (2013) Cohort profile: the 'children of the 90s'—the index offspring of the Avon longitudinal study of parents and children. *Int. J. Epidemiol.*, **42**, 111–127.
75. Golding, J., Pembrey, M., Jones, R. and The ALSPAC Study Team (2001) ALSPAC: the Avon longitudinal study of parents and children I. study methodology. *Paediatr. Perinat. Epidemiol.*, **15**, 74–87.

76. Fraser, A., Macdonald-Wallis, C., Tilling, K., Boyd, A., Golding, J., Davey Smith, G., Henderson, J., Macleod, J., Molloy, L., Ness, A. *et al.* (2013) Cohort profile: the Avon longitudinal study of parents and children: ALSPAC mothers cohort. *Int. J. Epidemiol.*, **42**, 97–110.
77. Houseman, E.A., Accomando, W.P., Koestler, D.C., Christensen, B.C., Marsit, C.J., Nelson, H.H., Wiencke, J.K. and Kelsey, K.T. (2012) DNA methylation arrays as surrogate measures of cell mixture distribution. *BMC Bioinform.*, **13**, 86.
78. Efron, B., Hastie, T., Johnstone, I. and Tibshirani, R. (2004) Least angle regression. *Ann. Stat.*, **32**, 407–499.
79. Tibshirani, R.J., Taylor, J., Lockhart, R. and Tibshirani, R. (2016) Exact post-selection inference for sequential regression procedures. *J. Am. Stat. Assoc.*, **111**, 600–620.
80. Zeilinger, S., Kühnel, B., Klopp, N., Baurecht, H., Kleinschmidt, A., Gieger, C., Weidinger, S., Lattka, E., Adamski, J., Peters, A. *et al.* (2013) Tobacco smoking leads to extensive genome-wide changes in DNA methylation. *PLoS One*, **8**, e63812.
81. Benjamini, Y. and Hochberg, Y. (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. Series B Stat. Methodol.*, **57**, 289–300.

Socioeconomic changes predict genome-wide DNA methylation in childhood

Jiaxuan Liu[1,2,†], Janine Cerutti[1,†], Alexandre A. Lussier[1,4,5], Yiwen Zhu[1,2], Brooke J. Smith[1], Andrew D.A.C. Smith[3], Erin C. Dunn[1,4,5,6,*]

[1] Center for Genomic Medicine, Massachusetts General Hospital, Boston, MA
[2] Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, MA
[3] Mathematics and Statistics Research Group, University of the West of England, Bristol, UK
[4] Department of Psychiatry, Harvard Medical School, Boston, MA
[5] Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge, MA
[6] Harvard Center on the Developing Child, Harvard University, Cambridge, MA

†Joint First Authors, who contributed equally to the manuscript.

[*]**Corresponding Author**: Erin C. Dunn, ScD, MPH, Psychiatric and Neurodevelopmental Genetics Unit, Center for Genomic Medicine, Massachusetts General Hospital, 185 Cambridge Street, Simches Research Building 6th Floor, Boston, MA 02114; Email: edunn2@mgh.Harvard.edu; Phone: 617-726-9387; Fax: 617-726-0830; Website: www.thedunnlab.com

SUPPELMENTAL DATA

Supplemental Methods

Figures S1 to S8

Titles for Table S1 to S8

SUPPLEMENTAL METHODS

Sample and procedures

This current study used data from the Accessible Resources for Integrated Epigenomics Studies (ARIES) (1), a subsample of 1,018 mother-child pairs from the Avon Longitudinal Study of Parents and Children (ALSPAC). ALSPAC is a prospective, longitudinal birth-cohort in the United Kingdom (UK) designed to investigate the genetic and environmental determinants of health across the lifespan (2-4). Women living in the county of Avon, UK with estimated delivery dates between April 1991 and December 1992 were invited to participate. Approximately 85% percent of eligible pregnant women (N=14,541) agreed to participate and were enrolled, which resulted in 14,062 live births and a sample size of 13,988 children alive at 1 year of age. Response rates to ALSPAC data collection have been good (75% have completed at least one follow-up).

Mother-child pairs in the ARIES were randomly selected from ALSPAC based on availability of DNA samples across five waves of data collection (1). We used data from 946 singletons in ARIES with blood-based DNA methylation (DNAm) profiles generated at age 7.

Ethical approval for the study was obtained from the ALSPAC Ethics and Law Committee and the Local Research Ethics Committee. Please note that the ALSPAC study website contains details of all the data that is available through a fully searchable data dictionary and variable search tool (http://www.bristol.ac.uk/alspac/researchers/our-data/).

Measures

<u>Early-life socioeconomic position (SEP)</u>

Considering the multidimensional nature of socioeconomic position (SEP) and that different indicators of SEP can have distinct relationships with DNAm (5), we analyzed six SEP indicators, spanning financial, occupational, and residential domains. Since repeated measurements were required for testing different life course hypotheses, these SEP indicators were chosen also because they were measured at least once in each of the following three developmental time periods: *very early childhood* (0-2 years), *early childhood* (3-5 years), and *middle childhood* (6-7 years). These time periods are consistent with previous ALSPAC and other research studies (6-9), roughly corresponding to three major developmental stages (infancy/toddlerhood, pre-school, and school-age) where different early-life policies and interventions could occur. For SEP indicators assessed more frequently, we collapsed the timepoints so children were classified as exposed to socioeconomic adversity within the time period if they were exposed on at least one occasion within that given time period (**Figure 1a**). Children were classified as exposed using criteria defined below for each construct. Additionally, we excluded the following SEP measures that were stable over time because of insufficient variation across childhood periods needed for the variable selection in SLCMA: maternal education, maternal marital status, maternal home ownership, maternal ever homelessness, and Townsend Deprivation Index. These SEP variables were, however, examined as additional covariates in a sensitivity analysis. The six SEP indicators included in our main are described below.

1. Job loss. Mothers indicated job loss for themselves and/or their partner and the extent to which it affected them on six occasions after birth (when the child was age 8 weeks, 8 months, 1.75, 2.75 years, 4 years, 5 years, and 6 years). This indicator was assessed on a Likert-type scale

ranging from: 1 = ‘yes & affected me a lot’; 2 = ‘yes, moderately affected’; 3 = ‘yes, mildly affected’; 4 = ‘yes, but did not affect me at all’; 5 = ‘no, did not happen’. At each time point, children were coded as exposed if their mothers reported yes to an event which at least mildly affected the family (corresponding to response options 1 to 3 on a 5-point scale, with a lower score reflecting a greater effect).

2. Income reduction. Mothers indicated the extent to which a reduction in income in the previous year affected them on seven occasions after birth (when the child was age 8 weeks, 8 months, 1.75 years, 2.75 years, 4 years, 5 years, and 6 years). This indicator was assessed and coded in the same way as job loss, as noted above.

3. Low family income. Mothers reported how much their average take home family income was per week (<£100, £100 - £199, £200 - £299, £300 - £399, £400+) at three occasions after birth (when the child was age 2.75 years, 4 years, and 7 years). Children were coded as exposed if their reported family income was in the two lowest income categories (<£100, £100 - £199). This cut-off is consistent with the official average weekly income threshold (i.e., £190 before housing costs) for households living in relative low income as published by the Department for Work and Pensions in its most recent Households Below Average Income (HBAI) report (10).

4. Financial hardship. Mothers indicated the extent to which the family had difficulty affording the following: 1) items for the child; 2) rent or mortgage; 3) heating; 4) clothing; 5) food. Each of the 5 items was coded on a Likert-type scale (1=not difficult; 2=slightly difficult; 3=fairly difficult; 4=very difficult). These items were completed at five occasions after birth (when the

child was age 8 months, 1.75 years, 2.75 years, 5 years, and 7 years). At 1.75 years, 2.75 years, and 5 years, a fifth response option was added to some of the items: 5=paid directly by social security. At 7 years, a fifth response option was added to all five items: 5=don't pay for this. In order to make categories consistent and comparable across occasions, we estimated the difficulty level on the 4-point scale for participants who selected the fifth option (indicating an undefined level of perceived financial hardship) using multiple imputation. We included all SEP indicators and covariates in the imputation, and assigned the rounded mean based on 51 imputation datasets as the imputed difficulty level. We chose an odd number of imputations in order to compare the imputation results between using the mean or the median as the imputed value, and found little difference between the two approaches. Children were coded as exposed if their mothers reported at least fairly difficult for three or more items at each time point (corresponding to response option 3 to 4 on a 4-point scale, with a higher score reflecting more difficulty).

5. Major financial problem. Mothers indicated the extent to which a major financial problem affected them on four occasions after birth (when the child was age 8 months, 2.75 years, 5 years, and 6 years). This indicator was assessed and coded in the same way as job loss, as noted above.

6. Neighborhood disadvantage. On four occasions after birth (when the child was age 1.75 years, 2.75 years, 5 years, and 7 years), mothers indicated the degree to which the following were problems in their neighborhood: 1) noise from other homes; 2) noise from the street; 3) garbage on the street; 4) dog dirt; 5) vandalism; 6) worry about burglary; 7) mugging; and 8) disturbance from youth. Response options to each item were: 0=not a problem or no opinion, 1=minor

problem, 2=serious problem. Items were summed, yielding scores ranging from 0-16. Children with scores of eight or greater, which generally corresponded to the 95th percentile, were classified as exposed to neighborhood disadvantage.

DNA methylation (DNAm)

As described elsewhere (1, 11), DNA extracted from peripheral blood (whole blood or buffy coat). DNAm was measured at 485,577 CpG dinucleotide sites across the genome using the Illumina Infinium HumanMethylation450 BeadChip microarray, which captures DNAm variation at 99% of RefSeq genes (17 CpG sites per gene, on average). Consent for biological samples was collected in accordance with the Human Tissue Act (2004). See Relton et al. (2015) for details about the laboratory procedures (1).

The proportion of molecules methylated at each interrogated CpG site on the array was detected using the microarray. The estimated level of DNAm at each CpG sites was expressed as a "beta" value (β), defined as the ratio of the intensity measured by the methylated probe and the sum of the overall intensity and a recommended offset value α = 100 (β = intensity of the Methylated allele (M) / (intensity of the Unmethylated allele (U) + intensity of the Methylated allele (M) + 100)). The β value ranges from 0 (no methylated dinucleotides observed) to 1 (all dinucleotides methylated). Background correction and functional normalization were applied to the raw methylation β-values using the R-package *meffil*, a pipeline developed by Min and colleagues to remove or minimize the effects of variation due to technical artifacts. Cross-hybridizing probes (both autosomal and XY-binding), polymorphic probes, and probes located in sex chromosomes were further excluded from analysis. A total of 412,956 CpGs on autosomal chromosomes passed QC and were included in this analysis. To limit the impact of extreme

values, we winsorized the beta values (i.e., values that represent % methylation) at each CpG site, setting the bottom 5% and top 5% of values to the 5th and 95th quantile, respectively.

Additionally, we estimated the proportions of the six white cells in the whole blood (CD8 T cells, CD4 T cells, NK cells, B cells, monocytes, and granulocytes) using Houseman's method (12). Estimated cell proportions were included in all analyses to correct for cell type heterogeneity.

The gene symbol of and distance to the nearest gene to each CpG were obtained from the *FDb.InfiniumMethylation.hg19 package* in R (13). Genomic features including genomic location, relation to CpG islands (CGIs), and enhancers were obtained from Illumina HumanMethylation450 v1.2 Manifest File (https://support.illumina.com/array/array_kits/infinium_humanmethylation450_beadchip_kit/downloads.html).

Covariates

To adjust for baseline demographic differences in the cohort, we controlled in the main analyses for the following variables measured at birth: *child race/ethnicity* (0=non-White, 1=White); *child sex* (0=male, 1=female); *child birth weight*; *maternal age* (0=ages 15-19, 1=ages 20-35, 2=age>35); *number of previous pregnancies*; and *sustained maternal smoking during pregnancy* (0=non-smoker, 1=smoker in two or more trimesters, including the third trimester). Because age is a strong predictor of DNAm (14) and the actual time of blood draw at the age 7 time-point varied across children, we also adjusted for child age in months at the time of blood draw (ranging from 85 to 109 months, median=89 months). We also adjusted for cell proportions estimated using the Houseman method (12) to account for cell type heterogeneity.

In secondary analyses, we additionally controlled for time-invariant SEP indicators, population substructure, cord blood DNAm, and genetic variation to evaluate the robustness of our findings (see *Additional covariate adjustments* section below).

Data analysis

We used a two-stage structured life course modeling approach (SLCMA) (15-17) to investigate the time-dependent relationship between socioeconomic adversity and DNAm across three life-course hypotheses. SLCMA was performed in two stages: *model selection* followed by *post-selection inference*.

In the first stage, a set of hypotheses were encoded and entered into a model selection procedure (**Figure 1b**). The three life-course hypotheses tested in this analysis were parameterized as follows.

1) *Accumulation* hypothesis, in which the effect of low-SEP increases with the number of periods exposed, regardless of timing. Formally, for the k^{th} period (k=1 for very early childhood, k=2 for early childhood, k=3 for middle childhood), X_k=1 if exposed to the specific adversity during the k^{th} period, and X_k=0 if not exposed; then the accumulation hypothesis was coded as $X_1 + X_2 + X_3$.
2) *Sensitive period* hypotheses, in which the effect of low-SEP depends on the developmental time period of the exposure. We tested three sensitive period hypotheses, one for each childhood period, and the encoded variable was X_k for the k^{th} period (k=1,2,3).
3) *Mobility* hypothesis, in which DNAm is associated with an upward or downward change in SEP between adjacent developmental time periods. We tested two improvement hypotheses and two worsening hypotheses: improvement in SEP

between k^{th} and $(k+1)^{th}$ period (k=1,2) was coded as $X_k*(1- X_{k+1})$; worsening in SEP between k^{th} and $(k+1)^{th}$ period (k=1,2) was coded as $(1- X_k)*X_{k+1}$.

For each SEP indicator, variables encoding the theoretical hypotheses were entered into the Least Angle Regression (LARS) variable selection procedure (18) to identify the variable explaining the most variability in DNAm at each CpG (R^2), meaning the hypothesis most supported by the data. We focused only on the first variable selected to maximize statistical power and prioritize parsimonious explanations (15). For each CpG, six unique hypotheses were selected, corresponding to each SEP indicator.

In the second stage, we performed post-selection inference to test the association between the selected hypothesis and DNAm as well as estimate confidence intervals. We used selective inference (19), which was recommended for high-throughput applications of the SLCMA (15).

We controlled for baseline demographic variables, child age at blood draw, and cell type proportions in both stages of the analysis. We implemented the Frisch-Waugh-Lovell (FWL) Theorem (20-22) to adjust for covariates: first, encoded hypotheses and DNAm were separately regressed on the covariates; then we applied LARS with residuals from the regressions. It has been shown that this approach gives the same effect estimates as a fully adjusted linear model where covariates are included in the regression model directly (23), while it can improve statistical power and overcome bias. We performed complete-case analysis, meaning that only samples with non-missing exposures and covariates were included in the analysis.

Sensitivity analyses

We conducted three sensitivity analyses to evaluate the robustness of our top results (R^2>3%) in the primary analysis.

Additional covariate adjustments

We reran the SLCMA analysis for the top CpGs adjusting for additional covariates described below to evaluate the possibility of remaining distortions in the identified SEP-DNAm associations (or residual bias). We did not include these variables in main analysis, because prior studies have shown that covariate adjustment may substantially shift the results of DNAm-based analyses (7, 24); thus, a stepped approach could enable better detection of signal when the role of the covariate in the SEP-DNAm association is unclear, and avoid reducing sample size based on missing covariate data.

1) Time-invariant SEP indicators. We examined how our findings were influenced by the adjustment of SEP indicators that were relatively stable throughout childhood. These time-invariant SEP indicators were moderately correlated to the time-varying SEP indicators examined in the main analysis (average correlation with SEP exposures ranged from 0.20 to 0.47 in absolute values). Further controlling for these indicators allowed us to evaluate the robustness of our findings in response to invariant aspects of the socioeconomic environment. These invariant SEP indicators included: *maternal education* (1=less than O-level, 2=O-level, 3=A-level, 4=Degree or above); *maternal marital status* (0=never married, 1=widowed/divorced/separated, 2=married); *maternal home ownership* (0=owning, 1=rented); *maternal ever homelessness* (0=no, 1=yes); *Townsend Deprivation Index* (an indicator of neighborhood deprivation via Census data, ranging from 1-5 corresponding to the five quintiles). We adjusted for each of the time-invariant SEP indicators in a separate model, and also ran a model adjusted for all of them.

2) Population substructure. To assess if our results were biased by population substructure in the sample, we further adjusted for the top four epigenetic principal components (PCs). Although children included in the ARIES sample are primarily White, subtle substructure in the sample may still possibly bias the analysis. Bias may arise if ancestry is associated with both DNAm pattern (e.g. through genetic factors) and exposure to socioeconomic adversities (e.g. through different geographic locations of ancestors) (25). Therefore, we inferred population structure based on epigenetic PCs estimated using the EPISTRUCTURE algorithm developed by Rahmani et al. (2017) (26), and further adjusted for the top four epigenetic PCs in a sensitivity analysis.

3) Cord blood DNAm. To account for differences in DNAm that might have been present at birth (thus not caused by factors occurring postnatal), we further adjusted for cord blood methylation in the model for the top CpGs. Sample collection, laboratory procedures, and quality control for cord blood DNAm are described elsewhere (1, 11). Methylation beta values were normalized, corrected for cell count heterogeneity, and winsorized to remove outliers following the quality control for age 7 DNAm as described above.

4) Genetic variation. For CpGs associated with any methylation quantitative trait loci (mQTLs), based on a database of mQTLs of the ARIES cohort (http://www.mqtldb.org/) (27), we further controlled for genetic var iation at mQTLs linked to our top sites. We downloaded the list of mQTLs at age 7, and filtered the data to our top CpG sites. Children were genotyped using the Illumina HumanHap550 quad chip; imputation was performed to the 1000 Genomes (phase 1, version 3,

release Dec 2013) reference population using IMPUTE v2.2.2. Variants were filtered by minor allele frequency (MAF>0.01), Hardy-Weinberg equilibrium (HWE>$5x10^{-7}$), and imputation quality (info>0.8); subjects were filtered by missing genotype rate (missingness<3%) and cryptic relatedness (r<0.1). For each top CpG with four or fewer associated SNPs, we included minor allele dosages as additional covariates. For each top CpG site with more than four associated SNPs, we filtered SNPs by call rate (>99%) and ran a principal components analysis among all SNPs associated with each CpG. The top four principal components were used as covariates to represent genetic variation in the sensitivity analysis.

5) Exposure to other time-varying SEP indicators. Considering the moderate correlations among the six time-varying SEP indicators analyzed in our main analyses, we further investigated the extent to which the identified association between each SEP indicator and DNAm was explained by exposure to the other main SEP indicators. Specifically, we reran the SLCMA analysis for each of the top CpGs, while additionally adjusting for exposure to any of the five SEP indicators other than the one associated with the DNAm at the CpG (0=unexposed to any other five SEP indicators during any childhood periods, 1=ever exposed to at least one of the other five SEP indicators during one or more childhood periods).

Analysis without mobility hypothesis

While *accumulation* and *sensitive periods* have been previously examined for the association between SEP and DNAm (7), *mobility* within childhood has never been tested on DNAm. When additional hypotheses are added to the SLCMA, there is a cost in terms of statistical power for a fixed sample size. To better understand the value of adding *mobility*

hypotheses in the model selection procedure, which could help guide future analyses, we therefore reran the analyses of top CpGs for low family income, financial hardship, major financial problem, and neighborhood disadvantage using only *accumulation* and *sensitive period* hypotheses. We compared the results for the four SEP indicators from analysis with and without *mobility* tested.

EWAS of exposed vs. unexposed to socioeconomic adversity

We also performed epigenome-wide association studies (EWAS) of *any exposure* before age 7 and DNAm to evaluate the loss (or gain) of information from the SLCMA compared to more conventional approaches. For each SEP indicator, children were coded as ever-exposed (versus never-exposed) if they met the exposure criteria at one or more timepoints by age 7. We compared the detected CpGs in EWAS and SLCMA to determine how the two approaches were different in their findings. A mathematical proof is provided elsewhere (7) showing that when the true relationship between exposure and outcome depends on the timing or amount of exposure, a standard EWAS of lifetime exposure is underpowered compared to SLCMA.

Secondary analyses

We also conducted two secondary analyses to interpret our findings and place them in the context of prior literature.

Compare to prior EWAS

We compared the result of our R^2>3% CpGs to those identified from seven previous epigenome-wide DNAm studies on socioeconomic position (28-34), utilizing individual CpG-level summary statistics reported in a recent review paper (5). In these seven studies, investigators analyzed indicators spanning five SEP domains: household assets, education, occupation, income, and aggregated composite measures. As some studies analyzed more than

one SEP domain or SEP at multiple life stages, we included a total of 17 EWASs from the seven studies. Focusing on the 62 CpGs with R^2>3%, we then compared our results to these prior EWASs to evaluate: 1) the extent to which our effect estimates were in the same direction; 2) whether they presented statistical evidence in those previous studies at the nominal significance level ($p<0.05$) and at FDR<0.05 (multiple testing correction was performed within each EWAS separately by the aforementioned review paper).

Exploring the biological significance of the top CpGs

DNAm correlation across blood and brain. To examine the relevance of SEP-related DNAm pattern identified in peripheral blood tissues to brain health, we used a publicly available database that compare DNAm in peripheral blood tissue and brain tissue. The Blood Brain DNA Methylation Comparison Tool (http://epigenetics.essex.ac.uk/bloodbrain/) (35) includes DNAm levels in whole blood and four brain regions (prefrontal cortex (PFC), entorhinal cortex (EC), superior temporal gyrus (STG), and cerebellum (CER)) in N = 71–75 matched samples from individuals archived in the MRC London Neurodegenerative Disease Brain Bank. This sample includes both neuropathologically unaffected controls and individuals with variable levels of neuropathology. Pearson correlation r values measuring the blood-brain correlations were retrieved from this database.

Enrichment of genomic features and regulatory elements. We examined whether the locations of the R^2>3% CpGs were enriched in certain genomic regions (e.g., gene body, UTR regions, etc.), CGI and CGI flanking regions (CGI shore, 0–2 kb from CGI; CGI shelf, 2-4 kb from CGI), and enhancers. Annotations of these features were obtained for all R^2>3% CpGs from the *FDb.InfiniumMethylation.hg19* package in R (13). We tested if the genomic features were more

296 common among the R^2>3% CpG sites as compared to all CpGs tested across the epigenome
297 using two-sided Chi-squared tests.

298 Enrichment of biological pathways. Gene set enrichment analyses were conducted using the
299 methylGSA package in R (36) to identify important biological pathways indicated by our
300 SLCMA results. Epigenome-wide SLCMA results ranked by p-values were used in the gene set
301 enrichment analyses, and Robust Rank Aggregation (37) was used to adjust for the number of
302 CpG sites in each gene. Gene Ontology (GO) terms were tested, and Bonferroni method was
303 used for multiple testing correction.

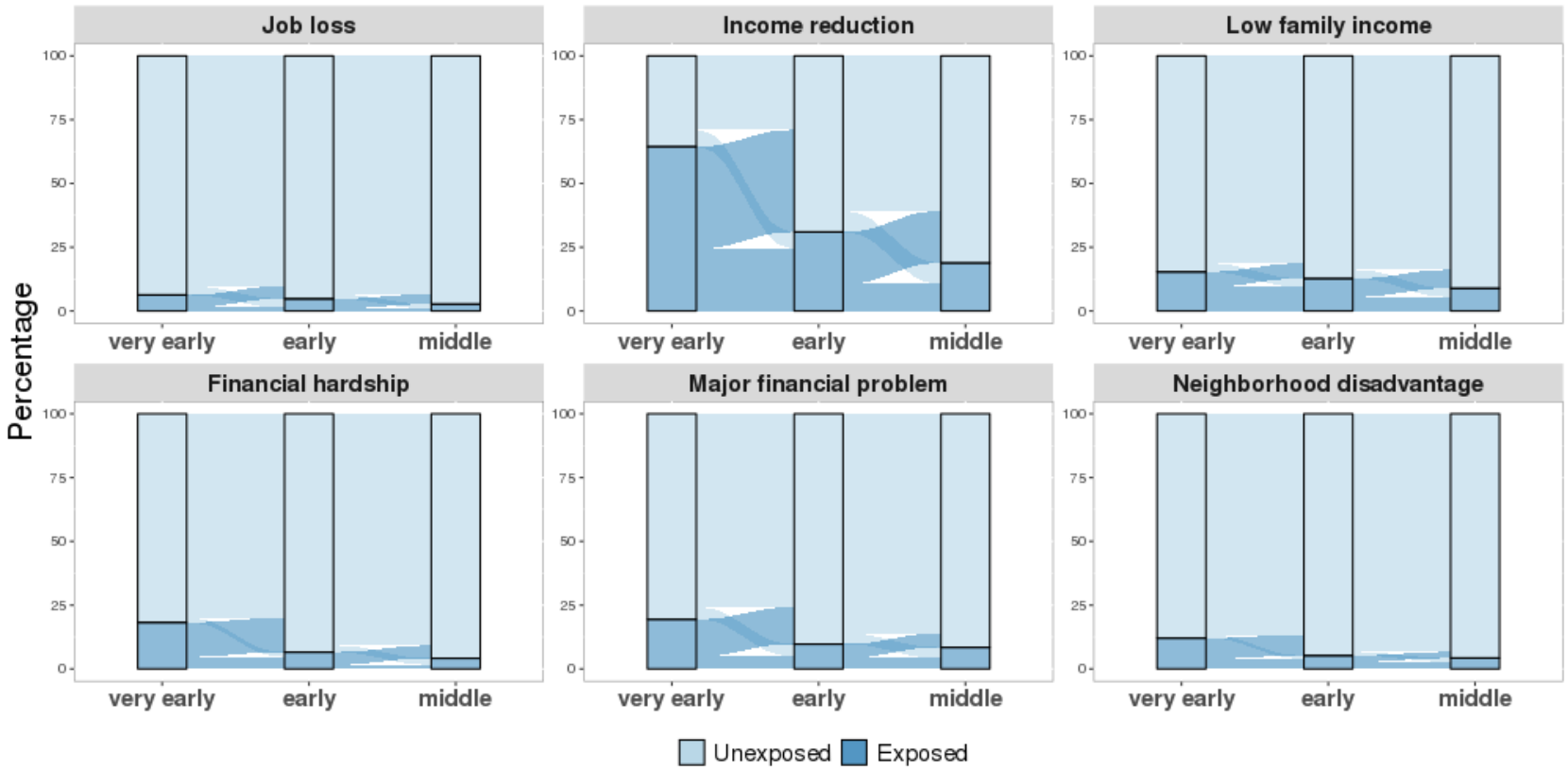

Figure S1. Prevalence of each indicator of socioeconomic position (SEP) or socioeconomic adversity, across the three developmental periods. Each panel shows for one of the six SEP indicators the percentage of exposed children (dark blue) and unexposed children (light blue) at three developmental periods as well as the change between periods. For all six SEP indicators, the prevalence of exposure decreased over time; the majority of unexposed children remained unexposed across periods, while exposed children tended to improve in their SEP over time.

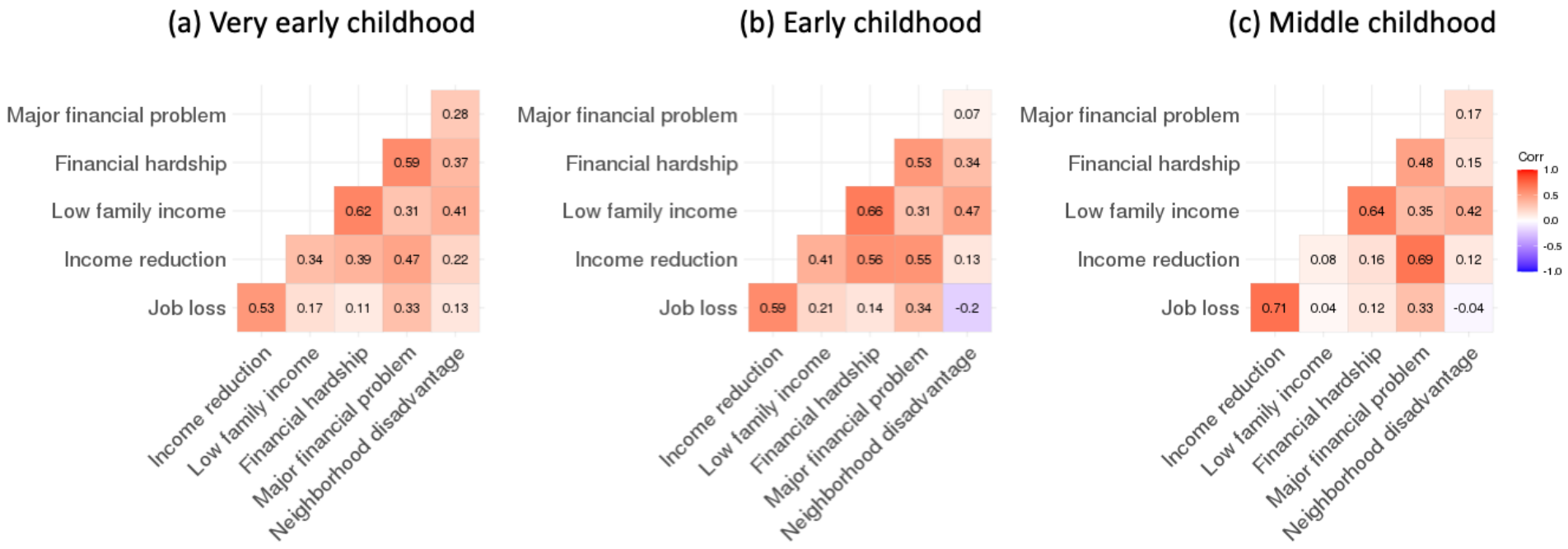

Figure S2. Polychoric correlation between socioeconomic adversity pairs during (a) very-early childhood (0-2 years), (b) early childhood (3-5 years), and (c) middle childhood (6-7 years). The six socioeconomic adversities were moderately correlated during all three childhood periods (r_{avg}=0.35 at very-early childhood, r_{avg}=0.34 at early childhood, r_{avg}=0.29 at middle childhood).

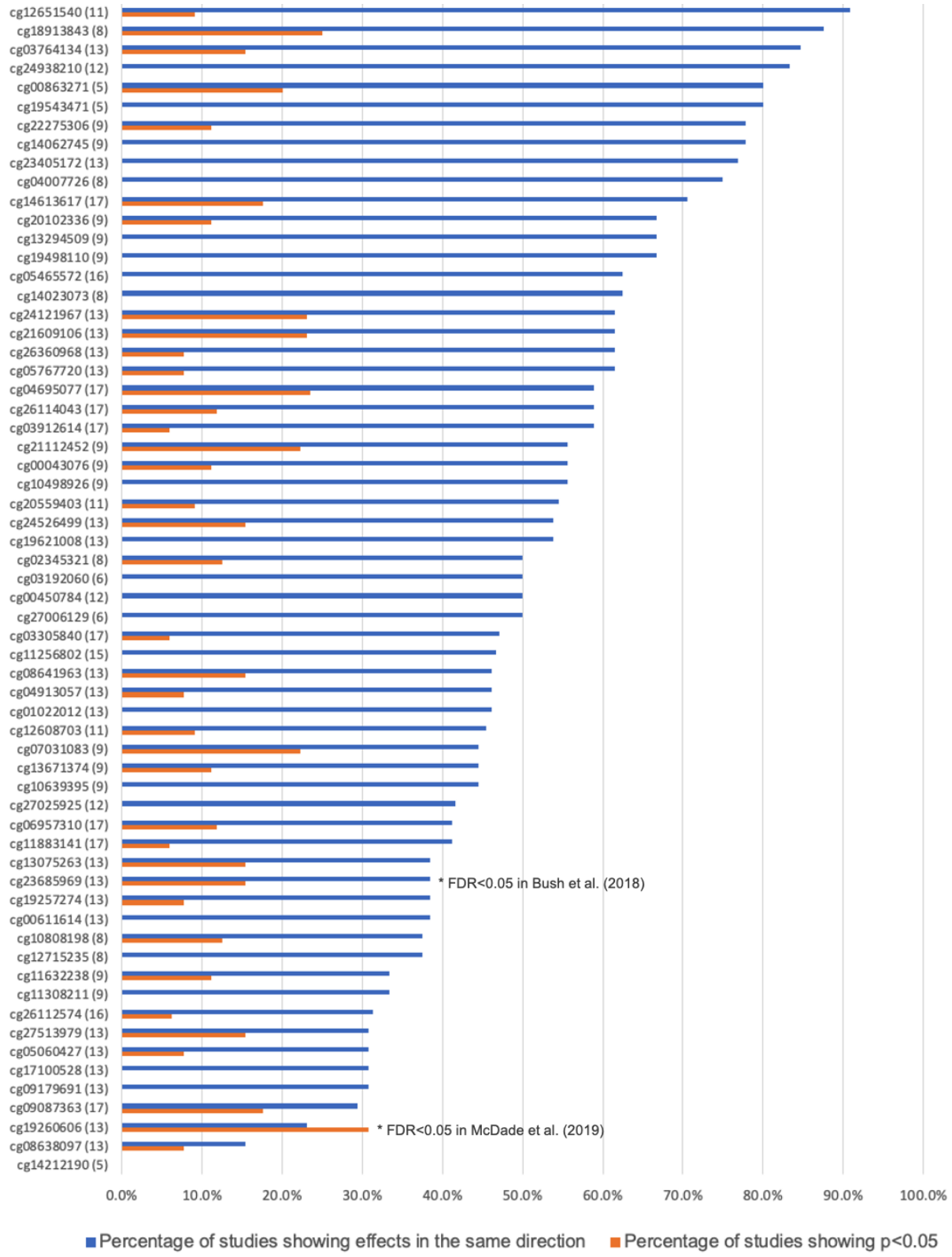

Figure S3. Comparison of our results to previous EWAS for the 62 CpGs with R^2>3%. The x-axis shows the percentage of previous studies showing effects in the same direct (blue) and the proportion of previous studies showing p<0.05 (orange) for each CpG site. The y-axis shows the

CpG names with the number of previous EWAS analyses being compared to in the parentheses. Two CpGs passed multiple testing correction at an FDR<0.05 in previous EWAS: cg23685969 was significantly associated with income in Bush et al. (2018) (in our analysis it was significantly associated with low family income); cg19260606 was significantly associated with education and an aggregated composite measure in McDade et al. (2019) (in our analysis it was significantly associated with major financial problem). See **Supplementary Methods** for more details about the studies included.

Figure S4. Bar plots showing the number of CpGs across the epigenome selected by each life-course hypothesis for each type of socioeconomic adversity. (a) For job loss and income reduction, we tested accumulation and sensitive period hypotheses. (b) For the other four socioeconomic adversities, we tested accumulation, sensitive period, and mobility hypotheses. *Very early*, *Early*, and *Middle* refer to sensitive period hypotheses related to the three childhood periods: very early (0-2 years), early (3-5 years), and middle childhood (6-7 years). *Early worsening/improvement* refer to mobility hypotheses for changes between very early and early childhood, and *later worsening/improvement* refer to mobility hypotheses for changes between early and middle childhood.

Figure S5. Scatterplots showing that the structured life course modeling approach (SLCMA) was more powerful than a standard epigenome-wide association study (EWAS) at identifying time-dependent effect of socioeconomic adversity on DNA methylation (DNAm). This plot compares the effect of neighborhood disadvantage estimated by the SLCMA approach (y-axis) versus those estimated by the standard approach of EWAS (x-axis) of ever-exposure, for the 62 CpGs associated with socioeconomic adversity explaining more than 3% variance in DNAm. The shaded area indicates where the effect estimates were in the same direction in SLCMA and EWAS, but larger in magnitude in SLCMA. The unshaded areas indicates where the effect estimates were greater in the EWAS or estimates were in opposite directions from two analyses. For 59 of the 62 CpGs (including the 4 FDR-significant CpGs), the estimated effects were stronger in SLCMA than in EWAS, regardless of the direction of effect.

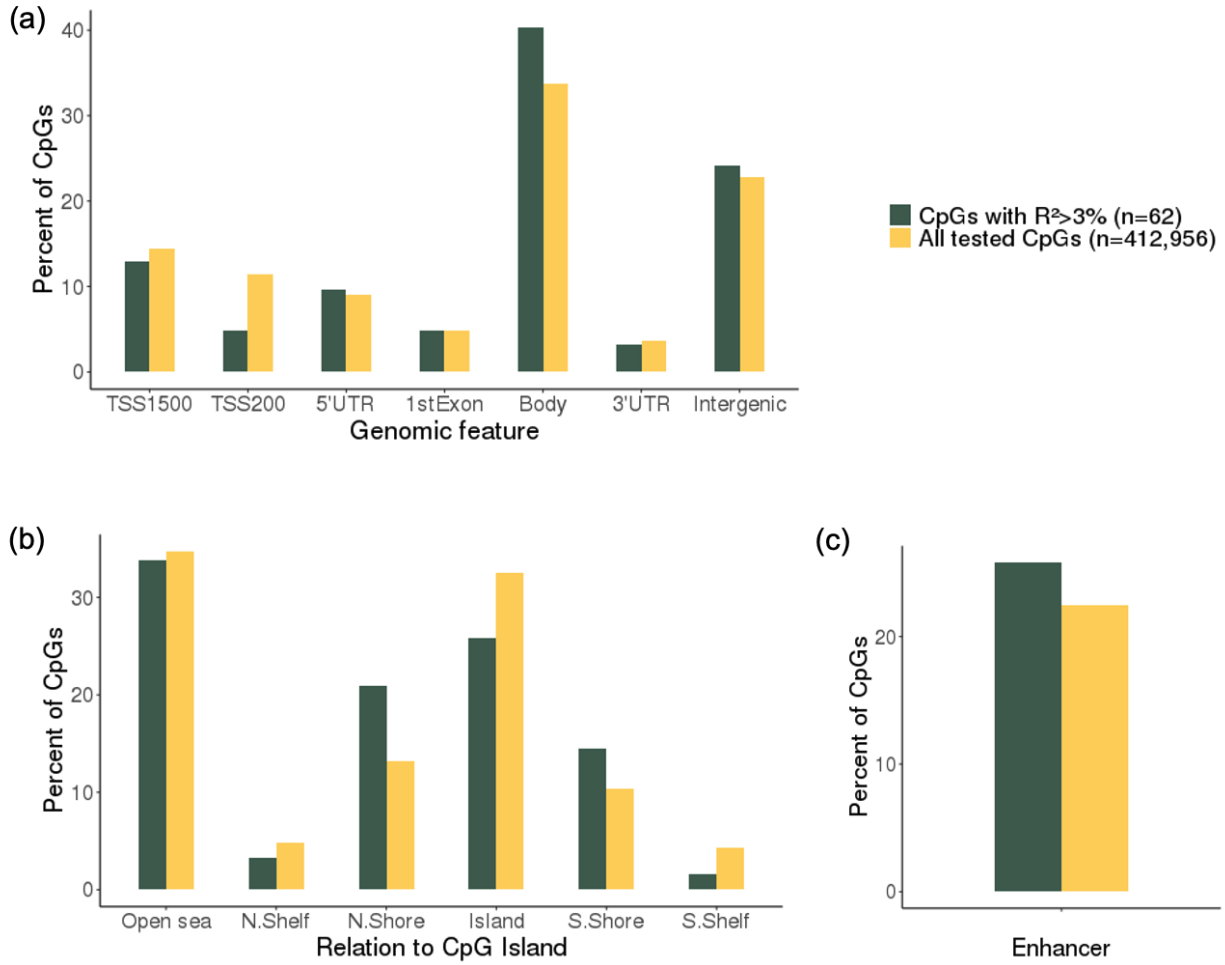

Figure S6. The distribution of genomic features (a), CpG island (CGI) locations (b), and enhancer (c) in the 62 CpGs associated with socioeconomic adversity explaining more than 3% variance in DNAm (R^2>3%, dark green) and all tested CpGs (n=412,956, yellow). TSS1500: within 1500 bp before the transcription start site of a gene. TSS200: within 200 bp before the transcription start site of a gene. CGI shore: 0–2 kb from CGI. CGI shelf: 2-4 kb from CGI.

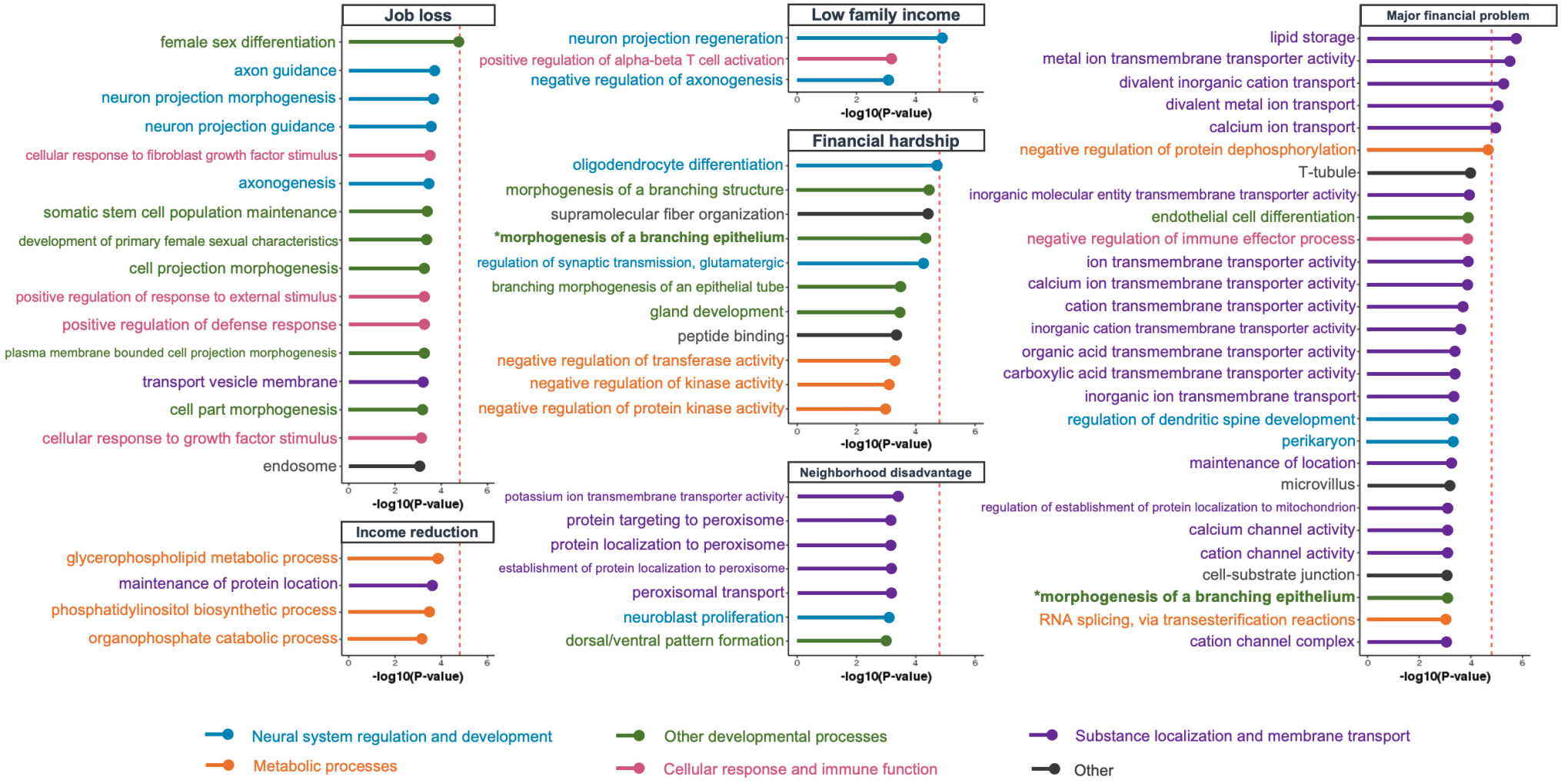

Figure S7. Results of gene set enrichment analysis. Gene Ontology (GO) terms that showed p-values<0.001 are shown on the y axis. GO terms were colored by pathway type. The red lines indicate p-value thresholds based on Bonferroni correction. Little overlap in the top pathways was observed across SEP indicators, except for *morphogenesis of a branching epithelium*, which emerged in the

enrichment analysis for both financial hardship and major financial problem. These findings suggest different socioeconomic adversities may lead to shifts in distinct biological pathways.

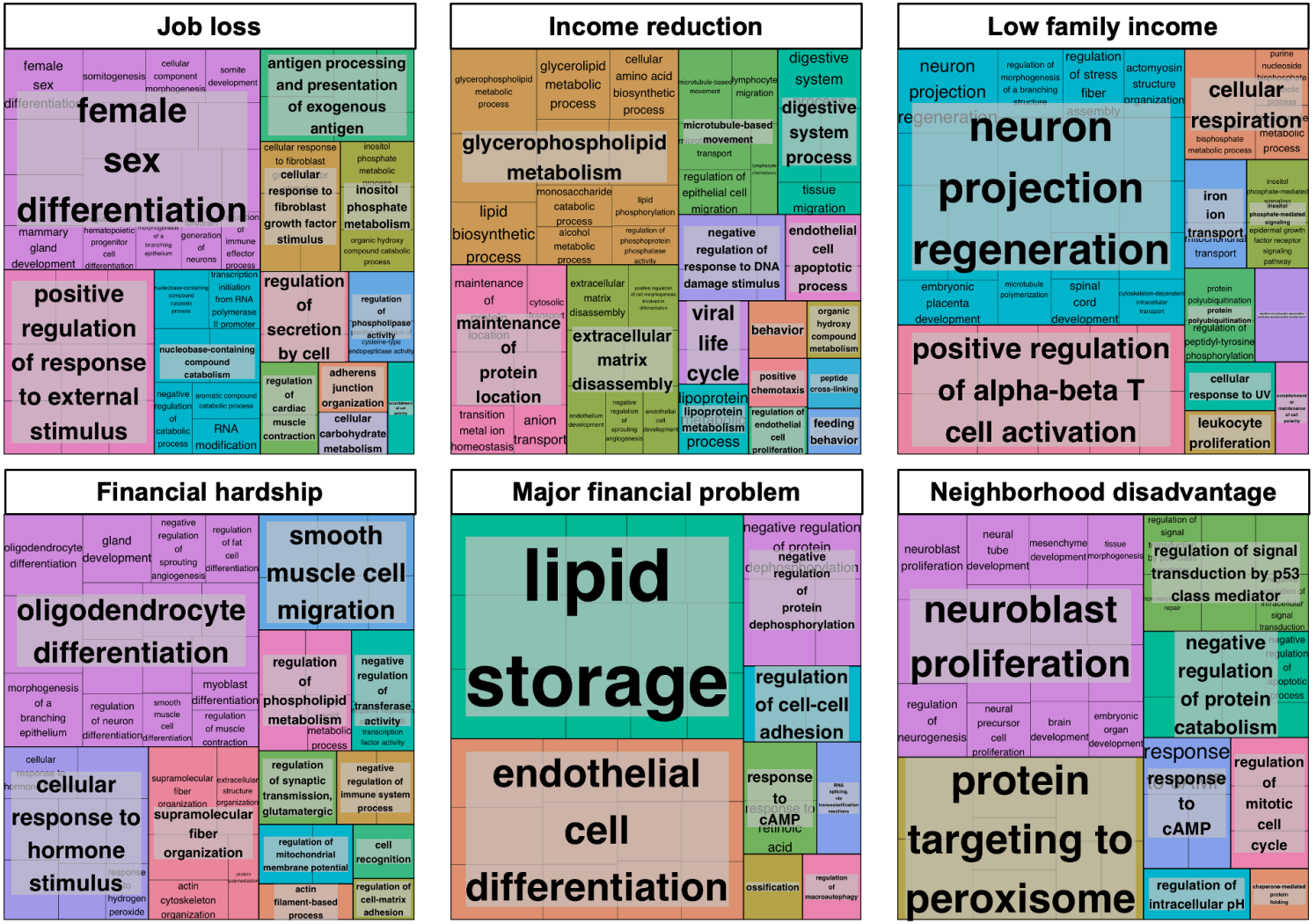

Figure S8. Clusters of biological pathways identified in the top 100 Gene Ontology (GO) terms of each SEP measure. GO terms are presented in cells whose size is proportional to the level of significance (-log(p)). Clusters of GO terms were determined by semantic similarity calculated by REVIGO (http://revigo.irb.hr). Clusters are labeled by the GO terms with the lowest p value within a cluster.

Table S1 (separate file). Distribution of covariates in the ALSPAC cohort (N=15,445) and in the ARIES subsample (N=946).

Table S2 (separate file). Results of the SLCMA with annotation to the closest gene, for the 62 CpGs linked to socioeconomic adversity explaining more than 3% variability in DNA methylation (R^2>3%).

Table S3 (separate file). Comparison to previous EWAS for the 62 CpG sites linked to socioeconomic adversity explaining more than 3% variability in DNA methylation (R^2>3%).

Table S4 (separate file). Summary of tested hypotheses by each socioeconomic adversity.

Table S5 (separate file). Results of sensitivity analyses adjusting for additional covariates for the 62 CpG sites identified by SLCMA analysis (R^2>3%).

Table S6 (separate file). Results of sensitivity analysis controlling for genetic variation for R^2>3% CpGs linked to mQTLs.

Table S7 (separate file). Results of the sensitivity analysis excluding mobility hypotheses in SLCMA, for the R^2>3% CpGs associated with four SEP indicators.

Table S8 (separate file). Correlation of methylation between blood and four brain regions for the 62 CpGs with R^2>3%

Supplementary References

1 Relton, C.L., Gaunt, T., McArdle, W., Ho, K., Duggirala, A., Shihab, H., Woodward, G., Lyttleton, O., Evans, D.M., Reik, W. *et al.* (2015) Data Resource Profile: Accessible Resource for Integrated Epigenomic Studies (ARIES). *Int. J. Epidemiol.*, **44**, 1181-1190.

2 Boyd, A., Golding, J., Macleod, J., Lawlor, D.A., Fraser, A., Henderson, J., Molloy, L., Ness, A., Ring, S. and Davey Smith, G. (2013) Cohort Profile: The 'Children of the 90s'—the index offspring of the Avon Longitudinal Study of Parents and Children. *Int. J. Epidemiol.*, **42**, 111-127.

3 Golding, J., Pembrey, M., Jones, R. and the ALSPAC Study Team. (2001) ALSPAC: The Avon Longitudinal Study of Parents and Children I. Study methodology. *Paediatr. Perinat. Epidemiol.*, **15**, 74-87.

4 Fraser, A., Macdonald-Wallis, C., Tilling, K., Boyd, A., Golding, J., Davey Smith, G., Henderson, J., Macleod, J., Molloy, L., Ness, A. *et al.* (2013) Cohort Profile: the Avon Longitudinal Study of Parents and Children: ALSPAC mothers cohort. *Int. J. Epidemiol.*, **42**, 97-110.

5 Cerutti, J.K., Lussier, A.A., Zhu, Y., Liu, J. and Dunn, E.C. (2021) Associations between indicators of socioeconomic position and DNA methylation: A scoping review. *Clin. Epigenetics*, **13**.

6 Dunn, E.C., Soare, T.W., Raffeld, M.R., Busso, D.S., Crawford, K.M., Davis, K.A., Fisher, V.A., Slopen, N., Smith, A., Tiemeier, H. *et al.* (2018) What life course theoretical models best explain the relationship between exposure to childhood adversity and psychopathology symptoms: recency, accumulation, or sensitive periods? *Psychol. Med.*, **48**, 2562-2572.

7 Dunn, E.C., Soare, T.W., Zhu, Y., Simpkin, A.J., Suderman, M.J., Klengel, T., Smith, A., Ressler, K.J. and Relton, C.L. (2019) Sensitive periods for the effect of childhood adversity on DNA methylation: results from a prospective, longitudinal study. *Biol. Psychiatry*, **85**, 838-849.

8 Andersen, S.L., Tomada, A., Vincow, E.S., Valente, E., Polcari, A. and Teicher, M.H. (2008) Preliminary evidence for sensitive periods in the effect of childhood sexual abuse on regional brain development. *J. Neuropsychiatry Clin. Neurosci.*, **20**, 292-301.

9 Slopen, N., Kubzansky, L.D., McLaughlin, K.A. and Koenen, K.C. (2013) Childhood adversity and inflammatory processes in youth: a prospective study. *Psychoneuroendocrino.*, **38**, 188-200.

10 Jenkins, S.P. (2016) The income distribution in the UK. in press.

11 Min, J.L., Hemani, G., Davey Smith, G., Relton, C. and Suderman, M. (2018) Meffil: efficient normalization and analysis of very large DNA methylation datasets. *Bioinform.*, **34**, 3983-3989.

12 Houseman, E.A., Accomando, W.P., Koestler, D.C., Christensen, B.C., Marsit, C.J., Nelson, H.H., Wiencke, J.K. and Kelsey, K.T. (2012) DNA methylation arrays as surrogate measures of cell mixture distribution. *BMC Bioinform.*, **13**, 86.

13 Triche, T. (2014), in press.

14 Pal, S. and Tyler, J.K. (2016) Epigenetics and aging. *Sci Adv*, **2**, e1600584.

15 Zhu, Y., Simpkin, A.J., Suderman, M.J., Lussier, A.A., Walton, E., Dunn, E.C. and Smith, A.D.A.C. (2021) A Structured Approach to Evaluating Life Course Hypotheses: Moving Beyond Analyses of Exposed Versus Unexposed in the Omics Context. *Am. J. Epidemiol.*, **190**, 1101-1112.

16 Smith, B.J., Smith, A.D.A.C. and Dunn, E.C. (2022) Geyer, M.A., Marsden, C.A., Ellenbroek, B.A., Barnes, T.R.E. and Andersen, S.L. (eds.), In *Curr. Top. Behav. Neurosci.* Springer, Berlin, Heidelberg, in press.
17 Smith, A.D., Heron, J., Mishra, G., Gilthorpe, M.S., Ben-Shlomo, Y. and Tilling, K. (2015) Model Selection of the Effect of Binary Exposures over the Life Course. *Epidemiol.*, **26**, 719-726.
18 Efron, B., Hastie, T., Johnstone, I. and Tibshirani, R. (2004) Least angle regression. *Ann. Stat.*, **32**, 407-499.
19 Tibshirani, R.J., Taylor, J., Lockhart, R. and Tibshirani, R. (2016) Exact Post-Selection Inference for Sequential Regression Procedures. *J. Am. Stat. Assoc.*, **111**, 600-620.
20 Frisch, R. and Waugh, V.F. (1933) Partial Time Regressions as Compared with Individual Trends. *Econometrica*, in press.
21 Lovell, M.C. (1963) Seasonal adjustment of economic time series and multiple regression analysis. *J. Am. Stat. Assoc.*, **58**, 993-1010.
22 Lovell, M.C. (2008) A simple proof of the FWL theorem. *J. Econ. Educ.*, **39**, 88-91.
23 Yamada, H. (2017) The Frisch–Waugh–Lovell theorem for the lasso and the ridge regression. *Commun. Stat.*, **46**, 10897-10902.
24 Lussier, A.A., Zhu, Y., Smith, B.J., Simpkin, A.J., Smith, A.D.A.C., Suderman, M.J., Walton, E., Ressler, K.J. and Dunn, E.C. (2022) Updates to data versions and analytic methods influence the reproducibility of results from epigenome-wide association studies. *Epigenetics*, in press.
25 Rakyan, V.K., Down, T.A., Balding, D.J. and Beck, S. (2011) Epigenome-wide association studies for common human diseases. *Nat. Rev. Genet.*, **12**, 529-541.
26 Rahmani, E., Shenhav, L., Schweiger, R., Yousefi, P., Huen, K., Eskenazi, B., Eng, C., Huntsman, S., Hu, D., Galanter, J. *et al.* (2017) Genome-wide methylation data mirror ancestry information. *Epigenetics Chromatin*, **10**, 1.
27 Gaunt, T.R., Shihab, H.A., Hemani, G., Min, J.L., Woodward, G., Lyttleton, O., Zheng, J., Duggirala, A., McArdle, W.L., Ho, K. *et al.* (2016) Systematic identification of genetic influences on methylation across the human life course. *Genome Biol.*, **17**, 61.
28 Suderman, M., Pappas, J.J., Borghol, N., Buxton, J.L., McArdle, W.L., Ring, S.M., Hertzman, C., Power, C., Szyf, M. and Pembrey, M. (2015) Lymphoblastoid cell lines reveal associations of adult DNA methylation with childhood and current adversity that are distinct from whole blood associations. *Int. J. Epidemiol.*, **44**, 1331-1340.
29 Karlsson Linnér, R., Marioni, R.E., Rietveld, C.A., Simpkin, A.J., Davies, N.M., Watanabe, K., Armstrong, N.J., Auro, K., Baumbach, C., Bonder, M.J. *et al.* (2017) An epigenome-wide association study meta-analysis of educational attainment. *Mol. Psychiatry*, **22**, 1680-1690.
30 van Dongen, J., Bonder, M.J., Dekkers, K.F., Nivard, M.G., van Iterson, M., Willemsen, G., Beekman, M., van der Spek, A., van Meurs, J.B.J., Franke, L. *et al.* (2018) DNA methylation signatures of educational attainment. *NPJ Sci. Learn.*, **3**, 7.
31 Bush, N.R., Edgar, R.D., Park, M., MacIsaac, J.L., McEwen, L.M., Adler, N.E., Essex, M.J., Kobor, M.S. and Boyce, W.T. (2018) The biological embedding of early-life socioeconomic status and family adversity in children's genome-wide DNA methylation. *Epigenomics*, in press.
32 Alfano, R., Guida, F., Galobardes, B., Chadeau-Hyam, M., Delpierre, C., Ghantous, A., Henderson, J., Herceg, Z., Jain, P., Nawrot, T.S. *et al.* (2019) Socioeconomic position during

pregnancy and DNA methylation signatures at three stages across early life: epigenome-wide association studies in the ALSPAC birth cohort. *Int. J. Epidemiol.*, **48**, 30-44.
33 McDade, T.W., Ryan, C.P., Jones, M.J., Hoke, M.K., Borja, J., Miller, G.E., Kuzawa, C.W. and Kobor, M.S. (2019) Genome-wide analysis of DNA methylation in relation to socioeconomic status during development and early adulthood. *Am. J. Phys. Anthropol.*, **169**, 3-11.
34 Laubach, Z.M., Perng, W., Cardenas, A., Rifas-Shiman, S.L., Oken, E., DeMeo, D., Litonjua, A.A., Duca, R.C., Godderis, L., Baccarelli, A. *et al.* (2019) Socioeconomic status and DNA methylation from birth through mid-childhood: a prospective study in Project Viva. *Epigenomics*, **11**, 1413-1427.
35 Hannon, E., Lunnon, K., Schalkwyk, L. and Mill, J. (2015) Interindividual methylomic variation across blood, cortex, and cerebellum: implications for epigenetic studies of neurological and neuropsychiatric phenotypes. *Epigenetics*, **10**, 1024-1032.
36 Ren, X. and Kuan, P.F. (2018) methylGSA: a Bioconductor package and Shiny app for DNA methylation data length bias adjustment in gene set testing. *Bioinform.*, **35**, 1958-1959.
37 Kolde, R., Laur, S., Adler, P. and Vilo, J. (2012) Robust rank aggregation for gene list integration and meta-analysis. *Bioinform.*, **28**, 573-580.