ASSOCIATIONS OF DNA METHYLATION WITH ADVERSITY THROUGHOUT THE LIFE COURSE

by

NICOLE GLADISH

BSc., University of British Columbia, 2012

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

in

THE FACULTY OF GRADUATE AND POSTDOCTORAL STUDIES

(Medical Genetics)

THE UNIVERSITY OF BRITISH COLUMBIA

(Vancouver)

August 2021

The following individuals certify that they have read, and recommend to the Faculty of Graduate and Postdoctoral Studies for acceptance, the dissertation entitled:

Associations of DNA methylation with adversity throughout the life course

submitted by	Nicole Gladish	in partial fulfillment of the requirements for
the degree of	Doctor of Philosophy	
in	Medical Genetics	

Examining Committee:

Dr. Michael S. Kobor, Medical Genetics, UBC
Supervisor

Dr. Martin Hirst, Microbiology and Immunology, UBC
Supervisory Committee Member

Dr. Sophia Frangou, Psychiatry, UBC
University Examiner

Dr. Louis Lefebvre, Medical Genetics, UBC
University Examiner

Dr. Connie J. Mulligan, Anthropology, UF
External Examiner

Additional Supervisory Committee Members:

Dr. Jan M. Friedman, Medical Genetics, UBC
Supervisory Committee Member

Dr. Fabio Rossi, Medical Genetics, UBC
Supervisory Committee Member

Abstract

Interpersonal and societal adversity experienced throughout life, such as abuse or low socio-economic status (SES), have been associated with negative health outcomes. Many of these diseases share the common thread of inflammation, inspiring the hypothesis that chronic stress resulting from adverse experiences over activates the hypothalamic-pituitary-adrenal (HPA) axis resulting in dysregulation of the immune system. DNA methylation (DNAm), a methyl group covalently bound to cytosine bases for the purposes of this thesis, is one of many epigenetic mechanisms involved in responding to environmental signals in the genomic context. As such, this modification may be particularly pertinent to understanding how adverse experiences can become embedded in a way that results in lifelong health disparities.

The overarching aim of this dissertation is to understand how various measures of adversity throughout the life-course could associate with DNAm differences between individuals. Initially, I compared whole blood DNAm patterns amongst elderly individuals with different years of education, household assets and self-reported measures of economic standing in childhood and older adulthood, in addition to a composite socioeconomic (SES) measure. I found there were significantly more DNAm associations with older adulthood relative to retrospective childhood SES measures, and the subjective SES measures displayed a dampened signal relative to objective ones. Next, I investigated the relationship between the inflammatory biomarker serum IL-6, lifetime SES measures and purified monocyte DNAm patterns amongst adults. Here, I found the relationship between some CpGs and IL-6 was partially explained by SES. Additionally, differences in SES-associated DNAm sites were seen predominantly amongst individuals who experienced low early life and high adulthood SES. Finally, I investigated how childhood abuse associated with DNAm in spermatozoa of adult men. Though this was a small pilot study, I found a subset of differentially methylated regions associated with childhood abuse. These associations were stable over a two-month period and survived adjusting for current life adversity measures. Overall, these findings provide evidence that the tissue, timing, and measure of adversity experienced are important considerations for social epigenetic studies and can yield unique DNAm patterns in a context-specific way.

Lay Summary

Health disparities are associated with an individual's interpersonal and societal adverse experiences. How adversity impacts lifelong health at the molecular level is not fully understood. Epigenetics is the study of how the genome can interact with and respond to the environment. DNA methylation (DNAm) is an epigenetic mark that is important for cellular identity and gene regulation. Epigenetics may be the link between experiences of adversity and long-term cellular programming resulting in negative health outcomes. This thesis investigates how DNAm relates to various measures of adversity including abuse and low SES, measured as occupational prestige, education, household assets, or measures of self-report throughout life in several tissues including blood, monocytes and sperm. Results support the conclusion that several measures of life-long adversity do associate with DNAm and may provide insight into relevant molecular mechanisms at play and how they integrate into the pathways linking adversity and health.

Preface

Data **Chapter 4** in this thesis is presented in manuscript format, as it is currently published. Data **Chapters 2 & 3** are original and unpublished.

Portions of **Chapter 1** (introduction) have been adapted from previously published work:

- Merrill SM, Gladish N, Kobor MS. (2019). Social Environment and Epigenetics. In: Binder E & Klengel T (Eds) *Behavioral Neurogenomics. Current Topics in Behavioral Neurosciences, vol 42*. Springer, Cham. Reprinted with permission of Springer, Cham (License Number: 4947321357680).

Chapter 2 is original and unpublished.

- Gladish N, Merrill SM, McEwen LM, MacIsaac JL, Lin DTS, Ramadori KE, Dow WH, Rosero-Bixby L, Kobor MS, Rehkopf DH. Multiple components of socioeconomic status measured in older adulthood and childhood differentially associated with DNA methylation.

Processing of all human samples used in this study were approved by the joint University of British Columbia and Children and Women's Hospital Ethics board (Certificate: H16-02570). I developed the design and research questions for this study alongside D. Rehkopf and M. Kobor. Sample and data collection were performed in Costa Rica by W. Dow and L. Rosero-Bixby. DNA methylation and genotyping arrays were performed by L. McEwen, J. MacIsaac, D. Lin and K. Ramadori. I was responsible for all bioinformatics analyses and manuscript preparation, with critical feedback from S. Merrill, D. Rehkopf and M. Kobor.

Chapter 3 is original and unpublished.

- Gladish N, Merrill SM, MacIsaac JL, Chen E, Miller GE[§], Kobor MS[§]. Lifetime trajectories of socioeconomic status partially mediated associations between DNA methylation and IL-6 levels. [§]Authors jointly supervised work.

Collection of all human samples used in this study were approved by the joint University of British Columbia and Children and Women's Hospital Ethics board (Certificate: H08-02773). The experimental and analytical design for this study was developed by myself, G. Miller and M. Kobor. Sample and data collection were performed in Vancouver, BC by G. Miller and E. Chen. DNA methylation arrays were run by myself and J. MacIsaac. I

performed all statistical and bioinformatics analyses of the DNA methylation array data. I wrote the manuscript and prepared all publication figures. S. Merrill, E. Chen, G. Miller and M. Kobor provided critical feedback for the analyses and manuscript preparation. G. Miller and M. Kobor supervised all steps of the process.

A version of **Chapter 4** has been published as:

- Roberts AL, Gladish N, Gatev E, Jones MJ, MacIsaac JL, Tworoger SS, Austin SB, Tanrikut C, Chavarro JE, Baccarelli AA, Kobor MS. (2018). Exposure to childhood abuse is associated with human sperm DNA methylation. *Translational Psychiatry*. Reprinted with permission of Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/).

Collection of all human samples used in this study were approved by the Brigham and Women's Hospital Institutional Review Board (Certificate: 2009-P-000116/11). This study was conceived and designed by A. Roberts and myself with input from M. Kobor. Recruitment of study participants and sample preparation was done by A. Roberts, J. Chavarro, S. Tworoger, S. Austin and C. Tanrikut. DNA methylation arrays were run by J. MacIsaac and M. Jones and pyrosequencing was performed by myself. I was responsible for statistical analysis of all data except the production of the predictor which was performed by E. Gatev. I generated most publication figures and helped write the manuscript with A. Roberts. A. Baccarelli and M. Kobor provided critical feedback at all stages of manuscript preparation.

Chapter 5 (discussion) contains adapted excerpts from the following publication:

- Merrill SM, Gladish N, Kobor MS. (2019). Social Environment and Epigenetics. In: Binder E & Klengel T (Eds) *Behavioral Neurogenomics. Current Topics in Behavioral Neurosciences, vol 42*. Springer, Cham. Reprinted with permission of Springer, Cham (License Number: 4947321357680).

Given that **Chapter 4** remains largely unchanged from the published version, I have retained the use of plural first-person pronouns in these sections. In the remainder of the dissertation, singular first-person pronouns are employed.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

27K	Illumina HumanMethylation27K Beadchip Array
450K	Illumina Infinium HumanMethylation450K Beadchip Array
ACTH	Adrenocorticotropic hormone
BECon	Blood-Brain Epigenetic Concordance
BMI	Body mass index
CpG	Cytosine-guanine dinucleotide
CRELES	Costa Rican Longevity and Healthy Aging Study
CRH	Corticotropin-releasing hormone
CRP	C-reactive protein
CTQ	Childhood Trauma Questionnaire
CTS	Conflict Tactics Scales
DHEAS	Dehydroepiandrosterone sulfate
DMR	Differentially methylated regions
DNA	Deoxyribose nucleic acid
DNAm	DNA methylation
DNMT	DNA methyltransferase
DOHaD	Developmental Origins of Health and Disease
EL	Early Life Study
EPIC	Illumina Infinium HumanMethylationEPIC Beadchip Array
EWAS	Epigenome-wide association study
FACS	Fluorescence-activated cell sorting
FDR	False discovery rate
GEO	Gene Expression Omnibus
GR	Glucocorticoid receptor
GSA	Illumina Infinium Global Screening Array
GUTS	Growing Up Today Study
GWAS	Genome-wide association study
GxE	Gene-by-environment

HPA	Hypothalamic-pituitary-adrenocortical
HWE	Hardy-Weinberg Equilibrium
IBD	Identity-by-descent
ICC	Interclass correlation coefficient
IL-1β	Interleukin-1 beta
IL-6	Interleukin-6
LD	Linkage disequilibrium
MAF	Minor allele frequency
MeDIP	Methylated DNA immunoprecipitation
MHC	Major histocompatibility complex
miRNA	Micro ribonucleic acid
mQTL	Methylation quantitative trait locus
mRNA	Messenger ribonucleic acid
PBMC	Peripheral blood mononuclear cells
PC(A)	Principal component (analysis)
PCR	Polymerase chain reaction
PTSD	Post-traumatic stress disorder
SAM	sympathetic-adrenomedullary
scBS-seq	Single-cell bisulfite sequencing
scRRBS	Single-cell reduced representation bisulfite sequencing
SES	Socioeconomic status
sncRNA	Small non-coding ribonucleic acid
SNP	Single nucleotide polymorphism
SNS	Sympathetic nervous system
SVA	Surrogate variable analysis
SWAN	Subset-quantile within array normalization
TLR	Toll-like receptor
TNFα	Tumor necrosis factor alpha
WHR	Waist-to-hip ratio
Δβ	Delta beta
27K	Illumina HumanMethylation27K Beadchip Array

Acknowledgements

The completion of this PhD dissertation could not have been possible without the support of numerous people. Firstly, my sincere gratitude Dr. Michael S. Kobor, for his mentorship and for deciding to take a risk and supervise me during a tumultuous time, the integrity you displayed by supporting my efforts to do the ethical thing will always inspire me to persevere. My genuine appreciation to my fantastic committee members, Dr. Jan Friedman, Dr. Martin Hirst and Dr. Fabio Rossi for your continuous guidance and support.

I would like to acknowledge all of my wonderful collaborators: Dr. Andrea Roberts, Dr. Greg Miller and Dr. David Rehkopf, you made this degree especially rewarding. I would also like to thank the members of the Medical Genetic graduate program, including my professors, fellow graduate students, and in particular, Cheryl Bishop, whose kindness and encouragement has meant a lot.

Thank you to former and current members of the Kobor lab, it is through many of you that I was able to learn and grow as a researcher. A special thank you to Dr. Sarah Merrill, Rachel Edgar, Dr. Lisa McEwen, Dr. Sumaiya Islam, and Dr. Chaini Konwar who not only provided critical feedback and support throughout my PhD but also inspire me. Drs. Daniel Radiloff and Sumaiya Islam deserve additional acknowledgment, they supported me through the most difficult time during my degree, prior to entering the Kobor lab, and it because of your courage and integrity that I was able to do the right thing and push through.

Finally, those I am most thankful for, and the main reason I could accomplish what I did, has been my family. To my two beautiful children, Luca and Elena, you were lights of sunshine throughout this process and wonderful reminders of why I do what I do. I hope these experiences will enrich your lives, even if only a modicum, of how you have enriched mine. Jay, my amazing husband and best friend, you have stuck with me and supported me financially and emotionally as I follow my passion. I am beyond grateful to have you in my life and I really, truly and honestly, could never have done this without you. I love you all with all my heart.

Dedication

To Jay, my amazing husband and Luca and Elena, my wonderful children, for all of your love, support and sacrifice. You have always been my safe haven; I could have never done this without you.

1 Introduction

1.1 Dissertation context and aims

Adversity experienced in both the interpersonal and societal context has been consistently associated with many negative health outcomes, with effects especially damaging when experienced during the vulnerable period of early life (1–13). All children develop within a dynamic social context of both interpersonal relationships and wider social structures, which can shape their cognitive, emotional, and biological processes for the remainder of their lives (14). From the first moments with parents, to city planning, to feeling accepted by the community, the multifaceted nature of the social environment provides ample opportunity for both advantages and hindrances to be embedded "under the skin" (15). Though the debate between solitary contributions of nature and nurture is an academic artifact, there are a variety of overlapping conceptualizations of this embedding process. This cascade of interactions between innate differences and environmental exposures is addressed in the developmental origins of health and disease (DOHaD) hypothesis (16,17). This hypothesis refers to the potential biological programming from environmental exposures that may cause some of the intrinsic predispositions on which later environments may act (18). Layered onto the foundation of early life embedding, subsequent experiences throughout life continue to shape an individual's health highlighting the need to investigate these relationships in a more holistic way.

The most general and often-used scientific term for these relationships is gene-by-environment (GxE) interactions. Ultimately, the distillation of these models equates to a basic understanding that the scaffolding of experience can be secured upon existing biological foundations. With poor foundations or inferior craftsmanship, the overall integrity of the structure may fail. One important aspect of how these pieces come together to build the human experience is through epigenetics (19).

This dissertation aimed at understanding how the social environment, specifically in the context of interpersonal and societal adversity, associates with interindividual epigenetic differences in a healthy population. To achieve this, I performed a series of epigenome-wide association studies (EWASs) interrogating DNA methylation (DNAm) profiles in several populations. DNAm is a particularly promising candidate as this is an epigenetic marker that has been demonstrated to play a role in gene expression regulation, cellular differentiation and the

preservation of cellular identity (20–22). Additionally, DNAm is dynamic, as it has been shown to fluctuate in response to various environmental influences (23,24). To examine this, I performed three studies to assess adversity-associated DNAm by a) comparing multiple measures of socioeconomic status (SES), b) investigating how an inflammatory biomarker may play a role and c) examining these associations in gametes. Specifically, I investigated DNAm patterns related to education, household assets and self-reported measures of economic standing in both retrospective childhood and concurrent older adulthood in the whole blood of elderly individuals (**Chapter 2**). Next, I wanted to determine what role, if any, the inflammatory biomarker interleukin-6 (IL-6) plays amongst SES-associated DNAm sites within purified monocytes (**Chapter 3**). Finally, to examine if signals of adversity could be detected in gametes, I investigated how childhood abuse associated to the DNAm profiles of adult spermatozoa samples (**Chapter 4**). Overall, these findings provide evidence that adversity experienced in various contexts is associated with DNAm throughout life, evidenced in multiple tissues, and highlights avenues for potential future research.

1.2 Epigenetics at the intersection between social environment and physiology

1.2.1 Environmental influence on epigenetic mechanisms and impact on cellular function

Epigenetics refers to genetic modifications which regulate the underlying, permanent deoxyribose nucleic acid (DNA) sequence, leading to altered gene expression and, ultimately, changes in phenotypes such as health and behaviors. The fundamental purpose of these epigenetic alterations is to achieve a diverse landscape of expression from a single DNA source (15). These changes are likely involved in the biological embedding of environmental influence because of both their dynamic nature and sensitivity to experiential feedback. The field of epigenetics is a natural accretion of biological reductionism as it provides evidence that while we may be a product of our biology, our biology is partially a product of our environment.

The degree to which a gene is transcribed and translated into protein depends, in part, on the chromatin state. This packaging of DNA and its associated proteins is determined by both underlying DNA sequence and external factors, and thus can potentially link genetic and environmental mechanisms. Examples of mechanisms that can change the chromatin state and potentially influence transcription include covalent modifications of histone proteins and DNA bases, such as cytosine DNAm (21).

Modifications in chromatin structure during cellular differentiation are of paramount importance to healthy development (25–27), and the role that chromatin state plays in mediating adaptive responses to environmental influences remains significant throughout life (15). For example, Angelman syndrome results from *in utero* aberrant methylation of the paternally-imprinted ubiquitin protein ligase E3A (*UBE3A*) gene, resulting in lifelong seizures and intellectual disability (28). Considering the importance of accurate epigenetic states for typical development, these mechanisms may explain how a short-duration environmental exposure could influence an individual's physiology decades later (15,29).

Animal studies have provided a growing body of evidence to support epigenetically mediated effects of prenatal environment on offspring phenotype (30,31). In particular, the Agouti mouse model provides a seminal example of prenatal environment influencing phenotype through an epigenetic mark, specifically cytosine DNAm. In such experiments, pregnant mice carrying offspring with an Agouti viable yellow (*AVY*) allele are fed a methyl-rich diet, which increases methylation of *Agouti* in the offspring and results in a brown coat color, decreased obesity, and improved health outcomes (32,33). While it is likely that numerous epigenetic marks interact with the environment and genetic landscape in association with various phenotypes, this thesis will center specifically on cytosine DNAm, specifically findings from human cohorts, as DNAm is by far the most extensively studied epigenetic mark investigated in human populations (34).

1.2.2 Epigenome-wide association studies (EWASs)

As there are a multitude of epigenetic marks that can be affected by many environmental exposures, attempts have been made to apply epidemiological approaches to investigate DNAm associations with a given phenotype, resulting in study designs referred to as epigenome-wide association studies (EWASs) (35). These are exploratory studies, similar to the more established genome-wide association studies (GWASs), where the relationship between a particular epigenetic mark and a variable of interest is performed on a site-by-site basis across the genome. Limitations that plague GWASs, such as harmonization of variables, functional relevance of findings, population stratification and sample size issues, also plague EWASs (35). However, EWASs have additional limitations, as epigenetic marks can be influenced by a wide variety of environmental factors in addition to underlying genetics, including age, sex, and tissue

heterogeneity, making study design extraordinarily important (35). The most common forms of epigenetic investigations in humans are correlational studies on DNAm (36,37). Due to its critical role in cell type differentiation, DNAm is highly affected by cell type differences, and thus tissue types, as well as age, ethnicity, genotype, sex, and disease state (38–46). Three of these factors are especially important to highlight: tissue, age, and genotype.

1.3 DNA methylation (DNAm)

1.3.1 Function and properties

DNAm consists of a methyl group covalently bound to nucleotide bases in the DNA, for the purposes of this thesis I will focus specifically on cytosine DNAm. The resulting nucleic base is 5-methyl-cytosine and occurs primarily in the context of cytosine-guanine dinucleotides (CpG) (20). CpGs, though located throughout the genome, are most frequently found in CpG islands, defined as regions around 1,000 base pairs in length with a high concentration of CpG sequences. These CpG islands are overrepresented in gene promoters, being present in 70% of them, in part due to the fact that they are ancestral remnants of the genome and so evolutionarily conserved (47). Approximately 2-5% of all cytosines in the genome are methylated, with 90% of methylated cytosines existing within repetitive elements such as Alu and Long Interspersed Nuclear Element-1 (*LINE-1*) transposons (48). There are various functions associated with DNAm, including chromatin remodelling, X-chromosome inactivation (49,50), imprinting (51), and transcription (52,53) — transcription being of particular interest as a mechanism mediating the effects of environmental factors (54). Somatic cells have a unique pattern of DNAm that is maintained throughout their lifetime via the action of DNA methyltransferase enzymes (DNMTs). However, while some DNAm patterns essential for cellular differentiation are more permanent, some remain dynamic throughout life (37). As a primary function of DNAm is in defining and maintaining cellular identity, several unique and important experimental considerations need to be accounted for (53). Tissues are heterogenous and contain various cell types, all with unique DNAm profiles, and individuals vary in the proportions of these cell types requiring additional correction in downstream analyses (53) and careful consideration when interpreting results (36).

1.3.2 Tissue specificity

While the genetic background of all cells in a given human body is identical, there are a vast number of cell types with their own acquired and unique functions. Epigenetics, and specifically DNAm, regulate gene expression within the context of genetic background, defining a given cell's functionality and therefore its identity (53). Cellular identity, being a primary function of DNAm, brings with it several key issues when being investigated as a marker in epidemiological studies (55,56). First, the tissue being interrogated has specific implications for results interpretation, as DNAm patterns between tissues are predominantly the primary source of variation (46,57,58). Typically, peripheral tissues, easily accessible and non-invasive, are most commonly interrogated in EWASs. As such, it is difficult to make inferences about epigenetic modifications in brain tissue when measuring more peripheral tissues, such as blood or saliva; however, there are resources correlating DNAm measures in these tissues that can assist in forming educated inferences (38,59). This same logic applies to interpreting the effect of differential DNAm occurring in a gene which is not normally expressed in the interrogated tissue (55,56).

The second issue is the cellular composition of the interrogated tissue, the second largest contributor to DNAm variation. Many tissues are heterogenous and the proportions of cell type which comprise them can vary across individuals (57). As this adds considerable variation to DNAm data and when varying cell type proportions are not the variable of interest, this must be accounted for in subsequent analyses (60,61). While the gold standard is to perform cell counts or fluorescence activated cell sorting (FACS) analysis to assay the proportion of cell types directly, it may not be feasible for larger cohorts. As such, several bioinformatics approaches have been produced to predict varying cell type proportions in many commonly assayed tissues (62–66). The most common approach includes using reference datasets, derived from assaying purified cell types from FACS-sorted tissues. The methylation profiles of these purified cells can be used to predict proportions of cell types in another sample. When no reference dataset exists, there are statistical methods, such as surrogate variable analysis (SVA), that can provide estimates based on a general knowledge of how many cell types exist and that this is the primary driver of variation in DNAm data (66,67). While differences in cell type proportions could be associated with the variable of interest and themself interesting, these measures are usually treated as covariates that must be accounted for in EWASs (55,56).

1.3.3 Age

Patterns of epigenetic modifications, including DNAm, change with age (37,68–73). There is a global loss of DNAm with aging, with about one third of DNAm sites being impacted (70). This hypomethylation occurs within all genomic regions, but especially within CpG-poor promoters and tissue specific genes (74). However, increases in DNAm with age are found in CpG dense regions, such as CpG islands (75,76). In general, there are two categories with DNAm and aging: epigenetic drift and epigenetic clocks.

Epigenetic drift refers to the increase in DNAm variability associated with aging. This variability occurs at specific genomic regions, but the directionality varies across individuals suggesting that the source of variation is the result of stochastic changes or differential environmental exposures (37,77,78). Monozygotic twin studies highlight this effect where methylome profiles are nearly identical at birth and become subsequently more divergent with age (79). Epigenetic drift has also been associated with various health outcomes, which could be a direct effect of variation or a proxy of progressive inefficiencies in genetic regulation seen with aging (80–82).

In contrast, there are age-associated DNAm sites that change in a predictable way across individuals. A benefit of these predictable differences is the ability to estimate age using DNAm "clocks" that can be either pan-tissue or tissue specific (83–88). By estimating age using DNAm, it is also possible to determine the acceleration of an individual's epigenetic, or biological, age from their chronological age. Generally, when an adult is predicted to be older than their chronological age (i.e., epigenetic age acceleration), this suggests increased cellular aging and is associated with increases in morbidity and mortality (70,89,90).

Interestingly, the relationships between age and DNAm are not linear where the rate of change differs significantly amongst children, adults and the elderly in addition to across tissue (91). Therefore, it is crucial that age be accounted for, selected for, or counterbalanced across groups in the variable of interest.

1.3.4 Genetic variation

Another factor that contributes greatly to DNAm is the genomic background, both as a confounder to be accounted for and as a variable of interest. GxE interactions, defined as

differential phenotypes resulting from a combination of genetic background and environmental exposures (92,93), have been an area of intensive research in which DNAm is increasingly investigated as a potential mediator (94). One example of this was found when investigating the link between a risk allele in *FKBP5* for post-traumatic stress disorder (PTSD) being mediated by childhood maltreatment through changes in methylation in the *FKBP5* intron responsible for glucocorticoid receptor (GR) binding (95). A genetic variant linked to the alteration of a DNAm site is termed a methylation quantitative trait locus (mQTL), a significant source of DNAm variation. mQTLs have been reported to be tissue, age, and population dependent, contributing to 20-80% of DNAm variation (74–77). The dynamic relationship between DNAm and genotype highlights not only the importance of accounting for population stratification, but also the pertinent role it may play in developmental trajectories and later life health (96).

1.3.5 DNAm measuring methods

At the advent of DNAm research, this epigenetic mark was first analyzed at the global methylation level using fluorescence microscopy and immunoblotting techniques (97,98). As technology became more advanced, interrogation at specific sites became possible using sequencing-based methods, with the eventual development of array-based methods allowing for high-throughput interrogation in population studies (99,100). The general technique used to interrogate DNAm using sequencing-based methods, is to perform bisulfite treatment of the samples converting non-methylated cytosine into uracil bases. Bisulfite converted DNA is then subjected to typical sequencing techniques referred to as whole-genome bisulfite sequencing.

While sequencing is now possible genome-wide, other more targeted sequencing methods have been developed. One such method sequences genomic fragments selected using antibodies which bind specifically to methylated DNA (MeDIP); however, this method is biased for only enriching for regions of high CpG density (101–103). Pyrosequencing is another targeted method, where a polymerase chain reaction (PCR) is performed on bisulfite converted DNA releasing light during each base pair extension which is measured to indicate the addition of a known nucleotide (104).

The existence of hydroxy-methylated cytosines (DNAhm) should also be noted, these modifications are the intermediary produced during ten-eleven translocation (TET) enzyme-mediated demethylation. DNAhm varies widely across tissues and can act as a stable

modification in a few instances. DNAhm is protected during bisulfite treatment, similar to DNAm, and so are counted in the pool of DNA molecules as methylated. There are techniques available to distinguish between these modifications; however, other than a few exceptions (e.g., brain, pluripotent stem cells), DNAhm levels are low enough in most tissues to not make this necessary for most EWASs (105–108).

The vast majority of EWASs performed on larger cohorts are conducted using the Illumina Infinium BeadChip arrays (109–112). There have been four iterations throughout the years with more sites being assayed with increasing technological advances. These include the GoldenGate, 27K, 450K and EPIC arrays which interrogate 1,536, 27,578, 485,557 and 866,895 sites throughout the genome, respectively. Briefly, this technology utilizes probes bound to a glass slide which bind to bisulfite converted DNA fragments. A single labelled base pair is then added and excited by a laser which emits a signal detected by a reader, conveyed as intensity measures using software. A given DNAm site on a single DNA molecule can be either methylated (0) or unmethylated (1), in haploid cells only these two states can exist; in diploid cells there can be the situation where only one allele is methylated, thus producing a 0.5 measure. However, when a pool of cells is interrogated, this measure takes on a continuous value (0-1), representing the proportion of methylated DNA molecules.

1.4 Transgenerational inheritance

Transgenerational epigenetic inheritance is the transmission of epigenetic markers from parent to offspring through the germline (113,114). This method of transmission implies an ability for lived experiences to become inherited in future generations, an idea which has been investigated since the early 1900s (114). While there is evidence that this mechanism exists in plants and nematodes, whether it does in humans remains uncertain (114). It has been established that DNAm is mitotically heritable; however, there are issues to consider around whether epigenetic modifications can be transmitted via gametes. DNAm is mostly erased during embryogenesis as there are regions of DNAm essential to reset, though there is evidence that some sites escape this process (e.g., parentally imprinted genes) (115).

In humans this phenomenon has yet to be definitively proven, that may be the result of human longevity in combination with genetic, cultural, and ecological inheritance which can confound any associations (113,114). One of the studies closest to addressing this issue in

humans is the investigation of individuals who had prenatal exposure to the Dutch Hunger Winter. This study showed that *in utero* exposure to famine resulted in *IGF2* DNAm differences which were associated with increased metabolic disease in adulthood relative to unexposed siblings (116). However, as these individuals were exposed *in utero*, it does not prove transgenerational inheritance but is instead an example of fetal programming. Beyond the inability to design human studies to ensure no cultural, genetic, or ecological inheritance is occurring, our generational time is long (113). The human lifespan ranges from 60-80 years and depending on whether the paternal or maternal line is being examined, differing sets of generations are required to establish transgenerational inheritance. For example, females pregnant with offspring have the gametes of the fetuses also exposed to the environment, a term referred to as "fetal programming". Therefore, while only F2 associations are required for paternal exposure, investigations into the F3 generation are required for maternal exposure (113).

Although we do not know whether this occurs in humans, transgenerational inheritance of DNAm states have been demonstrated in some rodent models. For example, sperm DNAm patterns resulting from streptozotocin-induced prediabetes were found in F1 and F2 offspring in mice (117). Also, DNAm differences in the olfactory gene, *Olfr151,* resulting from fear conditioning were found in the paternal F1 and F2 generations, in addition to behavioral sensitivity related to the condition without prior exposure in the offspring (118). This work highlights the potential functions that could be attributable to transgenerational epigenetic inheritance and that the effects of trauma or adversity could be passed to future generations, providing a quicker, more primed response and increased survival (119).

1.5 Adversity

Adversity, especially when experienced in early life, has been associated with many subsequent negative health outcomes (e.g., type 2 diabetes, cardiovascular disease, obesity, psychiatric disorders, and multiple sclerosis) (1–13). These associations are robust, reported across many diverse countries with varying levels of income, and have a dose-dependent relationship, as would be expected if the association were causal (120). There are many different kinds of adversity that might adversely affect health, but most can be assigned to one of two main overarching categories: interpersonal and societal.

1.5.1 Interpersonal

The first social environment a human infant is born into is the uniquely vital and interpersonal relationship with a caregiver. Human altricial young need constant care and affection which they hopefully receive from their primary caregiver, often a parent. Having a positive, loving relationship in infancy with this person, who responds immediately and effectively to the child's needs, has lifelong effects (121). Attachment theory, one of the most well-established social psychological theories, provides a foundation to understand both the impact and quality of early close relationships amid development.

During development, establishing successful relationships with adults and peers provides a foundation for capacities that children will use for a lifetime (121,122). Thus, Bowlby, father of attachment theory, referred to these attachment patterns as being "from cradle to grave" (123), possibly similar to the epigenetic modifications established in infancy and lasting throughout life. This attachment theoretical foundation provides a comprehensive account of the ontogeny and developmental sequelae of infant caregiver bonds, as well as a framework for investigating how perturbations of this system may result in individual differences.

In early care environments, the perception of safety is a critical component and breaching trust results in negative social experiences that can be both stressful and painful. There are many ways to experience pain, such as acutely, chronically, physically, and emotionally; in abusive early social environments, children are exposed to all four kinds of pain. Pain is such a powerful motivator for learning that is often used in animal models for fear conditioning, which is quick to establish and difficult to extinguish (124).

The literature establishes that neural reactions to physical and socioemotional pain are exceptionally similar, specifically in regards to the affective processing of pain (125,126). However, though pain is processed in the same brain regions, socioemotional pain, such as social rejection or isolation, is more potent on a chronic timescale because it is much more easily relived and remembered than physical pain once the original source of pain has subsided (127). Animal studies have also repeatedly found that any unpredicted reward devaluation, such as through a sudden or bewildering social rejection, triggers the brain circuits involved in pain and stress (126). Individuals who experience abuse, especially from a caregiver, may experience the physical pain of abuse, but most certainly experience the social pain of rejection and betrayal in that moment and for years later.

1.5.2 Societal

Chronic stress has clear physiological, psychological, neurological, and epigenetic effects. However, not all chronic stress stems from abuse or trauma. In fact, not all chronic stress stems from interpersonal relationships at all. As a social species, we have developed a society with biases, prejudices, and hierarchies. Our modern social environments remain embedded in historical power relationships. Though much work is being done to correct these injustices, racial, ethnic, gender, sexuality, or religious minorities, as well as those with a low SES and social position, face considerably more stress from societal pressures and inequities than others in their societies. This stress may then exacerbate these inequalities through cellular means, as well as societal.

The construction of our societies leads to social environments that influence the types of nutrients we can access, the configuration of neighborhoods in which we live, the services to which we have access, and the physical environments to which we are exposed. The social administration of our physical environments is yet another form of social environmental influence on what we are able to learn and adapt to psychologically and biologically, whether it is conscious or unconscious.

One major aspect in the construction of socially-administered environments is the spatial sorting of people based on their SES, race, or ethnicity. In addition to this divide, health differences amongst neighborhoods persist even after adjusting for SES and demographic factors, most likely due to the impact of broad environmental factors such as access to nutrition or exposure to pollution (128–131). Factors linked to differences in physical environment most likely contribute to and reinforce the detrimental effects of chronic societal stress on low SES and minority communities (132,133).

One example of how physical environments may perpetuate the biological differences amongst classes are food deserts. These are areas, either urban or rural, where fresh produce and other healthy foods are either not available or are too expensive to be purchased as an everyday source of caloric intake. Low SES neighborhoods are especially likely to be located in a food desert (134). Food availability and food advertising, which is different for lower SES neighborhoods, influence energy intake and the nutritional value of foods consumed (135,136).

Additionally, neighborhood conditions can both create stress (e.g., through feeling unsafe) and be a buffer against stress by providing social cohesion or integration into the neighborhood or environments like work or school (128,137–140). One possible reason some immigrant groups have better morbidity and mortality than other groups in the same city is the social support and cohesion within the community (141). Even in the midst of possible adaptation to a harmful environment, a positive and enriched social environment shows ecological rescue effects for health.

When these adverse events occur in relation to SES disparities in health, the effects are not equal. Early life is a particularly sensitive period during which an individual is more vulnerable to adversity. Doctors who experienced low SES in early life, were assessed for incidence of coronary heart disease and SES in childhood displayed the same disparities as seen amongst those who remain in poverty throughout their life (142). This observation suggests interventions should be implemented in childhood as attaining higher SES in later life will likely not alleviate the negative health outcomes attributed to early life poverty.

1.5.3 Impact on the stress response and immune function

The possibility of experiencing pain, because it is such a noxious experience, is highly motivating for mitigating or preventing injury whenever and wherever possible. This causes a combination of constant uncertainty and the need for hypervigilance – a recipe for chronic stress. The stress pathway is extraordinarily far-reaching and complex, rooted in the effects of uncertainty and requiring no unique or immediately relevant stimulus. The stress response is the collection of immune, neural, and homeostatic mechanisms that shift into a long-term state of hyperawareness with the anticipation of threats that could appear at any time. While this state can be lifesaving when triggered appropriately, being in a constant state of fearful uncertainty is not healthy and has lasting deleterious biological effects.

When there is uncertainty about a potentially aversive or harmful outcome, the downstream stress response is activated by the hypothalamus. The hypothalamus secretes corticotropin-releasing hormone (CRH), which stimulates release of adrenocorticotropic hormone (ACTH) from the pituitary, which in turn signals two different stress pathways, one fast and one slow (143). The two major stress pathways are the sympathetic-adrenomedullary (SAM) response and the hypothalamic-pituitary-adrenocortical (HPA) response (143). In the fast,

immediate SAM response, ACTH triggers the adrenal medulla to release norepinephrine and epinephrine, neurotransmitters required to cause a rapid and intense nervous system response and hypervigilant attention (144). In the slower, long-lasting HPA response, ACTH triggers the adrenal cortex to signal the release of glucocorticoid stress hormones, the most important of which is cortisol. This dysregulates metabolism, suppressing the immune system and disrupting homeostasis through glucocorticoid receptor binding systemically (143). During the stress response, homeostatic mechanisms attempt to maintain equilibrium over a wide range of adaptive circumstances in order to respond to any possible challenge. Stress is, in essence, a "ready" state from which a large, quick biological response is primed at a moment's notice and equipped with constant vigilance. Therefore, this cascade of biological effects both elicits a physiological and behavioral response and poises the requisite systems for future environmental reactivity.

Many of the negative health outcomes associated with adversity have a pro-inflammatory phenotype (5,145). Additionally, adversity associated chronic stress has been linked with increases in white blood cell counts and low-grade inflammation, indicative of immune system dysregulation (5,12,146–155). This is unsurprising considering all lymphocytes have receptors that bind stress hormones released from both the SAM and HPA responses (i.e., epinephrine, norepinephrine, and cortisol), impacting a diverse array of cellular functions (156).

Generally, the immune system has two arms, the innate (cellular Th1) and adaptive (humoral Th2) responses. The more evolutionarily-conserved innate response is the first line of defence, providing a broad response to pathogens, and is driven predominantly by natural killer (NK) cells, macrophages, dendritic cells and granulocytes. These innate immune cells have pattern recognition receptors (e.g., Toll-like receptors (TLRs)) that bind molecules common across many pathogens. The highly specialized adaptive response is activated by the innate arm and so is slower to initiate but is more specific, consisting of T and B lymphocytes. This system results in acquired immunity where antibodies are made specific to the target pathogen and memory cells are produced for faster subsequent responses. Based on a meta-analysis of over 300 empirical articles investigating this relationship, different measures of stress had differential associations to the two main arms of the immune system. Acute stress resulted in upregulation of innate but downregulation of the adaptive response, whereas brief stress exposures had the

opposite effect, chronic stress resulted in the suppression of both responses, and subjective measures of stress had no impact at all (157).

Innate immune function is especially important for chronic stress-induced immune dysregulation (12,120,151,158–161). Several proinflammatory biomarkers have been associated with stress, such as C-Reactive Protein (CRP) and the cytokines Tumor Necrosis Factor alpha (TNFα), IL-6 and IL-1β (162–166). There have also been several studies where exaggerated cytokine responses were seen from *ex vivo* stimulation of TLRs (151,167,168). Increases in inflammatory markers are one of several factors thought to contribute to allostatic load, a measure for general disease risk, which positively correlates with early life adversity (169–171). However, the molecular mechanisms underpinning the relationship between adversity and this proinflammatory phenotype are unknown. Due to the stability and environmental responsiveness of epigenetic mechanisms, investigating this as a potential mediator explaining how adversity could induce long-term changes to the genetic landscape, and subsequently the immune and stress response, is a promising research endeavor (12,120,172).

1.6 Adversity associations with DNAm

1.6.1 Interpersonal

Our past relationships and interactions with the social environment are stored in memory (123,173–177). From an epigenetic perspective, these mental representations may reflect the adaptations to early caregiver relationships and social environments that ultimately result in differing phenotypes. There is a wealth of literature exhibiting reported epigenetic differences correlated with early life environments, termed social epigenetics.

The clear, widespread effect of stress throughout all physiological systems makes it an accessible and unique candidate for understanding the epigenetics of social environment. Many adverse environments, both interpersonal, such as traumatic and abusive relationships with caregivers and peers, and societal, such as minority and socioeconomic status stress, can trigger these same underlying processes, as all are sources of aversive, potentially harmful uncertainty. It is then understandable why the most commonly researched gene in social epigenetics is *NR3C1*, the glucocorticoid receptor gene (178–180). Consistent DNAm changes have been found in exon 1F/17 of *NR3C1* regarding parental stress, but were inconsistent in other types of stress (178).

Stress effects relate to a poor versus positive early caregiver relationship. An often-cited study focused on the relationship between maternal licking and grooming in rats with DNAm differences in the promoter region of *NR3C1* (179). While the authors found that more licking and grooming by the mothers led to hypomethylation in the *NR3C1* promoter and DNAm patterns in a broader surrounding area (180), this study has yet to be replicated and is met with some skepticism in the field (36). However, more recent work in mice has also found different biological reactions to maternal caregiving in the hippocampus, a region associated with both learning and reactions to stress, and the dorsal raphe nucleus (*DRN*), the brain center for serotonin production and distribution (181,182). In *DRN* specifically, Araki and colleagues found hypomethylation affecting *GABBR* which codes for the GABA(B) receptor, a common pharmacological target for depression and anxiety relief (181,183).

Human studies have primarily focused on the DNAm of a few candidate genes of specific and special prominence in the research, namely *BDNF* (184–186), *NR3C1* (36,178,187,188), *SLC6A4* (189), and *OXTR* (186,190,191). All of the proteins these genes encode have many functions across neurodevelopment and stress. For example, *BDNF* codes for brain-derived neurotrophic factor, the canonical neuronal growth factor in the brain widely involved in the formation of any neuroarchitectural changes (185). Additionally, *NR3C1* is the most widely researched gene in regard to fMRI and stress-causing environments, such as poor maternal care, in both the animal and human literature (36,178). One recent study found that increased maternal responsiveness and touch were correlated with hypomethylation in *NR3C1* exon 1F in female children (187). This sex-dependent response has been replicated in other studies, finding the greatest DNAm differences amongst attachment styles in females (188). Additionally, across the sexes, attachment behavior patterns have been correlated to approximately 10% of global DNAm variability in a sample of at-risk children, suggesting biological responses to the sensitivity and consistency of the parental care environment (188).

Another nuanced aspect of the caregiver relationship is soft touch, which is incredibly important in healthy, normative infant development (192). There were significant DNAm differences between children who received high amounts of soft touch and those who did not as infants. Additionally, infants who were more distressed, yet received lower amounts of touch, were epigenetically younger, possibly indicating a biological developmental delay (193). This is most likely due to the social buffering effects of both mental representations and human touch.

For example, holding the hand of a stranger can reduce both the subjective experience of pain and its neural signature, but holding the hand of a close relationship partner reduces pain in these areas to an even greater extent (194).

Not all early relationships provide positive experiences. For example, both maternal and paternal stress during early life was correlated with adolescent differences in DNAm in humans (195). In rats, newborns exposed to a stress-abusive mother showed increased methylation in the promoter region and decreased expression of *BDNF* (196). This difference in *BDNF* levels for abused versus non-abused rats appears to persist through adulthood (197). In this same study, a different group of newborn rats was also exposed to positive caregiving mothers. Both the maltreatment and beneficial caregiving mothers initially caused an increase in *BDNF* mRNA levels in the hippocampus (197). Both experiences, regardless of valance, equitably guided the growth of new neuronal connections in the brain's memory center. Adaptationally, negative relationships and social environments are just as powerful as positive social learning experiences.

In addition to *NR3C1*, there is robust literature associating early-life adversity such as social, physical, or parental stressors, with epigenetic changes related to *BDNF* expression (198,199). These differences may correlate with a reduction in socioemotional learning and plasticity, and have shown an increased capacity for fear learning (200). Impacted learning has also been implicated in more specific associations than general stress, such as the several reported associations between trauma, abuse, and differences in epigenetic modifications (30,201–206), as well as associations specifically with posttraumatic stress disorder (207).

Though there is a range of epigenetic findings in regards to early life adversity from both candidate gene approaches and EWAS approaches, overlapping genes associated with stress, pain, learning, and the immune system were common. For example, one study found immune cell differences and accelerated epigenetic age associated with lifetime PTSD severity (208). Another found the fear of parental deportation after the 2016 US presidential election associate with decreased epigenetic age amongst children of Latinx immigrants (209). Several EWAS and candidate gene analyses also found associations with various forms of childhood abuse and trauma, patterns which were detected in later life (205,210–212) and at times associated with psychiatric outcomes (213–216). These variables were also correlated with increased epigenetic age acceleration, itself associated with several negative health outcomes (217,218). While

increased epigenetic age acceleration is associated with negative outcomes in adults, what this means in children is still relatively unknown and so must be considered when interpreting these findings.

It is possible that trauma and pain in early life lead to learning and adaptation to a harsher world that requires more vigilance instead of a conservative homeostasis, a molecular push toward fear conditioning instead of socioemotional development, and a greater sensitivity to pain in order to more quickly identify threats. Those who develop in positive, enriched environments, on the other hand, may thrive with reduced allosteric load and have resilience that seems especially prominent in stress coping and synaptic plasticity. Instead of being able to learn more and put their energy toward other endeavors, these adaptations in a world of agonizing uncertainty could be primed or activated by epigenetic modifications for the sole purpose of survival. A plausible model of chronic life stressors proposes a similar line of reasoning and theorizes with significant evidence that epigenetic modifications set into motion by the cascade of stress hormones both affect and prime a traumatized individual for accelerated aging and biological weathering (219). Supporting this hypothesis, a large number of sites used in DNAm epigenetic clocks are located within glucocorticoid response elements (220).

However, though it may not undo the harmful developmental environment, social support, especially touch, has been associated with stress buffering in many studies (141,194,221,222). This is most likely through the dual mechanisms of affect regulation through a social buffer so as to not trigger the stress hormone cascade, and through the mu-opiates that are released in social reward reducing affective motivational pain in the nociceptive pathways and the intensity of perceived threats (223). This may contribute to the difficulty in reproducing many social epigenetic findings, as adverse effects are often accounted for, but buffering and resiliency effects are not.

Ultimately, the research indicates that having a healthy, supportive early social environment leads to positive epigenetic, neurological, and psychological outcomes. This is most clear when examining the literature on the benefits of enriched environments both in buffering stress and in rescuing memory formation. In women, for example, methylation within the *NR3C1* promoter had a dose-dependent relationship with childhood abuse where the effect was attenuated with early life emotional support (224). Additionally, having a supportive family environment during development protected against harmful cellular and epigenetic aging due to

experiences of racial prejudice. However, individuals without a supportive family environment did experience biological weathering (225).

1.6.2 Societal

Both physical and social environments can affect our epigenetic modifications (226). There is evidence that chronic stress which accompanies being in a reduced societal position, whether through racism or classism, as well as the stress of deprivation, associates with epigenetic change. Most likely due to many social, physical, and biological factors encompassed in SES, there are a plethora of associations between SES and epigenetic changes. As SES is societally constructed, it is difficult to create a valid animal model with which to investigate epigenetic modifications; therefore, the majority of epigenetic research related to SES is focused on DNAm associations in humans.

Low SES, especially during youth, has a significant and robust association with age acceleration and DNAm sites connected to immune function, development, and age-related diseases (159,227–234). Even amongst low-SES youth, greater self-control associates with improved socioemotional functioning and general success, supporting the idea that increased allosteric load may contribute to worse health outcomes amongst the disadvantaged (235).

The literature does appear to indicate that SES associated chronic stress in early life has a greater impact on DNAm patterns than those that occur in later life, but more research in this area is needed to make a definitive statement (227,231,236). Work on cumulative stress, as opposed to early life or later life considered separately, also indicates an association with accelerated epigenetic aging (220). If this stronger association with stress and DNAm at younger ages is robust, it may be due to differences in the immune system's environmental sensitivity during early development (237). While it is true that a recent, large study in humans did not replicate findings from smaller studies that reported correlations between DNAm patterns and chronic social stress (238), this may be due to differences in sample size, population, and consistency amongst ecologically valid measures of the type, context, and experience of stress. More replications with standardized measures and large sample sizes are needed to make any definitive statements about detectable DNAm differences amongst those who have experienced chronic stress.

Another aspect of the relationship between SES and epigenetic modifications is that lower SES correlates with both smoking and drinking behavior (239,240). Not surprisingly, social stress also triggers the urge to smoke and drink (241,242). This relationship of SES with smoking and drinking, possibly as a way to deal with stress, may exacerbate epigenetic disparities due to the strong, reproducible effects of smoking and drinking on DNAm, especially on sites related to age, immune, and cardiovascular function (243–249).

In addition to smoking and drinking, it is important in every DNAm investigation to account for cell type proportions, as these are the primary drivers of variation, but this is especially true in explorations of SES due to the significant immune system effects of the chronic stress system. For example, one study found that leukocyte composition of peripheral blood covaried with patterns of DNAm at many sites and DNAm was strongly associated with the monocyte inflammatory response (231). Monocytes also epigenetically aged faster in those exposed to low SES in early life (227). These likely stress-related immune responses may contribute, at least in part, to the association of SES, especially in early life, with epigenetic age acceleration and aging-related disease risk, even controlling for related factors such as smoking and drinking (227,233).

SES can also greatly impact the nutrition available to an individual. The wealth of literature on the epigenetics of nutrition, especially prenatal nutrition, pales only in comparison to epigenetic work in cancer (250). The importance of a balanced and healthy diet, from conception and throughout life, on epigenetic modifications is an incredibly robust finding, as are similar results for morbidity and mortality (250–256). Along similar lines, the structure of a socially administrated physical environment can also be linked to differences in children's physical activity (257–260). Physical activity is often linked to morbidity and mortality, as well as epigenetic modifications, learning, and aging (261–265). Another example is physical proximity to hazardous sites and pollution, which tend to be more prevalent in low-income or minority neighborhoods (266–268). The effects of exposure to air pollution are well evidenced in both morbidity and epigenetics research (29,269–279). This is most likely due to immune responses to breathing in toxic exogenous factors (279).

When Simons et al. (2016) investigated the main environmental driver of epigenetic age acceleration in a low SES sample, they found that it was the stress of financial insecurity that drove the SES and accelerated aging association, providing further evidence for the link between

early life stress, immune response, epigenetic change, and health outcomes (12). However, similar to other early exposures to stress, low SES effects on immune response can be buffered though social support in the form of warm, positive caregiving (229). Our epigenetic mechanisms are modified by, and our psychological mechanisms learn from, our social environments.

1.7 Dissertation overview

This dissertation aims to obtain an overall understanding of how the social environment, in the context of interpersonal and societal adversity, associates with interindividual epigenetic differences in a healthy population. As outlined in Section 1.1, the empirical data for the studies comprising this thesis will be presented as three separate chapters. I performed three studies to assess adversity-associated DNAm by a) investigating multiple measures of SES, b) determining how inflammatory markers play a role and c) examining these associations in gametes. In **Chapter 2**, I investigated how DNAm patterns varied with educational, household asset and self-reported measures of economic standing in both childhood and adulthood in the whole blood of elderly individuals. Next, in **Chapter 3**, I determined if an inflammatory marker, serum IL-6, was related to SES-associated DNAm sites in purified monocytes. Finally, in **Chapter 4**, I investigated how childhood abuse associated with the DNAm profiles of adult spermatozoa samples. Overall, these findings provide evidence that adversity experienced in various contexts is associated with DNAm throughout life in various tissues, in addition to highlighting avenues of future research.

2 Multiple components of socioeconomic status measured in older adulthood and childhood differentially associated with DNA methylation in older adults

2.1 Background and rationale

Low socioeconomic status (SES) has been consistently associated with increased risk of mortality (280) and multiple diseases such as diabetes (281), cardiovascular disease (282), and depression (6). Commonalities amongst the health risks associated with SES, such as stress and immune dysregulation, are being leveraged to focus research. Pro-inflammatory cytokines, predominantly C-reactive protein (CRP) (168,283–287) and interleukin-6 (IL-6), (167,285,288,289) predict several of these negative health outcomes (290–292). These inflammatory responses are also upregulated in individuals with adverse health behaviours such as smoking (293) and obesity, (294) which are commonly seen within low SES populations. In a meta-analysis of studies observing the association between either CRP or IL-6 and SES found an overall effect of cytokine increases with lower educational and income levels (295).

Dysregulation of stress pathways involving the sympathetic nervous system (SNS) and hypothalamic-pituitary-adrenal (HPA) axis have also been implicated in relation to SES and health disparities (296,297). The negative effects of lower SES can result in increasing an individual's allostatic load, the wear and tear of body systems exposed to chronic stress, culminating in multi-system dysregulation (298–300). Increased levels of SNS hormones epinephrine and norepinephrine (301–307) and the HPA-axis hormone cortisol (297,308–311) are also negatively correlated with SES (312).

Though SES negatively correlates to several inflammatory and stress biomarkers, the underlying mechanisms are not fully understood. The role of epigenetics, the mitotic inheritance of molecular markers that can influence genetic expression without altering the genetic sequence, have been investigated as a potential moderator of these relationships (313). The epigenetic mark DNA methylation (DNAm), a methyl group covalently bound to a cytosine base pair typically as a cytosine-guanidine dinucleotide (CpG), has been interrogated as a potential link due to evidence of this mark being both environmentally responsive and stable, making it a suitable biomarker to help unravel the impact of SES throughout life (15,94,313,314).

DNAm associations to several SES measures across the lifespan have been reported, with differences commonly found in inflammatory and stress-related genes. Prenatal SES was associated with DNAm differences measured at birth and persisting throughout childhood (315). Childhood DNAm in genes related to immune and developmental pathways has been associated with several measures of SES (228). Amongst adolescents, low early life SES was associated with poorer health and DNAm differences in signaling regulation genes, with these effects being mediated by protective parenting (316). A candidate approach investigating CpGs in inflammatory genes found associations with SES amongst a cohort of young adults, where a subset of sites could also predict plasma cytokine IL-6 levels (226). An epigenome-wide association study (EWAS) conducted in this same cohort revealed additional SES-associated CpGs enriched in genes involved in immune function and nervous system development (159). Amongst adults, DNAm was associated to childhood SES independent of adulthood SES, highlighting the importance of taking into account the timing of the exposure (231). Finally, amongst an elderly cohort, a candidate study showed DNAm in stress and inflammatory genes to be associated with SES (317). Though DNAm associations with SES at various life stages have been found, there have been replication issues in the field which are likely a result, in large part, due to how SES is defined, when it is measured, and how the variables are constructed (296,318–322).

Importantly, when in life these socioeconomic disparities are experienced can result in varying trajectories of health outcomes. For example, minorities who grew up in poverty but attended prestigious universities in adulthood had better mental health outcomes but worse metabolic health than their childhood peers who remained in poverty. These seemingly contradictory findings imply a superficial, skin-deep, resiliency exemplifying the complexity amongst the pathways linking low SES and health (323). This may be partially explained by the Developmental Origins of Health and Disease (DOHaD) paradigm which posits the existence of sensitive windows during early life development where adverse exposures result in a greater impact to health outcomes than if exposures occurred at other life stages (18,324). This has been reflected in several social epigenetics studies investigating SES occurring at different life stages reporting differential DNAm associations (231,317,325). However, timing is not the only factor to consider when attempting to understand SES and health associations: how SES is measured, and the societal context in which it is investigated, are also integral.

SES can be generally defined as a measure of an individual's economic standing in society, which makes it multifactorial, allowing it to be measured through multiple aspects (326). Each SES subcategory encompasses features that can have differing impacts on individuals (326). SES research investigates several categories of SES components, including subjective (e.g., self-report), objective (e.g., education) and composite measures (326,327). Objective measures of SES allow for more comparable findings across studies but fail to capture other aspects of an individuals' subjective lived experience like self-report measures do (326,328). Both subjective and objective measures of SES, while assessing differing characteristics, have been shown to impact health (326,328). While composite SES measures can help make different populations more comparable and represent several SES components in a single variable, this approach obscures nuances captured with more specific measures (326). Variability in measurements may be one reason for the SES EWAS replication problem, as there is no singular, unifying measure used in this epigenetic field (318). However, even with a unifying measure, not all of these components are comparable to one another in different populations, as all are dependent on societal context (326).

Most studies investigating SES and health have been limited to predominantly Caucasian populations in developed countries. It is known that SES gradients of health, while in general negatively associated with morbidity and mortality, are not in every society. For example, Latin American populations seem to display reversed SES gradients for a subset of morbidity and mortality measures, such as life expectancy and self-rated health (329,330). This phenomenon is known as the Hispanic paradox, where Hispanic individuals appear to have smaller social disparities in health in comparison to non-Hispanic Whites. There is also evidence this buffering against the potential health effects of social adversity is even further reduced amongst the elderly (331,332), highlighting the complexity of SES and health associations.

Given these considerations, intertwined with the multifactorial variable of SES and limitations discussed above, I aimed to investigate how DNAm associated with several components of SES (i.e., self-report, education, wealth, composite) in retrospective child and contemporary older adulthood measures within an underrepresented population amongst SES research to date: Costa Rican individuals aged 60-106 years (n = 482). Additionally, I interrogated these associations by introducing several inflammatory (C-reactive protein (CRP)) and stress (epinephrine, norepinephrine, cortisol, dehydroepiandrosterone sulfate (DHEAS))

biomarkers and lifestyle factors (body mass index (BMI), drinking, smoking) known to relate to SES in order to provide insight on the larger mechanisms at play. By comparing DNAm associations and contributions across temporal and measurement aspects of SES, I aimed to better understand the similarities and differences of these factors as social determinants of health.

2.2 Materials and Methods

2.2.1 Costa Rican Longevity and Healthy Aging Study (CRELES) cohort

Participants in this study were sub-sampled from the CRELES cohort which consists of 2,827 participants aged 60 or older from across Costa Rica in an effort to investigate factors driving aging and longevity (333). This study sample consists of 482 individuals with on over-representation of those ≥ 95 years of age. Longitudinal data used for this study was taken at either wave 1 or 2 to correspond with when DNAm was measured. Demographics of sex and age were obtained as control variables. Summary statistics of all variables investigated are in Table 2.1.

2.2.2 Socioeconomic status (SES) measures

Seven SES measures were selected to incorporate several sub-categories and represented two life stages of exposure. Summary statistics of all variables investigated are in Table 2.1.

1. *SES PC:* The first principal component (PC) of a composite household SES variable produced by incorporating a participant's years of education, household assets, joint income and housing value, as previously reported (334). This variable is significantly associated with all other SES measures (Supplementary Figure 2.1).
2. *Older adulthood assets:* A scale ranging from 0 – 100 based on a measure of 12 household assets as previously reported (330).
3. *Education years:* Educational attainment based on the number of years approved ranging from 0 – 17.
4. *Older adulthood SES self-report:* A 5-point scale of self-rated economic wellbeing, described in detail elsewhere (335).
5. *Childhood assets:* A yes/no response to the questions "During the first 15 years of your life, did you wear shoes regularly?", as previously investigated (336,337).

6. *Maternal education:* A measure of if the participants' mother had any education or not.
7. *Childhood SES self-report:* A yes/no response to the question "During the first 15 years of your life, did your family have problems or economic hardships that did not allow you to eat regularly, dress adequately or get necessary medical care?".

2.2.3 Biomarker and lifestyle measures

Biomarkers were either obtained from overnight urine (cortisol, epinephrine, norepinephrine, DHEAS) or from early morning (7-9 am) fasting blood collected by venipuncture (CRP, genomics). Lifestyle measures (BMI, drinking and smoking status) and biomarker collection has been reported in detail elsewhere (333,336–338). Genomic DNA, used for DNAm and genotyping arrays was extracted at the University of Costa Rica using previously reported methods (337,339). Summary statistics of all variables investigated are in Table 2.1.

2.2.4 DNA methylation array

Genomic DNA was isolated from the whole blood cell fraction by performing cell lysis with Proteinase K and using phenol-chloroform for extraction. DNA quality was assessed using the absorption ratio A260/A280. Isolated genomic DNA was bisulfite converted using the EZ-96 DNA Methylation kits (Zymo Research, Irvine, CA) and then run on the Infinium MethylationEPIC (EPIC) BeadChips (Illumina) to interrogate methylation at 865,918 CpG sites using the standard protocol.

2.2.5 Genotyping array

Genotyping data were determined at 618,540 single nucleotide polymorphisms (SNP) using the Infinium Global Screening Array (GSA) BeadChips according to the Illumina's standard protocol (Illumina).

2.2.6 Statistical analysis

All statistical analyses were performed in RStudio (version 1.1.456) using the R computing language (version 3.6.1) unless otherwise stated.

2.2.6.1 Genotyping pre-processing and normalization

GenomeStudio 2.0 Genotyping software was used to transform the raw intensity files into clusters, and subsequently genotype calls, by producing cohort-specific clustering files and annotating the SNPs using the Illumina produced manifest GSA-24v1-0_C1 (Version 1 A2, Illumina). Summary of all genotyping quality control steps are summarized in Supplementary Table 2.1. As suggested by Illumina, prior to clustering non-autosomal probes and outlier samples were removed so as to not influence genotype calls. Outlier samples were removed if the call rate ≤ 0.97 or if the 10th percentile of the GenomeStudio generated clustering score (p10 GenCall Score) across all genotypes was ≤ 0.5. Samples which did not match across EPIC and GSA array platforms using six overlapping SNP probes and performing Pearson's correlation on comparable DNAm beta and genotyping B Allele frequency values for each sample ($r^2 \leq 0.8$) were removed.

Several criteria suggested by Illumina were used to remove poorly performing probes. Probes that did not cluster distinctly (Cluster Sep ≤ 0.45), had low heterozygote cluster intensities (AB R Mean ≤ 0.4), were too close to a homozygote cluster (AB T Mean ≥ 0.8 or ≤ 0.2), had homozygote calls clustered either too far from the axis (AA Freq = 1 & AA T Mean ≥ 0.2) or were too spread out (AA Freq = 1 & AA T Dev ≥ 0.04) and incorrectly classified as heterozygotes (AB Freq = 0 & minor allele frequency (MAF) > 0) were removed from the data set.

Further SNPs were removed based on standard methods commonly used in the field (340). SNPs in heterozygote excess relative to Hardy-Weinberg Equilibrium (HWE) expectations (Het Excess ≤ -0.3 or ≥ 0.2) or had a minor allele frequency (MAF) ≤ 1% were removed. PLINK (version 1.07) was used to remove redundant SNPs that are highly correlated using linkage disequilibrium (LD) calculations with the -indep function (size = 50kB, window = 5 SNPs, variance inflation factor (VIF) = 2) (341). SNPs pre-determined to be in high-range LD, within an 8Mb region spanning the extended MHC region on chromosome (Chr.) 6 or within a 4Mb inversion on Chr. 8 were removed (340,342), resulting in 420,483 SNPs remaining for downstream analysis.

2.2.6.2 DNAm pre-processing and normalization

Raw IDAT files were imported to RStudio where beta values (ranging 0 – 1, representing the proportion of methylated DNA molecules in the sample) were produced from the intensity measures using the minfi package (343). Background subtraction, color correction and between-array normalization were performed using preprocessFunnorm (minfi) (343,344).

Outlier samples were removed if detected by any of the following three methods: (1) detectOutlier (minfi) - calculates if a sample mean is too far from the cohort mean (345). (2) outlyx (wateRmelon) - utilizes principal component analysis (PCA) to observe the distance and scatter of a sample from the cohort average (100). (3) locfdr (locfdr) - applies PCA to assign a z-score for each sample relative to the cohort (false discovery rate (FDR) ≤ 0.2) (80,346,347). Sex chromosome probes were used to determine sample sex mismatches by comparing the reported sex with DNAm predicted sex, performed using k-means clustering. Technical replicate samples and samples of related individuals determined by genetic pre-processing previously described were removed prior to analysis.

Poor performing probes defined as having a low detection *p*-value ($\leq 1x10^{-16}$) or fewer than three reads or an NA in more than 1% of samples were removed. Probes measuring SNPs or sex chromosomes (348) were removed. Poorly designed probes, defined as measuring a site pre-determined to contain a SNP (349) with a MAF ≥ 1% or predicted to non-specifically bind (99,348) to other genomic regions were removed.

Technical variation resulting from batch effects during the array run (plate, chip, chip position) were removed exploiting empirical Bayes methods using the ComBat function (SVA) (350,351). Invariable probes defined as having ≤ 5% beta value range between the 5^{th} and 95^{th} percentile across individuals were removed, resulting in 396,798 CpGs evaluated amongst all individuals (Supplementary Figure 2.2).

2.2.6.3 Determination of predicted genetic ancestry principal components

Relatedness was investigated using identity-by-descent (IBD) to ensure the independence assumptions for population stratification analysis was met. IBD was calculated within the cohort using the -genome function (Plink version 1.07) and one of the related pair of samples with PI_HAT ≥ 0.1875 were removed (352). To reduce noise, SNPs with a MAF ≤ 5% were removed prior to performing PCA using pca (PCAtools). Horn's analysis was used to determine how

many PCs to retain (n = 2) using the paran function (MASS) (353). Participants were ascribed the rotated PC1 and PC2 loadings to represent and control for genetic ancestry differences in subsequent analyses.

2.2.6.4 Estimation of whole blood cell-type proportions

Inter-individual differences in cell-type proportions found in the whole blood of participants were predicted using the DNAm data. Using reference methylation profiles from FACS sorted whole blood data (354), the Houseman method was applied (64) while incorporating the IDOL probes and adjustments developed by Koestler et al. (2016). These predicted proportions consisting of CD8 T cells, CD4 T cells, natural killer cells, B cells, monocytes and neutrophils, were adjusted for in downstream analyses.

2.2.6.5 Principal components analysis of DNAm

To assess any relationship between the DNAm variation and variables of interest, PCA was conducted on all measured DNAm sites (n = 396,798) using the prcomp function (stats). PCA loadings attributable to samples were associated with variables of interest using either Pearson's correlation or an ANOVA test where appropriate. Resultant *p*-values were multiple-test corrected using the Benjamini-Hochberg procedure.

2.2.6.6 SES associations with DNAm

EWASs were performed by using robust linear regression on each CpG site for the seven SES measures using the rlm function (MASS) (353). M-estimation was performed with Huber weighting. The model was adjusted for age, sex, genetic ancestry PCs, and predicted cell type proportions. Biological effect sizes (Δβ) were calculated by taking the difference between the means of the highest and lowest category for a categorical variable or by taking the beta coefficient of a linear model (DNAm ~ SES) and multiplying that by the range of the variable between the 5^{th} and 95^{th} percentile. Top sites were defined as the 1,000 CpGs with the lowest *p*-values for a given model.

2.2.6.7 Permutation analysis

To ensure results of top site analyses were not due to chance, for each SES variable, 500 randomized permutations were run, where SES values were randomly shuffled across samples, the model re-run and the top sites pulled. Permutation *p*-values were calculated by dividing the number of instances a result of equal or greater measures was found amongst the permutations by the total number of permutations.

2.2.6.8 Global distributions and *p*-value comparisons resulting from SES EWASs

The distribution of the resultant *p*-values obtained from the various SES EWASs were plotted against the random permutation *p*-value distributions. The true distributions were deemed significant if the number of CpGs in a given EWAS with a *p*-value $\leq 1x10^{-6}$ was greater than one would expect by chance using permutation results (*p*-value ≤ 0.05).

2.2.6.9 Comparison of SES associations with DNAm

The 1,000 DNAm sites with the lowest *p*-values for a given SES variable association were compared and overlapped with the others. Where overlapping did occur, permutation analysis was used to ensure the number of overlapping sites detected was not by chance (*p*-value ≤ 0.05).

2.2.6.10 Genomic feature enrichment of SES associated DNAm sites

Three sets of genomic features (genomic, CpG Island region, chromatin state) were selected to investigate whether there was enrichment of top CpGs in a given feature. The background was produced by subsetting the Illumina annotation file down to the 396,798 CpGs analysed, fold enrichment was calculated for each feature which was then compared to the fold enrichment found by chance using permutation analysis (*p*-value ≤ 0.05). Multiple testing was performed on the *p*-values for the number of features tested in a given set using the Benjamini-Hochberg procedure and reporting enrichments with FDR ≤ 0.2.

1 *Genomic region:* To determine enrichment for particular genomic regions annotation obtained from the Illumina manifest was used. The defined genomic regions were obtained from the UCSC database (355,356) using human genome build GRCh37/hg19, variable UCSC_RefGene_Group.

2 *CpG Island region:* This same procedure was performed for CpG site locations within pre-defined CpG Island regions (357) using the UCSC (355,356) variable Relation_to_UCSC_CpG_Island.

3 *Chromatin state:* Chromatin state was assessed by using the NIH RoadMap Epigenomics Consortium primary mononuclear data (358) and the 18-state ChromeHMM algorithm (359) for determining various active and repressive chromatin states based on the combinations of six histone modifications (H3K27ac, H3K4me3, H3K4me1, H3K36me3, H3K27me3 and H3K9me3).

2.2.6.11 Association of SES measures with biomarkers and lifestyle measures

To determine if various inflammatory (CRP) and stress (cortisol, epinephrine, norepinephrine, DHEAS) biomarkers or lifestyle behaviours (smoking, drinking, BMI) were associated with any of the SES measures, Pearson's correlations were performed where significance was determined if the *p*-value ≤ 0.05 and $|r^2 \geq 0.1|$.

2.2.6.12 Biomarker and lifestyle behaviour contribution to SES and DNAm associations

To determine if the SES-associated DNAm signal was impacted by the various biomarkers and lifestyle behaviours, a contribution sensitivity analysis was performed. First the beta coefficient of SES from the base model was calculated:

$$CpG \sim \boldsymbol{\beta_{base}}\, SES + Genetic\ Ancestry + Predicted\ Cell\ Proportions + Age + Sex$$

The beta coefficient of SES from the same model was then adjusted for a given biomarker or lifestyle measure:

$$CpG \sim \boldsymbol{\beta_{adj}}\, SES + Biomarker + Genetic\ Ancestry + Predicted\ Cell\ Proportions + Age + Sex$$

The following contribution calculation was conducted to find the percentage of change to the effect made when incorporating a given biomarker or lifestyle measure into the base model:

$$\%\ Contribution = 100\ [(\beta\_base\ -\ \beta\_adj)/\beta\_base\ ]$$

The CpG and SES associations where ≥ 10% effect size was contributed by a given biomarker or lifestyle variable were counted and a permutation analysis performed to determine if the number of sites with a contribution of this size was more than expected by chance (*p*-value ≤ 0.05).

2.3 Results

2.3.1 All SES measures were associated with global DNAm.

PCA was performed on the variable beta values of the DNAm data where the resultant PCs were then associated with variables of interest. All SES variables are associated with the first 15 PCs (Table 2.1, Figure 2.1), suggesting a global DNAm signal attributable in part to SES. There is a distinct trend of the self-report SES measure in both older adult- and childhood not being as strongly associated with as many PCs as other SES measures.

	N	Percent	Mean (SD)	Median [Min, Max]
Age (years)	482		79.3 (*10.8*)	79 [61, 107]
% Female	482	56.8		
% Nicoyan	482	17.8		
% Genotyped	465	96.4		
SES PC	481		-0.81 (*1.20*)	-0.95 [-3.57, 6.27]
Older Adulthood Assets	482		66.76 (*18.32*)	67 [0, 100]
Education Years	482		3.79 (*3.74*)	3 [0, 17]
Older Adulthood SES Self-Report	482		2.41 (*0.90*)	2 [1, 5]
Childhood Assets: % Low SES	461	26.0		
Maternal Education: % Low SES	461	39.0		
Childhood SES Self-Report: % Low SES	461	58.1		
Cortisol (ug/g)	217		27.26 (*25.52*)	22.32 [3.20, 171.73]
C-reactive Protein (mg/l)	470		6.76 (*8.55*)	3.48 [0.13, 49.38]
Epinephrine (ug/g)	461		8.16 (*8.79*)	4.99 [0.73, 118.80]
Norepinephrine (ug/g)	461		65.08 (*34.59*)	69.67 [4.94, 438.68]
DHEAS (ug/dl)	474		44.38 (*40.53*)	33.45 [10.00, 293.00]
Smoking: % Smoke Ever	480	43.1		
Drinking	461		3.89 (*1.30*)	4.00 [1.00, 5.00]
Body Mass Index	474		25.36 (*5.91*)	24.83 [11.10, 69.45]

Table 2.1 CRELES demographics, biomarker and behaviour summary statistics.
DHEAS – Dehydroepiandrosterone sulfate.

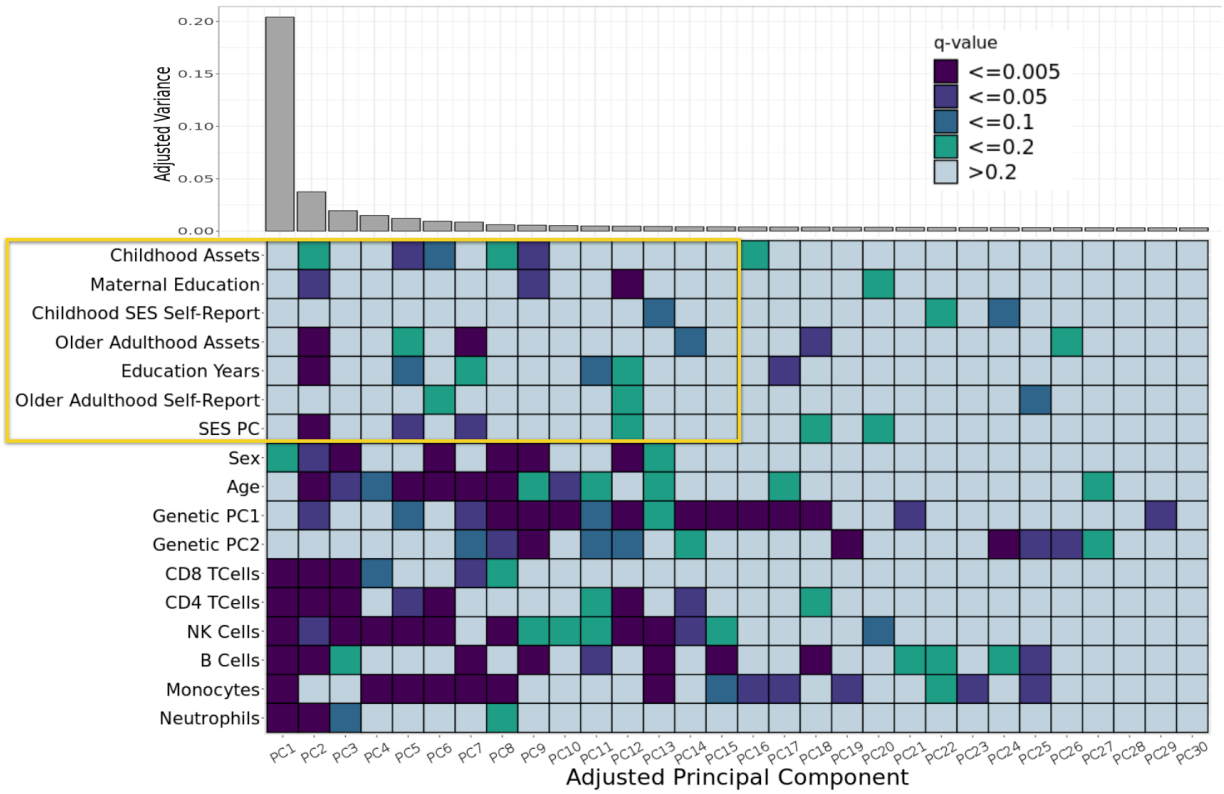


Figure 2.1 All SES measures were associated with at least one of the first 15 PCs resulting from PCA of the DNAm data.
PCA was performed on the variable DNAm beta values (n = 396,798). Loadings assigned to individuals for each PC was then associated with variables of interest, where Pearson's correlation was used for continuous variables and ANOVAs for categorical. All *p*-values were corrected for multiple testing using the Benjamini-Hochberg procedure. The yellow box highlights the SES measures. Variance of the DNAm data accounted for in each PC depicted in the bar graph above the heat map.

2.3.2 Current life SES measures showed stronger associations with DNAm than retrospective measures taken in childhood.

Robust linear regression was performed for each of the variable CpG sites with each of the SES measures where age, sex, genetic ancestry and predicted cell type proportions were all adjusted for. The *p*-value distributions for each of the seven EWASs showed the general trend where childhood SES measures have a flatter distribution in comparison to the older adulthood measures and SES PC, which have a right skew suggestive of potential signal (Figure 2.2). To determine if any of these *p*-value distributions were skewed more than one would expect by chance, randomized permutation analysis was performed where the given SES variable was shuffled and the robust linear regression re-run 500 times. The number of sites which resulted in a *p*-value $\leq 1x10^{-6}$ was counted and compared to the number of sites at the same threshold

expected by chance. The *p*-value distributions from the EWASs of two measures, education years and older adulthood assets, were skewed more than one would expect by chance (Supplementary Figure 2.3). Delta beta values were calculated by finding the change in beta measures between individuals with highest and lowest SES measures for each CpG site. The distributions of these delta beta values also displayed a similar trend where large differences in the beta values were observed predominantly in the older adulthood SES measures (Figure 2.2).

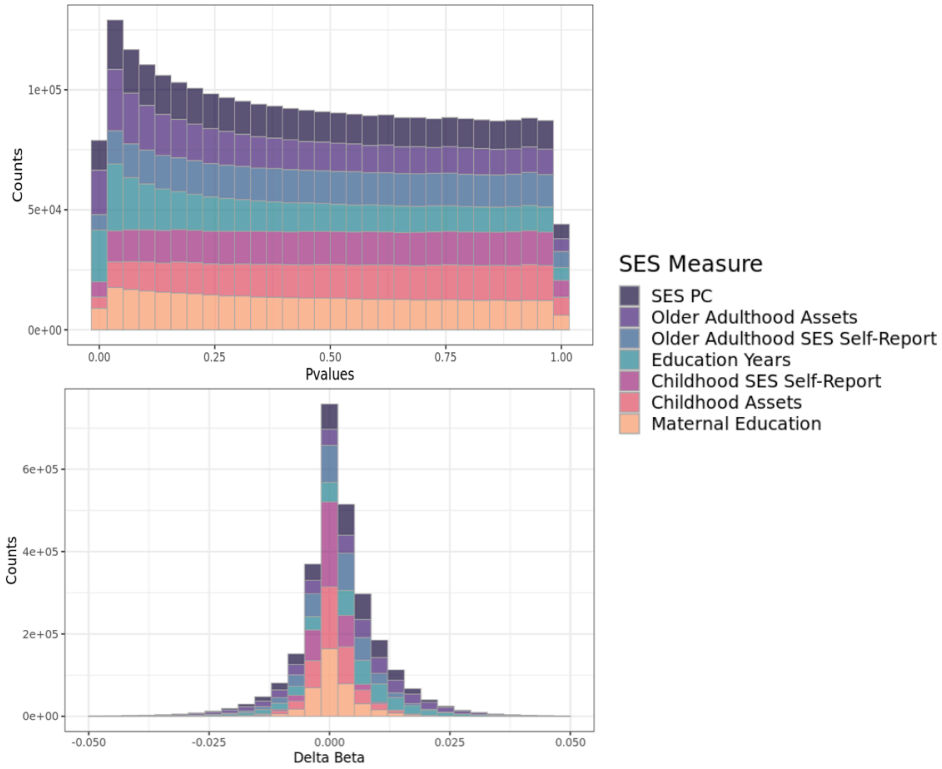


Figure 2.2 Older adulthood SES measures had significantly stronger associations with DNAm than in childhood.
The top panel is a stacked histogram of the *p*-value distributions for each SES variable resulting from a robust linear regression (n = 396,798) where sex, age, genetic ancestry and predicted cell type proportions were adjusted for. Below is a stacked histogram of the delta beta values which represent the change in methylation at a given site across the range of the SES variable measured.

2.3.3 There was significant overlap between the most significant CpG sites associated with multiple SES measures.

In order to compare the DNAm signal associated with each SES measure, top sites (1,000 CpGs with the lowest *p*-values) were interrogated. Initially, these sites were compared between all the SES variables and the number of overlapping sites was tested for the probability of obtaining this value by chance using permutation analysis. The number of sites obtained from all

overlaps of three SES measures was statistically significant based on 500 permutations (p-value ≤ 0.05) along with several of the overlaps observed between two categories (Figure 2.3). Most (13/15) of the significant overlaps contained at least one older adulthood SES variable, with the overlap between childhood assets and maternal education being the only significant set containing childhood measures alone.

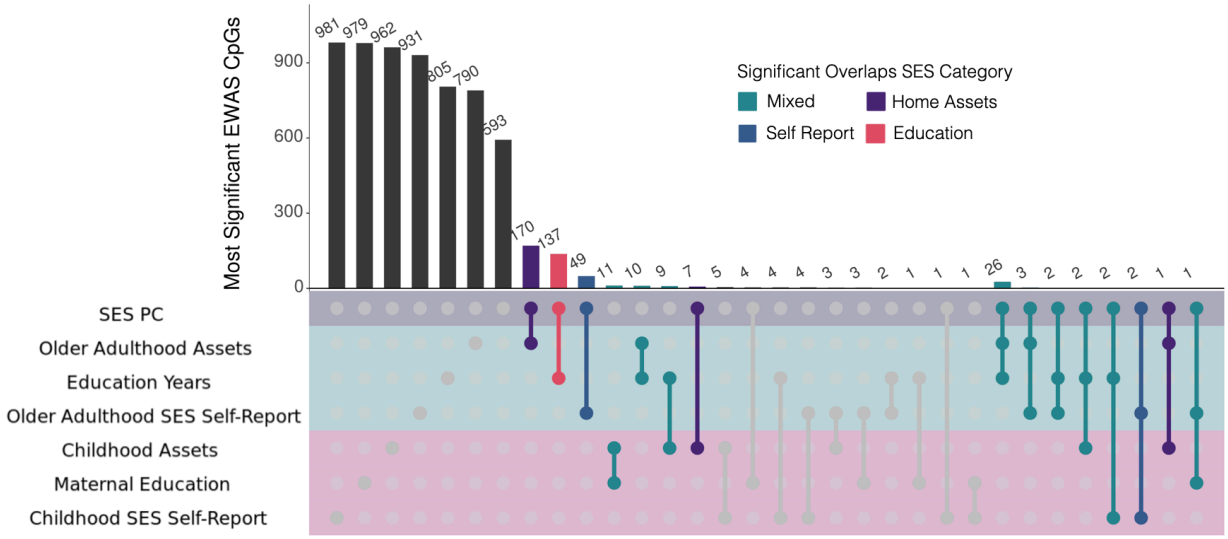


Figure 2.3 Upset plot of top 1,000 CpG sites displaying several significant overlaps across measures of SES throughout the life course.
This upset plot, depicting a simplified version of multiple Venn diagrams, shows the amount of overlap of the top thousand sites across all of the SES measures. Row color signifies SES PC composite (dark purple), older adulthood measures (green) and childhood measures (pink). The overlaps that are coloured represent those which are higher than would be expected by chance based on 500 random permutations (p-value ≤ 0.05). Significant overlaps are coloured based on similarity in SES category (green = mixed, purple = home assets, red = education, blue = self-report measures, grey = non-significant overlaps).

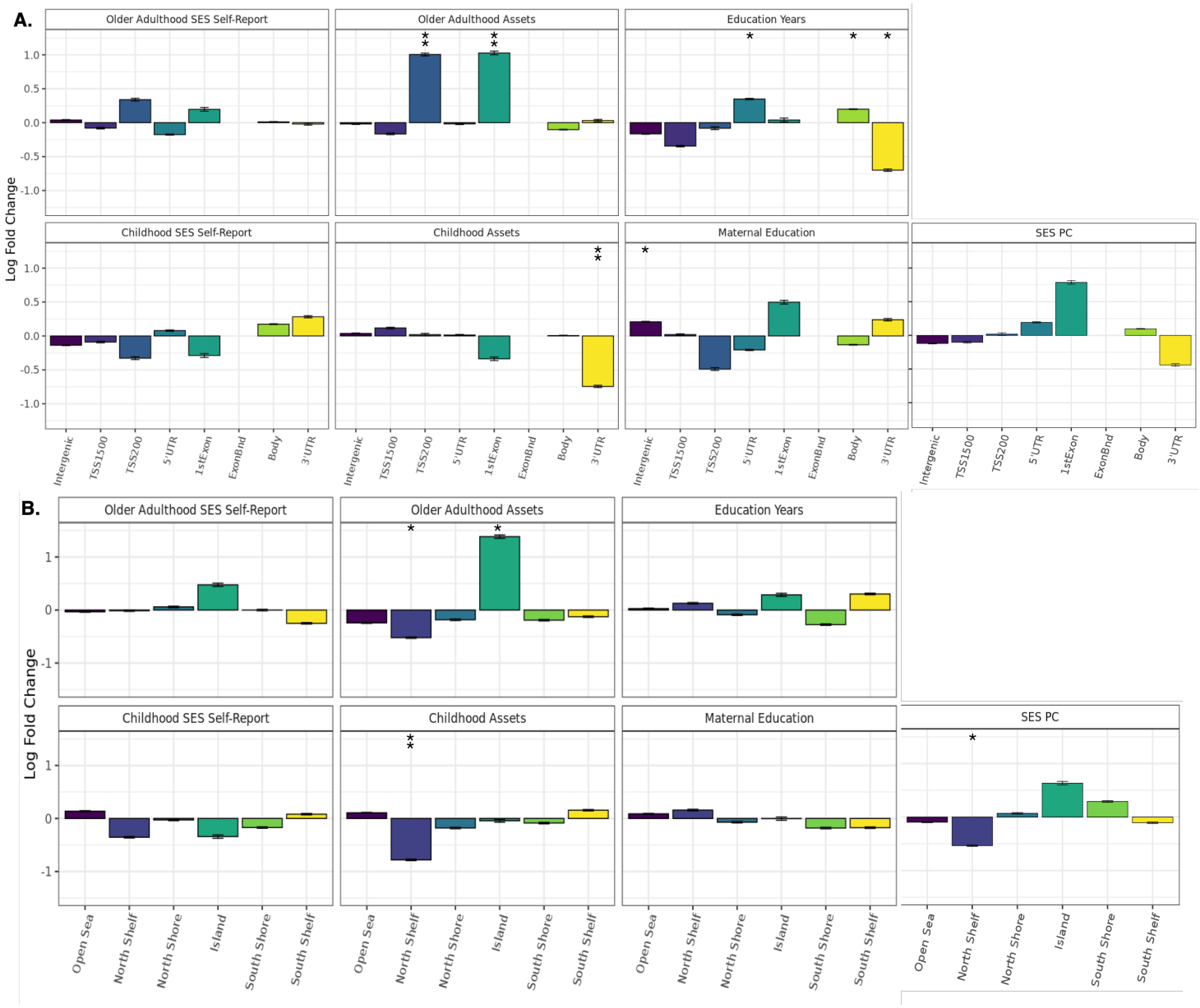
2.3.4 There was significant enrichment for the genomic location and chromatin enrichment of the top CpG sites associated with several SES measures.

The CpG sites most associated with SES were interrogated on their location based on genomic region, CpG island and predicted chromatin states. Fold enrichment was calculated by taking the number of SES associated sites located in a given genomic feature and dividing it by the number of background sites located in the same feature. The background used was the set of CpG sites investigated for the analysis. Whether the enrichment itself was significant was determined by permutation analysis. There were some genomic regions that were significantly enriched amongst sites associated with older adult and childhood assets measures (Figure

2.4(A)). Moderately significant enrichment was also detected in CpG sites associated with educational years and maternal education. Interestingly, no significant enrichment was found in CpGs associated with either self-report measure or the SES PC composite.

Enrichment within the north shelf regions of CpG islands was found amongst CpGs associated with older adult and childhood assets and the SES PC measure (Figure 2.4(B)). CpG sites located in CpG island shelf regions were generally reported to be tissue specific, perhaps reflecting differential proportions of cell sub-types not accounted for in the analysis (357,360). No CpG island enrichment was found in any education or self-report measures. Modest enrichment of CpGs associated with older adulthood assets was found in CpG islands themselves.

Amongst predicted chromatin states measured in mononuclear cells from the NIH RoadMap Epigenomics Consortium, significant depletion of CpGs associated with older adulthood assets was found in the zinc finger (ZNF) genes and repeats state in addition to modest depletion in heterochromatin and enrichment in bivalent transcription start sites (TSS) and enhancers (Figure 2.4(C)). This represents an enrichment of CpGs associated with older adulthood assets being located in regions predicted to be poised for expression and depleted in genomic regions typically silenced. There was also moderate enrichment in predicted TSSs and enhancers amongst all other SES measures except childhood assets.


A.
Older Adulthood SES Self-Report
Older Adulthood Assets
Education Years
Childhood SES Self-Report
Childhood Assets
Maternal Education
SES PC
Log Fold Change
Intergenic
TSS1500
TSS200
5'UTR
1stExon
ExonBnd
Body
3'UTR
B.
Open Sea
North Shelf
North Shore
Island
South Shore
South Shelf

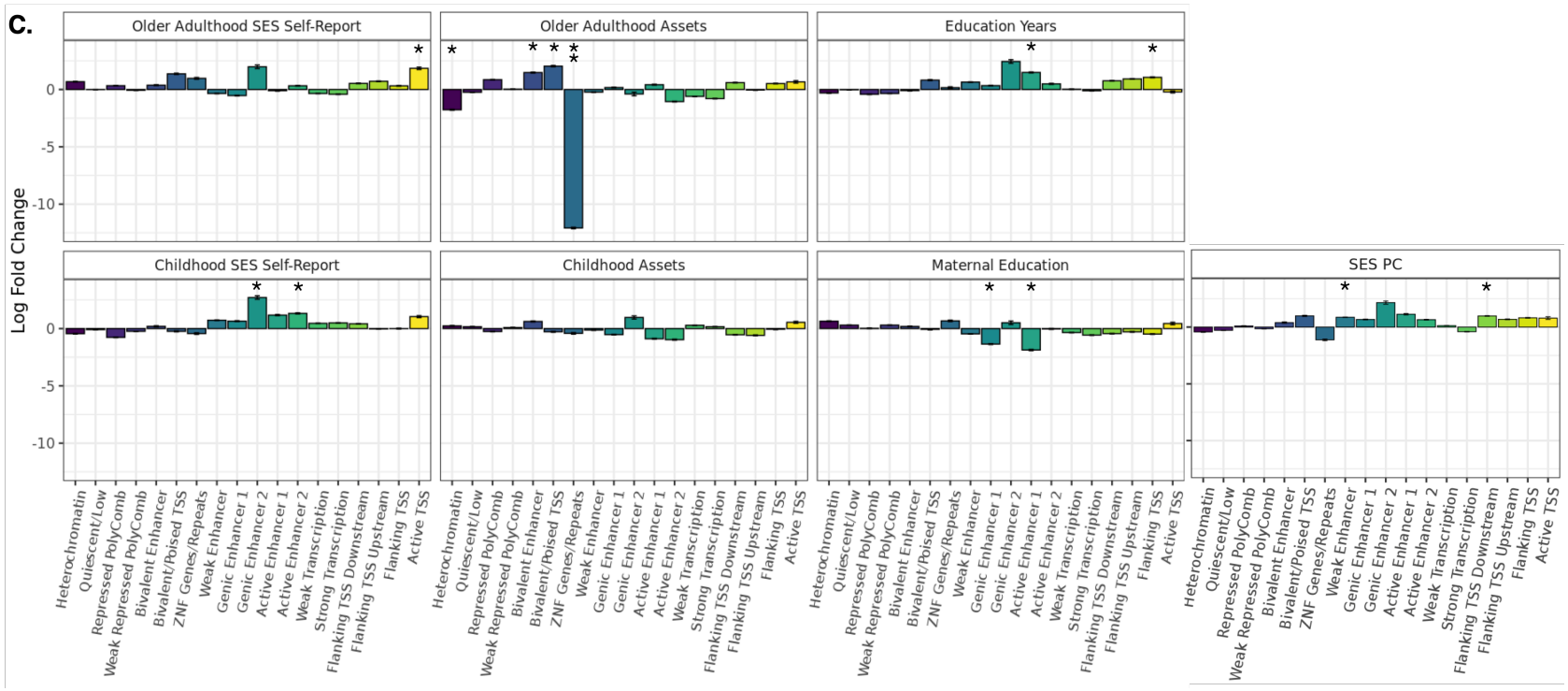


Figure 2.4 Genomic feature enrichment for top CpGs associated with each SES variable.
The number of top sites which were in a given region was compared with annotation background of the array to obtain a fold change and then compared with 500 permutations to determine if the enrichment or depletion was greater or less than by chance (*p-value ≤ 0.05, **FDR ≤ 0.2). (A) Enrichment for a genomic region a given CpG site was located in, if it was assigned to a gene. TSS1500 – 200-1,500 bases upstream of the transcription start site (TSS), TSS200 – 0-200 bases upstream of the TSS, 5'UTR – within the 5' untranslated region (UTR) between the TSS and ATG start site, 1stExon – within the first exon, ExonBnd – exon boundaries, Body – between the ATG and stop codon irrespective of other genomic elements, 3'UTR – between the stop codon and poly A signal. (B) Enrichment for CpG island regions a given CpG site was located in, if it was assigned to a CpG island. (C) Enrichment for SES associated CpG sites to be located in various chromatin states measured from NIH RoadMap Epigenomics Consortium mononuclear cells using ChromHMM methods.

2.3.5 Each SES measure was significantly associated with at least one biomarker or health lifestyle measure.

Associations were calculated between each SES variable and the biomarkers and lifestyle measures of interest using Pearson's correlation where significance is assessed if the *p*-value ≤ 0.05 and the $|r^2| \geq 0.1$. CRP, cortisol, epinephrine and norepinephrine were not significantly associated with any SES measure (Supplementary Table 2.2). DHEAS was negatively associated with both childhood assets and SES self-report. Smoking was significantly associated with older adulthood SES self-report and childhood assets, drinking was associated with both education measures and BMI was associated with both asset measures and the SES PC.

2.3.6 Drinking and multiple stress biomarkers significantly contributed to the distribution of effect sizes amongst the most significant CpG sites associated with several SES measures.

Contribution analysis was performed to determine how much of the effect seen in the relationship between SES and DNAm was attributable to inflammatory and stress biomarkers and lifestyle measures hypothesized to also relate to both variables. This was done by finding the change in the beta coefficient from the base model with that of the base model adjusted for a given covariate of interest. Comparing the overall distributions of all beta coefficients, significant shifts in the distribution were reported in the models of both education measures and older adulthood SES self-report when adjusted for cortisol (Figure 2.5). Additionally, the distributions of beta coefficients for the education years model were also significantly different when adjusting for epinephrine, norepinephrine and drinking status. Interestingly, when adjusting for cortisol the coefficient distribution was widened in the education years model but narrowed in the older adulthood SES self-report model, suggesting that cortisol levels were confounding for the former and correlated with the latter's associations with DNAm.

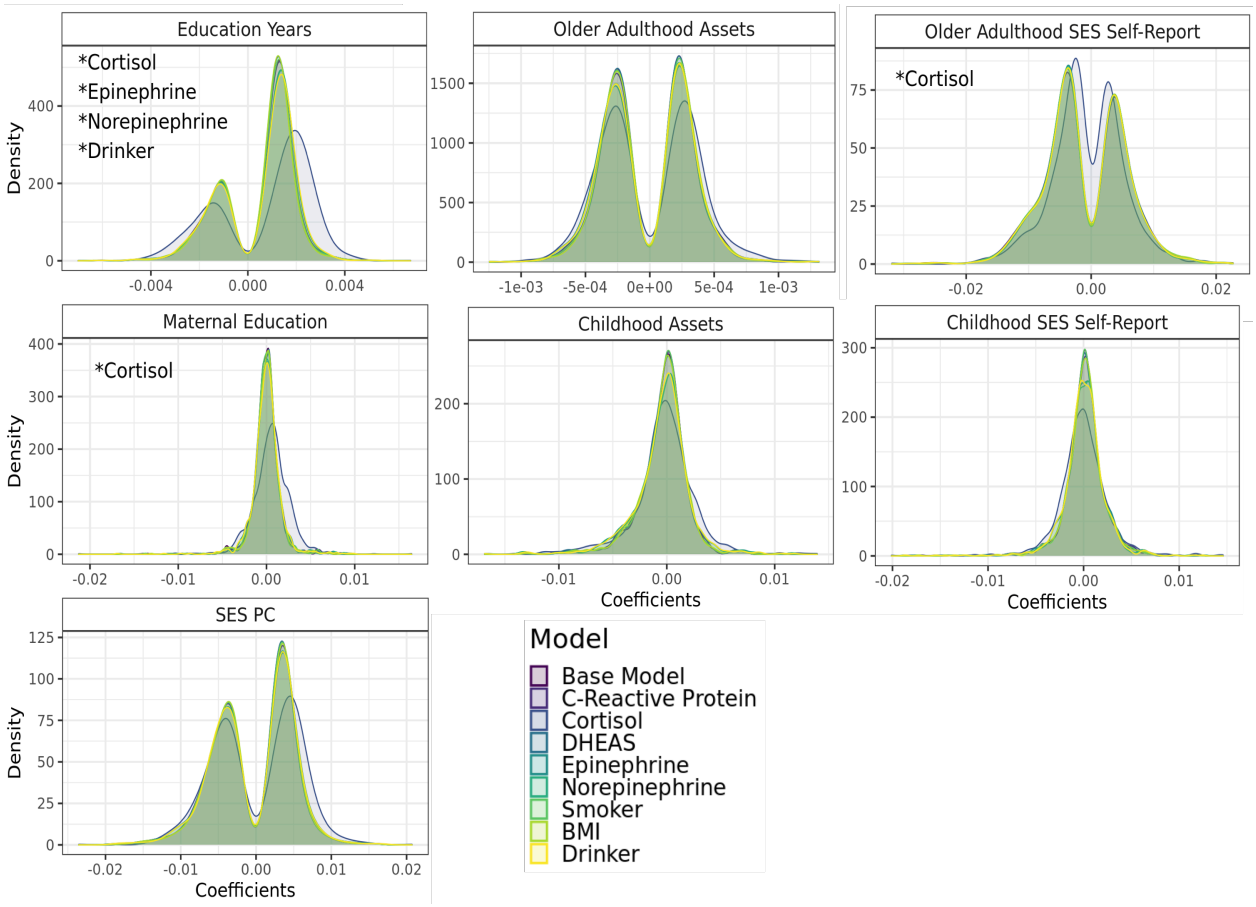


Figure 2.5 Global distributions of SES effect sizes from regression models adjusted for several biomarkers and health behaviours.

Beta coefficients from the base linear regression model used to find SES associated CpG sites were compared to the SES coefficients when the model was adjusted for several inflammatory and stress biomarkers in addition to several lifestyle factors known to be associated with SES. Significance was determined by performing 500 random permutations. *p-value ≤ 0.05 for particular covariate. DHEAS – dehydroepiandrosterone sulphate, BMI – body mass index.

2.3.7 Childhood SES associated CpGs are influenced by biomarkers and lifestyle factors while older adulthood SES CpGs are not.

The number of top CpGs most associated to a given SES variable was analyzed in terms of the proportion of relationships for which the effects were explained by 10% or more with a given biomarker or lifestyle measure. Significance was attributed to this proportion by using permutation analysis (p-value ≤ 0.05). There was significant enrichment in the number of associations whose effects were attributable to all biomarkers in every childhood SES measure, except for cortisol and maternal education (Table 2.2). Conversely, there was either no enrichment or a significant depletion in the number of associations impacted by the inclusion of any biomarker in every older adulthood SES measure and SES PC, with the exception of cortisol and SES self-report. Overall, biomarkers and lifestyle behaviours appeared to contribute more to the associations between DNAm and measures of childhood SES and generally confound the associations between DNAm and measures of older adulthood SES.

		Childhood SES			Older Adulthood SES		
	SES PC	Self-Report	Assets	Maternal Education	Self-Report	Assets	Education Years
Biomarkers							
C-reactive Protein	0	17.2*	20.3*	22.2*	0.2	0+	0
Cortisol	21.1	52.5*	52.2*	44.7	62.7*	17.9+	7.9+
DHEAS	0	19.6*	19.8*	20.6*	0	0.3	0.1
Epinephrine	12.9	41.9*	34.4*	42.3*	12.4	8.5	0.1+
Norepinephrine	14.6	41.1*	34.2*	42.6*	6.3	10.5	0.3+
Lifestyle Behaviours							
Smoker	0	10.6*	13.1*	11.8*	0.2	0	0+
Drinker	14.4	40.8*	36.9*	41.8*	6.9	9.8	0.6+
Body Mass Index	0.3	21.5*	23.4*	26.2*	0.2	2.2	1.9

Table 2.2 Summary of the percentage of the top 1,000 DNAm sites associated for SES which had a covariate contribute to the relationship by at least 10%.

Comparing the number of sites amongst the top thousand where the relationship between methylation and SES is explained by at least 10% by a given covariate of interest. Randomization permutation analysis (n = 500) was used to determine the probability that there was an enrichment (*) or a depletion (+) in the number of sites impacted by these covaries than one would expect by chance (p-value ≤ 0.05). DHEAS – Dehydroepiandrosterone sulfate.

2.4 Discussion

The link between DNAm and SES was explored by comparing the associations between multiple measures of SES occurring in either older adulthood or childhood. Measures of education, household assets and self-report of poverty were compared, in addition to a composite SES PC variable, which associated with all other SES measures. More significant associations were found between DNAm and older adulthood than childhood SES, while self-report measures either had reduced signal or a dissimilar trend compared to other SES categories. EWASs conducted on older adulthood measures of education and assets resulted in a top set of CpG sites where significant genomic enrichment analyses suggested an increased presence in genomic regions where active transcription may be taking place. When assessing the effect of including inflammatory and stress biomarkers in the DNAm and SES models, cortisol had the largest impact — contributing to the signal of both self-report SES measures and childhood assets while confounding the signal of the remaining older adulthood SES measures. The remaining biomarkers and lifestyle behaviours contributed significantly to the effect detected in the top CpGs associated with childhood SES measures where there was either no significant effect or they acted as confounders for DNAm and older adulthood SES associations. Subjective SES measures of self-report behaved differently from objective measures in relation to DNAm in most analyses. This finding may support the DOHaD hypothesis stating that the early life environment may impact lifelong trajectories of physiological responses, specifically inflammatory and stress pathways, as in the case of this study (18).

2.4.1 Subjective and objective SES components behave distinctly relative to associations with DNAm.

While several studies have reported associations between SES and DNAm, an issue remains in the field: replication of these findings is inconsistent (296). There could be many reasons for this, but one major issue is the complex nature of SES and the variability of how this is measured across studies (326). While not all-encompassing, three components of SES were investigated in the current study -- household assets, education, and self-reported economic standing -- in addition to a composite SES measure that associated with all others. When investigating overall DNAm variation using PCA, I observed correlations between all SES measures and the top PCs, though both self-report measures displayed fewer overall associations

relative to the other measures, suggesting that this metric may be distinctly impacting DNAm (361–364). The current study shows that, while there are some overlaps amongst CpGs most associated with a given SES variable, significant overlaps between any given set of variables were never greater than 20%. This suggests there are unique impacts of SES components to DNAm. Though it is important to note that though there was no overlap does not mean that the non-overlapping sites are not related to the same biological pathway. Furthermore, one objective measure, household assets, was found to have significant enrichment or depletion amongst three genomic features, suggesting localization of associated CpGs in areas of the genome where active transcription is potentially occurring, while self-report measures did not. Most revealingly, self-report measures of SES in both older adulthood and childhood behaved distinctly from other measures with an overall weaker DNAm signal. These findings are not surprising as it has been previously found that self-report measures can be distinct from more objective measures because self-report incorporates an individual's relativistic interpretation of their situation (335). As such, self-report measures are generally more closely linked to mental health, whereas objective measures are more closely linked to physical health (335,365,366).

Moreover, the contribution of both assessed biomarkers and lifestyle measures further solidified the interpretation that the DNAm signal of objective, subjective, and composite SES measures were distinct. It has been frequently reported that poverty is associated with levels of inflammatory and stress biomarkers. Additionally, lifestyle measures such as smoking, drinking and BMI are frequently higher amongst those with lower SES. What impact these measures may have on the relationship between DNAm and SES was investigated, and there were significant differences between the coefficients resulting from the base DNAm and SES models and those that were adjusted for several covariates. Cortisol is a particularly interesting biomarker, as the distributions amongst both education and the older adulthood self-report variables were significantly shifted, but in opposing directions, showing a distinct difference between objective and subjective SES measures. Though, it should be highlighted that cortisol was measured at a single time point in urine and not multiple times in the saliva, which better represents the diurnal patterns of individuals. The association between DNAm and older adulthood self-report became weaker when accounting for cortisol, suggesting cortisol as a potential driver for DNAm at these CpGs. Conversely, adjusting for cortisol in the models for both education measures appeared to enhance the relationship between these measures and DNAm, suggesting that cortisol levels are

adding noise to the association. This is supported by the fact that subjective measures of stress have been correlated more strongly to cortisol output compared with objective measures (367–369). The educational years EWAS was also significantly impacted by two other stress biomarkers (epinephrine and norepinephrine) in addition to drinking status. Interestingly, the composite SES measure, SES PC, had neither an enrichment nor depletion in the numbers of CpGs affected by these measures, the result of an apparent averaged effect between the older adult- and childhood measures. This is somewhat expected as the SES PC is significantly correlated with each SES variable to varying degrees, and so, the oppositional signal observed between childhood and older adulthood may average out, resulting in no significant enrichment in either direction when analyzing the composite variable alone. The subjective SES measure generally produced weaker associations with DNAm relative to objective measures, where the composite SES measure lost the nuance of how health biomarkers related to the SES components. Together, these findings highlight the need for social epigenetic studies to be carefully designed with consideration for the type and consistency of measurements (370).

2.4.2 Differential impact between older adult and childhood SES associations with DNAm.

An individual's environmental and subjective experiences can have differential impact on health, depending upon their developmental stage. Expanding on this concept, exposures occurring during developmentally sensitive periods in early life can set an individual on a potentially irreversible health trajectory leading to a variety of negative health outcomes, a concept known as the DOHaD hypothesis (18,324,371). Poverty in childhood may have differential impacts than in adulthood, so the timing of a given exposure should be considered when studying SES. In the current study, large differences in overall DNAm signal were seen between older adulthood and childhood measures of SES. Childhood measures resulted in flat *p*-value distributions, indicating little to no significant DNAm association, while an obvious skew was observed amongst older adulthood measures, in addition to the effect sizes being generally larger, indicating a stronger overall significant association.

Another clear distinction between older adult and childhood SES measures in relation to DNAm is the impact various biomarkers and lifestyle factors had. Generally, childhood SES measures were associated to significantly more CpGs related to these covariates than one would

expect by chance. Conversely, the older adulthood measures had either no effect or a significant depletion in the number of CpGs affected by these covariates.

Older adulthood SES measures displayed a stronger association with DNAm overall, where childhood SES measures were heavily confounded by various biomarkers of health, stress, and inflammation. This could be due to DNAm being measured in older adulthood, thus associating more strongly to measures more proximally timed. Another intriguing interpretation could be that childhood SES associated DNAm is heavily intertwined with the physiological state of an individual's stress and immune state. These findings support the DOHaD hypothesis that proposes early life events set the stage and trajectory for key biological systems that are maintained throughout life because it is markers of these systems which would more strongly explain the relationship between childhood SES and DNAm.

2.4.3 Limitations and considerations

There are several important considerations when interpreting the results of these analyses. The DNAm data were taken cross-sectionally, and so results of associations with childhood SES are not comparable to studies in which childhood DNAm and SES associations were taken concurrently. Additionally, when comparing variables it is best to compare measures on a similar scale, here the childhood measures were dichotomous relative to the more continuous measures of adulthood SES, and so may be part of the reason for the subdued DNAm signal in childhood relative to adulthood SES. DNAm is integral to defining cell type and varies greatly within an individual, but also varies substantially across different tissues. DNAm was measured in whole blood in this study, and while some other studies investigating SES and DNAm were reported in the same tissue, many were not. As such, trends discovered here cannot be applied to represent general relationships.

DNAm has also been shown to vary with age, both stochastically and specifically, and so age needs to be taken into account in result interpretations. This cohort is enriched for individuals older than 90 years, which may result in a selection bias for those whose underlying genetics, environment, and behaviour allowed for their longevity, selecting against those with factors which could predispose them to adverse health outcomes and earlier mortality. Additionally, research into SES gradients and varying impacts on health are predominantly

studied in younger populations, resulting in very few studies being conducted in cohorts with this age range.

Beyond how and when in an individual's lifetime SES is measured, where they live can also have a large impact on analysis interpretation. This study was conducted in Costa Rica amongst a population where the typical associations of SES gradients with various measures of mortality and morbidity do not fully apply (329,330,334,338). For example, those who are wealthier were found to have increased incidences of cardiovascular disease, a relationship which opposes commonly reported trends seen in both developed and developing countries. In this study, several measures of SES were positively correlated with BMI, contrary to what is commonly observed, as such the findings from this study need to be interpreted accordingly (372).

SES is a complex and multifaceted measure associated to many health disparities. DNAm is an epigenetic marker that may be able to elucidate the link between the two, with a wealth of EWAS studies finding associations between DNAm and SES. However, social epigenetic studies suffer from a lack of reproducibility, likely due to the same issues that plague social epidemiology: inconsistency in the SES measures themselves and the timing with which these various exposures occur in an individual's life (370). Specifically, self-report SES measures throughout life had a weaker and distinct signal in DNAm relative to the more objective measures, while the SES PC composite CpGs were not significantly impacted by biomarkers or health measures, possibly due to the opposing trends seen between older adult and childhood measures. Finally, I found a stronger association between concurrent older adulthood measures and DNAm, though, unlike older adulthood, DNAm associated with retrospective childhood measures were significantly confounded by stress and inflammatory biomarkers. These results highlight the importance of considering the type and the life stage of SES measures when designing a social epigenetic study.

3 Lifetime trajectories of socioeconomic status partially mediated associations between DNA methylation and IL-6 levels

3.1 Background and rationale

A significant contributor to health inequalities throughout the lifespan is the myriad of social and environmental risk factors associated with socioeconomic status (SES), which generally refers to a composite measure of the occupational, economic and social standing of an individual or group, relative to others in society (373). Several risk factors that low-SES individuals are exposed to impact many aspects of life and can include poor nutrition, air pollution exposure, food insecurity, and neighborhood and family violence (374). In turn, individuals living in low-SES conditions tend to have increased incidence of a number of chronic health conditions (1,3,4,142,375–377) with exacerbating effects stemming from low-SES experienced in early childhood (2,5,378–381). Interestingly, health associations with early childhood low-SES have been found to persist later in life, even amongst those whose SES has improved after childhood, indicating that these experiences may be getting "under the skin" to have long-lasting impacts on health (142,382–385).

At a biological level, the mechanisms by which a low early-life SES predisposes an individual to a variety of health conditions remains poorly understood but likely involves several biological processes and systems that might, in part, intersect. One phenotype commonly, but not exclusively, associated with both a low early-life SES and chronic health conditions is inflammation (5,15,145,167,376,386–388). Stress, in general, results in the release of the hormone cortisol, which can suppress immune function in part by reducing production of the pro-inflammatory cytokine interleukin-6 (IL-6), linking the stress and innate immune pathways (163,389). However, long-term release of cortisol, as seen during chronic stress that might originate from multiple sources including, but not limited to, low-SES, has been linked to increased levels of pro-inflammatory cytokines (390–392). For example, individuals who experienced low-SES throughout life revealed increases in basal IL-6 and C-reactive protein (CRP) levels, another marker of inflammation (12,295). In addition, many of the demographic and environmental variables often associated with low SES, such as air pollution (393) and poor nutrition (394), have also been associated with persistent low-grade inflammation, suggesting that SES might impact the immune system through multiple avenues.

The molecular mechanisms by which low-SES predisposes to a pro-inflammatory phenotype are still unclear. Epigenetic mechanisms, which are mitotically inheritable and tightly linked to gene transcription, may have potential importance in this association reported throughout life (314,395). In human populations, the most studied epigenetic mark is DNA methylation (DNAm), which refers to a methyl group bound to a cytosine base, typically existing in the context of cytosine-phosphate-guanosine (CpG) di-nucleotides (395). At the most fundamental level, DNAm has important roles in the establishment and maintenance of cellular identity during the development of an organism, and is important for the life-long maintenance of body systems, including immune function. On the more macroscopic level, DNAm associates with many variables that constitute the social environment in both human populations and animal models, including parental care, abuse exposure and educational attainment (313,314). As such, DNAm might be a potential mechanistic bridge linking environmental and social exposures with immune gene expression, and a possible mechanism by which early-life experiences become biologically embedded to influence health across the lifespan (15,151,313,314).

At least two different, yet complimentary lines of empirical evidence substantiate a link between DNAm and components involved in pro-inflammatory processes– first, the identification of cellular DNAm marks associated with key pro-inflammatory proteins and second, the identification of DNAm changes in pro-inflammatory genes associated with conditions that have an underlying pro-inflammatory phenotype. IL-6 and CRP measures have been central to the first approach, with several studies testing associations using either an epigenome-wide approach (EWAS) or at the candidate gene level (396,397). IL-6 is particularly of interest as it exhibits pleiotropic activity in the immune system and is expressed from many tissues (163). In a small pilot EWAS, serum IL-6 was associated with DNAm levels at two CpGs in peripheral blood mononuclear cells (PBMCs) (398). CRP, another pro-inflammatory marker downstream of IL-6 signaling (12,399), has also been linked with blood DNAm (400–403). In one study, associations between CRP levels and DNAm were found at 218 CpGs amongst individuals of European ancestry, 58 of which were validated amongst African Americans, hinting at possible trans-ethnic signatures in DNAm associated with pro-inflammatory markers (404). There have been substantially more candidate gene studies performed on cytokine related DNAm analyses, typically including genes involved in innate immune responses (396). DNAm changes within the *IL6* gene itself are associated with several diseases, including coronary heart

disease (405), schizophrenia (406), depression (407), and rheumatoid arthritis (408). Additionally, through activation of the JAK2/STAT3 pathway, IL-6 impacts DNAm genome-wide by decreasing expression of DNA (cytosine-5)-methyltransferase 1 (*DNMT1*), a gene whose protein product is integral to the deposition of methyl groups across the genome (409–411). Changes in global DNAm resulting from IL-6 impact specifically on DNMT1 levels has been investigated as a potential mechanism in several diseases including type 2 diabetes (409), lupus (412), neurogenesis (410) and Autism Spectrum Disorder (411). Moving from group differences to possible predictors at the level of an individual, blood DNAm patterns can also be used to bioinformatically determine levels of serum IL-6, with this prediction relating to cognitive functioning in older adults (413). The relation between pro-inflammatory cytokines and DNAm emphasizes the interconnectedness of these molecular pathways highlighting the need to integrate measures of each in subsequent analyses.

While less proximal than associations between components of the inflammatory response and immune cell DNAm, a growing body of literature reports on a link between SES throughout the life course and DNAm. Prenatal SES has been associated with DNAm changes in cord blood at birth, with some changes persisting into mid-childhood (315). In buccal epithelial cell (BEC) DNAm from children aged 5-6 years, CpGs in genes regulating immune function and developmental pathways were associated with early-life SES (228). Amongst young adults, an EWAS conducted on whole blood DNAm was found to be a significant mediator between early-life SES, parenting and later life health (316). Consistent with these findings, white blood cell DNAm in various immune genes was associated with early-life SES in adults (226), in addition to DNAm at genes involved with nervous system development (159). In PBMCs, DNAm was broadly associated with early-life SES irrespective of current-life SES, suggesting that at least in part, DNAm patterns in adults might be a vestige of their early life experiences (231). However, work focusing on a small subset of DNAm candidates in immune function and stress reactivity pathway related genes found both overlapping and unique associations of monocyte DNAm with early-life and current-life SES trajectories, suggesting possible differential impacts of SES status throughout life in this more focused analysis (317). In addition to identifying associated genes and pathways as in the studies described above, the information content of the DNA methylome can also be exploited to predict cellular or biological age (84). Specifically, when compared to chronological age, the predicted epigenetic age can inform on several adverse health outcomes

(37,88,414). The epigenetic age measure has been applied to life course studies of SES, in general finding that low SES associated with accelerated epigenetic age, which is associated with reduced longevity and increases in morbidity (227,415). Interestingly, one study found DNAm age accelerated in individuals with low-SES in early life but with a reduction in effect amongst those whose SES improved in later-life suggesting a potential reversibility of these biomarkers (230).

Building on the existing body of research that points to a relationship between DNAm and SES throughout the life course I have hypothesized that DNAm is associated with both IL-6 and SES within some of the same CpG sites. Specifically, I explored the relationship between the lifetime trajectory of SES, serum IL-6 levels, and DNAm using a stepwise approach. First, significant EWAS associations between monocyte DNAm and IL-6 were assessed with SES incorporated in the model to determine what, if any, impact this social variable may have on the relationship between DNAm and IL-6. Subsequently, associations between SES and serum levels of IL-6, a pro-inflammatory cytokine produced primarily by monocytes were tested (162,163). Finally, DNAm sites not associated with IL-6 were related with SES trajectory to determine any IL-6-independent associations between SES and DNAm. Collectively, this approach allowed for examination into the complex interaction between SES and inflammation through the lens of epigenetics by concurrently exploring the relationship of DNAm to proximal biological (serum IL-6 levels) and distal sociological (lifetime SES trajectory) measures.

3.2 Materials and Methods

3.2.1 Early Life cohort

The Early Life cohort consisted of 333 individuals aged 15-70 years recruited from the greater Vancouver area. SES, recruitment, and exclusion criteria have been previously described (227). Sample demographics are summarized in Supplementary Table 3.1.

3.2.2 CareGiver cohort

The independent validation cohort was selected from the CareGiver study, previously described (416). A propensity age-matched subset of 43 participants aged 22-54 years (recruited from the Vancouver area) was used. Sample demographics are summarized in Supplementary Table 3.1.

3.2.3 Childhood and adulthood socioeconomic status questionnaire

SES was defined as occupational prestige using the thoroughly validated United Kingdom's National Statistics Socioeconomic Classification, which is defined as the participants' parents' occupational prestige in early-life (0-5 years) and the maximum household occupational status within five years of the participants themselves, as previously described (227). Individuals were categorized as those who remained in high SES throughout their life (High/High, Early/Current, HH), individuals with downward mobility (High/Low, HL), individuals with upward mobility (Low/High, LH), and those who remained in low SES throughout their life (Low/Low, LL). Additionally, participants were selected such that there were roughly equal groups of individuals in every given category (LL 28%, HH 26%, LH 26%, HL 20%).

3.2.4 Biological measures

3.2.4.1 Serum IL-6

Interleukin-6 (IL-6) was measured from participants' serum using identical procedures in both cohorts. Serum was isolated from the centrifugation (2200-2500 rpm) of clotted whole blood samples and applied to a commercial ELISA kit with IL-6 specific antibodies where titers were read out in pg/ml. IL-6 levels were natural-log transformed to minimize the impact of extreme values.

3.2.4.2 Blood collection, monocyte isolation, genomic DNA extraction and DNAm measurement

Whole blood was collected for participants from both cohorts, and PBMCs were isolated. Monocytes, defined as $CD14^{+}$, were separated using immune-magnetic capturing methods (Miltenyi AutoMACS) as previously reported (227,416). For both cohorts, DNA was isolated from the purified monocytes using a column-based method and assessed for quality using NanoDrop, bisulfite converted using EZ-96 DNA Methylation kits (Zymo Research, Irvine, CA) and run on the Infinium HumanMethylation450 (450K) BeadChips (Illumina) (227). Samples were randomly assorted to ensure there was no confounding between variables of interest and processing batches.

3.2.5 Statistical analysis

3.2.5.1 DNAm pre-processing and normalization

For both cohorts, raw intensity IDAT files were imported and processed in RStudio, where intensity measures were converted to beta values (Methylated + Unmethylated/Methylated) ranging on a continuous scale from 0 (Unmethylated) to 1 (Methylated) for each of the 450,557 CpGs.

Background subtraction and color correction were performed using the preprocessNoob function (417) from the minfi package (343). Prior to normalizing the data, samples were removed if the median for both Methylated and Unmethylated signals were below 10.5 in order to ensure extreme outliers did not impact the normalization process (343). Control and out-of-band probes were used to perform normalization on the samples using the preprocessFunnorm function (344) from the minfi package (343) to remove technical variation.

After normalization, three additional methods were used to determine if a sample was considered an outlier. Samples were removed if detected by at least two methods. The first method, detectOutlier from the lumi package, (345) labeled a sample as an outlier if it was significantly distant from the center of the mean of other samples. The second outlier method, pcout from the wateRmelon package (100), used principal component analysis to calculate distance and scatter for each observation relative to the cohort average for outlier identification. The third method employed the locFDR package (80,346,347). Briefly, principal component analysis was used to determine if the z-score statistic of a given sample was significantly separated from the rest of the cohort (FDR ≤ 0.2) (80). Early Life had one sample and CareGiver had nine samples removed based on the criteria of being selected in any outlier method. In addition to detecting outliers, K-means clustering was performed on the methylation values of the sex chromosome probes to produce a predicted sex measure, where samples whose reported and predicted sex did not match were removed (Early Life n = 11, CareGiver n = 5). To test the purity of the monocyte samples the Houseman method was used to bioinformatically predict whole blood cell types using the Idol reference dataset (64,354). Removing outlier, sex-mismatched and technical replicate samples (Early Life n = 4, CareGiver n = 3) resulted in n = 333 samples in Early Life and n = 43 in CareGiver.

Poor performing probes, defined as those which give NA readings in more than 1% of samples, those with low detection p-values $\leq 1x10^{-16}$, those predicted to non-specifically bind in-

silico, and because of the imbalance of sex and the trimodal distribution of X chromosome probes in females, those which bind the X or Y chromosome (348) were removed, leaving n = 436,788 probes in Early Life and n = 434,560 probes in CareGiver.

Known technical variation resulting from run, chip, and chip position were removed with empirical Bayes methods using the ComBat function (350) from the SVA package (351). As I am not interested in CpG sites which do not vary across participants, invariable probes were filtered out (defined by having less than 5% beta value range across samples in the 5th and 95th percentile), leaving n = 136,435 probes remaining in Early Life for analysis.

3.2.5.2 Serum IL-6 and DNAm

The main effect of associations between DNAm and IL-6 (pg/ml) levels was examined with linear regression via the function lm (418,419) in the base R stats package (DNAm ~ log (IL-6 pg/ml) + age + ethnicity + sex). Multiple testing was corrected by adjusting the nominal *p-value*s using the Benjamini-Hochberg (BH) procedure to calculate the false discovery rate (FDR) (420). CpGs were considered significant at a fairly lenient FDR ≤ 0.20 and a change in methylation state of at least 0.03 in either direction, calculated at each site by taking the beta coefficient from the model DNAm ~ log (serum IL-6 pg/ml) and multiplying it by the cohort range of log (IL-6 pg/ml) levels. The threshold of $|\Delta\beta \geq 0.03|$ was selected because this value is higher than the root mean squared error of 0.025 calculated from the technical replicates after pre-processing and normalization. This approach provides some confidence that DNAm differences found are unlikely due to residual uncorrected technical variation (421).

3.2.5.3 Independent cohort validation

Sample collection, preparations, DNAm preprocessing, normalization, and serum IL-6 effect size calculations were performed identically in Early Life and the validation cohort, CareGiver. Propensity matching for age was done to ensure comparable cohorts, resulting in 43 samples from CareGiver being used for validation, as is summarized in Supplemental Table 3.1. The Early Life CpGs that passed both the statistical and effect size thresholds were evaluated in CareGiver. A CpG site was considered replicated if it had at least an $|\Delta\beta \geq 0.03|$ in the same direction of effect as was measured in the Early Life cohort based on previously reported

methods (422). Pearson's correlation was performed on both all and only validated IL-6 associated CpGs to determine how similar these CpGs behaved in an independent cohort.

3.2.5.4 Association between *DNMT1* expression and serum IL-6 levels

Assessment of previously reported findings of a skew towards decreased methylation and *DNMT1* expression related to increasing IL-6 levels was performed. Skewness in the distribution of the IL-6 Δβ values in the variable sites within Early Life (n = 136,435) and CareGiver (n = 131,321) were performed using the Shapiro-Wilk's test of normality with 1,000 Monte Carlo simulations (423). In addition to DNAm, transcriptome profiling was performed for the CareGiver cohort using Illumina Human HT-12 v4.0 beadchips as previously described (416). Expression levels of *DNMT1* from CareGiver were tested for association with serum levels of IL-6 across samples with all data available (n = 76) using a Pearson's correlation (p-value ≤ 0.05). In order to ensure extreme values of DNMT1 transcript levels were not driving any associations, values were winsorized (424).

3.2.5.5 SES and serum IL-6 measures

An ANCOVA was performed using the aov function from the stats package (418), to determine any association between lifetime SES trajectory measures and IL-6 levels while adjusting for age, ethnicity and sex. To determine which comparisons, if any, contributed to a significant finding a Tukey test was performed using the TukeyHSD function from the stats package (418).

3.2.5.6 SES contribution to IL-6 and DNAm associations

The DNAm changes associated with serum IL-6 levels at validated CpGs were investigated for potential contribution from lifetime SES status. Lifetime SES trajectory measures from Early Life were added to the base model, and the beta coefficients of the main effect of serum IL-6 were compared between the base and SES-adjusted models:

I. Base model: DNAm ~ log (IL-6) + Age + Sex + Ethnicity
II. Adjusted model: DNAm ~ log (IL-6) + SES Trajectory + Age + Sex + Ethnicity

CpGs where the effect of SES trajectory was statistically significant (FDR ≤ 0.2) in the adjusted model were further investigated. Percentage contribution was calculated as

$$[\boldsymbol{\beta}_{\text{IL-6, base model}} - \boldsymbol{\beta}_{\text{IL-6, adjusted model}}/\boldsymbol{\beta}_{\text{IL-6, base model}}] * 100$$

for every validated site as previously reported (210).

3.2.5.7 SES and DNAm

Of the remaining CpGs not associated with serum IL-6 levels, ANCOVAs were performed to investigate more broadly associations between SES and DNAm. Sex, age, and ethnicity were adjusted for in the model. Multiple test correction was applied to the nominal *p*-values using the BH procedure to calculate the FDR. As before, CpGs were only considered significant at a fairly liberal FDR ≤ 0.20 and change in methylation state by at least $|\Delta\beta \geq 0.03|$ as calculated at each site by taking the difference between the median of the groups. The four categories for SES resulted in six Δβ measures representing each comparison. To determine if there was an uneven distribution amongst the number of CpGs which were associated to a particular SES trajectory, 1,000 permutations were performed where SES was randomized and the analysis re-run. The *p*-value was determined by counting the number of times a random result was equal to or more extreme than the true result.

3.2.5.8 DNAm and gene expression

Only CareGiver samples from the full cohort that had matching DNAm and gene expression data were used for analysis (n = 77). Expression levels of genes containing model-significant CpGs were compared with methylation levels using Pearson's correlations.

3.2.5.9 Chromatin state and histone modification enrichment analysis

Using the online resource EWAS Atlas (425) and primary monocyte data from the Roadmap Epigenomics Consortium (358), enrichment analyses for chromatin states and post-translational histone modifications were performed on all significantly associated CpGs using hypergeometric tests to compute *p*-values and odds ratios.

3.2.5.10 Correlation between whole blood and brain sample DNAm

The application BECon (https://redgar598.shinyapps.io/BECon/) was used to interrogate CpGs located in genes which function in brain to determine if whole blood methylation correlated with brain post-mortem tissue methylation (38). The list of CpG IDs is submitted and the output is returned in the format of a table describing the correlation between blood and each of the measured three brain regions (Brodmann areas 7, 10, and 20) in addition to the variability seen across samples and within each tissue (38).

3.2.5.11 Exploratory analysis of other contributing lifestyle factors

As waist-to-hip ratio (WHR) and smoking have been previously associated with DNAm (426,427), IL-6 levels (428,429) and SES (430,431), I determined if these covariates were associated with these measures in the Early Life cohort. For associations that were significant (FDR $\leq$ 0.20), contribution analyses described in section 2.4.5 above were performed to determine if these factors contributed to the relationships between DNAm and IL-6 (dependent on or independent of SES).

3.3 Results

3.3.1 Serum IL-6 levels were related to DNAm

To determine if IL-6 levels were associated with monocyte DNAm, linear regression was performed to investigate the relationship between IL-6 and DNAm values adjusting for sex, age and reported ethnicity. 338 CpGs were associated with IL-6 levels at a fairly lenient stringency threshold (FDR $\leq$ 0.2) and showed an effect size of $|\Delta\beta \geq 0.03|$ (Figure 3.1, Supplementary Table 3.2).

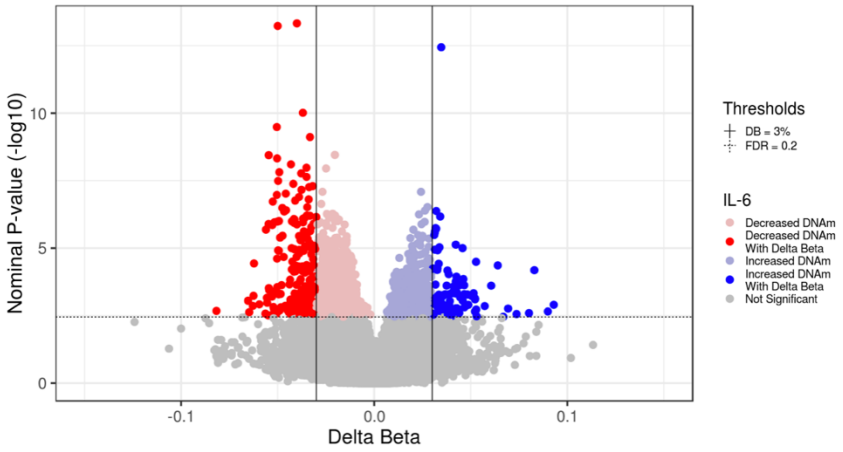


Figure 3.1 Serum IL-6 levels were associated with monocyte DNAm.
Volcano plot of the IL-6 EWAS on variable probes (136,435). The effect size is on the x-axis and the log of nominal *p*-values is on the y-axis, where the thresholds are drawn at the nominal threshold associated with an FDR = 0.2 and the effect size, delta beta (Δβ), thresholds of ±0.03. 338 sites (244 decrease and 94 increase in methylation) met these thresholds.

3.3.2 A substantial fraction of IL-6 associated CpGs validated in an independent cohort

Next, the 338 CpGs modestly associated with serum IL-6 levels were tested to determine validation in an independent cohort. CareGiver is a healthy population cohort described above where purified monocytes were analyzed from 43 samples which were propensity matched to ensure balanced demographics (Supplementary Table 3.1). The validation cohort contained 96% (324) of the significantly associated CpGs from Early Life; the remaining 4% (14) were removed during stringent data pre-processing. CpGs were considered to be validated if the $|\Delta\beta \geq 0.03|$ and was in the same direction as seen in Early Life as per previously published methods (422). Using Pearson's correlation, there was a significant relation between the cohort's Δβ values for all 324 CpGs ($r(322) = 0.53$, $p \leq 2.2\text{x}10^{-16}$, Figure 3.2A) and amongst the 80 CpGs (25%, 80/324) that met the validation criteria of $|\Delta\beta \geq 0.03|$ in the same direction ($r(78) = 0.95$, $p \leq 2.2\text{x}10^{-16}$, Figure 3.2B).

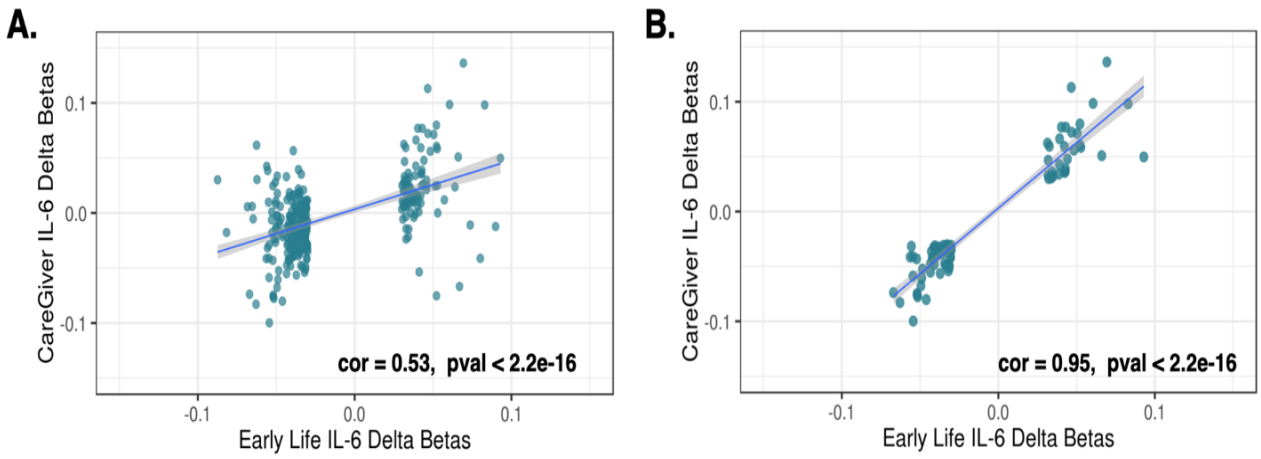


Figure 3.2 Twenty-five percent of DNAm sites significantly associated with serum IL-6 levels were validated in a secondary cohort.
(A) Correlation between Early Life and CareGiver cohorts of the effect size, delta beta (Δβ), values from the 324 sties were significantly associated between serum IL-6 and DNA methylation in Early Life. CareGiver was missing 14 sites of the original 338 sites. (B) Correlation of the delta beta values of the 80 validated sites, defined as significant Early Life IL-6 sites which also had an absolute $\Delta\beta \geq 0.03$ in the same direction in CareGiver.

3.3.3 *DNMT1* expression was inversely correlated to serum IL-6 levels

There was a significant trend observed amongst the IL-6 associations where there were more CpGs with decreased methylation than with increased methylation in both and CareGiver (T = -0.35, p-value $\leq 2.2 \times 10^{-16}$) and Early Life (T = -0.34, p-value $\leq 2.2 \times 10^{-16}$, Figure 3.1). This observation seemed consistent with reports of IL-6 decreasing *DNMT1* transcript levels resulting in global demethylation (409–411), and thus, I tested directly the relationship between serum IL-6 and *DNMT1* transcript levels in one of our cohorts. To investigate if this association persisted in the study, *DNMT1* expression and serum IL-6 levels from all available CareGiver samples (n = 76) were assessed. Pearson's correlation was performed and determined a significant negative correlation between serum IL-6 and *DNMT1* expression which was robust to winsorization (95%, $t(74) = -3.36$, $r = -0.36$, $p = 1.25 \times 10^{-3}$, Supplementary Figure 3.1).

3.3.4 Serum IL-6 levels were linked with SES trajectories

Increased levels of serum cytokines have been previously reported to associate with SES, as such, I also assessed the association of serum IL-6 and SES in this cohort. IL-6 levels were significantly associated with lifetime SES trajectory in the Early Life cohort ($F(3,320) = 3.439$, $p = 0.0172$, Figure 3.3) by performing an ANCOVA between the natural log of serum IL-6 and

lifetime SES trajectory adjusting for sex, age, and reported ethnicity. A post-hoc Tukey test determined the ANCOVA was driven by differences between the High/High and Low/Low SES trajectory groups (*p*-value = 0.021).

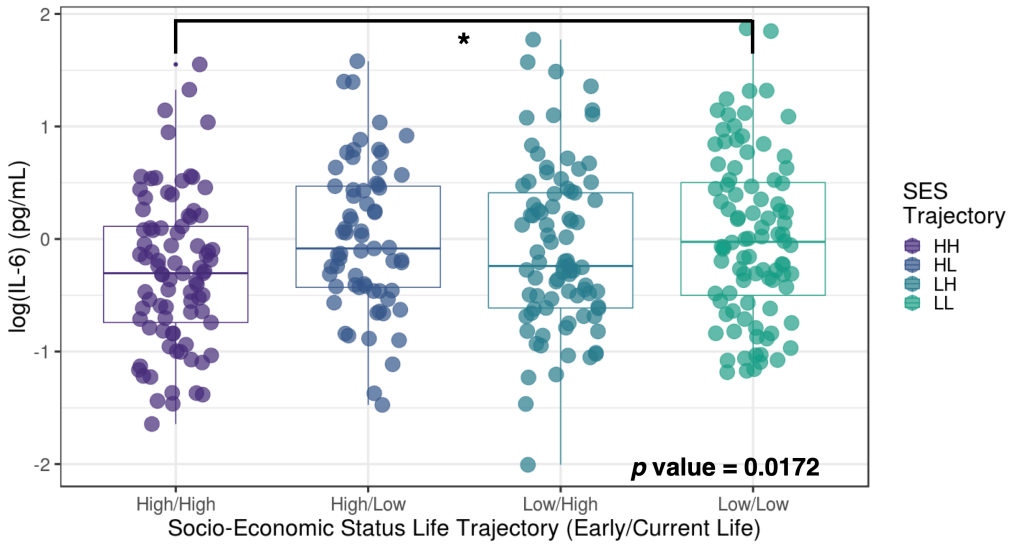


Figure 3.3 Lifetime trajectory of SES was significantly associated with adulthood serum IL-6 measures. ANCOVA was used to obtain the reported *p*-value and was adjusted for age, sex and reported ethnicity. Lifetime SES trajectory was measured based on whether someone was high or low in early life (0-5yrs) in combination with their current SES status measured as occupational prestige. A Tukey test determined that the IL-6 levels of the High/High relative to the Low/Low group was driving the association (**p*-value = 0.021).

3.3.5 SES contributed to the correlation between IL-6 and DNAm

Given the predictions of my theoretical framework stipulating a tight connection between IL6, DNAm and SES, I next tested whether methylation levels of IL-6 associated CpGs were influenced by SES trajectories. I employed a contribution analysis with two linear regression models - the base model used to discover the 80 IL-6 associations above, and this base model adjusted for SES. The beta coefficients for IL-6 were compared between the two models, where the percentage of change was interpreted as the amount SES contributed to the original relationship. Using this approach, I found that of the 80 CpGs, 15 were statistically associated with SES trajectory with varying percentages of contribution occurring ranging from -2.0% to 15.9% (Table 3.1).

CpG	*p* value	FDR	IL-6 Δβ	Max SES Trajectory Δβ	Gene Name	SES Trajectory Adjusted *p* value	SES Trajectory % Contribution
cg00277397	9.55e-04	0.1052	-3.30%	2.1% (HH-LH)	CALN1	0.099	3.84
cg00378510	1.36e-03	0.1246	-4.60%	-4.2% (LL-LH)	LINGO3	0.099	7.60
cg01294327	2.59e-03	0.1655	-4.20%	-3.6% (LL-LH)		0.169	6.53
cg06766960	3.93e-03	0.1975	6.61%	-5.0% (HH-LL)		0.169	15.87
cg09465703	1.41e-03	0.1263	-3.50%	-2.1% (HL-LH)	JMJD8	0.169	1.01
cg12073466	4.99e-04	0.0809	-4.85%	-3.2% (HL-LH)		0.181	-0.74
cg16287740	2.37e-04	0.0577	4.70%	3.3% (HL-LH)	TNPO3	0.181	-1.97
cg20146241	1.00e-03	0.1076	3.14%	-1.7% (HH-HL)	RCAN3	0.181	12.08
cg23651889	5.10e-04	0.0817	3.90%	-3.5% (HH-HL)	LPP/miR-28-5p	0.099	14.75
cg23842572	1.91e-06	0.0036	3.20%	-2.0% (HH-HL)	MPRIP	0.099	6.72
cg24333621	6.92e-04	0.0928	-5.60%	2.9% (HH-LH)	SLC35B4	0.169	1.52
cg26053840	6.88e-04	0.0925	3.20%	2.1% (HH-HL)	GLO1	0.175	11.77
cg26562691	5.27e-04	0.0829	4.80%	3.1% (HH-HL)	PRKCB	0.181	7.05
cg26754184	6.18e-04	0.0888	3.30%	1.9% (HH-HL)	OR3A2	0.169	12.96
cg27209729	4.03e-04	0.0727	-5.20%	2.5% (HH-LH)	NRXN2	0.169	-0.17

Table 3.1 The change in methylation across serum IL-6 levels at 15 CpGs was contributed in part by lifetime SES trajectory status.
Medium-confidence IL-6 associated CpGs were characterized by their significance level, effect size, genomic association, and SES contribution. UCSC was used to determine the gene name the CpG site was located in using human genome build GRCh37/hg19. The max SES trajectory delta beta value reports the highest effect size of the six SES comparisons.

3.3.6 SES associated with CpGs independently of IL-6 levels

IL-6, while influential, is only one factor potentially involved with SES effects on health. Therefore, to investigate other biological pathways potentially involved in the relationship between SES and DNAm, an EWAS of SES associations with DNAm independent of IL-6 was performed. This exploratory analysis was conducted on all but the 338 IL-6 associated CpGs reported above (n = 136,097 probes determined to be above my variability threshold). Using an ANCOVA approach comparing all four SES groups (Low/Low, Low/High, High/Low, High/High), 51 unique CpGs that passed the statistical and effect size thresholds (1,000 permutations, *p*-value = 0.013), with the majority of these changes being found in comparisons containing the LH, or upward SES mobility, group (1,000 permutations, *p*-value = 0.040, Figure 3.4, Supplementary Table 3.3).

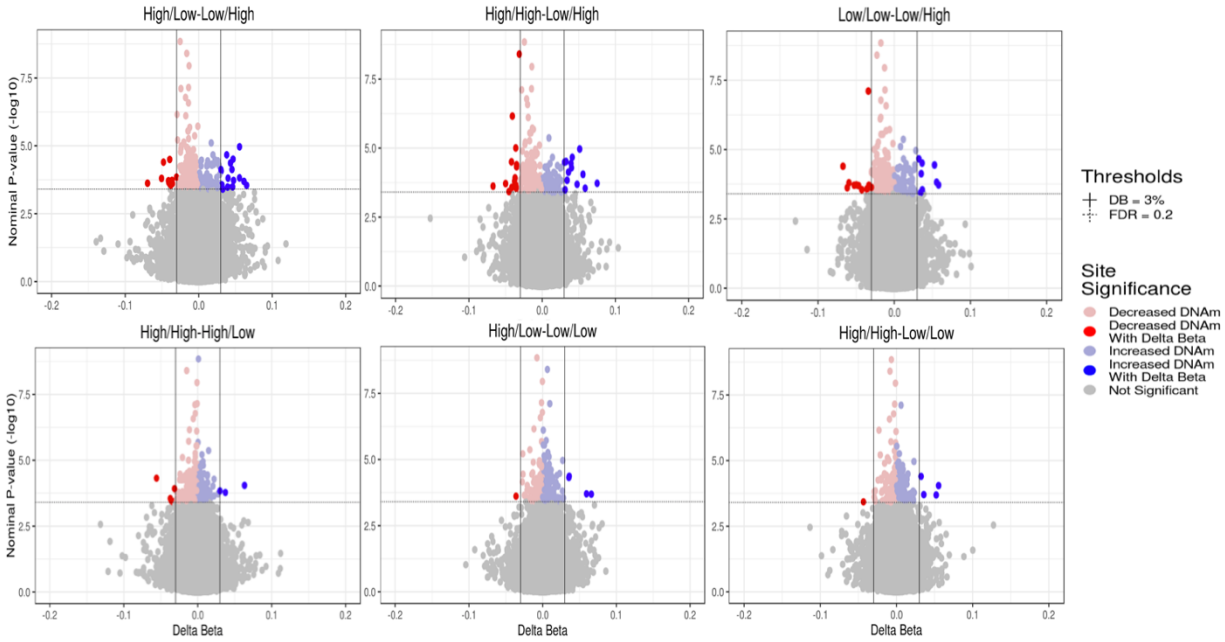


Figure 3.4 SES trajectory was associated with DNA methylation independent of serum IL-6 levels. Volcano plot of SES trajectory comparisons on variable probes (n = 136,435). The y-axis is log of the nominal *p*-values obtained from the ANCOVA model of DNAm ~ SES Trajectory + Age + Ethnicity + Sex and the x-axis is the delta betas calculated for the specific SES trajectory comparison by calculating the difference between the groups specified on the respective panel title.

3.3.7 Gene expression was correlated with DNA methylation for a subset of CpGs

One important role ascribed to DNAm is potential associations with gene expression, specifically mRNA levels. Thus, the CareGiver cohort, for which both DNAm and gene

expression were measured from the same monocyte preparations, was used to assess the relations between these two modalities of genomic expression of all model-significant CpGs. Pearson's correlations were conducted between the mRNA levels of the genes containing significant CpGs and the methylation levels of the CpGs (FDR ≤ 0.2, Supplementary Table 3.4). There were comparable numbers of CpG sites where the methylation levels of annotated genes were correlated to gene expression; 20% (12/60), 23% (3/13) and 12% (4/34) of IL-6, IL-6 and SES and IL-6 independent SES associated CpGs, respectively (Supplementary Figure 3.2).

3.3.8 Enhancers and H3K4me1-marked regions were enriched in IL-6-related CpGs

Regulation of gene expression by DNAm may involve regulatory regions not located in the immediate vicinity of the target gene, for example in enhancers. To assess if any IL-6- and SES-associated CpGs may function in such regulatory regions, publicly available data from the Roadmap Epigenomics Consortium, accessible through the EWAS Atlas Toolkit platform, was utilized (425). Specifically, the multiple chromatin marks measured in single peripheral blood primary monocyte samples were used. Enrichment analysis for chromatin state and histone modifications was performed separately on each set of significant CpGs; 80 CpGs from the IL-6 EWAS, which included the 15 CpGs from the joint SES and IL-6 analysis, and 51 CpGs from the IL-6 independent SES EWAS. The SES-associated IL-6 and SES EWAS CpGs were not enriched for any chromatin markers. The 80 IL-6 associated CpGs exhibited significant enrichment for active and primed enhancers (OR = 6.79, $p = 1.16 \times 10^{-12}$) as well as H3K4me1, another epigentic marker of these (OR = 3.47, $p = 5.08 \times 10^{-8}$) (432). Enrichment of both these epigenetic marks were suggestive of an active transcription state, meaning CpGs associated with IL-6 levels may be in actively transcribed genomic regions (432).

3.3.9 CpGs within brain related genes had correlated methylation levels between whole blood and brain tissue in a reference dataset

A subset of 17 IL-6 and/or SES associated CpGs located in 12 genes with neurological function, were interrogated using the online tool BECon (https://redgar598.shinyapps.io/BECon/) to determine whether the methylation levels at these CpGs in blood might report on the methylation state in the brain where the likely mode of gene action occurs (38). Overall, 4/8 IL-6 associated, 1/4 IL-6 and SES associated, and 3/5 IL-6

independent SES associated CpGs were correlated between whole blood and at least one of three brain regions ($r \geq 0.3$, Table 3.1, Supplementary Figure 3.3).

3.3.10 Waist-to-hip ratio and smoking contributed to IL-6 associations with DNAm and variables of interest

Previous associations have been found between smoking status and waist-to-hip ratio (WHR) with all three of my main effects: SES (430,431), serum IL-6 levels (428,429), and DNAm (426,427). As such, these associations were first tested for replication in this cohort, and secondly, as to whether they mediated the relationships discovered between IL-6 and DNAm and/or SES and DNAm. I found that serum IL-6 levels were significantly associated with WHR ($F(4,321) = 40.1$, $p = 1.31 \times 10^{-5}$, Supplementary Figure 3.4A) and smoking status ($F(1,309) = 6.2$, $p = 0.013$, Supplementary Figure 3.4B). Of the 80 validated CpGs associated with IL-6, 17 were found to be mediated by WHR and 15 by smoking with percentage contribution ranging between -19% to 20% and -16% to 37%, respectively (Supplementary Table 3.5A). Of the 15 CpGs associated with both IL-6 and SES with DNAm, nine were impacted by smoking by more than 10% and two were by WHR (Supplementary Table 3.6). The trajectory of SES was found to be significantly associated with both WHR ($F(3,325) = 4.452$, $p = 0.0044$, Supplementary Figure 3.4C) and smoking status ($X^2(3, N = 319) = 17.06$, $p = 0.0007$, Supplemental Figure 3.4D). Of the unique 51 CpGs found to be associated with SES trajectory, essentially none were mediated by WHR or smoking (Supplemental Table 3.5B).

3.4 Discussion

We investigated the physiological impacts of lifetime SES trajectory and IL-6 levels through the lens of DNAm, a relatively stable molecular mark that in part resides at the interface of the environment and genetics. Integrating multiple mechanistic avenues, such as inflammation and epigenetic markers, may help to elucidate the relationship between low SES and possible negative health outcomes observed throughout the literature (12,15,151,313,314). In the current study, this was addressed by measuring genome-wide monocyte DNAm and serum IL-6 from a community cohort of healthy individuals with varying lifetime trajectories of SES. SES was measured retrospectively in childhood and concurrently in adulthood with participants selected to ensure a balanced design between the four trajectories of high and low SES in early and

current life to aid in avoiding confounding. DNAm associated with serum IL-6 displayed a trend of negative correlation, reflective of previously reported impacts of IL-6 on global DNAm (409–411). Amongst IL-6 associated CpGs, 19% were also associated with SES trajectory, with many influenced by smoking status but not waist-to-hip ratio. CpGs associated with IL-6 were frequently located within genes involved in the innate immune response, while CpGs associated with SES independently of IL-6 were located in adaptive immune genes. Changes in methylation within SES associated CpGs were not evenly distributed across the four trajectories – the majority of identified CpGs were differentially methylated in comparisons with Low/High SES individuals. This study highlights that while there was overlap in DNAm signal related to both SES and IL-6, unique associations were more abundant. Additionally, at least in this cohort, varying life course trajectories of SES had a disproportionate association with DNAm, highlighting the need for more nuanced investigations into SES and adverse health outcomes in the context of DNAm.

3.4.1 SES upward mobility drove DNAm associations independent of IL-6 and lifestyle behaviours

SES is associated with numerous adverse health outcomes, many of which have a distinct pro-inflammatory phenotype at the molecular level (5,15,145,167,376,386–388). Specifically, those who experienced low SES tend to have elevated serum levels of inflammatory cytokines such as IL-6 and CRP, both also associated to adverse health outcomes (163,433–437). In addition to this overlap of health outcomes, negative correlations between serum IL-6 and SES have been established for childhood (166,438,439) and adulthood SES (285,289,438,440,441), in addition to SES trajectories (376), findings replicated in the current study. Specific trends of SES and IL-6 were observed in this study based on the balanced study design across the four trajectories, showing stronger associations in order of increasing mean levels of serum IL-6: High/High, Low/High, High/Low, Low/Low (Early SES/Current SES) with the significance primarily related to differences between the High/High and Low/Low groups. Those with low current SES displayed the highest levels of IL-6 regardless of early-life status. These findings differ from previous research associating early-life SES with IL-6 levels in adulthood controlling for current SES (166,442). However, this result is also supported by previous work finding that childhood SES influences IL-6 levels within the context of recent life events (443). This

incongruency may be the result of cohort differences, such as SES measurement and population demographics. Concordant amongst this research is the reproducible association between low SES with increases in serum IL-6, regardless of which time point low SES is experienced.

DNAm has been previously reported to associated with inflammatory cytokines, such as IL-6, supporting the hypothesis of it possibly being at the nexus of SES and inflammation. Many types of associations between DNAm and IL-6 have been reported, two of which, IL-6 suppressing *DNMT1* expression (409–411) and DNAm changes in relation to serum levels of IL-6 (398), were also observed in this study. There was an overall trend of decreasing methylation with increasing IL-6 levels amongst IL-6 associated CpGs, making it tempting to speculate that this skewed distribution may result from the suppression of *DNMT1* expression by IL-6 as previously reported (409–411). Consistent with these data, I observed an inverse relationship between IL-6 and *DNMT1* mRNA transcript levels in monocytes from the CareGiver cohort, supporting IL-6 mediated down-regulation of *DNMT1* gene transcription. However, while there is evidence in this sample for potential global effects of *DNMT1*, only half of the 15 IL-6 and SES associated CpGs displayed a negative correlation between IL-6 and DNAm levels and only three CpGs were driven by the Low/Low group, which has the highest levels of IL-6, suggesting the relationship between these three variables is not driven by global methylation effects of IL-6 and instead is more specific.

While IL-6 is one of the most influential cytokines, it is only one of numerous immune system players, therefore it is likely that there are other biological effects to investigate in the remaining CpGs that did not associate with IL-6 levels. Amongst these associations, all comparisons containing the Low/High SES trajectory had 2-fold more associations than the other groups. At a more nuanced, and perhaps speculative level, these findings mirror recent research suggesting that individuals belonging to this Low/High group had better mental health outcomes, but worse metabolic health relative to their peers who remained in poverty (323). It is also worth noting that previous research investigating epigenetic age in this cohort found the greatest epigenetic age acceleration amongst the Low/High SES trajectory (227). Additionally, increases in epigenetic age acceleration was found amongst low-SES youth who had higher self-control, a measure associated with higher education and SES later in life (235).

Both SES and IL-6 measures are associated with demographic factors such as WHR (429,430) and smoking (428,431), contributing to the multifaceted impact low SES can have on

health making interpretations of findings difficult. Smoking increases expression of the detoxification protein Aryl hydrocarbon receptor (AhR) which results in reducing inflammation in part by directly suppressing *IL6* transcription (163,444,445). Considering this interconnected relationship, it is understandable that CpGs significantly associated with IL-6 were mediated by smoking status and WHR. WHR associations are likely due to the positive correlations reported between adiposity and IL-6, as adipocytes are also a significant source of serum IL-6 (429,446–448). When assessing the contribution of these variables to IL-6- and SES-associated CpGs, most were impacted by smoking status but few by WHR**.** In contrast to IL-6 related CpGs, changes in DNAm associated with SES trajectory were not mediated by WHR or smoking status. As this EWAS was performed on the vast majority of genomic CpGs not associated with IL-6, these findings suggested associations between SES and WHR or smoking may potentially relate to DNAm primarily through pro-inflammation. Though modest associations, these findings were based on the integration of several measures along the hypothesized SES-pro-inflammatory pathway suggesting that SES associations with IL-6 levels may be affecting innate immune function in part through changes in DNAm.

3.4.2 IL-6 independent CpGs associated with SES within adaptive immune genes

Site-specific associations between DNAm, IL-6 and SES were investigated to allow for a more complex and nuanced look at underlying genetic pathways potentially involved. Two categories amongst the genes containing many of the differentially methylated CpGs were found, those with immune or neurological function. As DNAm was measured in purified monocytes in this study, it was reassuring that differential DNAm occurred within genes related to immune pathways.

SES trajectory contributed to the relationship between serum IL-6 and DNAm levels amongst 15 CpGs which were located in 12 genes, with roughly half being involved in immune function such as B-cell activation (*PRKCB*), HIV infection (*TNPO3*) and MHC function (*GLO1*). Methylation of three of these CpGs were negatively associated with expression of the related gene (*PRKCB*, *TNPO3* and *MPRIP*), driven specifically by the High/Low SES group who tended to have higher DNAm levels. These relationships were independent of life style behaviours, with only *MPRIP* being impacted by smoking status.

When characterizing SES-associated CpGs, several were in genes with various immune functions such as immune cell maturation (*EBF4*, *TMEM176B*), homing (*HAS1*) and MHC function (*HLA-B*, *ZNRF2*). Though monocytes primarily function in the innate arm of the immune response, they also have roles in the adaptive pathways where several SES associated CpGs were located in. Adaptive immunity has also been implicated in the relationship between health disparities and SES. Previous studies found that low SES individuals had higher seropositivity for cytomegalovirus (CMV), a proxy for immune health (150). Additionally, a large meta-analysis of over 300 empirical articles found the source and duration of the stressor impacted several arms of the immune system, where acute stressors suppressed cell-mediated immunity, part of the adaptive system, and chronic stressors suppressed both innate and cell-mediated immune functions (157). Interestingly, two SES associated CpGs were located in Notch signaling (*NCOR2*, *HEYL*) genes, important for adaptive immune function (449,450). Additionally, several were in genes that function in the major histocompatibility complex (MHC) class I mediated antigen presentation (*ZNRF2, FBXO27, HLA-B, LTBP4*), necessary for triggering the cell-mediated immune response to infection (451,452). Summarily, while IL-6 related and independent CpGs, associated with SES, were located in genes with innate immune function, adaptive immune genes were found mostly in IL-6 independent sites. These differences highlight the more nuanced impact SES could have on immune function, beyond IL-6 signaling, an important consideration for future research into this paradigm.

3.4.3 CpGs associated with SES were located within genes of neurological function

SES, across the life course, has been associated to mental health and behavioural outcomes such as depression, anxiety, cognitive function and working memory (453). It is also known that IL-6 can cross the blood-brain barrier (163) and has been associated to neurological disorders (411,454) suggesting a possible link between SES, inflammation and brain DNAm. Genes related to neurological function annotated to the 15 IL-6 and SES associated CpGs were related to learning (*CALN1*), pre-synaptic signaling (*NRXN2*) and the stress response (*LINGO3*) (319,455). The CpG located in *NRXN2* was documented as being correlated between peripheral blood and Brodmann 10 area (BA10) within the prefrontal cortex of the brain ($r \geq 0.3$) in an independent dataset (38). Interestingly, methylation of three IL-6 associated CpGs located in *NRXN1* were also positively correlated between blood and BA10. Previous research has

associated dysregulation of *NRXN1* with ASD (456–458) and elevated IL-6 levels (454). Additional caution needs to be taken with these findings, however, as smoking status contributed to more than 10% in all of these relationships, making the driver of these associations difficult to untangle.

When observing IL-6 independent, SES associated CpGs, six were related to neurological function with four previously reported to show correlation of DNAm levels between brain and blood ($r \geq 0.3$) highlighting the possibility to discover potentially biologically meaningful findings (38). One was located in the promoter of *DDO*, where blood DNAm was previously reported to be positively correlated with BA10 and BA7 (38) and negatively correlated to gene expression with decreased methylation amongst those with the Low/High SES trajectory. Interestingly, D-aspartate oxidase acts as an agonist to the NMDA receptor in the brain where decreased promoter methylation resulted in increased expression which were previously associated with improved age-related neurodegeneration in mice (459). Methylation of another CpG located in *SLC6A5* had a positive correlation to gene expression and negative correlation between methylation of blood and BA10, this relationship also driven by decreased methylation amongst the Low/High SES group. This gene encodes for the glycine transporter 2 protein integral to the transport and storage of glycine, and thus has been found to shape inhibitory neurotransmission in the central nervous system (460). Differences in levels of glycine transporter 2 protein have been found to be crucial considerations in the treatment of pain due to its relationship with NMDA receptor activity (461–465). While it is interesting to find SES, IL-6 and DNAm associations within neurological-related genes, it must be emphasized that these findings were documented in purified monocytes. However, external resources can be leveraged to help support these relationships, such as comparing the correlation between DNAm measured in matched blood and brain samples (38). Overall, these findings highlight a possible link between SES and inflammation with potential influence in neurological pathways.

3.4.4 Considerations and limitations

It is important to highlight that all relations found between DNAm, IL-6 and SES in this study were associations, without any evidence for causation. Determining causality requires *in vitro* experiments where IL-6 or site-specific DNAm changes could be introduced and responses measured. One of the main functions of DNAm is in regulating gene expression, making it

tempting to assume differences in DNAm signals between groups result in changes in expression. However, studies have found relatively low correlation between expression and DNAm (40,466–469), reflected in this current study. There may be many reasons for this, one of which is the regulation of gene expression involves poising for a specific environmental response, without this there is no transcript present to correlate. Additionally, the stability of mRNA is highly variable and dependent on transcript function.

While investigating DNAm in purified tissues is helpful to understand where a given association is coming from, the molecular markers used to purify and define immune cells, including monocyte subtypes, are constantly evolving. CD14 and CD16 markers can be used in combination to further divide monocytes into classical ($CD14^+CD16^-$), non-classical ($CD14^-CD16^+$), and intermediary ($CD14^+CD16^+$) subtypes (470,471). Emerging research has shown that non-classical and intermediary monocytes are associated with inflammatory diseases such as rheumatoid arthritis (470,472,473). As such, while measuring DNAm in relevant purified cell lineages is an important step in understanding underlying mechanisms, more granular cellular subtypes should be taken into consideration for future studies, possibly in conjunction with single cell analyses.

Another limitation related to differences in epigenetic profiles across cell types is how to interpret DNAm changes in a given tissue when the mechanism under investigation primarily occurs in a completely different body system. For example, while there are some CpGs with concordant DNAm levels between blood and brain tissue, many do not display this pattern (38,59,474). While in the current study I focused on the immune related implications of SES throughout life however, there is a plethora of research highlighting the significant impacts of SES on psychological health as well (313). While brain tissue for this cohort is not available, there are online resources, such as the BECon application (38), which provide insight into the likelihood of peripheral blood reporting on DNAm in the brain. In this study, I found that three out of four IL-6-associated CpGs located in *NRXN1* were correlated by more than 0.3 between whole blood and Brodmann Area 10 (BA10), a region including the ventromedial prefrontal cortex, in the brain. This area, in particular, has been associated with processing contextual judgments about the self and others (475). These findings are alluring, but because no psychological measures were assessed, correlations are between brain and whole blood (not purified monocytes), and most processes of interest occur in a tissue which was not measured,

caution must be taken in interpretation. These results are presented with the aim of informing future research to confidently and appropriately design a study able to answer these queries.

Additional limitations to consider are that no genetics were measured in this cohort even though we know genetic background can also influence DNAm levels (40) and while this study controls for reported ethnicity, genetic measures would better account for population stratification. DNAm can also change throughout development (68), and while SES throughout the life course was investigated, this study was conducted cross-sectionally with early life SES measures taken retrospectively and DNAm and expression measured only in adulthood.

Investigating several aspects of the hypothesized SES-immune function pathway simultaneously allows for an understanding of how lifetime SES experiences could get "under the skin" (15) and elucidates pathways for further investigation. This study suggested associations amongst SES, serum IL-6 levels, DNAm, and SES throughout the life course. CpGs associated with IL-6 and whose effect was modulated by SES trajectory were located in immune and neurological gene pathways. We also found DNAm differences independent from serum IL-6 levels that associated with the lifetime trajectory of SES an individual experienced, with an abundant signature present amongst individuals with an upward mobility of SES throughout their life. This highlights the benefit of measuring SES throughout life to help reveal relationships of vulnerable trajectories often lost in the health research milieu. Together these findings provide evidence of relations between SES at both the molecular and physiological level with implications for pro-inflammatory phenotypes and later life health. Overall, they highlight the importance of integrating multiple biological measures when evaluating social variables such as SES, and address the first steps to understanding its potentially multifaceted role in precipitating diseased states. Multidimensional approaches to understand unique SES trajectories will provide more targeted solutions to alleviate health disparities.

4 Exposure to childhood abuse is associated with human sperm DNA methylation

4.1 Background and Rationale

Childhood abuse is associated with adverse mental and physical health outcomes across the life course (7,476,477). Childhood abuse has also been associated with altered function of multiple biological systems (8–11), with differences persisting into adulthood (12,13). Changes in epigenetic marks have been proposed as a mechanism by which childhood abuse increases risk of neuropsychiatric and cardiometabolic disease (478,479). Differences in epigenetic marks have been found in DNA methylation (DNAm) of blood (203,205), saliva (206), and brain tissue (480) by experience of childhood abuse (30). The association of childhood abuse with DNAm in gametes is of particular interest, both because the patterns of DNAm in gametes have been associated with fertility (481,482) and the possibility that gamete DNAm may affect the healthy development of the offspring (483,484).

In animal models, a variety of exposures has been shown to affect sperm DNAm, including nutritional status (485), endocrine-disrupting hormones (486), and other pollutants (487). Animal experiments have also indicated that paternal stressors can affect DNAm (118), gene expression (488,489), and behavior (117,118) in the offspring. In mice, exposure to social instability early in life leads to anxiety and defective social interactions, behaviors that are transmitted to three generations of offspring through the paternal line, though this work remains controversial (490). Transmission of paternal experiences of psychological trauma through gametes has also been documented (489) and corresponds with alterations in paternal sperm DNAm (118).

To our knowledge, no studies in humans have examined the effects of psychosocial stressors on sperm DNAm; however, psychological stress in humans has been associated with poorer semen quality, including lower motile sperm concentration, lower percentage of progressively motile sperm, and reduced lateral head displacement (491,492). Evidence suggests that environmental exposures such as cigarette smoke (493) and health status indicators, such as age (494) and obesity (495,496), are associated with sperm epigenetics in humans. Additionally, relevant to abuse in childhood, the pre-pubertal period has been identified as a potential window of sensitivity of the sperm epigenome to environmental influences (497). Thus, it is possible that

psychosocial stressors, including childhood abuse, affect the human sperm epigenome, including DNAm.

In the present study, we assessed the differences in genome-wide sperm DNAm in association with childhood abuse in a healthy longitudinal cohort of men. We calculated the principal components (PCs) of methylation values for all probes and examined the association of childhood abuse with PCs. DNAm sites typically function in concert with neighboring sites to affect gene expression (498); thus, it may be more meaningful to investigate DNAm within genomic regions as opposed to at individual sites. We therefore examined differentially methylated regions (DMRs) for association with childhood abuse. Finally, we used machine learning methods in order to identify sites indicative of childhood abuse from all sites and construct a parsimonious indicator of child abuse status. As childhood abuse has been associated with higher prevalence of adulthood health risk behaviors (499,500), mental disorders (501,502), and trauma exposure (503,504), we conducted exploratory analyses to examine whether body mass index (BMI), smoking, depressive symptoms, posttraumatic stress symptoms, and trauma exposure accounted for a possible association of childhood abuse with sperm DNAm.

4.2 Materials and Methods

4.2.1 Sample

The Growing Up Today Study (GUTS) is a US longitudinal cohort of 16,882 offspring of women participating in the Nurses' Health Study II, enrolled in 1996 at ages 9–14 years and followed annually or biennially (505). In 2010, male participants were asked whether they would be willing to donate a semen sample. Nearly two-thirds (64%) were willing. Age, BMI, and race did not differ between men willing and unwilling to donate. In 2012, we contacted 66 men to request a sample; 54 men (82%) returned the sample by mail. We further invited the first 28 men who returned the sample to send a second one; 24 men (86%) returned the second sample. Men were asked to abstain from ejaculation for at least 48 h prior to producing the sample by masturbation into a collection container (Thermo Scientific Nalgene Jars). Samples were shipped overnight, with four gel refrigerant packs surrounding the sample, to the Massachusetts General Hospital Fertility Center where sperm concentration and morphology were measured. Remaining semen was aliquoted and flash frozen in liquid nitrogen. Informed consent was obtained from all participants. The Institutional Review Board of Partners Healthcare approved this study.

We conducted DNAm assays on 48 samples from 34 men. Of these, 20 men contributed single samples, 12 men contributed two samples each, produced ~three months apart, and two men's samples were assayed twice as technical replicates, for a total of 48 samples assayed. We oversampled men who had suffered high levels of abuse, such that the samples that were assayed included 17 men who suffered high, five men medium, and 12 men no childhood abuse.

4.2.2 Measures

Experiences of physical, emotional, and sexual abuse before age 18 were reported in 2007 when participants were aged 18–23 years. Physical and emotional abuse were measured with four items from the Childhood Trauma Questionnaire (CTQ), querying frequency that an adult in the family yelled, insulted, punished cruelly, and hit so hard that it left bruises (506). Responses to the CTQ were summed (506) and then divided into quartiles based on their distribution in the entire cohort (lowest quartile = 0 points, highest quartile = 3 points). Physical and emotional abuse were also measured with three items from the Conflict Tactics Scales (CTS), querying frequency that an adult in the family shoved; threatened to punch, kick, or hit with something; actually punched, kicked, or hit with something; or physically attacked (507). Response options for the CTQ and the CTS ranged from "never" to "very often". Responses to the CTS were skewed, with most respondents reporting none of these experiences. We therefore divided this scale into 0: lowest 50%, 1: next 25%, and 2: highest 25%.

Sexual abuse was queried in each time period with two questions regarding unwanted sexual experiences with an adult or older child (e.g., "Did an adult or an older child force you into any sexual activity by threatening you or hurting you in some way?") (507). Response options included: no; once; or > once.

To oversample men exposed to high levels of abuse, we created an overall measure of childhood abuse in three levels: none, moderate, and high. Men with "no abuse" ($N = 12$) were in the lowest category of both measures of physical and emotional abuse and had not experienced sexual abuse. Respondents with "high abuse" ($N = 17$) were either in the highest level of the CTS or the highest level of the CTQ, or had a mixture of elevated responses across both questionnaires. All or nearly all men in this group had experienced punishments that seemed cruel, were yelled and screamed at, and had hurtful and insulting things said to them. All had been shoved, grabbed, hit, or physically attacked in some other way, and most had also been

threatened with violence. Two men in this group had been sexually abused. Five participants fell between the "no abuse" and "high abuse" groups and were considered to have experienced "medium" abuse. We also summed the CTQ, CTS, and sexual abuse measures to create a continuous measure of abuse severity (range, 0–7) and dichotomized participants as none-to-medium (0–2) vs. high abuse (7–11).

4.2.3 Covariates

We examined the characteristics of the semen sample, including ejaculate volume, sperm concentration, percent normal morphology, collection date, collection time, and abstinence interval, as well as characteristics of the participant, including age at collection, month of birth, and race/ethnicity as possible covariates. Additionally, we included information reported by the participants' mothers, Nurses' Health Study II cohort members, regarding her ancestry as well as participants' childhood socioeconomic status, an index of family income, maternal social standing, and paternal education, reported in 1999–2001.

4.2.4 Hypothesized mediators

Childhood abuse increases risk for adulthood health risk behaviors, mental disorders (508,509), and trauma exposure (504), factors that may explain an association of childhood abuse with adulthood sperm DNAm. We examined smoking, BMI (by self-report in 2010 and 2007), depressive symptoms (measured with the Center for Epidemiologic Studies Depression Scale-10 (510) in 2010), posttraumatic stress symptoms (measured with the 7-item Short Screening Scale for DSM-PTSD (511) in 2007), and trauma exposure (measured in 2007 with 13 items adapted from the Brief Trauma Questionnaire (512) e.g., physical assault, intimate partner violence, and serious illness) as potential mediators.

4.2.5 DNAm assay

A differential lysis method involving a series of six washes was performed to separate sperm cells from epithelial and round cells. We then conducted DNAm assays with Infinium HumanMethylation450 (450 K) BeadChips (Illumina) using bisulfite-treated DNA (EZ-96 DNA Methylation kit, Zymo Research, Irvine, CA). These assays produce 485,577 data points encompassing 482,421 CpG and 3091 non-CpG (CpN) sites. Raw intensity scores were color

corrected and background was subtracted using GenomeStudio Software (Illumina). Methylation β value for each probe represents a continuous ratio between 0 (0% methylated DNA molecules) and 1 (100% methylated DNA molecules). Probes were excluded from further analysis if they had a detection p-value < 0.01 (n = 2144 probes) or if > 5% of samples were missing a β value (n = 12,353 probes). Probes which bound in silico to the X and Y chromosome in addition to the specified targets were excluded (348), leaving N = 439,746 probes available for subsequent analysis. Inter-sample normalization was performed using quantile normalization (345). To account for the two probe types on the Illumina BeadChip, normalization was performed using subset-quantile within array normalization (SWAN) (513). To determine if there were batch effects, PCA was performed on the normalized data followed by Spearman's correlations of the PCs with all technical variables. A slight batch effect associated with chip number and position was removed using empirical Bayes methods (R package SVA, ComBat function, Supplementary Figure 4.1) (350).

To evaluate the purity of our washed sperm samples, we compared DNAm in our sample with DNAm from an independent study of contaminated and purified sperm samples (Gene Expression Omnibus (GEO) (514) GSE108058, Supplementary Figure 4.2). We merged the GEO dataset with our own data and performed PCA. The vast majority of variation in methylation is associated with tissue heterogeneity, therefore the first few PCs should be correlated with the purity of the semen samples. Plotting PC1 against PC2 (for visualization purposes), our samples clustered with the pure semen, providing evidence that we had successfully purified our samples. We additionally examined the methylation status of two imprinting control regions (*HYMAI* and *GNAS-AS*). These regions are paternally expressed, and therefore we would anticipate that these regions would be fully unmethylated if our samples contained purified haploid gametes (as opposed to hemi-methylated in somatic tissue). We calculated the median DNAm β value for each probe underlying these regions (130 probes) for each sample in our study. The vast majority of samples had median β < 0.05, suggesting good purity (Supplementary Table 4.1, Supplementary Figure 4.3).

4.2.6 Analyses

To characterize the study sample, we compared age, race, and abuse exposure of study participants with all GUTS men. Next, for study participants, we calculated prevalence for categorical variables and mean for continuous variables for covariates by childhood abuse status.

4.2.7 Principal components analysis

To investigate whether childhood abuse and our covariates were associated with variation in DNAm, we conducted PCA with all probes ($N = 439{,}746$) using one randomly selected sample per subject. PCA reduces the dimensionality of the data by identifying orthogonal components from methylation values of all individual probes, with PC1 explaining the most variance. We examined the association of both the continuous and categorical childhood abuse variables and the covariates with centered PCs, using one-way ANOVAs for ordinal and categorical variables and Spearman's correlations for continuous variables. For PCs that were statistically significantly associated with childhood abuse, we investigated which specific probes contributed most to the PC by first identifying individual probes with the largest PC scores (the 1% of probes with the largest positive scores and 1% with the most negative scores) and then, to increase the likelihood of biological relevance, selected only probes with methylation $\Delta\beta \geq 5\%$, where $\Delta\beta = \bar{\beta}_{\text{high abuse}} - \bar{\beta}_{\text{no abuse}}$. *P*-values were not adjusted for multiple testing, as this was an exploratory analysis to determine associations with DNAm.

4.2.8 DMRs analysis

We next investigated whether childhood abuse was associated with patterns of DNAm in spatially clustered probes. We investigated DMRs by childhood abuse exposure using the R package DMRcate (515) (Bioconductor, http://www.bioconductor.org), using the same randomly selected sample per subject used in the PCA. DMRcate first assesses the association of the exposure (childhood abuse) with methylation at each individual CpG site, then groups the probes into DMRs based on the similarity of effect size and directionality with distances of ≤ 1000 bp between them. DMRs are then corrected for multiple testing by calculating the false discovery rate (FDR) for each DMR. DMRs that do not meet an FDR ≤ 0.05 and a fold change ≥ 0.05 are dropped. We considered regions to be DMRs if they were statistically significant at an FDR ≤ 0.05, contained ≥ 3 probes, and had a difference in DNAm β ($\Delta\beta$) $\geq 5\%$. We conducted

these analyses with the ordinal childhood abuse variable to reduce the effects of outliers, then checked that results were similar in analyses using the continuous childhood abuse variable. We verified our findings by replacing the sample used in the primary analyses with the replicate sample from each man who contributed two samples ($N = 12$) and re-running the DMR analyses using original samples from 22 men and replicate samples from 12 men. Finally, in sites located in identified DMRs, we calculated the interclass correlation coefficient (ICC) between the first and second sample in DNAm β values.

To examine the concordance of our two methods of identifying probes differentially methylated by childhood abuse, we compared the overlap in probes identified using PCA and probes identified in DMR analysis.

4.2.9 Machine learning analysis

Finally, we used machine learning to identify sites predictive of childhood abuse and: (1) compare them with the sites identified in the DMR analysis and (2) construct a parsimonious predictor of child abuse status. We fit a penalized linear regression ("elastic net") to select informative probes from the set of all probes using the dichotomized childhood abuse variable (none/medium vs. high abuse, mixing parameter α set to 0.5, the default). The penalized regression begins by fitting a single linear model including all probes, then selects a subset of relevant probes by shrinking the linear coefficients and setting to zero coefficients below a given threshold (516). The selected probes are those with non-zero coefficients. We estimated the penalty parameter λ with tenfold cross-validation and set it to 0.095. We applied the resulting predictor to three independent datasets (Gene Expression Omnibus (514) GSE108058, GSE102970 (517), and GSE64096 (518)**)** to ascertain whether the prevalence of abuse estimated with this predictor was approximately the same as the prevalence in the whole GUTS cohort (high abuse prevalence = 30%). As no datasets of sperm DNAm were available with childhood abuse measured, we could not test its ability to predict abuse status.

4.2.10 Pyrosequencing methylation confirmation

To confirm findings from the 450 K array, we performed pyrosequencing with bisulfite-converted DNA. We selected five sites for confirmation, prioritizing sites within DMRs and sites with a low FDR. We calculated Spearman correlations between β values obtained from

pyrosequencing and the 450 K array and performed linear regression to ascertain the association of pyrosequencing β values with childhood abuse.

4.2.11 Exploratory mediation analysis

To examine whether adulthood health risk factors might explain a possible association between childhood abuse and DNAm, we conducted two analyses. First, we examined whether these risk factors loaded on DNAm PCs, using one-way ANOVAs for ordinal variables and Spearman's correlations for continuous variables. Next, we examined probes identified in DMR analyses. For each probe in a childhood abuse DMR, we compared the association of childhood abuse with DNAm in linear models adjusted only for age and semen volume (base model) and in models further adjusted for: (1) health risk behaviors (smoking and BMI); (2) mental health (depressive and posttraumatic stress symptoms); and (3) trauma exposure. We calculated % mediation as:

$$[(\beta_{\text{child abuse, base model}} - \beta_{\text{child abuse, adjusted model}})/\beta_{\text{child abuse, base model}}]*100$$

for each probe and calculated the mean mediation across all probes within each DMR for each set of hypothesized mediators. We did not include all hypothesized mediators in a single model to avoid overfitting.

4.2.12 Probes associated with childhood abuse in prior studies

We examined the association of 1,667 probes previously identified as associated with childhood abuse (206,479–481). We considered probes with FDR < 0.05 as statistically significant, accounting for multiple testing within this set of 1,667 probes.

4.3 Results

4.3.1 Sample

Study participants were similar to all GUTS participants in age (participants, mean = 25.7 years, range = 23–29 years; GUTS, mean = 25.8 years, range = 23–31 years) and race/ethnicity (participants, 91.2% white; GUTS, 93.2% white), and had a higher prevalence of high levels of childhood abuse (participants, no abuse = 35.3%, high = 50%; GUTS, no abuse = 26.3%, high = 28.8%). Characteristics of study participants and semen samples were similar across levels of childhood abuse (all $p > 0.05$, Table 4.1).

		Experience of childhood abuse		
		None (*N* = 12)	**Medium (*N* = 5)**	**High (*N* = 17)**
Covariates				
Age, years	Mean (range)	26 (24-28)	25 (23-27)	25 (23-29)
Race/ethnicity				
White	% (*N*)	92 (11)	100 (5)	88 (15)
Nonwhite	% (*N*)	8.3 (1)	0.0 (0)	12 (2)
Maternal ancestry				
Scandinavian	% (*N*)	0 (0)	20 (1)	12 (2)
Southern European	% (*N*)	42 (5)	20 (1)	18 (3)
Other Caucasian	% (*N*)	83 (10)	60 (3)	71 (12)
Hispanic	% (*N*)	0 (0)	0 (0)	5.9 (1)
Childhood socioeconomic status	Mean (SD)	7.3 (1.5)	7.6 (1.5)	7.0 (1.9)
Semen volume, ml	Mean (SD)	2.5 (1.3)	3.9 (1.7)	2.7 (1.8)
Sperm concentration, M/ml	Mean (SD)	56 (27)	56 (19)	53 (29)
Normal sperm morphology	% (SD)	7.8 (3.5)	7.8 (5.5)	6.5 (2.9)
Collection time, morning	% (*N*)	92 (11)	60 (3)	65 (11)
Abstinence time, hours	Mean (SD)	93 (18)	97 (22)	83 (12)
Hypothesized mediators				
Smoking				
Current	% (*N*)	8.3 (1)	0 (0)	24 (4)
Past	% (*N*)	17 (2)	20 (1)	5.9 (1)
BMI	Mean (SD)	24 (3.2)	24 (2.9)	24 (4.7)
Depressive symptoms	Mean (SD)	5.7 (4.7)	5.5 (5.6)	7.7 (5.0)
Posttraumatic stress symptoms	Mean (SD)	1.2 (0.3)	1.3 (0.5)	2.0 (1.2)
Traumatic events	Mean (SD)	0.2 (0.4)	1.6 (1.5)	1.4 (1.7)

Table 4.1 Participant and semen sample characteristics by experience of childhood abuse (N = 34).
SD, standard deviation. Maternal ancestry by maternal self-report in 1989. Ancestry percentages do not sum to 100, as women could endorse more than one ancestry. No mothers reported African, American, Asian, or "other" ancestry. Childhood socioeconomic status is an index of family income in 2001, paternal educational attainment in 1999, and maternal perceived social standing in the US in 2001. Normal sperm morphology ascertained according to the World Health Organization (519).

4.3.2 Principal components analysis

PC4 was correlated with childhood abuse (Spearman's correlation $p \leq 0.05$) and explained 6.2% of the variation in DNAm (Figure 4.1). Participant's age was also correlated with PC4 and adjusted in DMR analyses. To identify probes that were both strongly associated with PC4 and were related to childhood abuse exposure, we selected probes with the largest PC4 scores ($N = 8{,}795$) and then from these selected probes with DNAm $\Delta\beta \geq 5\%$ between high and

no abuse, resulting in over 1000 probes ($N = 1{,}137$, Supplementary Table 4.2). The two men who had experienced sexual abuse were not outliers amongst men who experienced abuse (Supplementary Figure 4.4).

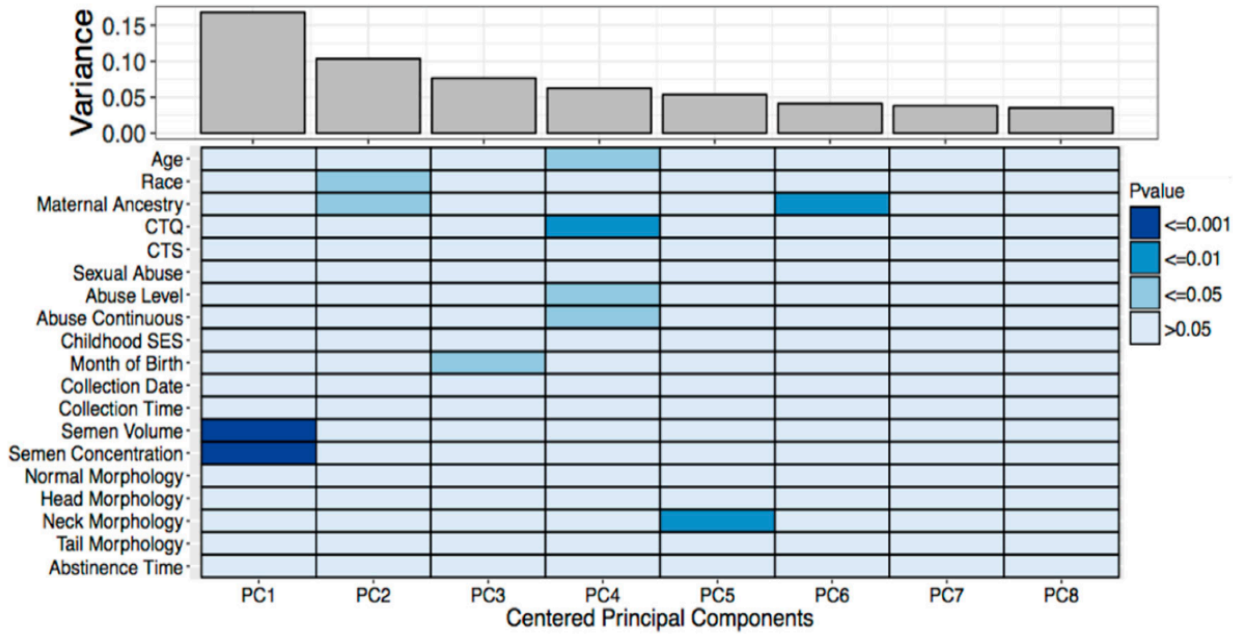


Figure 4.1 Principal component 4 (PC4) was associated with childhood abuse exposure (one sample per participant, N = 34).
PC4, representing 6.24% of the variance present in the methylation data, was significantly correlated ($p < 0.05$) with childhood abuse exposure. Darker regions signify stronger correlations between variables and principal components (N probes = 439,746). Normal sperm morphology is characterized, beginning at the head and moving toward the tail. Thus, "head morphology" is the % of sperm in a sample with normal heads, "neck morphology" is the % with normal heads and necks, and "tail morphology" is % with normal head, neck, and tail. Abstinence time is the time between the sperm donation and the most recent preceding ejaculation. PC principal component, CTQ Childhood Trauma Questionnaire, CTS Conflict Tactic Scale.

4.3.3 DMRs analysis

We identified 13 DMRs meeting our criteria: (1) FDR ≤ 0.05; (2) mean $\Delta\beta \geq 5\%$; and (3) contained ≥ 3 probes. Of these 13 DMRs, 12 met these three criteria in analyses using original samples from 22 men and replicate samples from 12 men ($N = 34$). These 12 DMRs contained 64 probes (Table 4.2, Figure 4.2, and Supplementary Figure 4.5), three of which were located in enhancers, two in transcription start sites, six in CpG islands, and three in gene bodies (Supplementary Table 4.3). The ICC between replicate samples ($N = 12$) for the 63 CpG sites comprising these 12 DMRs was greater than 0.7 for 90% of sites (Supplementary Figure 4.6). Results were similar with childhood abuse coded as a continuous variable.

Sites identified in the DMR analysis overlapped considerably with sites identified in the PC analysis. Thirty-five of the 63 CpG sites in the DMRs were amongst the sites loading most strongly on PC4.

Cluster name	Number of significant probes	*p*-value	FDR	Average Δβ	Max Δβ
ARL17A	3	1.54E-10	2.43E-07	−0.29	−0.35
MAPT	8	7.66E-10	7.99E-07	0.132	0.173
CLU	11	9.82E-05	1.04E-02	0.08	0.139
LRRK1	3	1.03E-17	1.19E-13	0.103	0.12
PRDM16	7	4.13E-05	6.95E-03	0.094	0.148
TCERG1L	3	1.60E-04	2.26E-02	0.131	0.147
CFAP46	5	2.09E-04	2.61E-02	−0.108	−0.122
MIR5093	4	2.52E-07	1.49E-04	0.108	0.128
TAF1B	3	6.47E-05	1.19E-02	0.148	0.194
DLL1	5	4.13E-05	8.52E-03	0.115	0.135
SYCE1	3	1.14E-09	1.31E-06	0.083	0.114
NDFUA10	3	1.60E-06	6.80E-04	0.119	0.138
SDK1	8	1.60E-04	1.93E-02	−0.091	−0.12

Statistically significant DMRs were discovered using DMRcate (FDR ≤ 0.05), had a mean Δβ ≥ 5%, and were verified using replicates. *P*-value, FDR, and mean Δβ for each DMR are the mean across all probes within the DMR. Δβ values were calculated as the difference between the mean β for high and no childhood abuse.

Table 4.2 Differentially methylated regions (DMRs) associated with childhood abuse exposure.

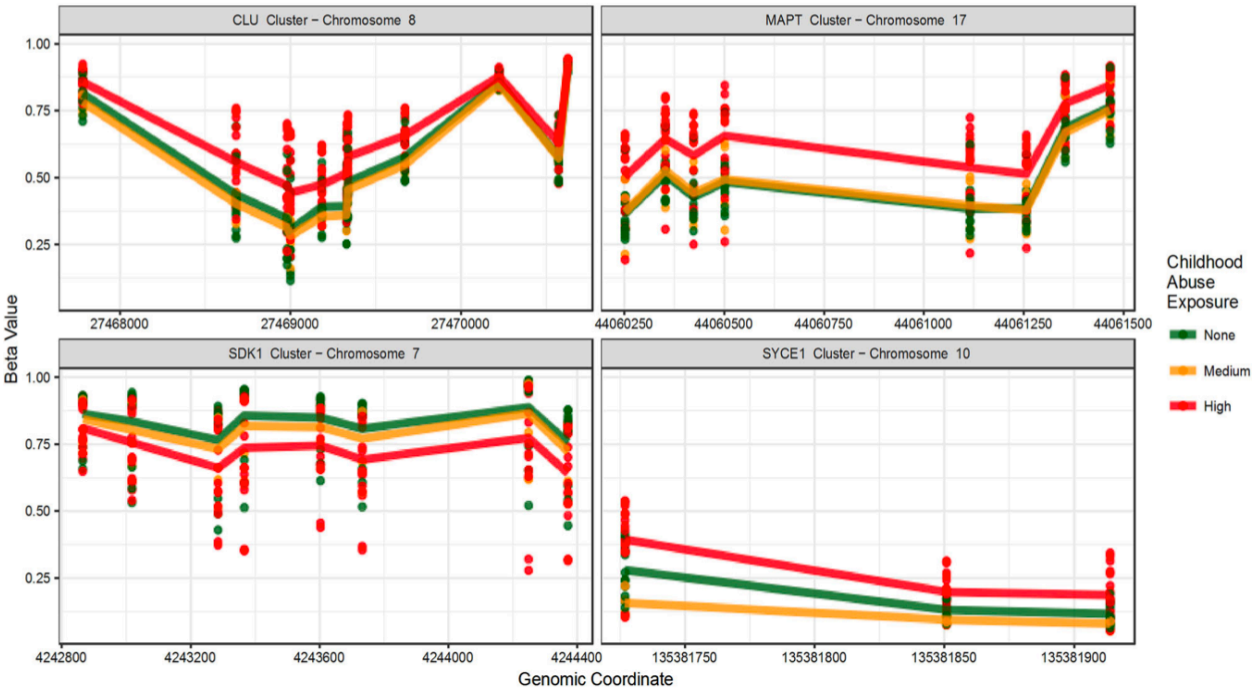


Figure 4.2 Four genomic regions differentially methylated by childhood abuse.
Differentially methylated regions (DMRs) were defined as regions that differed statistically by abuse exposure at an FDR ≤ 0.05, had a mean $\Delta\beta \geq 5\%$ across probes, and were confirmed using replicates. The "CLU cluster" includes the 5' UTR transcription start site and part of the gene body spanning 2.8 kb. The "MAPT cluster" is located in the gene body and spans 1.2 kb. The "SDK1 cluster" is located in the gene body and spans 1.5 kb. The "SYCE1 cluster" is located in the 5' UTR and spans 200 bp.

4.3.4 Pyrosequencing

For pyrosequencing confirmation of 450K array results, we selected four CpG sites contained in childhood abuse DMRs: the *ARL17A* (cg04703951), the *MAPT* (cg00438222) and the *LRRK1* (cg09926099 and cg00293616) clusters, and one additional site (cg08780220) based on its low FDR. All sites had significantly high correlations between measurements obtained by 450K and pyrosequencing (Spearman's rank $\rho \geq 0.74$, p-value $\leq 4.0 \times 10^{-7}$, Supplementary Figure 4.7), and were significantly associated with childhood abuse in linear regressions after correction for multiple testing (Supplementary Figure 4.8, Supplementary Table 4.4). The pyrosequencing assay for cg04703951 additionally measured DNAm at four CpG sites not represented on the 450K array. These four additional sites were highly correlated with neighboring sites measured on the 450K array ($\rho \geq 0.88$) and differed significantly by childhood abuse (p-value $\leq 3.9 \times 10^{-8}$, Figure 4.3, Supplementary Table 4.4).

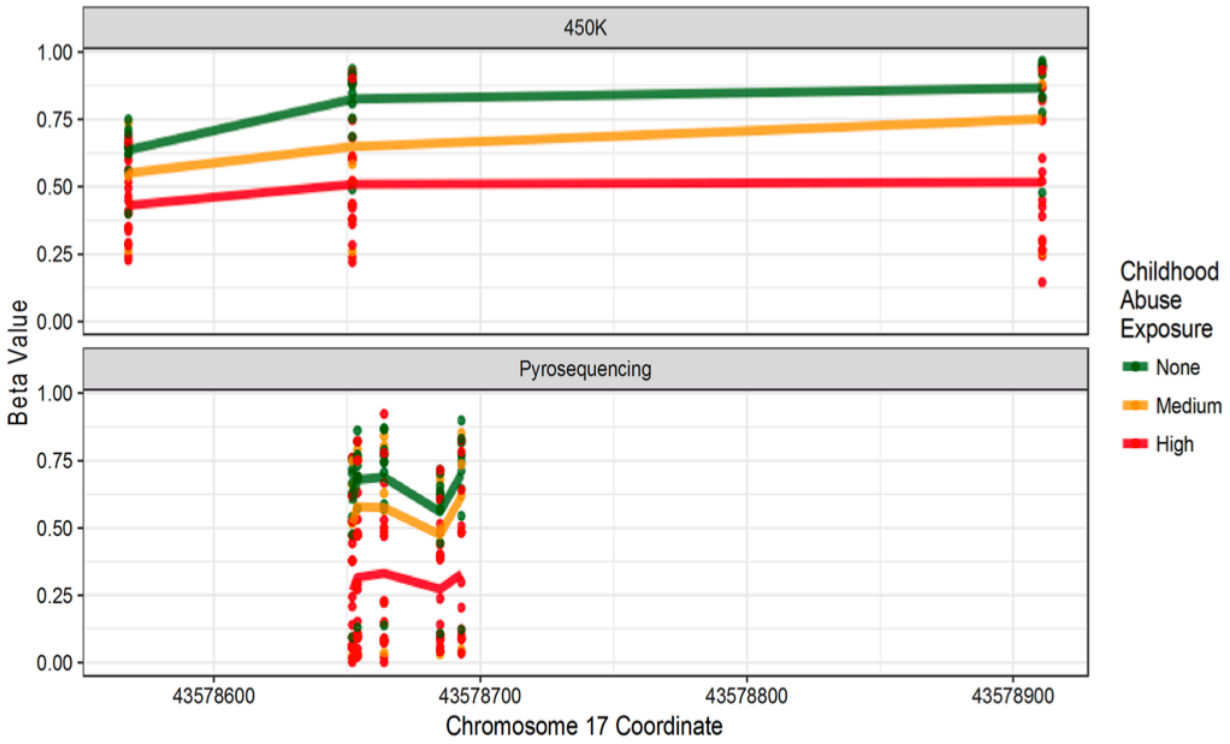


Figure 4.3 Additional sites measured during pyrosequencing of "ARL17A cluster" correlated significantly with 450K sites in relation to childhood abuse.
The "ARL17A cluster" found using DMRcate is located 30 kb away from ARL17A and spans 344 bp. The 450K methylation measurements of site cg04703951 (top panel) was confirmed using pyrosequencing techniques (bottom panel). The pyrosequencing assay measured DNAm at four additional sites not represented on the 450K array (bottom panel). SD - standard deviation, DMR - differentially methylated region, FDR - false discovery rate.

4.3.5 Machine learning analyses

The machine learning approach identified three probes (cg02622647, cg04703951, and cg17369694) as most useful for classifying participants as none or medium vs. high abuse exposure. These probes correctly classified 71% of participants (12 true positives, five false positives, 15 true negatives, and two false negatives). Two of these three probes were also identified in the DMR and PC analyses (cg02622647 and cg04703951, *ARL17A* cluster), showing the concordance of these methods. In three independent datasets (NCBI GEO accession numbers GSE108058 (514), GSE102970 (517), and GSE64096 (518)) this three-probe predictor predicted abuse prevalence of 30%, 35%, and 25%, respectively, similar to the 29% found in the GUTS cohort.

4.3.6 Mediation analyses

None of our hypothesized mediators was associated with PC4, the PC that was significantly associated with childhood abuse. For probes in child abuse DMRs, the association of childhood abuse with DNAm was somewhat attenuated in models also including depressive and posttraumatic stress symptoms (two of 12 DMRs, mean mediation = 11.2 and 13.6%) and in models including lifetime trauma exposure (four of 12 DMRs, mean mediation range = 14.0–23.7%), but not in models including smoking and BMI (mean mediation < 5.7% for all DMRs). The association of childhood abuse with DNAm was somewhat *stronger* after adjustment for mental health in two DMRs (*DLL1* and *SYCE1*) and after adjustment for lifetime trauma in three DMRs (*MAPT*, *DLL1*, and *NDFUA10*, Supplementary Tables 4.5-4.7).

We did not find a statistically significant association of childhood abuse with any of the candidate probes identified in prior studies of childhood abuse, likely due to these EWASs being conducted in other tissues (30,206,479,480).

4.4 Discussion

Childhood abuse has been associated with alterations to multiple biological systems in adulthood (13), and several studies have found differences in DNAm in somatic tissue by childhood abuse (30). We examined whether childhood abuse was associated with sperm DNAm in adulthood and found evidence that sperm DNAm varies by experiences of childhood abuse. The three approaches we used to identify differences in DNAm associated with childhood abuse, PCA, DMR analysis, and machine learning, found significantly overlapping sites. Moreover, pyrosequencing assays identified additional sites proximate to and correlated with sites measured by the 450K array that were also differentially methylated by childhood abuse. Together these findings suggest that our results reflect differences in DNAm associated with abuse.

Several DMRs we identified were located within genes, although it is unknown whether these specific sites are associated with future gene expression or, if so, whether the $\Delta\beta$ we found by abuse level has biological significance. Clusterin (CLU) is an extracellular molecular chaperone expressed in the brain and embryonic tissues that responds to stress conditions and has been implicated in neurodegenerative disorders, including Alzheimer's and Parkinson's disease (520). Additionally, Clusterin RNA transcripts pass from human sperm to the oocyte at fertilization (521). MAPT is hypothesized to be involved in neuronal migration and in

establishing neuronal polarity (522) and has been implicated in neuroticism (523) and neurodegenerative disorders. PRDM16 is a transcriptional regulator involved in the regulation of fat cells (524,525). *SDK1* encodes a protein in the immunoglobulin superfamily (522). Thus, DMRs were found on genes coding for proteins with a variety of functions, consistent with the documented effects of childhood abuse on the brain, body weight, and immune system. The DMRs we found did not overlap with prior DMRs identified in a study of paternal sperm and offspring symptoms of autism spectrum disorder (526) nor with probes in brain tissue, saliva, and peripheral blood identified in prior studies of childhood maltreatment (30,206,479,480).

We found that higher trauma exposure and higher prevalence of depressive and posttraumatic stress symptoms in men who experienced childhood abuse compared with men who did not accounted for some of the association between childhood abuse and sperm DNAm in five DMRs. Childhood abuse and other types of traumatic events have common biological effects, e.g., on the HPA-axis (527,528) and systemic inflammation (529), thus it is plausible that abuse and other trauma types share effects on DNAm as well (530,531). However, the association of childhood abuse with DNAm was also stronger in five DMRs after further adjustment for mental health and lifetime trauma.

Our findings should be considered in light of important limitations. First, our sample size was small. Therefore, our identification of DMRs associated with childhood abuse should be interpreted cautiously and be used primarily as a starting point for further research. Due to our small sample size, our examination of adulthood sequelae of childhood abuse that might mediate a relation between abuse and DNAm must be considered exploratory. Second, our sample was predominantly white, thus, our findings may not apply to men of other races.

Animal studies have indicated that psychosocial stressors can affect both epigenetic patterns in sperm and offspring phenotype. Male mice conditioned to odor-related fear exhibited differences in sperm DNAm at a locus related to the odor receptor. These fear-conditioned mice produced two generations of offspring with the same odor-related fear response as well as corresponding alterations to neuronal structures, results that were robust to cross-fostering and *in vitro* fertilization (118). Mice exposed to chronic stressors showed a greater concentration of nine sperm micro RNAs (miRNAs) and HPA-axis alterations in offspring (489). Injection of these nine miRNAs in zygotes produced similarly altered HPA-axis function, suggesting a causal role for the miRNAs in offspring biology (532). In another experiment, mice exposed early in

life to unpredictable maternal separation had altered patterns of small noncoding RNAs (sncRNAs) in sperm and had offspring with behavioral differences compared with control offspring. Injection of RNA from sperm into fertilized oocytes reproduced these behavioral differences (117). Thus, robust experiments have indicated that stressors may affect murine sperm epigenetics, including DNAm, and offspring biology.

Evidence that psychosocial stressors affect human sperm epigenetics remains limited. To our knowledge, our study is the first to document an association of psychosocial stressors and sperm epigenetics in humans. Indirect evidence that stressors could affect sperm epigenetics in humans is suggested by studies finding reduced sperm quality in men exposed to psychosocial stressors (491,533,534), as well as the association of other kinds of environmental exposures with human sperm epigenetic patterns (493,496,535–537).

While most mammalian paternal epigenetic marks are erased at fertilization and again during preimplantation development (538), some loci are resistant to demethylation (115), and these preserved epigenetic marks may be biologically important. Several pieces of evidence suggest that human sperm epigenetics influences both fertility and embryogenesis: (1) sperm cells are transcriptionally silent yet have epigenetic marks characteristic of transcription (539,540); (2) sperm chromatin has patterns of histone modifications at loci related to embryo development (539,540); (3) sperm mRNA, produced prior to transcriptional arrest, are transferred to the oocyte (521); and (4) sperm epigenetic marks are associated with fertility (540).

Childhood abuse greatly precedes the time period in which the ejaculated sperm were dividing and maturing, thus could not directly affect sperm DNAm at this stage. Instead, childhood exposures may affect the epigenome of spermatogonia, which then gets propagated during spermatogenesis in adulthood (497,541). Additionally, our results suggest that childhood abuse may lead to adulthood exposures that affect the sperm epigenome during spermatogenesis (487). Regardless of their origin, it is tempting to speculate that these DNAm marks are somehow propagated to the offspring. However, research in human developmental biology has not yet provided strong evidence for this possibility (540). Moreover, we note that offspring inherit the material from a single sperm, for which each CpG site is either methylated or unmethylated. If differences in DNAm associated with child abuse render affected sperm less likely to fertilize an egg, then the potential impact of these changes on offspring would likewise

be reduced. Studies in humans have documented adverse neurodevelopmental outcomes in offspring of persons exposed to severe psychosocial stressors, in particular, to childhood abuse (500,542–549). The hypothesis that the experience of stress may affect offspring through the parental epigenome has been raised as a potential mechanism for these associations (550,551). While this possibility is intriguing, molecular evidence from human germ cells remains sparse. Our results recommend further consideration of this promising hypothesis.

5 Conclusion

5.1 Dissertation summary and overarching conclusions

This dissertation represents a body of work aimed at analyzing the complex aspects of how adversity may associate with genome-wide DNAm throughout life in the context of markers of stress and inflammation, encompassing a comprehensive set of analyses interrogating unknown or problematic aspects of social epigenetic studies. One key issue in the field of social epidemiology and epigenetics is that of variable inconsistency (370,552). Adversity can be experienced in many ways, with SES being a common measure investigated. However, the definition of SES varies across disciplines and societies and typically encompasses a myriad of sub-categories, which could each have unique and differential impacts on human health and social determinants (326,552).

In **Chapter 2**, I analyzed a richly phenotyped Costa Rican cohort (n = 482) to attempt to disentangle four sub-categories of SES by determining if these measures exhibited differential associations with whole blood DNAm. EWASs were performed on retrospective childhood and concurrent older adulthood measures of wealth (household assets), education (self and maternal) and self-reported economic standing, in addition to a composite SES variable. Post hoc sensitivity analyses were done to determine how much of the relationship could be explained by inflammatory (CRP) and stress (epinephrine, norepinephrine, DHEAS, cortisol) biomarkers, and lifestyle factors (BMI, drinking, smoking). Dissecting SES sub-categories, the self-report measure behaved distinctly and could be interpreted as dampened relative to more objective SES measures. Temporally, most of the effect was seen with older adulthood SES measures, while a large proportion of the childhood-associated sites was confounded by stress and inflammatory biomarkers.

In the sensitivity analysis, where childhood measures were highly impacted and older adulthood measures were not, the composite variable combining these opposing trends displayed neither enrichment nor depletion in CpGs impacted. This finding, that childhood SES measures were especially influenced by stress and inflammation biomarkers, supports the developmental origins of health and disease (DOHaD) hypothesis that early life environments set the stage for the metabolic and physiological phenotype in later life (18,324). Notably, my work highlights the difference in how subjective versus objective measures associate with DNAm. Additionally,

while composite measures aim to encompass many complex attributes of SES, there may be cases where null results occur because of opposing trends of sub-categories cancelling out signal.

When investigating how a molecular mark, such as DNAm, associates with a relatively distal variable, such as a social environment, it may be important to incorporate other measures that may be involved in the pathway of effect to fully unravel what is happening (313). One hypothesis on how SES is associated with negative health disparities later in life is that experiences of poverty can equate to experiences of chronic stress, which itself can result in dysregulation of important body systems (12). Specifically, immune system dysregulation has been a focus, as many of the diseases linked to poverty have a proinflammatory phenotype, with affected individuals commonly having increased basal cytokine levels, such as IL-6 (295).

To investigate the link between these three variables, I investigated purified monocyte DNAm amongst a cohort (n = 333) balanced for having varying trajectories of SES throughout life (Early/Current Life: Low/Low, Low/High, High/Low, High/High). Initially, I found there was a subset of CpGs associated with serum IL-6 levels, which were validated in an independent cohort. Amongst these sites, a sensitivity analysis was done incorporating SES trajectories to determine if there was any relationship between SES, DNAm, and IL-6. An EWAS was also performed on all CpGs not associated with IL-6 to determine any independent associations between DNAm and SES trajectory alone. CpGs associated with serum IL-6 levels, where the effect was partially explained by SES trajectory, were located in extracellular and intracellular signaling integration genes. Differences amongst the CpG sites associated with SES trajectory, but not IL-6, were driven predominantly by the individuals who experienced low early life SES but high adulthood SES. This finding is in line with previous work where increased metabolic syndrome was experienced amongst minorities who grew up in poverty even though they ended up with high adulthood SES, relative to their peers who remained in poverty throughout life (323).

Finally, there has been substantive interest into the idea that adverse experiences could be transgenerationally inherited through epigenetic mechanisms (113,114,117,118). As described in the introduction, proving this has been extraordinarily difficult, and studies in humans are still in their infancy. However, if this truly is a potential mechanism of inheritance, then it stands to reason one should be able to detect signatures of adversity in the gametes, the cells responsible for producing future offspring. As such, I interrogated DNAm in the sperm of adult men who had

experienced a range of physical and sexual abuse in childhood (204). This was a small, pilot, proof-of-concept study using a sub-sample from the GUTS cohort (n = 34). I compared DNAm amongst those who experienced high, medium, and low childhood abuse and found 12 regions with differential methylation between the groups. These regions were located in genes with neuronal function (*CLU*, *MAPT*), immune function (*SDK1*) and fat regulation (*PRDM16*). The delta betas (Δβs) amongst these regions were large and verified using repeat samples taken from a subset of the individuals months apart, supporting that the signal detected was not highly variable across this period of time. Pyrosequencing was performed on a subset of regions to verify that differences were not platform specific, where I was able to interrogate additional CpGs not measured on the array and concluded that these sites also trended significantly in the same direction, adding confidence that some of these DMRs may be true findings. Finally, to ensure there was no confounding between adversity experienced in early life and in adulthood, a sensitivity analysis was performed to account for lifetime trauma experienced, depression, and PTSD amongst other related variables. While there were regions where the signal was confounded by these additional variables, most were not impacted, with some resulting in an even stronger associations once current adversity measures were accounted for.

While this was a pilot study and no concrete conclusions should be made regarding our specific findings, it is encouraging that a signal was picked up within such a small cohort, that these signals were stable over several months, and that additional non-array CpGs measured on another platform were comparably different between the groups. All of this suggests the potential for a DNAm signal of childhood adversity to be detected decades later in the gametes of adults, which I believe warrants further investigation in larger cohorts with deep phenotyping and appropriate study designs to address legitimate concerns pertaining to this research.

The common thread woven throughout the studies contained in this dissertation is in understanding how adversity, in all of its complexity and context-dependent effects, can potentially associate with DNAm to help answer how these experiences can get “under the skin”, resulting in health disparities in later life (15,195). This was accomplished by taking a deeper look into some of the key issues that may prevent much of this research from being replicated, how specific aspects of SES relate to DNAm, how these associations could be influenced by proinflammatory cytokines, and finally, how far this adversity signal can be biologically carried.

Firstly, breaking down the components of SES and dissecting how each associate with DNAm, as I did in **Chapter 2**, reflects concerns already reported in the social epidemiology field: how replication is low in large part because of the inconsistency in how SES is defined across studies (326,370). I was able to show that DNAm associations with SES differed based on life stage and whether the measure was subjective or objective which may help guide future research and study design. **Chapter 3** took a closer look at incorporating additional measures hypothesized to be involved in the pathway between SES and altered DNAm to help better understand what these associations are biologically representing (12). I was able to show that, while there was a subset of CpGs associated to both SES and IL-6, more CpGs had independent associations with SES, highlighting that IL-6 levels contribute to only a part of this relationship. Finally, **Chapter 4** delves into the exciting, yet controversial space of transgenerational inheritance. While not addressing if this mode of inheritance is an actual phenomenon in humans, the presence of a DNAm signal of childhood adversity in the gametes of adults is an intriguing find. This signal was present in most of the regions tested even when accounting for several adulthood measures that could reasonably confound childhood abuse exposures, though this was done in a very small pilot study and needs to be replicated in a much larger cohort to make any definitive statements (113,114,117,118).

5.2 Limitations and considerations for social epigenetics

While the field is acutely aware of the importance of accounting for potential confounders, many early foundational studies in social epigenetics did not account for these differences and are primarily correlational, observational designs. Additionally, due to the difficulty, cost, and technological limitations in conducting epigenome-wide association studies (EWAS), much of the initial work focused on candidate gene approaches. The previous literature on DNAm in candidate genes should not be disregarded; however, new appreciation of the interconnectedness between sites of DNAm indicates the benefit of EWASs due to accounting for differences amongst many correlated sites or regions simultaneously (193). While there is crucial work to be done in understanding how social environments affect the underlying biology of developmental sequelae, a healthy dose of skepticism and a critical eye must be maintained both when evaluating past and current literature, as well as developing new experimental designs

(36). Below, I will highlight some of the major limitations with social epigenetic studies in addition to some suggestions on how to potentially circumvent these issues.

5.2.1 The reproducibility problem

Reproducibility of findings in an independent cohort is the gold standard in EWASs. Many findings have been reproduced in the EWAS space, such as the CpGs associated with aging in *ELOVL2* (553,554) and smoking in *AHRR* (555). Social epigenetics, studying associations of DNAm and various social environments such as SES, acculturation, abuse, and parenting styles, have particularly suffered from the problem of failure of replication (296,318). There is no canonical gene or region one can cite as being a consistently reproduced, differentially methylated CpG associated with a given social variable. Most replicated studies were candidate analyses using an *a priori* hypothesis assumption resulting in a significantly smaller testing space and lowering the burden of multiple testing. These hypotheses commonly involve investigating stress and neurological related genes, such as those in the HPA-axis (e.g., *NR3C1* and *FKBP5*), neuroendocrine (*OXTR)*, neurotransmitter (*SLC6A4*), and neurotrophin (*BDNF*) pathways (178,190,556–558). However, little to no such replication has been found amongst CpGs discovered from unsupervised EWASs conducted on similar variables, including CpGs within commonly investigated candidate genes (296,318). Aside from individual CpG sites, this includes general trends of those studies which support the DOHaD hypothesis showing early life SES is more associated to DNAm despite current SES status and those with findings similar to **Chapter 2** where current life SES is more highly related to DNAm in adulthood. There are many factors that can produce a lack of reproducibility, including differing demographics of cohorts, variable measurements, tissue sources and statistical analyses, which I will discuss in more detail below.

5.2.2 Statistical power

5.2.2.1 Sample sizes

Sample sizes for EWASs have always been an issue (421,559). Similar to issues plaguing genome-wide association studies (GWASs), EWAS power is influenced by population stratification, testing space, effect, sample size and study design. However, as this epigenetics field evolves, it is evident that there are several additional factors, such as tissue heterogeneity

and environmental influence on DNAm that further complicate power in EWASs. Statistical power impacts the ability of an analysis to detect an effect, which means the smaller the sample size, the more likely one is to miss true positives, increasing type II error. This is likely one of the major reasons why there is a lack of reproducibility for many EWASs. Where the effect is small, larger sample sizes are required to detect differences, resulting in only the strongest DNAm associations being reproducible. The only other way to increase statistical power without increasing sample size is to relax statistical thresholds for significance, though this increases the risk of discovering false positives, inflating type I error (560).

Regardless of the issues, it is clear EWASs would greatly benefit from larger sample sizes, though obtaining more participants has significant practical limitations. The larger the sample size, the more funding and manpower is required, and often to obtain this level of funding, smaller cohort pilot studies are first required as proof of concept. While findings from smaller studies may eventually become obsolete, they are necessary to advance the field and ensure resources are not squandered for larger EWASs that are unlikely to be informative.

5.2.2.2 Effect sizes

A common issue encountered when trying to calculate sample sizes necessary to achieve appropriate statistical power is understanding what effect size to expect. A single CpG site can either be methylated or unmethylated, yet for sequencing and array measurements of DNAm, it is typical to obtain a continuous measure for a given site. This is because a sample contains a pool of DNA molecules, which can each carry a methylated or unmethylated state at a particular site. As such, beta values analyzed for DNAm array data is in a range of 0-1, which represents the proportion of methylated DNAm molecules in a given sample. The preferred effect sizes of these measures are the change in the pooled DNAm for a site between groups, frequently termed delta betas (Δβs). These are calculated by subtracting the mean of beta values between groups, if categorical, or by multiplying the regression coefficient by the variables range, if continuous. Incorporating Δβs in a given analysis provides more robust findings as it is not merely the statistical threshold but also an effect size threshold that is required to report a given association (421).

Unfortunately, for DNAm there are no clear-cut answers as to what effect sizes are biologically meaningful (421). One of the most reproducible EWAS findings, promoter

methylation in *ELOVL2* and age, has been commonly reported to have an effect size of only around 2% across a 10-year age range, a relatively small effect with no known function related to aging (68,553,554). As such, what effect size threshold to use has mostly been arbitrary and commonly reported as 5%. To avoid picking a biologically unjustified threshold, in this dissertation I utilized other criteria I think could be valuable to the community moving forward.

For cohorts with samples processed in our lab, we typically split samples with enough DNA and randomly distribute them across multiple batches. This practice allows us to obtain a collection of technical replicate samples from the same DNA source to compare during our pre-processing of the DNAm data. In **Chapter 3**, the root-mean-square error (RSME), the standard deviation of the residuals, between technical replicates was calculated across all beta values and rounded to the nearest percent, to be used as the effect size threshold. My scientific reasoning for using this measure for my effect size threshold is that significantly associated CpGs that differed in methylation by at least this value were likely not due to some random, unmeasured noise as the RMSE between technical replicates represents unaccounted for technical variation. While this solves the issue of picking an arbitrary threshold, it does not address what differences are biologically meaningful. As we further understand the complexity of how context-specific DNAm changes are as they relate to gene expression and function, we will be able to utilize this information to inform future sample size calculations and interpretations of findings.

5.2.3 Consistency in variable measurements

5.2.3.1 Type

As we are beginning to understand how interconnected the world is, research is becoming progressively more interdisciplinary. While important to uncover the intricate way everything is connected, interdisciplinary studies also compound the limitations and caveats unique to each field. Where EWASs have learned from the limitations of GWASs, so should social epigenetics take advantage of lessons from social epidemiology. One significant problem plaguing this field of research is the inconsistency with how complex variables are measured, an issue that is likely a significant contributor to the replication problems seen in social epigenetics (296,326,370,552). For example, SES in both **Chapters 2 & 3** were measured differently, in part due to the difference in cultural relevance of the characteristics, reflecting the diversity of how SES can be measured. For a variable to be consistent means that various definitions and measurements of it

will not result in a different effect on the outcome of interest (370). While there are many variables that have a clear definition such that the variation is minimal, for example chronological age or smoking status, many do not, especially in the field of social epigenetics.

SES is a prime example; it is a complex measure historically assessed in multiple ways and includes objective and subjective components. It has been commonly reported that there are different health associations related to objective outcomes (e.g., household wealth) compared with self-report measures (e.g., self-rated socioeconomic standing). Objective SES measures generally associate with more physical ailments, such as cardiovascular disease, while subjective measures correlate with mental health outcomes, like depression (361–364). Additionally, composite measures are commonly used to simultaneously capture many components related to the breadth of a variable (326). However, ensuring composite variables are consistent is even more challenging because the underlying assumptions are that every component encompassed must individually meet these criteria (370). In **Chapter 2**, I was able to show how different measures of SES varied in their associations with DNAm. In general, the subjective measures behaved either distinctly or with a diminished association to DNAm relative to objective measures. Though the SES composite measure was correlated with all other SES measures, it predominantly related to DNAm most similarly to the older adulthood objective measures. Seeing the variability of effect between different SES measures within the same sample highlights the issues likely plaguing replication across social epigenetic EWASs using different SES measures in different populations (296).

5.2.3.2 Timing

When in a participant's life variables are measured can impact both the outcome and the interpretation of a study. DNAm varies predictably throughout life (561), where DNAm decreases globally with increasing age, likely due to insufficient maintenance of DNAm deposition during subsequent cell divisions (562). Most DNAm changes occur during embryogenesis, where there is a global erasure of DNAm before it is re-established (563), an important consideration for interpreting gamete DNAm patterns, as was cautioned in **Chapter 4**. Though DNAm changes predictably throughout life, it does not occur at the same rate, with the first years of life appearing to have the largest changes and variation (564–566). This may be due to the existence of sensitive periods in early life – limited windows of receptivity during

development when an individual is particularly responsive to certain exposures, potentially resulting in adaptations that may be experienced through other life stages (15). While the role of epigenetics in sensitive period timing and responsivity has been theorized and is logical based on existing evidence, further work is needed to establish what the association is between known windows of developmental timing and long-term DNAm differences (567).

Additionally, EWASs investigating older adults need to consider the possibility of survivor bias, enrichment for healthy individuals who survived to old age, as findings may not be as generalizable to those most greatly affected by the variable of interest who are less likely to live as long (568). There was a range of developmental stages assessed in this dissertation: pre-conception (**Chapter 4**), mid-life (**Chapter 3**), and older adulthood (**Chapter 2**). Because of how dynamic DNAm is throughout the life course, even within this dissertation it is difficult to make cross-study comparisons, let alone in the field at large. The best way to control for these changes would be to obtain longitudinal samples across the life course and consider what age or ages are most appropriate to answer the research question (569–571).

5.2.3.3 Location

Investigating societal factors requires addressing the context of the society being assessed. Some socioeconomic differentials observed in high-income countries are not seen in several mid- to lower-income countries (330,572–577). Obesity and low physical activity, for example, is prevalent in low SES groups from high-income countries, but in many low-income countries, it is the high SES groups displaying this trend (572). This issue predominantly impacts objective measures, whereas subjective measures, by definition, are how an individual interprets their own experiences, making intersocietal comparisons more robust (330). Additionally, most research investigating social health disparities was done in predominantly western populations that will inevitably embed assumptions of Western values, diets and social structures, which may not be appropriate for non-Western populations (552). As such, care is needed to interpret findings related to social variables, with the goal that conclusions should reflect the societal context the population was sampled from.

Such considerations are mentioned in **Chapter 2**, as there are known differences amongst social health disparity relationships seen within Costa Ricans (330). One solution utilized to help compare across different societies is in the construction of composite variables where a given

metric is weighted allowing it to be compared across populations. This was done for the larger CRELES cohort, from which the **Chapter 2** participants were sub-sampled, where the SES composite was produced to allow for comparisons with a US cohort (333). However, using these more generalizable composite measures can result in a loss of nuance, as was seen in **Chapter 2** where opposing trends of childhood and older adulthood measures were essentially cancelled out when observing the composite measure results in isolation. To help generalize findings, obtaining as diverse a dataset as possible is ideal, though more practically, producing weighted composite measures can address this concern.

5.2.4 Tissue

5.2.4.1 Cellular heterogeneity

In EWASs, many assayed tissues are heterogenous, comprised of multiple cell types that vary in proportion across individuals. While reporting differences in predicted proportions of cell types between groups could be an interesting finding in and of itself, many EWASs are interested in changes not associated with tissue composition. **Chapter 2** consisted of whole blood samples in older adults, thus the adult blood reference from the identifying optimal DNAm libraries (IDOL) package was used, and the Housman algorithm applied, to bioinformatically predict varying cell type proportions, to be incorporated in the linear models (64,354). **Chapter 3** investigated a relatively purified cell type, monocytes, which was confirmed after running the Houseman algorithm predicting monocytes to consist of ≥60% of the cellular population in every sample, and as such, no adjustments were needed. **Chapter 4** also assayed a purified cell type, spermatozoa, which is often contaminated with epithelial cells (578). To assess sample purity two analyses were performed: first, a publicly-available sperm methylation dataset (578) consisting of both contaminated and purified samples was combined with my samples and run through principal component analysis (PCA) to determine which group they clustered with. Second, utilizing knowledge of paternally imprinted genes and the haploid nature of sperm, *HYMAI* and *GNAS-AS* methylation levels were assessed to ensure that they were fully unmethylated (28).

When a reference dataset does not exist and a purified tissue cannot be obtained, there are several reference-free surrogate variable analysis-based (SVA) methods that can be used to help alleviate confounding due to heterogeneity (66,67). It is also important to note that reference

datasets are only as specific as the methods used to define the cell types, and even within the classes identified there are likely multiple subtypes of cells not accounted for. Not accounting for interindividual differences in cellular composition have resulted in confounding and spurious correlations, highlighting the necessity of adjusting for this variation in EWASs (55,56,61,579).

5.2.4.2 Tissue Source

Often social epigenetic studies focus on complex traits, such as mental health and disease, which can impact the function of many body systems and complicate EWAS interpretations. Most samples assayed for DNAm are collected from peripheral tissues such as blood, buccal epithelial cells, and saliva; yet, as DNAm variability is predominantly driven by cellular identity, DNAm profiles vary widely across tissue (41,46,57,358,580). Due to this, caution is needed in interpreting DNAm associations measured in a tissue that differs from the target tissue involved in the physiological processes being investigated. Therefore, it is imperative to address the issue of DNAm concordance between tissues to support any findings. There are some tools available to help with this, such as the Blood-Brain Epigenetic Concordance (BECon) tool (38), which was used in **Chapter 3** to show correlations of the methylation pattern between blood and Brodmann area 10 in a subset of IL-6 associated CpGs located in neurologically related genes. While these findings are supportive of the potential for the DNAm pattern to be present in brain tissue, the correlations were still assessed in a different cohort, in whole blood instead of purified monocytes, and contained large differences in overall cohort demographics (38).

The tissues selected for the studies reported in this dissertation were hypothesis driven. In **Chapter 4**, male gametes were assayed to specifically address the initial step along the route to investigating transgenerational epigenetic inheritance in association with early life adversity. In both **Chapters 2 and 3**, the driving hypothesis centered around how adversity results in negative health outcomes concerns dysregulation of the stress and immune systems. Stress hormones are systemically released into the blood stream, affecting immune responses (12), and since the immune system is largely comprised of cells residing in the blood, this tissue, measured in **Chapter 2**, was an appropriate target to assay. Another approach is to investigate more purified cell types, which avoids the confounding inherent in estimating and correcting for multiple cell types in one sample. This was the approach taken for **Chapter 3** where purified monocytes from

blood were assayed. Monocytes were chosen as the tissue of interest due to their integral role in the innate immune response, a key player in my hypothesis on DNAm as a link between adversity and negative health outcomes (12). Incorporating the biology inherent within the study's driving hypothesis is invaluable for determining which tissue would be most appropriate to assay.

5.2.5 Genotype

5.2.5.1 Population stratification

Variation in DNAm has been frequently associated with differences in ethnicity, which itself can represent a combination of differences in genetic ancestry, diet, and both social and physical environment (36,40,581–585). As such, the same considerations of population stratification accounted for in GWASs also needs to be accounted for in EWASs. Genetic differences in an individual can impact DNAm in several ways: there could be a difference in the genetic sequence at the CpG site, potentially making methylation at that site impossible in a population if cytosine was changed to a thymine. For example, there could be a polymorphic variant in one of the methyltransferase enzymes, resulting in global differences in DNAm, or a variant could alter a protein or a DNA binding site that impacts downstream molecular pathways, resulting in DNAm changes (96,586,587). It is, therefore, important to address the genetic background of individuals in EWASs.

Commonly, reported ethnicity is used as a covariate in EWAS, due to its non-invasiveness and relative ease to obtain. However, using this can lead to self-report error and a conflation of ethnic and racial identity with genetic ancestry, leading to false classification in epigenetic studies. In **Chapters 3 and 4**, reported ethnicity measures were used as genetic data were unavailable, likely resulting in adjustments removing variation associated with a combination of genetic and environmental differences. The best method to account for genetic variation in EWASs is to genotype participants and calculate genetic ancestry, using the same techniques employed by GWASs, and adjusting for it in subsequent analyses as was done in **Chapter 2**. When designing future studies, care needs to be taken both when selecting the population of interest and choosing the measures used to account for genetic differences. Furthermore, if reported ethnicity is used to adjust for differences, it should be acknowledged that this social construct represents more than just genetic admixture.

5.2.5.2 Methylation quantitative trait loci (mQTLs)

Even when accounting for population stratification, there can still be single nucleotide polymorphisms (SNPs) that associate with DNAm independent of genetic ancestry. mQTLs, SNPs associated with DNAm, can differ in association across tissues and age, making these regions of particular interest to investigate (41,588–590). When there is no genotype information for a given sample, there are publicly accessible mQTL databases in which CpGs of interest can be investigated to determine if they are associated with a SNP. Such was done in **Chapter 3** where I investigated the significant CpGs using the mQTLdb database, consisting of 1,000 longitudinally measured mother-child pairs from the ARIES cohort (590). Though none of the CpGs I found were reported to be mQTLs, this may be due to the ARIES cohort being assayed in blood, where **Chapter 3** measured DNAm in purified monocytes. Unfortunately, this is only an option for tissues that have such a database; in **Chapter 4** this analysis could not be done as there is no such sperm dataset. In **Chapter 2**, I performed a post hoc analysis amongst the top CpGs of interest using genomic data measured in the same sample to determine if there was any enrichment of mQTLs present amongst the comparisons. Even when genetic ancestry is accounted for, it is still advisable to address the interindividual differences in genotype amongst CpG sites of interest, considering the amount of variability contributed to DNAm by genetic background and gene-by-environment interactions.

5.2.6 Accounting for confounders

There is an important balance to strike between selecting a more homogenous population to reduce noise and confounding and obtaining as diverse a population as possible to increase the generalizability of findings (579). For example, it is known that age, sex, ethnicity and tissue heterogeneity all add a significant amount of variability to DNAm -- this is why they are typically accounted for in EWAS analyses (34,36,579). If a sample is fairly homogenous for these confounders, the chance of finding replicable associations increases as a result of reduced noise in the data, but this occurs at the cost of the generalizability of the findings. Any results discovered in a homogenous population might only be applicable to this very specific demographic. This was the case for **Chapter 4**, where I had over 90% Caucasian males aged 20-25 years. Having such a homogenous sample was likely the reason I was able to find regions of

differential methylation with the large effect sizes reported; however, these results cannot be applied to men of other ethnicities or ages, and so further studies need be conducted to determine how generalizable these findings are. In general, it is ideal to recruit as large and diverse samples as possible to reduce the chances of unintentional confounding. However, if there are limitations to sample size, the more homogenous the population, the better chance of finding true associations, even if they limit interpretability.

5.3 Future directions

Social epigenetic EWASs are expanding, in part because of the alluring mechanisms and hypotheses surrounding how adverse environments cause many health disparities. Sample sizes in studies are becoming increasingly larger, and study designs have been vastly improved since the field's inception. There has been an expansion of technologies available to improve upon both the number of DNAm sites interrogated and some limitations described above. For example, single-cell reduced representation bi-sulfite sequencing (scRRBS) and single-cell bi-sulfite sequencing (scBS-seq) studies are increasing with regularity, solving the issue of tissue heterogeneity, as only the absolute methylation status in a single cell is reported (591). While these technologies have not yet been widely applied in large cohorts, this will likely be the way of the future.

Part of the reason so little is known about the functional changes of DNAm at the majority of genomic sites is because of technical challenges involved in experimentally altering methylation at a single, specific site. While there are compounds available that can cause pleiotropic changes to global methylation states, such as cytidine analogues like 5-azacytidine that inhibit DNA methyltransferases (592,593), alterations are not targeted and so not useful to interrogate the impact of specific genomic regions. However, the advent of clustered regularly interspaced short palindromic repeats (CRISPR)/Cas9-based technology provide a means to specifically edit the epigenome (594). CRISPR/Cas9 technology can be appropriated to bind either DNMT3A to induce methylation or ten-eleven translocation (TET) to remove it at specific genomic regions, resulting in more unambiguous methylation changes and allowing for interrogation of CpGs of interest in the future (595). This will allow us to test our hypotheses and determine the impact of differential methylation at a particular site.

There is increasing interest in transgenerational inheritance, though it is exceedingly difficult to determine what DNAm associations are the result of inheritance or simply an artifact of accidental exposure or generationally shared early life environments, like behavioural differences in parenting (113,114). In order to best ensure these are not the true contributors to the associations, several generations of participants in multiple contexts would need to be sampled. The closest data to supporting this phenomenon in humans are found in studies investigating individuals prenatally exposed to famine during the Dutch Hunger Winter in 1944-45. Here studies compared exposed and unexposed sibling pairs and found differential methylation in the *IGF2, INSR* and *CPT1A* genes amongst the exposed individuals 60 years after the exposure (116,596). While exposure occurred prenatally, it still occurred directly to the individuals being tested while *in utero*. As EWASs are relatively new, occurring during a single generation thus far, it has not been possible to address this issue. However, as time passes and more data are collected from longitudinal cohorts, it will be possible to obtain at least three generations using cross-generational superstudies (552). The Growing Up Today Study (GUTS) is one such cohort, a subsample of which comprised the cohort in **Chapter 4**. GUTS participants are the offspring of the Nurses' Health Study II participants recruited by Brigham and Women's Hospital and Harvard School of Public Health with the intention of continuing the study for successive generations (505,597,598). Even if the methylomes of various generations are able to be compared, there are still challenges to overcome, but these studies will be able to answer some integral questions about how adaptive to the environment these mechanisms are.

Finally, as molecular mechanisms are being individually investigated over time, it is becoming increasingly apparent that none operates in a vacuum. DNAm is one of many epigenetic mechanisms, which are, in turn, only one of several cellular mechanisms by which the genome and environment interact. All of these mechanisms are pieces of a larger puzzle that eventually encompass the whole individual and their external environment. Analogous to the metaphor of the blind men touching an elephant, in order to properly assess the whole picture experts across disparate fields of molecular biology, amongst many others, need to join forces merging various layers of data together.

These efforts are being vigorously applied, with many 'omics being combined in analyses using machine learning and clustering techniques. In **Chapter 2** genotyping and DNAm were both assayed and analyzed together, strengthening the analysis by allowing me to potentially

identify CpGs of interest that were under genetic control. In **Chapter 3**, mRNA expression was measured in the validation cohort, which allowed me to assess the number of IL-6-associated CpGs correlated to gene expression. Additionally, these data allowed me to support the hypothesis that the observed global decrease in DNAm with increasing IL-6 levels may be due in part to decreased *DNMT1* expression, as has been previously reported (409–411). Though merging 'omics will likely be a fruitful and worthwhile endeavor, many of the caveats and limitations described above, such as power and effect size issues, are likely to be amplified with this approach and require appropriate consideration.

Until these exciting technologies become more commonplace, we can apply the considerations mentioned in the previous section to data and technology currently available to us, vastly improving social epigenetic research. One of the best ways to improve any EWAS is to procure as many samples as possible. With large enough samples, more diverse demographics can be incorporated without risk of losing signal. Where it is not possible to obtain large samples, increasing the homogeneity in the sample, though limiting the generalizability of the findings, will allow for increased detection of true findings. When possible, obtaining longitudinal measures of individuals adds strength to the study by making each participant their own control. Additionally, increasing collaborative relationships between researchers and establishing consortiums may be a way to obtain larger cohorts and ensure more consistency in variable measurements. Most importantly, thorough research into all aspects of the proposed hypothesis and measuring additional variables integral to the proposed pathway of effect, will bolster interpretations and add confidence to findings.

Overall, the field of social epigenetics is still in its infancy, and there is plenty of room to improve and grow. I believe the analyses currently being conducted, including the work presented here, build upon extant literature to create a foundation from which to support this evolution. There is a clear desire by both the academy and the public to understand the questions the field of social epigenetics aims to answer, further emphasizing the need for responsible development. With careful thought, consideration, and consultation with field experts, proper variable production and selection can result in more revealing and meaningful EWAS results, which is especially important for interdisciplinary work. While there is a substantial replication problem, there are also several ways to potentially alleviate these issues. I believe that with

appropriate considerations, we can get closer to unraveling how adversity can get "under the skin" and alter the trajectory of our health for decades to come.

Bibliography

1. Adler, N. E., Boyce, W. T., Chesney, M. A., Folkman, S. & Syme, S. L. Socioeconomic inequalities in health. No easy solution. *JAMA : the journal of the American Medical Association* **269,** 3140 3145 (1993).

2. Cohen, S., Janicki-Deverts, D., Chen, E. & Matthews, K. A. Childhood socioeconomic status and adult health: Childhood socioeconomic status and adult health. *Ann Ny Acad Sci* **1186,** 37–55 (2010).

3. Haan, M. N., Kaplan, G. A. & Syme, S. L. in *Socioeconomic Status and Health: Old Observations and New Thoughts* (eds. Bunker, Gomby & Kehrer) (Hentry J Kaiser Family Foundation, 1989).

4. Marmot, M. G., Kogevinas, M. & Elston, M. A. Social/Economic Status and Disease. *Annu Rev Publ Health* **8,** 111–135 (1987).

5. Ziol-Guest, K. M., Duncan, G. J., Kalil, A. & Boyce, W. T. Early childhood poverty, immune-mediated disease processes, and adult productivity. *P Natl Acad Sci Usa* **109 Suppl 2,** 17289–93 (2012).

6. Lorant, V. *et al.* Socioeconomic Inequalities in Depression: A Meta-Analysis. *Am J Epidemiol* **157,** 98–112 (2003).

7. Heim, C. & Nemeroff, C. B. The role of childhood trauma in the neurobiology of mood and anxiety disorders: preclinical and clinical studies. *Biol Psychiat* **49,** 1023–1039 (2001).

8. Danese, A. *et al.* Elevated Inflammation Levels in Depressed Adults With a History of Childhood Maltreatment. *Arch Gen Psychiat* **65,** 409–415 (2008).

9. Danese, A., Pariante, C. M., Caspi, A., Taylor, A. & Poulton, R. Childhood maltreatment predicts adult inflammation in a life-course study. *Proc National Acad Sci* **104,** 1319 1324 (2007).

10. Heim, C., Newport, D. J., Bonsall, R., Miller, A. H. & Nemeroff, C. B. Altered pituitary-adrenal axis responses to provocative challenge tests in adult survivors of childhood abuse. *Am J Psychiat* **158,** 575 581 (2001).

11. Heim, C. *et al.* Pituitary-adrenal and autonomic responses to stress in women after sexual and physical abuse in childhood. *JAMA : the journal of the American Medical Association* **284,** 592 597 (2000).

12. Miller, G. E., Chen, E. & Parker, K. J. Psychological stress in childhood and susceptibility to the chronic diseases of aging: moving toward a model of behavioral and biological mechanisms. *Psychol Bull* **137,** 959 997 (2011).

13. Bellis, M. D. D. & Zisk, A. The Biological Effects of Childhood Trauma. *Child Adol Psych Cl* **23,** 185 222 (2014).

14. Barnett, E. & Casper, M. A definition of "social environment." *Am J Public Health* **91,** 465a–4465 (2001).

15. Boyce, W. T. & Kobor, M. S. Development and the epigenome: the "synapse" of gene-environment interplay. *Developmental Sci* **18,** 1 23 (2015).

16. Mandy, M. & Nyirenda, M. Developmental Origins of Health and Disease: the relevance to developing nations. *Int Health* **10,** 66–70 (2018).

17. Suzuki, K. The developing world of DOHaD. *J Dev Orig Hlth Dis* **9,** 266–269 (2018).

18. Wadhwa, P. D., Buss, C., Entringer, S. & Swanson, J. M. Developmental Origins of Health and Disease: Brief History of the Approach and Current Focus on Epigenetic Mechanisms. *Semin Reprod Med* **27,** 358 368 (2009).

19. Meaney, M. J. in **35,** 59 94 (Springer, Cham, 2017).

20. Bird, A. DNA methylation patterns and epigenetic memory. *Gene Dev* **16,** 6 21 (2002).

21. Henikoff, S. & Greally, J. M. Epigenetics, cellular memory and gene regulation. *Curr Biol* **26,** R644 8 (2016).

22. Kim, M. & Costello, J. DNA methylation: an epigenetic mark of cellular memory. *Exp Mol Medicine* **49,** e322–e322 (2017).

23. Feil, R. & Fraga, M. F. Epigenetics and the environment: emerging patterns and implications. *Nat Rev Genet* **13,** 97 109 (2012).

24. Marsit, C. J. Influence of environmental exposure on human epigenetic regulation. *J Exp Biology* **218,** 71 79 (2015).

25. Marcho, C., Cui, W. & Mager, J. Epigenetic dynamics during preimplantation development. *Reproduction* **150,** R109 20 (2015).

26. Guo, H. *et al.* The DNA methylation landscape of human early embryos. *Nature* **511,** 606 610 (2014).

27. Smith, Z. D. *et al.* A unique regulatory phase of DNA methylation in the early mammalian embryo. *Nature* **484,** 339 344 (2012).

28. Robertson, K. D. DNA methylation and human disease. *Nat Rev Genet* **6,** 597 610 (2005).

29. Barouki, R. *et al.* Epigenetics as a mechanism linking developmental exposures to long-term toxicity. *Environ Int* **114,** 77 86 (2018).

30. Lutz, P.-E. & Turecki, G. DNA methylation and childhood maltreatment: from animal models to human studies. *Neuroscience* **264,** 142 156 (2014).

31. Godfrey, K. M., Lillycrop, K. A., Burdge, G. C., Gluckman, P. D. & Hanson, M. A. Epigenetic mechanisms and the mismatch concept of the developmental origins of health and disease. *Pediatr Res* **61,** 5R 10R (2007).

32. Morgan, H. D., Sutherland, H. G., Martin, D. I. & Whitelaw, E. Epigenetic inheritance at the agouti locus in the mouse. *Nat Genet* **23,** 314 318 (1999).

33. Wolff, G. L., Kodell, R. L., Moore, S. R. & Cooney, C. A. Maternal epigenetics and methyl supplements affect agouti gene expression in Avy/a mice. *FASEB journal : official publication of the Federation of American Societies for Experimental Biology* **12,** 949 957 (1998).

34. Michels, K. B. *et al.* Recommendations for the design and analysis of epigenome-wide association studies. *Nat Methods* **10,** 949 955 (2013).

35. Mill, J. & Heijmans, B. T. From promises to practical strategies in epigenetic epidemiology. *Nat Rev Genet* **14,** 585–594 (2013).

36. Jones, M. J., Moore, S. R. & Kobor, M. S. Principles and Challenges of Applying Epigenetic Epidemiology to Psychology. *Annu Rev Psychol* **69,** 459 485 (2018).

37. Jones, M. J., Goodman, S. J. & Kobor, M. S. DNA methylation and healthy human aging. *Aging Cell* **14,** 924 932 (2015).

38. Edgar, R. D., Jones, M. J., Meaney, M. J., Turecki, G. & Kobor, M. S. BECon: a tool for interpreting DNA methylation findings from blood in the context of brain. *Transl Psychiat* **7,** e1187 (2017).

39. Hannon, E., Lunnon, K., Schalkwyk, L. & Mill, J. Interindividual methylomic variation across blood, cortex, and cerebellum: implications for epigenetic studies of neurological and neuropsychiatric phenotypes. *Epigenetics* **10,** 1024 1032 (2015).

40. Husquin, L. T. *et al.* Exploring the genetic basis of human population differences in DNA methylation and their causal impact on immune gene regulation. *Genome Biol* **19,** 222 (2018).

41. Islam, S. A. *et al.* Integration of DNA methylation patterns and genetic variation in human pediatric tissues help inform EWAS design and interpretation. *Epigenet Chromatin* **12,** 1 (2019).

42. Lienert, F. *et al.* Identification of genetic elements that autonomously determine DNA methylation states. *Nat Genet* **43,** 1091–1097 (2011).

43. Turinsky, A. L., Butcher, D. T., Choufani, S., Weksberg, R. & Brudno, M. Don't brush off buccal data heterogeneity. *Epigenetics* **14,** 109–117 (2019).

44. Wagner, M. *et al.* Age-Dependent Levels of 5-Methyl-, 5-Hydroxymethyl-, and 5-Formylcytosine in Human and Mouse Brain Tissues. *Angewandte Chemie Int Ed* **54,** 12511–12514 (2015).

45. Yousefi, P. *et al.* Sex differences in DNA methylation assessed by 450 K BeadChip in newborns. *Bmc Genomics* **16,** 148 (2015).

46. Ziller, M. J. *et al.* Charting a dynamic DNA methylation landscape of the human genome. *Nature* **500,** 477–481 (2013).

47. Antequera, F. & Bird, A. Number of CpG islands and genes in human and mouse. *Proc National Acad Sci* **90,** 11995–11999 (1993).

48. Yang, A. S. *et al.* A simple method for estimating global DNA methylation using bisulfite PCR of repetitive DNA elements. *Nucleic Acids Res* **32,** e38 (2004).

49. Lyon, M. F. Sex chromatin and gene action in the mammalian X-chromosome. *American journal of human genetics* **14,** 135 148 (1962).

50. Ohhata, T. & Wutz, A. Reactivation of the inactive X chromosome in development and reprogramming. *Cell Mol Life Sci* **70,** 2443 2461 (2013).

51. Barlow, D. P. & Bartolomei, M. S. Genomic imprinting in mammals. *Csh Perspect Biol* **6,** a018382 a018382 (2014).

52. Baccarelli, A. & Bollati, V. Epigenetics and environmental chemicals. *Environ Health Persp* **21,** 243 251 (2009).

53. Smith, Z. D. & Meissner, A. DNA methylation: roles in mammalian development. *Nat Rev Genet* **14,** 204 220 (2013).

54. Leenen, F. A. D., Muller, C. P. & Turner, J. D. DNA methylation: conducting the orchestra from exposure to phenotype? *Clin Epigenetics* **8,** 92 (2016).

55. Lappalainen, T. & Greally, J. M. Associating cellular epigenetic models with human phenotypes. *Nat Rev Genet* **18,** 441 451 (2017).

56. Michels, K. B. & Binder, A. M. Considerations for Design and Analysis of DNA Methylation Studies. *Methods Mol Biology Clifton N J* **1708,** 31 46 (2018).

57. Farré, P. *et al.* Concordant and discordant DNA methylation signatures of aging in human blood and brain. *Epigenet Chromatin* **8,** 19 (2015).

58. Jiang, R. *et al.* Discordance of DNA Methylation Variance Between two Accessible Human Tissues. *Sci Rep-uk* **5,** 8257 (2015).

59. Braun, P. R. *et al.* Genome-wide DNA methylation comparison between live human brain and peripheral tissues within individuals. *Transl Psychiat* **9,** 47 (2019).

60. Teschendorff, A. E. & Relton, C. L. Statistical and integrative system-level analysis of DNA methylation data. *Nat Rev Genet* **19,** 1 19 (1AD).

61. Birney, E., Smith, G. D. & Greally, J. M. Epigenome-wide Association Studies and the Interpretation of Disease -Omics. *Plos Genet* **12,** e1006105 (2016).

62. McGregor, K. *et al.* An evaluation of methods correcting for cell-type heterogeneity in DNA methylation studies. *Genome Biol* **17,** 84 (2016).

63. Kaushal, A. *et al.* Comparison of different cell type correction methods for genome-scale epigenetics studies. *Bmc Bioinformatics* **18,** 216 (2017).

64. Houseman, E. A. *et al.* DNA methylation arrays as surrogate measures of cell mixture distribution. *Bmc Bioinformatics* **13,** 86 (2012).

65. Houseman, E. A., Molitor, J. & Marsit, C. J. Reference-free cell mixture adjustments in analysis of DNA methylation data. *Bioinformatics* **30,** (2014).

66. Scherer, M. *et al.* Reference-free deconvolution, visualization and interpretation of complex DNA methylation data using DecompPipeline, MeDeCom and FactorViz. *Nat Protoc* **15,** 3240–3263 (2020).

67. Houseman, E. A. *et al.* Reference-free deconvolution of DNA methylation data and mediation by cell composition effects. *Bmc Bioinformatics* **17,** 1 15 (2016).

68. Gopalan, S. *et al.* Trends in DNA Methylation with Age Replicate Across Diverse Human Populations. *Genetics* **206,** 1659–1674 (2017).

69. Koch, C. M. & Wagner, W. Epigenetic-aging-signature to determine age in different tissues. *Aging* **3,** 1018–1027 (2011).

70. McEwen, L. M., Goodman, S. J., Kobor, M. S. & Jones, M. J. The Ageing Immune System and Health. 35–52 (2016). doi:10.1007/978-3-319-43365-3_3

71. McEwen, L. M. *et al.* Differential DNA methylation and lymphocyte proportions in a Costa Rican high longevity region. *Epigenet Chromatin* **10,** 21 (2017).

72. Sen, P., Shah, P. P., Nativio, R. & Berger, S. L. Epigenetic Mechanisms of Longevity and Aging. *Cell* **166,** 822–839 (2016).

73. Mulder, R. H. *et al.* Epigenome-wide change and variation in DNA methylation in childhood: Trajectories from birth to late adolescence. *Hum Mol Genet* **30,** ddaa280- (2021).

74. Heyn, H. *et al.* Distinct DNA methylomes of newborns and centenarians. *Proc National Acad Sci* **109,** 10522–10527 (2012).

75. Johansson, Å., Enroth, S. & Gyllensten, U. Continuous Aging of the Human DNA Methylome Throughout the Human Lifespan. *Plos One* **8,** e67378 (2013).

76. Christensen, B. C. *et al.* Aging and Environmental Exposures Alter Tissue-Specific DNA Methylation Dependent upon CpG Island Context. *Plos Genet* **5,** e1000602 (2009).

77. Boks, M. P. *et al.* Longitudinal changes of telomere length and epigenetic age related to traumatic stress and post-traumatic stress disorder. *Psychoneuroendocrino* **51,** 506–512 (2015).

78. Teschendorff, A. E., West, J. & Beck, S. Age-associated epigenetic drift: implications, and a case of epigenetic thrift? *Hum Mol Genet* **22,** R7–R15 (2013).

79. Fraga, M. F. *et al.* Epigenetic differences arise during the lifetime of monozygotic twins. *P Natl Acad Sci Usa* **102,** 10604–10609 (2005).

80. Hannum, G. *et al.* Genome-wide methylation profiles reveal quantitative views of human aging rates. *Mol Cell* **49,** 359 367 (2013).

81. Córdova-Palomera, A. *et al.* Genome-wide methylation study on depression: differential methylation and variable methylation in monozygotic twins. *Transl Psychiat* **5,** e557–e557 (2015).

82. Winnefeld, M. & Lyko, F. The aging epigenome: DNA methylation from the cradle to the grave. *Genome Biol* **13,** 165 (2012).

83. Horvath, S. *et al.* An epigenetic clock analysis of race/ethnicity, sex, and coronary heart disease. *Genome Biol* **17,** 171 (2016).

84. Horvath, S. & Raj, K. DNA methylation-based biomarkers and the epigenetic clock theory of ageing. *Nat Rev Genet* **19,** 371–384 (2018).

85. Liu, Z., Leung, D. & Levine, M. Comparative analysis of epigenetic aging clocks from CpG characteristics to functional associations. *Biorxiv* 512483 (2019). doi:10.1101/512483

86. Marioni, R. E. *et al.* The epigenetic clock is correlated with physical and cognitive fitness in the Lothian Birth Cohort 1936. *Int J Epidemiol* **44,** 1388–1396 (2015).

87. Wagner, W. Epigenetic aging clocks in mice and men. *Genome Biol* **18,** 107 (2017).

88. Horvath, S. DNA methylation age of human tissues and cell types. *Genome Biol* **14,** R115 (2013).

89. Fransquet, P. D., Wrigglesworth, J., Woods, R. L., Ernst, M. E. & Ryan, J. The epigenetic clock as a predictor of disease and mortality risk: a systematic review and meta-analysis. *Clin Epigenetics* **11,** 62 (2019).

90. Horvath, S. *et al.* Accelerated epigenetic aging in Down syndrome. *Aging Cell* **14,** 491–495 (2015).

91. Bell, C. G. *et al.* DNA methylation aging clocks: challenges and recommendations. *Genome Biol* **20,** 249 (2019).

92. Favé, M.-J. *et al.* Gene-by-environment interactions in urban populations modulate risk phenotypes. *Nat Commun* **9,** 827 (2018).

93. Rutter, M. Gene–environment interplay. *Depress Anxiety* **27,** 1–4 (2010).

94. Meaney, M. J. Epigenetics and the Biological Definition of Gene × Environment Interactions. *Child Dev* **81,** 41–79 (2010).

95. Klengel, T. T. *et al.* Allele-specific FKBP5 DNA demethylation mediates gene-childhood trauma interactions. *Nat Neurosci* **16,** 33 41 (2013).

96. Teh, A. L. *et al.* The effect of genotype and in utero environment on interindividual variation in neonate DNA methylomes. *Genome Res* **24,** 1064 1074 (2014).

97. Harrison, A. & Parle-McDermott, A. DNA Methylation: A Timeline of Methods and Applications. *Frontiers Genetics* **2,** 74 (2011).

98. Sun, Z., Cunningham, J., Slager, S. & Kocher, J.-P. Base resolution methylome profiling: considerations in platform selection, data preprocessing and analysis. *Epigenomics-uk* **7,** 813–828 (2015).

99. Pidsley, R. *et al.* Critical evaluation of the Illumina MethylationEPIC BeadChip microarray for whole-genome DNA methylation profiling. *Genome Biol* **17,** 208 (2016).

100. Pidsley, R. *et al.* A data-driven approach to preprocessing Illumina 450K methylation array data. *Bmc Genomics* **14,** 293 (2013).

101. Pälmke, N., Santacruz, D. & Walter, J. Comprehensive analysis of DNA-methylation in mammalian tissues using MeDIP-chip. *Methods* **53,** 175–184 (2011).

102. Taiwo, O. *et al.* Methylome analysis using MeDIP-seq with low DNA concentrations. *Nat Protoc* **7,** 617–636 (2012).

103. Nair, S. S. *et al.* Comparison of methyl-DNA immunoprecipitation (MeDIP) and methyl-CpG binding domain (MBD) protein capture for genome-wide DNA methylation analysis reveal CpG sequence coverage bias. *Epigenetics* **6,** 34–44 (2011).

104. Kurdyukov, S. & Bullock, M. DNA Methylation Analysis: Choosing the Right Method. *Biology* **5,** 3 (2016).

105. Tahiliani, M. *et al.* Conversion of 5-Methylcytosine to 5-Hydroxymethylcytosine in Mammalian DNA by MLL Partner TET1. *Science* **324,** 930–935 (2009).

106. Santiago, M., Antunes, C., Guedes, M., Sousa, N. & Marques, C. J. TET enzymes and DNA hydroxymethylation in neural development and function — How critical are they? *Genomics* **104,** 334–340 (2014).

107. Ito, S. *et al.* Role of Tet proteins in 5mC to 5hmC conversion, ES-cell self-renewal and inner cell mass specification. *Nature* **466,** 1129–1133 (2010).

108. Kriaucionis, S. & Heintz, N. The Nuclear DNA Base 5-Hydroxymethylcytosine Is Present in Purkinje Neurons and the Brain. *Science* **324,** 929–930 (2009).

109. Bibikova, M. *et al.* High-throughput DNA methylation profiling using universal bead arrays. *Genome Res* **16,** 383–393 (2006).

110. Bibikova, M. *et al.* Genome-wide DNA methylation profiling using Infinium assay. *Epigenomics-uk* **1,** 177–200 (2009).

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

112. Moran, S., Arribas, C. & Esteller, M. Validation of a DNA methylation microarray for 850,000 CpG sites of the human genome enriched in enhancer sequences. *Epigenomics-uk* **8,** 389–399 (2016).

113. Horsthemke, B. A critical view on transgenerational epigenetic inheritance in humans. *Nat Commun* **9,** 2973 (2018).

114. Heard, E. & Martienssen, R. A. Transgenerational Epigenetic Inheritance: Myths and Mechanisms. *Cell* **157,** 95–109 (2014).

115. Tang, W. W. C. *et al.* A Unique Gene Regulatory Network Resets the Human Germline Epigenome for Development. *Cell* **161,** 1453–1467 (2015).

116. Heijmans, B. T. *et al.* Persistent epigenetic differences associated with prenatal exposure to famine in humans. *Proc National Acad Sci* **105,** 17046 17049 (2008).

117. Gapp, K. *et al.* Implication of sperm RNAs in transgenerational inheritance of the effects of early trauma in mice. *Nat Neurosci* **17,** 667–669 (2014).

118. Dias, B. G. & Ressler, K. J. Parental olfactory experience influences behavior and neural structure in subsequent generations. *Nat Neurosci* **17,** 89 96 (2014).

119. Gröger, N. *et al.* The transgenerational transmission of childhood adversity: behavioral, cellular, and epigenetic correlates. *J Neural Transm* **123,** 1037–1052 (2016).

120. Elwenspoek, M. M. C., Kuehn, A., Muller, C. P. & Turner, J. D. The effects of early life adversity on the immune system. *Psychoneuroendocrino* **82,** 140–154 (2017).

121. Zayas, V., Günaydin, G. & Shoda, Y. Bases of Adult Attachment, Linking Brain, Mind and Behavior. 157–183 (2015). doi:10.1007/978-1-4614-9622-9_8

122. Thompson, R. A. The Legacy of Early Attachments. *Child Dev* **71,** 145–152 (2000).

123. Bowlby, J. The Making and Breaking of Affectional Bonds. *Brit J Psychiat* **130,** 201–210 (1977).

124. Hermans, D., Craske, M. G., Mineka, S. & Lovibond, P. F. Extinction in Human Fear Conditioning. *Biol Psychiat* **60,** 361–368 (2006).

125. Eisenberger, N. I. & Lieberman, M. D. Why rejection hurts: a common neural alarm system for physical and social pain. *Trends Cogn Sci* **8,** 294–300 (2004).

126. Papini, M. R., Fuchs, P. N. & Torres, C. Behavioral neuroscience of psychological pain. *Neurosci Biobehav Rev* **48,** 53–69 (2015).

127. Meyer, M. L., Williams, K. D. & Eisenberger, N. I. Why Social Pain Can Live on: Different Neural Mechanisms Are Associated with Reliving Social and Physical Pain. *Plos One* **10,** e0128294 (2015).

128. Mair, C., Roux, A. V. D. & Galea, S. Are neighbourhood characteristics associated with depressive symptoms? A review of evidence. *J Epidemiol Commun H* **62,** 940 (2008).

129. Paczkowski, M. M. & Galea, S. Sociodemographic characteristics of the neighborhood and depressive symptoms. *Curr Opin Psychiatr* **23,** 337–341 (2010).

130. Pickett, K. E. & Pearl, M. Multilevel analyses of neighbourhood socioeconomic context and health outcomes: a critical review. *J Epidemiol Commun H* **55,** 111 (2001).

131. Roux, A. V. D. & Mair, C. Neighborhoods and health. *Ann Ny Acad Sci* **1186,** 125–145 (2010).

132. Bleich, S. N., Jarlenski, M. P., Bell, C. N. & LaVeist, T. A. Health Inequalities: Trends, Progress, and Policy. *Annu Rev Publ Health* **33,** 7–40 (2012).

133. LaVeist, T., Pollack, K., Jr., R. T., Fesahazion, R. & Gaskin, D. Place, Not Race: Disparities Dissipate In Southwest Baltimore When Blacks And Whites Live Under Similar Conditions. *Health Affair* **30,** 1880–1887 (2017).

134. Ghosh-Dastidar, B. *et al.* Distance to Store, Food Prices, and Obesity in Urban Food Deserts. *Am J Prev Med* **47,** 587–595 (2014).

135. Grier, S. A. & Kumanyika, S. K. The Context for Choice: Health Implications of Targeted Food and Beverage Marketing to African Americans. *Am J Public Health* **98,** 1616–1629 (2011).

136. Harris, J. L., Pomeranz, J. L., Lobstein, T. & Brownell, K. D. A Crisis in the Marketplace: How Food Marketing Contributes to Childhood Obesity and What Can Be Done. *Annu Rev Publ Health* **30,** 211–225 (2009).

137. Cutrona, C. E., Wallace, G. & Wesner, K. A. Neighborhood Characteristics and Depression. *Curr Dir Psychol Sci* **15,** 188–192 (2006).

138. Silva, M. J. D., McKenzie, K., Harpham, T. & Huttly, S. R. A. Social capital and mental illness: a systematic review. *J Epidemiol Commun H* **59,** 619 (2005).

139. Do, D. P. *et al.* Circadian rhythm of cortisol and neighborhood characteristics in a population-based sample: The Multi-Ethnic Study of Atherosclerosis. *Health Place* **17,** 625–632 (2011).

140. Merkin, S. S. *et al.* Neighborhoods and Cumulative Biological Risk Profiles by Race/Ethnicity in a National Sample of U.S. Adults: NHANES III. *Ann Epidemiol* **19,** 194–201 (2009).

141. Matthews, K. A., Gallo, L. C. & Taylor, S. E. Are psychosocial factors mediators of socioeconomic status and health connections? *Ann Ny Acad Sci* **1186,** 146–173 (2010).

142. Kittleson, M. M. *et al.* Association of Childhood Socioeconomic Status With Subsequent Coronary Heart Disease in Physicians. *Arch Intern Med* **166,** 2356 (2006).

143. Gunnar, M. & Quevedo, K. The Neurobiology of Stress and Development. *Annu Rev Psychol* **58,** 145–173 (2007).

144. Benarroch, E. E. Enteric nervous system. *Neurology* **69,** 1953–1957 (2007).

145. Scrivo, R., Vasile, M., Bartosiewicz, I. & Valesini, G. Inflammation as “common soil” of the multifactorial diseases. *Autoimmun Rev* **10,** 369–74 (2010).

146. Black, P. H. Central nervous system-immune system interactions: psychoneuroendocrinology of stress and its immune consequences. *Antimicrob Agents Ch* **38,** 1–6 (1994).

147. Braveman, P. & Gottlieb, L. The Social Determinants of Health: It’s Time to Consider the Causes of the Causes. *Public Health Rep* **129,** 19–31 (2014).

148. Cain, D. W. & Cidlowski, J. A. Immune regulation by glucocorticoids. *Nat Rev Immunol* **17,** 233–247 (2017).

149. Cohen, S., Doyle, W. J., Turner, R. B., Alper, C. M. & Skoner, D. P. Childhood Socioeconomic Status and Host Resistance to Infectious Illness in Adulthood. *Psychosom Med* **66,** 553 558 (2004).

150. Dowd, J. B. & Aiello, A. E. Socioeconomic differentials in immune response. *Epidemiology Camb Mass* **20,** 902–8 (2009).

151. Ehrlich, K. B., Ross, K. M., Chen, E. & Miller, G. E. Testing the biological embedding hypothesis: Is early life adversity associated with a later proinflammatory phenotype? *Dev Psychopathol* **28,** 1273–1283 (2016).

152. Franchimont, D. Overview of the Actions of Glucocorticoids on the Immune Response: A Good Model to Characterize New Pathways of Immunosuppression for New Treatment Strategies. *Ann Ny Acad Sci* **1024,** 124–137 (2004).

153. Jensen, S. K. G., Berens, A. E. & Nelson, C. A. Effects of poverty on interacting biological systems underlying child development. *Lancet Child Adolesc Heal* **1,** 225–239 (2017).

154. Liu, R. S. *et al.* Childhood Infections, Socioeconomic Status, and Adult Cardiometabolic Risk. *Pediatrics* **137,** e20160236–e20160236 (2016).

155. Surtees, P. *et al.* Association of depression with peripheral leukocyte counts in EPIC-Norfolk—role of sex and cigarette smoking. *J Psychosom Res* **54,** 303–306 (2003).

156. Ader, R., Felten, D. L. & Cohen, N. *Psychoneuroimmunology, Vols. 1-2, 3rd ed.* (Academic Press, 2001).

157. Segerstrom, S. C. & Miller, G. E. Psychological Stress and the Human Immune System: A Meta-Analytic Study of 30 Years of Inquiry. *Psychol Bull* **130,** 601–630 (2004).

158. Azad, M. B. *et al.* Influence of socioeconomic status trajectories on innate immune responsiveness in children. *Plos One* **7,** e38669 (2012).

159. McDade, T. W. *et al.* Genome-wide analysis of DNA methylation in relation to socioeconomic status during development and early adulthood. *Am J Phys Anthropol* **169,** 3–11 (2019).

160. Miller, G. & Chen, E. Unfavorable socioeconomic conditions in early life presage expression of proinflammatory phenotype in adolescence. *Psychosom Med* **69,** 402 409 (2007).

161. Miller, G. E., Gaudin, A., Zysk, E. & Chen, E. Parental support and cytokine activity in childhood asthma: the role of glucocorticoid sensitivity. *J Allergy Clin Immun* **123,** 824 830 (2009).

162. Miller, G. E. *et al.* A functional genomic fingerprint of chronic stress in humans: blunted glucocorticoid and increased NF-kappaB signaling. *Biol Psychiat* **64,** 266 272 (2008).

163. Tanaka, T., Narazaki, M. & Kishimoto, T. IL-6 in Inflammation, Immunity, and Disease. *Csh Perspect Biol* **6,** a016295–a016295 (2014).

164. Tursich, M. *et al.* Association of trauma exposure with proinflammatory activity: a transdiagnostic meta-analysis. *Transl Psychiat* **4,** e413–e413 (2014).

165. Boeck, C. *et al.* Inflammation in adult women with a history of child maltreatment: The involvement of mitochondrial alterations and oxidative stress. *Mitochondrion* **30,** 197–207 (2016).

166. Carroll, J. E., Cohen, S. & Marsland, A. L. Early childhood socioeconomic status is associated with circulating interleukin-6 among mid-life adults. *Brain Behav Immun* **25,** 1468–74 (2011).

167. Miller, G. E. & Chen, E. Harsh family climate in early life presages the emergence of a proinflammatory phenotype in adolescence. *Psychol Sci* **21,** 848 856 (2010).

168. Miller, G. E. *et al.* Low early-life social class leaves a biological residue manifested by decreased glucocorticoid and increased proinflammatory signaling. *Proc National Acad Sci* **106,** 14716 14721 (2009).

169. Carroll, J. E. *et al.* Childhood abuse, parental warmth, and adult multisystem biological risk in the Coronary Artery Risk Development in Young Adults study. *Proc National Acad Sci* **110,** 17149–17153 (2013).

170. Friedman, E. M., Karlamangla, A. S., Gruenewald, T. L., Koretz, B. & Seeman, T. E. Early Life Adversity and Adult Biological Risk Profiles. *Psychosom Med* **77,** 176–185 (2015).

171. Dich, N. *et al.* Early life adversity potentiates the effects of later life stress on cumulative physiological dysregulation. *Anxiety Stress Coping* **28,** 372–390 (2014).

172. Danese, A. & McEwen, B. S. Adverse childhood experiences, allostasis, allostatic load, and age-related disease. *Physiol Behav* **106,** 29–39 (2012).

173. Bowlby, J. Attachment and Loss: Retrospect and Prospect. *Am J Orthopsychiat* **52,** 664–678 (1982).

174. Bretherton, I. & Munholland, K. A. in 102–127 (The Guilford Press, 2008).

175. Collins, N. L., Guichard, A. C., Ford, M. B. & Feeney, B. C. in 196–239 (Guilford Publications, 2004).

176. Pietromonaco, P. R. & Barrett, L. F. The internal working models concept: What do we really know about the self in relation to others? *Rev Gen Psychol* **4,** 155–175 (2000).

177. Zayas, V., Mischel, W., Shoda, Y. & Aber, J. L. Roots of Adult Attachment. *Soc Psychol Pers Sci* **2,** 289–297 (2011).

178. Turecki, G. & Meaney, M. J. Effects of the Social Environment and Stress on Glucocorticoid Receptor Gene Methylation: A Systematic Review. *Biol Psychiat* **79,** 87 96 (2016).

179. Weaver, I. C. G. *et al.* Epigenetic programming by maternal behavior. *Nat Neurosci* **7,** 847 854 (2004).

180. McGowan, P. O. *et al.* Broad Epigenetic Signature of Maternal Care in the Brain of Adult Rats. *Plos One* **6,** e14739 (2011).

181. Araki, R. *et al.* Epigenetic regulation of dorsal raphe GABAB1a associated with isolation-induced abnormal responses to social stimulation in mice. *Neuropharmacology* **101,** 1–12 (2016).

182. Bedrosian, T. A., Quayle, C., Novaresi, N. & Gage, Fred. H. Early life experience drives structural variation of neural genomes in mice. *Science* **359,** 1395–1399 (2018).

183. Felice, D., O'Leary, O. F. & Cryan, J. F. GABAB Receptor. 219–250 (2016). doi:10.1007/978-3-319-46044-4_12

184. Zheleznyakova, G. Y., Cao, H. & Schiöth, H. B. BDNF DNA methylation changes as a biomarker of psychiatric disorders: literature review and open access database analysis. *Behav Brain Funct* **12,** 17 (2016).

185. Miranda, M., Morici, J. F., Zanoni, M. B. & Bekinschtein, P. Brain-Derived Neurotrophic Factor: A Key Molecule for Memory in the Healthy and the Pathological Brain. *Front Cell Neurosci* **13,** 363 (2019).

186. Chagnon, Y. C., Potvin, O., Hudon, C. & Préville, M. DNA methylation and single nucleotide variants in the brain-derived neurotrophic factor (BDNF) and oxytocin receptor (OXTR) genes are associated with anxiety/depression in older women. *Frontiers Genetics* **6,** 230 (2015).

187. Ostlund, B. D. *et al.* Prenatal Stress, Fearfulness, and the Epigenome: Exploratory Analysis of Sex Differences in DNA Methylation of the Glucocorticoid Receptor Gene. *Front Behav Neurosci* **10,** 147 (2016).

188. Garg, E. *et al.* The early care environment and DNA methylome variation in childhood. *Dev Psychopathol* **30,** 891–903 (2018).

189. Moore, S. R. & Kobor, M. S. Variability in DNA methylation at the serotonin transporter gene promoter: epigenetic mechanism or cell-type artifact? *Mol Psychiatr* **14,** 1 (2018).

190. Maud, C., Ryan, J., McIntosh, J. E. & Olsson, C. A. The role of oxytocin receptor gene (OXTR) DNA methylation (DNAm) in human social and emotional functioning: a systematic narrative review. *Bmc Psychiatry* **18,** 154 (2018).

191. Kusui, C. *et al.* DNA Methylation of the Human Oxytocin Receptor Gene Promoter Regulates Tissue-Specific Gene Suppression. *Biochem Bioph Res Co* **289,** 681–686 (2001).

192. Barnett, L. Keep in touch: The importance of touch in infant development. *Infant Observation* **8,** 115–123 (2005).

193. Moore, S. R. Commentary: What is the case for candidate gene approaches in the era of high-throughput genomics? A response to Border and Keller (2017). *J Child Psychol Psyc* **58,** 331–334 (2017).

194. Coan, J. A., Schaefer, H. S. & Davidson, R. J. Lending a Hand. *Psychol Sci* **17,** 1032–1039 (2006).

195. Essex, M. J. *et al.* Epigenetic vestiges of early developmental adversity: childhood stress exposure and DNA methylation in adolescence. *Child Dev* **84,** 58 75 (2013).

196. Huang, E. J. & Reichardt, L. F. NEUROTROPHINS: Roles in Neuronal Development and Function1. *Annu Rev Neurosci* **24,** 677–736 (2001).

197. Roth, T. L., Lubin, F. D., Funk, A. J. & Sweatt, J. D. Lasting epigenetic influence of early-life adversity on the BDNF gene. *Biol Psychiat* **65,** 760 769 (2009).

198. Roth, T. L. & Sweatt, J. D. Annual Research Review: Epigenetic mechanisms and environmental shaping of the brain during sensitive periods of development. *J Child Psychol Psyc* **52,** 398–408 (2011).

199. Roth, T. L. & Sweatt, J. D. Epigenetic marking of the BDNF gene by early-life adverse experiences. *Horm Behav* **59,** 315–320 (2011).

200. Dias, B. G., Maddox, S. A., Klengel, T. & Ressler, K. J. Epigenetic mechanisms underlying learning and the inheritance of learned behaviors. *Trends Neurosci* **38,** 96–107 (2015).

201. Dickson, D. A. *et al.* Reduced levels of miRNAs 449 and 34 in sperm of mice and men exposed to early life stress. *Transl Psychiat* **8,** 101 (2018).

202. Klengel, T., Pape, J., Binder, E. B. & Mehta, D. The role of DNA methylation in stress-related psychiatric disorders. *Neuropharmacology* **80,** 115 132 (2014).

203. Mehta, D. *et al.* Childhood maltreatment is associated with distinct genomic and epigenetic profiles in posttraumatic stress disorder. *Proc National Acad Sci* **110,** 8302 8307 (2013).

204. Roberts, A. L. *et al.* Exposure to childhood abuse is associated with human sperm DNA methylation. *Translational psychiatry* **8,** 194 (2018).

205. Suderman, M. *et al.* Childhood abuse is associated with methylation of multiple loci in adult DNA. *Bmc Med Genomics* **7,** 13 (2014).

206. Weder, N. *et al.* Child abuse, depression, and methylation in genes involved with stress, neural plasticity, and brain circuitry. *J Am Acad Child Adolesc Psychiatry* **53,** 417 24.e5 (2014).

207. Zannas, A. S., Provençal, N. & Binder, E. B. Epigenetics of Posttraumatic Stress Disorder: Current Evidence, Challenges, and Future Directions. *Biol Psychiat* **78,** 327 335 (2015).

208. Wolf, E. J. *et al.* Traumatic stress and accelerated DNA methylation age: A meta-analysis. *Psychoneuroendocrino* **92,** 123–134 (2018).

209. Clausing, E. S., Binder, A. M. & Non, A. L. Epigenetic age associates with psychosocial stress and resilience in children of Latinx immigrants. *Epigenomics-uk* **0,** (2020).

210. Roberts, A. L. *et al.* Exposure to childhood abuse is associated with human sperm DNA methylation. *Transl Psychiat* **8,** 194 (2018).

211. Kumsta, R. *et al.* Severe psychosocial deprivation in early childhood is associated with increased DNA methylation across a region spanning the transcription start site of CYP2E1. *Transl Psychiat* **6,** e830–e830 (2016).

212. Cecil, C. A. M. *et al.* Epigenetic signatures of childhood abuse and neglect: Implications for psychiatric vulnerability. *J Psychiatr Res* **83,** 184–194 (2016).

213. Smearman, E. L. *et al.* Oxytocin Receptor Genetic and Epigenetic Variations: Association With Child Abuse and Adult Psychiatric Symptoms. *Child Dev* **87,** 122–34 (2016).

214. Labonte, B. *et al.* Differential Glucocorticoid Receptor Exon 1B, 1C, and 1H Expression and Methylation in Suicide Completers with a History of Childhood Abuse. *Biol Psychiat* **72,** 41–48 (2012).

215. Martín-Blanco, A. *et al.* Association between methylation of the glucocorticoid receptor gene, childhood maltreatment, and clinical severity in borderline personality disorder. *J Psychiatr Res* **57,** 34–40 (2014).

216. Mainali, P. *et al.* From Child Abuse to Developing Borderline Personality Disorder Into Adulthood: Exploring the Neuromorphological and Epigenetic Pathway. *Cureus J Medical Sci* **12,** e9474 (2020).

217. Tang, R. *et al.* Adverse childhood experiences, DNA methylation age acceleration, and cortisol in UK children: a prospective population-based cohort study. *Clin Epigenetics* **12,** 55 (2020).

218. Marini, S. *et al.* Adversity exposure during sensitive periods predicts accelerated epigenetic aging in children. *Psychoneuroendocrino* **113,** 104484 (2020).

219. Gassen, N. C., Chrousos, G. P., Binder, E. B. & Zannas, A. S. Life stress, glucocorticoid signaling, and the aging epigenome: Implications for aging-related diseases. *Neurosci Biobehav Rev* **74,** (2016).

220. Zannas, A. S. *et al.* Lifetime stress accelerates epigenetic aging in an urban, African American cohort: relevance of glucocorticoid signaling. *Genome Biol* **16,** 266 (2015).

221. Ozbay, F. *et al.* Social support and resilience to stress: from neurobiology to clinical practice. *Psychiatry Edgmont Pa Townsh* **4,** 35–40 (2007).

222. Ozbay, F., Fitterling, H., Charney, D. & Southwick, S. Social support and resilience to stress across the life span: A neurobiologic framework. *Curr Psychiat Rep* **10,** 304 (2008).

223. LaGraize, S. C., Borzan, J., Peng, Y. B. & Fuchs, P. N. Selective regulation of pain affect following activation of the opioid anterior cingulate cortex system. *Exp Neurol* **197,** 22–30 (2006).

224. Shields, A. E. *et al.* Childhood abuse, promoter methylation of leukocyte NR3C1 and the potential modifying effect of emotional support. *Epigenomics-uk* **8,** 1507–1517 (2016).

225. Brody, G. H., Miller, G. E., Yu, T., Beach, S. R. H. & Chen, E. Supportive Family Environments Ameliorate the Link Between Racial Discrimination and Epigenetic Aging: A Replication Across Two Longitudinal Cohorts. *Psychol Sci* **27,** 530 541 (2016).

226. McDade, T. W. *et al.* Social and physical environments early in development predict DNA methylation of inflammatory genes in young adulthood. *Proc National Acad Sci* **114,** 7611–7616 (2017).

227. Austin, M. K. *et al.* Early-life socioeconomic disadvantage, not current, predicts accelerated epigenetic aging of monocytes. *Psychoneuroendocrino* **97,** 131 134 (2018).

228. Bush, N. R. *et al.* The biological embedding of early-life socioeconomic status and family adversity in children's genome-wide DNA methylation. *Epigenomics-uk* **10,** 1445–1461 (2018).

229. Chen, E., Miller, G. E., Kobor, M. S. & Cole, S. W. Maternal warmth buffers the effects of low early-life socioeconomic status on pro-inflammatory signaling in adulthood. *Mol Psychiatr* **16,** 729 737 (2011).

230. Fiorito, G. *et al.* Social adversity and epigenetic aging: a multi-cohort study on socioeconomic differences in peripheral blood DNA methylation. *Sci Rep-uk* **7,** 16266 (2017).

231. Lam, L. L. *et al.* Factors underlying variable DNA methylation in a human community cohort. *Proc National Acad Sci* **109 Suppl 2,** 17253 17260 (2012).

232. McCrory, C. *et al.* How does Socio-Economic Position (SEP) get biologically embedded? A Comparison of Allostatic Load and the Epigenetic Clock(s). *Psychoneuroendocrino* **104,** 64–73 (2019).

233. Simons, R. L. *et al.* Economic hardship and biological weathering: The epigenetics of aging in a U.S. sample of black women. *Soc Sci Med* **150,** 192–200 (2016).

234. Tehranifar, P. *et al.* Early life socioeconomic factors and genomic DNA methylation in mid-life. *Epigenetics* **8,** 23–27 (2013).

235. Miller, G. E., Yu, T., Chen, E. & Brody, G. H. Self-control forecasts better psychosocial outcomes but faster epigenetic aging in low-SES youth. *P Natl Acad Sci Usa* **112,** 10325–30 (2015).

236. Esposito, E. A. *et al.* Differential DNA methylation in peripheral blood mononuclear cells in adolescents exposed to significant early but not later childhood adversity. *Dev Psychopathol* **28,** 1385–1399 (2016).

237. Miller, G. E. *et al.* Pathways to Resilience. *Psychol Sci* **22,** 1591–1599 (2011).

238. Marzi, S. J. *et al.* Analysis of DNA Methylation in Young People: Limited Evidence for an Association Between Victimization Stress and Epigenetic Variation in Blood. *Am J Psychiat* **175,** 517–529 (2018).

239. Sweeting, H. & Hunt, K. Adolescent Socioeconomic and School-Based Social Status, Smoking, and Drinking. *J Adolescent Health* **57,** 37–45 (2015).

240. Oers, J. A. van, Bongers, I. M., Goor, L. A. van de & Garretsen, H. F. Alcohol consumption, alcohol-related problems, problem drinking, and socioeconomic status. *Alcohol Alcoholism* **34,** 78–88 (1999).

241. Fouquereau, E., Fernandez, A., Mullet, E. & Sorum, P. C. Stress and the urge to drink. *Addict Behav* **28,** 669–685 (2003).

242. Niaura, R., Shadel, W. G., Britt, D. M. & Abrams, D. B. Response to social stress, urge to smoke, and smoking cessation. *Addict Behav* **27,** 241–250 (2002).

243. Brückmann, C. *et al.* DNA methylation signatures of chronic alcohol dependence in purified CD3+ T-cells of patients undergoing alcohol treatment. *Sci Rep-uk* **7,** 6605 (2017).

244. Goldowitz, D. *et al.* Molecular pathways underpinning ethanol-induced neurodegeneration. *Frontiers Genetics* **5,** 203 (2014).

245. Hillemacher, T. *et al.* Global DNA methylation is influenced by smoking behaviour. *Eur Neuropsychopharm* **18,** 295–298 (2008).

246. Mahnke, A. H., Miranda, R. C. & Homanics, G. E. Epigenetic mediators and consequences of excessive alcohol consumption. *Alcohol* **60,** 1–6 (2017).

247. McCartney, D. L. *et al.* Epigenetic signatures of starting and stopping smoking. *Ebiomedicine* **37,** 214–220 (2018).

248. Ponomarev, I. *et al.* Mechanistic insights into epigenetic modulation of ethanol consumption. *Alcohol* **60,** 95–101 (2017).

249. Tulisiak, C. T., Harris, R. A. & Ponomarev, I. DNA modifications in models of alcohol use disorders. *Alcohol* **60,** 19–30 (2017).

250. Anderson, O. S., Sant, K. E. & Dolinoy, D. C. Nutrition and epigenetics: an interplay of dietary methyl donors, one-carbon metabolism and DNA methylation. *J Nutritional Biochem* **23,** 853–859 (2012).

251. Gabbianelli, R. & Damiani, E. Epigenetics and neurodegeneration: Role of early-life nutrition. *J Nutritional Biochem* **57,** 1–13 (2018).

252. Lillycrop, K. A. & Burdge, G. C. Epigenetic mechanisms linking early nutrition to long term health. *Best Pract Res Cl En* **26,** 667–676 (2012).

253. Mathers, J. C. Nutritional modulation of ageing: Genomic and epigenetic approaches. *Mech Ageing Dev* **127,** 584–589 (2006).

254. Milagro, F. I., Mansego, M. L., Miguel, C. D. & Martínez, J. A. Dietary factors, epigenetic modifications and obesity outcomes: Progresses and perspectives. *Mol Aspects Med* **34,** 782–812 (2013).

255. Navarro, E., Funtikova, A. N., Fíto, M. & Schröder, H. Prenatal nutrition and the risk of adult obesity: Long-term effects of nutrition on epigenetic mechanisms regulating gene expression. *J Nutritional Biochem* **39,** 1 14 (2017).

256. Zhang, W. *et al.* Functional Variants in Notch Pathway Genes NCOR2, NCSTN, and MAML2 Predict Survival of Patients with Cutaneous Melanoma. *Cancer Epidemiology Biomarkers Prev Publ Am Assoc Cancer Res Cosponsored Am Soc Prev Oncol* **24,** 1101–10 (2015).

257. Bringolf-Isler, B. *et al.* Built environment, parents' perception, and children's vigorous outdoor play. *Prev Med* **50,** 251–256 (2010).

258. Davison, K. K. & Lawson, C. T. Do attributes in the physical environment influence children's physical activity? A review of the literature. *Int J Behav Nutr Phy* **3,** 19 (2006).

259. Galvez, M. P., Pearl, M. & Yen, I. H. Childhood obesity and the built environment. *Curr Opin Pediatr* **22,** 202–207 (2010).

260. Glanz, K. & Sallis, J. F. The Role of Built Environments in Physical Activity, Eating, and Obesity in Childhood. *Futur Child* **16,** 89–108 (2006).

261. Kaliman, P. *et al.* Neurophysiological and epigenetic effects of physical exercise on the aging process. *Ageing Res Rev* **10,** 475–486 (2011).

262. Kashimoto, R. K. *et al.* Physical exercise affects the epigenetic programming of rat brain and modulates the adaptive response evoked by repeated restraint stress. *Behav Brain Res* **296,** 286–289 (2016).

263. Kirchner, H., Osler, M. E., Krook, A. & Zierath, J. R. Epigenetic flexibility in metabolic regulation: disease cause and prevention? *Trends Cell Biol* **23,** 203–209 (2013).

264. Ling, C. & Rönn, T. Epigenetic adaptation to regular exercise in humans. *Drug Discov Today* **19,** 1015–1018 (2014).

265. Mikkelsen, K., Stojanovska, L., Polenakovic, M., Bosevski, M. & Apostolopoulos, V. Exercise and mental health. *Maturitas* **106,** 48–56 (2017).

266. Brulle, R. J. & Pellow, D. N. ENVIRONMENTAL JUSTICE: Human Health and Environmental Inequalities. *Annu Rev Publ Health* **27,** 103–124 (2006).

267. Mohai, P., Lantz, P. M., Morenoff, J., House, J. S. & Mero, R. P. Racial and Socioeconomic Disparities in Residential Proximity to Polluting Industrial Facilities: Evidence From the Americans' Changing Lives Study. *Am J Public Health* **99,** S649–S656 (2009).

268. Morello-Frosch, R., Zuk, M., Jerrett, M., Shamasunder, B. & Kyle, A. D. Understanding The Cumulative Impacts Of Inequalities In Environmental Health: Implications For Policy. *Health Affair* **30,** 879–887 (2017).

269. Brook, R. D. *et al.* Particulate Matter Air Pollution and Cardiovascular Disease. *Circulation* **121,** 2331–2378 (2010).

270. Chen, R. *et al.* DNA hypomethylation and its mediation in the effects of fine particulate air pollution on cardiovascular biomarkers: A randomized crossover trial. *Environ Int* **94,** 614–619 (2016).

271. Clifford, R. L. *et al.* Inhalation of diesel exhaust and allergen alters human bronchial epithelium DNA methylation. *J Allergy Clin Immun* **139,** 112 121 (2017).

272. Gref, A. *et al.* Genome-Wide Interaction Analysis of Air Pollution Exposure and Childhood Asthma with Functional Follow-up. *Am J Resp Crit Care* **195,** 1373–1383 (2017).

273. Laumbach, R. J. & Kipen, H. M. Respiratory health effects of air pollution: Update on biomass smoke and traffic pollution. *J Allergy Clin Immun* **129,** 3–11 (2012).

274. Rabinovitch, N. *et al.* Methylation of cysteinyl leukotriene receptor 1 genes associates with lung function in asthmatics exposed to traffic-related air pollution. *Epigenetics* 1–9 (2020). doi:10.1080/15592294.2020.1790802

275. Luyten, L. J. *et al.* Air pollution and the fetal origin of disease: A systematic review of the molecular signatures of air pollution exposure in human placenta. *Environ Res* **166,** 310–323 (2018).

276. Mustafić, H. *et al.* Main Air Pollutants and Myocardial Infarction: A Systematic Review and Meta-analysis. *Jama* **307,** 713–721 (2012).

277. Rider, C. F. *et al.* Controlled diesel exhaust and allergen coexposure modulates microRNA and gene expression in humans: Effects on inflammatory lung markers. *J Allergy Clin Immun* **138,** 1690–1700 (2016).

278. Somineni, H. K. *et al.* Ten-eleven translocation 1 (TET1) methylation is associated with childhood asthma and traffic-related air pollution. *J Allergy Clin Immun* **137,** 797-805.e5 (2016).

279. Tzivian, L. Outdoor Air Pollution and Asthma in Children. *J Asthma* **48,** 470–481 (2011).

280. Shishehbor, M. H., Litaker, D., Pothier, C. E. & Lauer, M. S. Association of Socioeconomic Status With Functional Capacity, Heart Rate Recovery, and All-Cause Mortality. *Jama* **295,** 784–792 (2006).

281. Everson, S. A., Maty, S. C., Lynch, J. W. & Kaplan, G. A. Epidemiologic evidence for the relation between socioeconomic status and depression, obesity, and diabetes. *J Psychosom Res* **53,** 891–895 (2002).

282. Clark, A. M., DesMeules, M., Luo, W., Duncan, A. S. & Wielgosz, A. Socioeconomic status and cardiovascular disease: risks and implications for care. *Nat Rev Cardiol* **6,** 712–722 (2009).

283. Taylor, S. E., Lehman, B. J., Kiefe, C. I. & Seeman, T. E. Relationship of Early Life Stress and Psychological Functioning to Adult C-Reactive Protein in the Coronary Artery Risk Development in Young Adults Study. *Biol Psychiat* **60,** 819–824 (2006).

284. Dowd, J. B., Zajacova, A. & Aiello, A. E. Predictors of Inflammation in U.S. Children Aged 3–16 Years. *Am J Prev Med* **39,** 314–320 (2010).

285. Koster, A. *et al.* Association of Inflammatory Markers With Socioeconomic Status. *Journals Gerontology Ser Biological Sci Medical Sci* **61,** 284–290 (2006).

286. Elliot, A. J. & Chapman, B. P. Socioeconomic Status, Psychological Resources, and Inflammatory Markers: Results From the MIDUS Study. *Health Psychol* **35,** 1205–1213 (2016).

287. McDade, T. W., Hawkley, L. C. & Cacioppo, J. T. Psychosocial and Behavioral Predictors of Inflammation in Middle-Aged and Older Adults: The Chicago Health, Aging, and Social Relations Study. *Psychosom Med* **68,** 376–381 (2006).

288. Sin, N. L., Graham-Engeland, J. E., Ong, A. D. & Almeida, D. M. Affective Reactivity to Daily Stressors Is Associated With Elevated Inflammation. *Health Psychol* **34,** 1154–1165 (2015).

289. Hostinar, C. E., Ross, K. M., Chen, E. & Miller, G. E. Modeling the association between lifecourse socioeconomic disadvantage and systemic inflammation in healthy adults: The role of self-control. *Heal Psychology Official J Div Heal Psychology Am Psychological Assoc* **34,** 580–90 (2014).

290. Gimeno, D. *et al.* Associations of C-reactive protein and interleukin-6 with cognitive symptoms of depression: 12-year follow-up of the Whitehall II study. *Psychol Med* **39,** 413–423 (2008).

291. Davidson, K. W. *et al.* Relation of Inflammation to Depression and Incident Coronary Heart Disease (from the Canadian Nova Scotia Health Survey [NSHS95] Prospective Population Study). *Am J Cardiol* **103,** 755–761 (2009).

292. Möhlenkamp, S. *et al.* Quantification of Coronary Atherosclerosis and Inflammation to Predict Coronary Events and All-Cause Mortality. *J Am Coll Cardiol* **57,** 1455–1464 (2011).

293. Shiels, M. S. *et al.* Cigarette Smoking and Variations in Systemic Immune and Inflammation Markers. *Jnci J National Cancer Inst* **106,** (2014).

294. Mora, S., Lee, I.-M., Buring, J. E. & Ridker, P. M. Association of Physical Activity and Body Mass Index With Novel and Traditional Cardiovascular Biomarkers in Women. *Jama* **295,** 1412–1419 (2006).

295. Muscatell, K. A., Brosso, S. N. & Humphreys, K. L. Socioeconomic status and inflammation: a meta-analysis. *Mol Psychiatr* **25,** 2189–2199 (2020).

296. Wood, N. M., Trebilco, T. & Cohen-Woods, S. Scars of childhood socioeconomic stress: A Systematic Review. *Neurosci Biobehav Rev* **118,** 397–410 (2020).

297. Tarullo, A. R., Tuladhar, C. T., Kao, K., Drury, E. B. & Meyer, J. Cortisol and socioeconomic status in early childhood: A multidimensional assessment. *Dev Psychopathol* **32,** 1876–1887 (2020).

298. McEwen, B. S. Allostasis and Allostatic Load: Implications for Neuropsychopharmacology. *Neuropsychopharmacol* **22,** 108–124 (2000).

299. McEwen, B. S. & Stellar, E. Stress and the Individual. *Arch Intern Med* **153,** 2093 (1993).

300. Edes, A. N. & Crews, D. E. Allostatic load and biological anthropology. *Am J Phys Anthropol* **162,** 44–70 (2017).

301. Janicki-Deverts, D. *et al.* Socioeconomic Status is Related to Urinary Catecholamines in the Coronary Artery Risk Development in Young Adults (CARDIA) Study. *Psychosom Med* **69,** 514–520 (2007).

302. Evans, G. W. & English, K. The Environment of Poverty: Multiple Stressor Exposure, Psychophysiological Stress, and Socioemotional Adjustment. *Child Dev* **73,** 1238–1248 (2002).

303. Singer, B. & Ryff, C. D. Hierarchies of life histories and associated health risks. *Ann Ny Acad Sci* **896,** 96 115 (1999).

304. Seeman, T. E. *et al.* Cumulative biological risk and socio-economic differences in mortality: MacArthur Studies of Successful Aging. *Soc Sci Med* **58,** 1985–1997 (2004).

305. Weinstein, M., Goldman, N., Hedley, A., Yu-Hsuan, L. & Seeman, T. Social linkages to biological markers of health among the elderly. *J Biosoc Sci* **35,** 433–453 (2003).

306. Kubzansky, L. D., Kawachi, I. & Sparrow, D. Socioeconomic status, hostility, and risk factor clustering in the normative aging study: Any help from the concept of allostatic load? *Ann Behav Med* **21,** 330–338 (1999).

307. Castro-Diehl, C. *et al.* Associations of socioeconomic and psychosocial factors with urinary measures of cortisol and catecholamines in the Multi-Ethnic Study of Atherosclerosis (MESA). *Psychoneuroendocrino* **41,** 132–141 (2014).

308. Vliegenthart, J. *et al.* Socioeconomic status in children is associated with hair cortisol levels as a biological measure of chronic stress. *Psychoneuroendocrino* **65,** 9–14 (2016).

309. Zhu, Y. *et al.* Socioeconomic status disparities affect children's anxiety and stress-sensitive cortisol awakening response through parental anxiety. *Psychoneuroendocrino* **103,** 96–103 (2019).

310. Desantis, A. S., Kuzawa, C. W. & Adam, E. K. Developmental origins of flatter cortisol rhythms: socioeconomic status and adult cortisol activity. *Am J Hum Biol* **27,** 458–467 (2015).

311. Clearfield, M. W., Carter-Rodriguez, A., Merali, A.-R. & Shober, R. The effects of SES on infant and maternal diurnal salivary cortisol output. *Infant Behav Dev* **37,** 298–304 (2014).

312. Cohen, S., Doyle, W. J. & Baum, A. Socioeconomic Status Is Associated With Stress Hormones. *Psychosom Med* **68,** 414–420 (2006).

313. Merrill, S. M., Gladish, N. & Kobor, M. S. in (eds. Binder", ["Elisabeth B. & Klengel"], "Torsten) 83–126 (Springer International Publishing, 2019). doi:10.1007/7854_2019_114

314. Aristizabal, M. J. *et al.* Biological embedding of experience: A primer on epigenetics. *Proc National Acad Sci* 201820838 (2019). doi:10.1073/pnas.1820838116

315. Laubach, Z. M. *et al.* Socioeconomic status and DNA methylation from birth through mid-childhood: a prospective study in Project Viva. *Epigenomics-uk* **11,** 1413–1427 (2019).

316. Beach, S. R. H. *et al.* Parenting, Socioeconomic Status Risk, and Later Young Adult Health: Exploration of Opposing Indirect Effects via DNA Methylation. *Child Dev* **87,** 111–121 (2016).

317. Needham, B. L. *et al.* Life course socioeconomic status and DNA methylation in genes related to stress reactivity and inflammation: The multi-ethnic study of atherosclerosis. *Epigenetics* **10,** 958–69 (2015).

318. Cecil, C. A. M., Zhang, Y. & Nolte, T. Childhood maltreatment and DNA methylation: A systematic review. *Neurosci Biobehav Rev* **112,** 392–409 (2020).

319. Park, C. *et al.* Stress, epigenetics and depression: A systematic review. *Neurosci Biobehav Rev* **102,** 139–152 (2019).

320. Sosnowski, D. W., Booth, C., York, T. P., Amstadter, A. B. & Kliewer, W. Maternal prenatal stress and infant DNA methylation: A systematic review. *Dev Psychobiol* **60,** 127–139 (2018).

321. Simanek, A. M. & Auer, P. L. Early Life Socioeconomic Disadvantage and Epigenetic Programming of a Pro-inflammatory Phenotype: a Review of Recent Evidence. *Curr Epidemiology Reports* **5,** 407–417 (2018).

322. Evans, L., Engelman, M., Mikulas, A. & Malecki, K. How are Social Determinants of Health Integrated into Epigenetic Research? A Systematic Review ,. *Soc Sci Med* **273,** 113738 (2021).

323. Gaydosh, L., Schorpp, K. M., Chen, E., Miller, G. E. & Harris, K. M. College completion predicts lower depression but higher metabolic syndrome among disadvantaged minorities in young adulthood. *P Natl Acad Sci Usa* **115,** 109–114 (2017).

324. Barker, D. J. P. The origins of the developmental origins theory. *J Intern Med* **261,** 412 417 (2007).

325. Stringhini, S. *et al.* Life-course socioeconomic status and DNA methylation of genes regulating inflammation. *Int J Epidemiol* **44,** 1320–1330 (2015).

326. Braveman, P. A. *et al.* Socioeconomic Status in Health Research: One Size Does Not Fit All. *Jama* **294,** 2879–2888 (2005).

327. Danese, A. *et al.* Biological embedding of stress through inflammation processes in childhood. *Mol Psychiatr* **16,** 244–6 (2010).

328. Danese, A. & Widom, C. S. Objective and subjective experiences of child maltreatment and their relationships with psychopathology. *Nat Hum Behav* **4,** 811–818 (2020).

329. Rosero-Bixby, L. High life expectancy and reversed socioeconomic gradients of elderly people in Mexico and Costa Rica. *Demogr Res* **38,** 95–108 (2018).

330. Rosero-Bixby, L. & Dow, W. H. Surprising SES Gradients in mortality, health, and biomarkers in a Latin American population of adults. *Journals Gerontology Ser B Psychological Sci Soc Sci* **64,** 105–17 (2009).

331. Turra, C. M. & Goldman, N. Socioeconomic Differences in Mortality Among U.S. Adults: Insights Into the Hispanic Paradox. *Journals Gerontology Ser B* **62,** S184–S192 (2007).

332. Elo, I. T. & Preston, S. H. Educational differentials in mortality: United States, 1979–1985. *Soc Sci Med* **42,** 47–57 (1996).

333. Rosero-Bixby, L., Fernández, X. & Dow, W. H. CRELES: Costa Rican Longevity and Healthy Aging Study, 2005 (Costa Rica Estudio de Longevidad y Envejecimiento Saludable). (2013). doi:10.3886/icpsr26681.v2

334. Rosero-Bixby, L. & Dow, W. H. Exploring why Costa Rica outperforms the United States in life expectancy: A tale of two inequality gradients. *Proc National Acad Sci* **113,** 1130–1137 (2016).

335. Fernández, B. R., Rosero-Bixby, L. & Koivumaa-Honkanen, H. Effects of Self-Rated Health and Self-Rated Economic Situation on Depressed Mood Via Life Satisfaction Among Older Adults in Costa Rica. *J Aging Health* **28,** 225–243 (2016).

336. Gersten, O., Dow, W. H., Dow, W. D. & Rosero-Bixby, L. Stressors Over the Life Course and Neuroendocrine System Dysregulation in Costa Rica. *J Aging Health* **22,** 748–771 (2010).

337. Rosero-Bixby, L. *et al.* Correlates of longitudinal leukocyte telomere length in the Costa Rican Longevity Study of Healthy Aging (CRELES): On the importance of DNA collection and storage procedures. *Plos One* **14,** e0223766 (2019).

338. Rosero-Bixby, L. & Dow, W. H. Predicting mortality with biomarkers: a population-based prospective cohort study for elderly Costa Ricans. *Popul Health Metr* **10,** 11 (2012).

339. Herrmann, B. G. & Frischauf, A.-M. Isolation of genomic DNA. *Methods Enzymol* **152,** 180–183 (1987).

340. Weale, M. E. Genetic Variation, Methods and Protocols. *Methods Mol Biology* **628,** 341–372 (2010).

341. Purcell, S. *et al.* PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. *Am J Hum Genetics* **81,** 559–575 (2007).

342. Price, A. L. *et al.* Long-Range LD Can Confound Genome Scans in Admixed Populations. *Am J Hum Genetics* **83,** 132–135 (2008).

343. Aryee, M. J. *et al.* Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. *Bioinform Oxf Engl* **30,** 1363–9 (2014).

344. Fortin, J.-P. *et al.* Functional normalization of 450k methylation array data improves replication in large cancer studies. *Genome Biol* **15,** 503 (2014).

345. Du, P., Kibbe, W. A. & Lin, S. M. lumi: a pipeline for processing Illumina microarray. *Bioinformatics* **24,** 1547 1548 (2008).

346. Efron, B. Large-Scale Simultaneous Hypothesis Testing: The Choice of a Null Hypothesis. *J Am Stat Assoc* **99,** 96–104 (2004).

347. Efron, B. Correlation and Large-Scale Simultaneous Significance Testing. *J Am Stat Assoc* **102,** 93–103 (2007).

348. Price, M. E. *et al.* Additional annotation enhances potential for biologically-relevant analysis of the Illumina Infinium HumanMethylation450 BeadChip array. *Epigenet Chromatin* **6,** 1 15 (2013).

349. Fortin, J.-P., Triche, T. & Hansen, K. *Preprocessing, normalization and integration of the Illumina HumanMethylationEPIC array*. (2016).

350. Johnson, W. E., Li, C. & Rabinovic, A. Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics* **8,** 118 127 (2007).

351. Leek, J. T., Johnson, W. E., Parker, H. S., Jaffe, A. E. & Storey, J. D. The sva package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinform Oxf Engl* **28,** 882–3 (2012).

352. Ellingson, S. R. & Fardo, D. W. Automated quality control for genome wide association studies. *F1000research* **5,** 1889 (2016).

353. Venables, W. N. & Ripley, B. D. Modern Applied Statistics with S. *Stat Comput* (2002). doi:10.1007/978-0-387-21706-2

354. Salas, L. A. *et al.* An optimized library for reference-based deconvolution of whole-blood biospecimens assayed using the Illumina HumanMethylationEPIC BeadArray. *Genome Biol* **19,** 64 (2018).

355. Karolchik, D. *et al.* The UCSC Genome Browser Database. *Nucleic Acids Res* **31,** 51–54 (2003).

356. Kent, W. J. *et al.* The Human Genome Browser at UCSC. *Genome Res* **12,** 996–1006 (2002).

357. Visone, R. *et al.* DNA methylation of shelf, shore and open sea CpG positions distinguish high microsatellite instability from low or stable microsatellite status colon cancer stem cells. *Epigenomics-uk* **11,** 587–604 (2019).

358. Kundaje, A. *et al.* Integrative analysis of 111 reference human epigenomes. *Nature* **518,** 317 330 (2015).

359. Ernst, J. & Kellis, M. Chromatin-state discovery and genome annotation with ChromHMM. *Nat Protoc* **12,** 2478–2492 (2017).

360. Irizarry, R. A. *et al.* The human colon cancer methylome shows similar hypo- and hypermethylation at conserved tissue-specific CpG island shores. *Nat Genet* **41,** 178–186 (2009).

361. Hoog, N. de, Dinther, S. van & Bakker, E. Socioeconomic Status and Health-Compromising Behaviour: Is it All About Perception? *Europe S J Psychology* **16,** 498–513 (2020).

362. Nobles, J., Weintraub, M. R. & Adler, N. E. Subjective socioeconomic status and health: Relationships reconsidered. *Soc Sci Med* **82,** 58–66 (2013).

363. Bradshaw, M., Kent, B. V., Henderson, W. M. & Setar, A. C. Subjective Social Status, Life Course SES, and BMI in Young Adulthood. *Health Psychol* **36,** 682–694 (2017).

364. Ghaed, S. G. & Gallo, L. C. Subjective Social Status, Objective Socioeconomic Status, and Cardiovascular Risk in Women. *Health Psychol* **26,** 668–674 (2007).

365. Bridges, S. & Disney, R. Debt and depression. *J Health Econ* **29,** 388–403 (2010).

366. Senik, C. Direct evidence on income comparisons and their welfare effects. *J Econ Behav Organ* **72,** 408–424 (2009).

367. Wright, C. E. & Steptoe, A. Subjective socioeconomic position, gender and cortisol responses to waking in an elderly population. *Psychoneuroendocrino* **30,** 582–590 (2005).

368. Weckesser, L. J. *et al.* The psychometric properties and temporal dynamics of subjective stress, retrospectively assessed by different informants and questionnaires, and hair cortisol concentrations. *Sci Rep-uk* **9,** 1098 (2019).

369. Metzenthin, P. *et al.* A one-item subjective work stress assessment tool is associated with cortisol secretion levels in critical care nurses. *Prev Med* **48,** 462–466 (2009).

370. Rehkopf, D. H., Glymour, M. M. & Osypuk, T. L. The Consistency Assumption for Causal Inference in Social Epidemiology: When a Rose Is Not a Rose. *Curr Epidemiology Reports* **3,** 63–71 (2016).

371. Barker, D. J. The fetal and infant origins of adult disease. *Brit Med J* **301,** 1111 (1990).

372. Newton, S., Braithwaite, D. & Akinyemiju, T. F. Socio-economic status over the life course and obesity: Systematic review and meta-analysis. *Plos One* **12,** e0177151 (2017).

373. Mueller, C. W. & Parcel, T. L. Measures of Socioeconomic Status: Alternatives and Recommendations. *Child Dev* **52,** 13–30 (1981).

374. Baum, A., Garofalo, J. P. & Yali, A. M. Socioeconomic Status and Chronic Stress: Does Stress Account for SES Effects on Health? *Ann Ny Acad Sci* **896,** 131–144 (2017).

375. Winkleby, M. A., Jatulis, D. E., Frank, E. & Fortmann, S. P. Socioeconomic status and health: how education, income, and occupation contribute to risk factors for cardiovascular disease. *Am J Public Health* **82,** 816–820 (1992).

376. Stringhini, S. *et al.* Association of Lifecourse Socioeconomic Status with Chronic Inflammation and Type 2 Diabetes Risk: The Whitehall II Prospective Cohort Study. *Plos Med* **10,** e1001479 (2013).

377. Farah, M. J. Socioeconomic status and the brain: prospects for neuroscience-informed policy. *Nat Rev Neurosci* **19,** 428–438 (2018).

378. Puolakka, E. *et al.* Childhood Socioeconomic Status in Predicting Metabolic Syndrome and Glucose Abnormalities in Adulthood: The Cardiovascular Risk in Young Finns Study. *Diabetes Care* **39,** 2311–2317 (2016).

379. Maty, S. C., Lynch, J. W., Raghunathan, T. E. & Kaplan, G. A. Childhood socioeconomic position, gender, adult body mass index, and incidence of type 2 diabetes mellitus over 34 years in the Alameda County Study. *Am J Public Health* **98,** 1486–94 (2008).

380. Clegg, L. X. *et al.* Impact of socioeconomic status on cancer incidence and stage at diagnosis: selected findings from the surveillance, epidemiology, and end results: National Longitudinal Mortality Study. *Cancer Causes Control Ccc* **20,** 417–35 (2008).

381. Mao, Y. *et al.* Socioeconomic status and lung cancer risk in Canada. *Int J Epidemiol* **30,** 809–817 (2001).

382. Wood, E., Sallar, A. M., Schechter, M. T. & Hogg, R. S. Social inequalities in male mortality amenable to medical intervention in British Columbia. *Soc Sci Med* **48,** 1751–1758 (1999).

383. Smith, G. D., Bartley, M. & Blane, D. The Black report on socioeconomic inequalities in health 10 years on. *Brit Med J* **301,** 373 (1990).

384. Hertzman, C. The Biological Embedding of Early Experience and Its Effects on Health in Adulthood. *Ann Ny Acad Sci* **896,** 85–95 (1999).

385. Hertzman, C. & Boyce, T. How experience gets under the skin to create gradients in developmental health. *Annu Rev Publ Health* **31,** 329 47 3p following 347 (2010).

386. Snyder-Mackler, N. *et al.* Social determinants of health and survival in humans and other animals. *Sci New York N Y* **368,** eaax9553 (2020).

387. Rea, I. M. *et al.* Age and Age-Related Diseases: Role of Inflammation Triggers and Cytokines. *Front Immunol* **9,** 586 (2018).

388. Furman, D. *et al.* Chronic inflammation in the etiology of disease across the life span. *Nat Med* **25,** 1822–1832 (2019).

389. Liberman, A. C. *et al.* Regulatory and Mechanistic Actions of Glucocorticoids on T and Inflammatory Cells. *Front Endocrinol* **9,** 235 (2018).

390. Cohen, S. *et al.* Chronic stress, glucocorticoid receptor resistance, inflammation, and disease risk. *Proc National Acad Sci* **109,** 5995 5999 (2012).

391. Miller, G. E., Cohen, S. & Ritchey, A. K. Chronic psychological stress and the regulation of pro-inflammatory cytokines: A glucocorticoid-resistance model. *Health Psychol* **21,** 531 (2002).

392. Kiecolt-Glaser, J. K. *et al.* Chronic stress and age-related increases in the proinflammatory cytokine IL-6. *Proc National Acad Sci* **100,** 9090–9095 (2003).

393. Alvarez, H. A. O., Kubzansky, L. D., Campen, M. J. & Slavich, G. M. Early Life Stress, Air Pollution, Inflammation, and Disease: An Integrative Review and Immunologic Model of Social-Environmental Adversity and Lifespan Health. *Neurosci Biobehav Rev* **92,** 226–242 (2018).

394. Harrison, C. A. & Taren, D. How poverty affects diet to shape the microbiota and chronic disease. *Nat Rev Immunol* **18,** 279–287 (2018).

395. Bird, A. Perceptions of epigenetics. *Nature* **447,** 396 398 (2007).

396. Gonzalez-Jaramillo, V. *et al.* Epigenetics and Inflammatory Markers: A Systematic Review of the Current Evidence. *Int J Inflamm* **2019,** 1–14 (2019).

397. Ahsan, M. *et al.* The relative contribution of DNA methylation and genetic variants on protein biomarkers for human diseases. *Plos Genet* **13,** e1007005 (2017).

398. Verschoor, C. P. *et al.* The relation between DNA methylation patterns and serum cytokine levels in community-dwelling adults: a preliminary study. *Bmc Genet* **18,** 57 (2017).

399. Giudice, M. D. & Gangestad, S. W. Rethinking IL-6 and CRP: Why They Are More Than Inflammatory Biomarkers, and Why It Matters. *Brain Behav Immun* **70,** 61–75 (2018).

400. Guénard, F. *et al.* Methylation and Expression of Immune and Inflammatory Genes in the Offspring of Bariatric Bypass Surgery Patients. *J Obes* **2013,** 1–9 (2013).

401. Sun, Y. V. *et al.* Gene-Specific DNA Methylation Association with Serum Levels of C-Reactive Protein in African Americans. *Plos One* **8,** e73480 (2013).

402. Marzi, C. *et al.* Epigenetic Signatures at AQP3 and SOCS3 Engage in Low-Grade Inflammation across Different Tissues. *Plos One* **11,** e0166015 (2016).

403. Verschoor, C. P., McEwen, L. M., Kobor, M. S., Loeb, M. B. & Bowdish, D. M. E. DNA methylation patterns are related to co-morbidity status and circulating C-reactive protein levels in the nursing home elderly. *Exp Gerontol* **105,** 47–52 (2018).

404. Ligthart, S. *et al.* DNA methylation signatures of chronic low-grade inflammation are associated with complex diseases. *Genome Biol* **17,** 255 (2016).

405. Fernández-Sanlés, A., Sayols-Baixeras, S., Subirana, I., Degano, I. R. & Elosua, R. Association between DNA methylation and coronary heart disease or other atherosclerotic events: A systematic review. *Atherosclerosis* **263,** 325–333 (2017).

406. Venugopal, D. *et al.* Impact of antipsychotic treatment on methylation status of Interleukin-6 [IL-6] gene in Schizophrenia. *J Psychiatr Res* **104,** 88–95 (2018).

407. Ryan, J. *et al.* Investigating the epigenetic profile of the inflammatory gene IL-6 in late-life depression. *Bmc Psychiatry* **17,** 354 (2017).

408. Nile, C. J., Read, R. C., Akil, M., Duff, G. W. & Wilson, A. G. Methylation status of a single CpG site in the IL6 promoter is related to IL6 messenger RNA levels and rheumatoid arthritis. *Arthritis Rheumatism* **58,** 2686 2693 (2008).

409. Balakrishnan, A., Guruprasad, K. P., Satyamoorthy, K. & Joshi, M. B. Interleukin-6 determines protein stabilization of DNA methyltransferases and alters DNA promoter methylation of genes associated with insulin signaling and angiogenesis. *Laboratory Investigation J Technical Methods Pathology* **98,** 1143–1158 (2018).

410. Kong, X. *et al.* JAK2/STAT3 signaling mediates IL-6-inhibited neurogenesis of neural stem cells through DNA demethylation/methylation. *Brain Behav Immun* **79,** 159–173 (2019).

411. Kumari, E. *et al.* Developmental IL-6 Exposure Favors Production of PDGF-Responsive Multipotential Progenitors at the Expense of Neural Stem Cells and Other Progenitors. *Stem Cell Rep* **14,** 861–875 (2020).

412. Garaud, S. *et al.* IL-6 Modulates CD5 Expression in B Cells from Patients with Lupus by Regulating DNA Methylation. *J Immunol* **182,** 5623–5632 (2009).

413. Stevenson, A. J. *et al.* Creating and Validating a DNA Methylation-Based Proxy for Interleukin-6. *Journals Gerontology Ser* (2021). doi:10.1093/gerona/glab046

414. Ryan, J., Wrigglesworth, J., Loong, J., Fransquet, P. D. & Woods, R. L. A Systematic Review and Meta-analysis of Environmental, Lifestyle, and Health Factors Associated With DNA Methylation Age. *Journals Gerontology Ser* **75,** 481–494 (2019).

415. Hughes, A. *et al.* Socioeconomic Position and DNA Methylation Age Acceleration Across the Life Course. *Am J Epidemiol* **187,** 2346–2354 (2018).

416. Miller, G. E. *et al.* Greater inflammatory activity and blunted glucocorticoid signaling in monocytes of chronically stressed caregivers. *Brain Behav Immun* **41,** 191 199 (2014).

417. Triche, T. J., Weisenberger, D. J., Berg, D. V. D., Laird, P. W. & Siegmund, K. D. Low-level processing of Illumina Infinium DNA Methylation BeadArrays. *Nucleic Acids Res* **41,** e90 (2013).

418. Chambers, J. M. & Hastie, T. J. Statistical Models in S. 1–12 (2017). doi:10.1201/9780203738535-1

419. Wilkinson, G. N. & Rogers, C. E. Symbolic Description of Factorial Models for Analysis of Variance. *Appl Statistics* **22,** 392 (1973).

420. Benjamini, Y. & Hochberg, Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J Royal Statistical Soc Ser B Methodol* **57,** 289–300 (1995).

421. Tsai, P.-C. & Bell, J. T. Power and sample size estimation for epigenome-wide association scans to detect differential DNA methylation. *Int J Epidemiol* **44,** 1429 1441 (2015).

422. Wilson, S. L., Leavey, K., Cox, B. J. & Robinson, W. P. Mining DNA methylation alterations towards a classification of placental pathologies. *Hum Mol Genet* **27,** 135–146 (2017).

423. Shapiro, S. S., Wilk, M. B. & Chen, H. J. A Comparative Study of Various Tests for Normality. *J Am Stat Assoc* **63,** 1343 (1968).

424. Hastings, C., Mosteller, F., Tukey, J. W. & Winsor, C. P. Low Moments for Small Samples: A Comparative Study of Order Statistics. *Ann Math Statistics* **18,** 413–426 (1947).

425. Li, M. *et al.* EWAS Atlas: a curated knowledgebase of epigenome-wide association studies. *Nucleic Acids Res* **47,** gky1027- (2018).

426. Lee, K. W. K. & Pausova, Z. Cigarette smoking and DNA methylation. *Frontiers Genetics* **4,** 132 (2013).

427. Dick, K. J. *et al.* DNA methylation and body-mass index: a genome-wide analysis. *Lancet* **383,** 1990–1998 (2014).

428. Gonçalves, R. B. *et al.* Impact of smoking on inflammation: overview of molecular mechanisms. *Inflamm Res* **60,** 409–424 (2011).

429. Papanicolaou, D. A., Wilder, R. L., Manolagas, S. C. & Chrousos, G. P. The Pathophysiologic Roles of Interleukin-6 in Human Disease. *Ann Intern Med* **128,** 127 (1998).

430. Ball, K. & Crawford, D. Socioeconomic status and weight change in adults: a review. *Soc Sci Med* **60,** 1987–2010 (2005).

431. Hiscock, R., Bauld, L., Amos, A., Fidler, J. A. & Munafò, M. Socioeconomic status and smoking: a review: Hiscock et al. *Ann Ny Acad Sci* **1248,** 107–123 (2011).

432. Rada-Iglesias, A. Is H3K4me1 at enhancers correlative or causative? *Nat Genet* **50,** 4–5 (2018).

433. Schnabel, R. B. *et al.* Multiple inflammatory biomarkers in relation to cardiovascular events and mortality in the community. *Arteriosclerosis Thrombosis Vasc Biology* **33,** 1728–33 (2013).

434. Akbari, M. & Hassan-Zadeh, V. IL-6 signalling pathways and the development of type 2 diabetes. *Inflammopharmacology* **26,** 685–698 (2018).

435. Kristiansen, O. P. & Mandrup-Poulsen, T. Interleukin-6 and Diabetes The Good, the Bad, or the Indifferent? *Diabetes* **54,** S114–S124 (2005).

436. Nowell, M. A. *et al.* Therapeutic Targeting of IL-6 Trans Signaling Counteracts STAT3 Control of Experimental Inflammatory Arthritis. *J Immunol* **182,** 613–622 (2009).

437. Emery, P. *et al.* IL-6 receptor inhibition with tocilizumab improves treatment outcomes in patients with rheumatoid arthritis refractory to anti-tumour necrosis factor biologicals: results from a 24-week multicentre randomised placebo-controlled trial. *Ann Rheum Dis* **67,** 1516 (2008).

438. Loucks, E. B. *et al.* Life course socioeconomic position is associated with inflammatory markers: the Framingham Offspring Study. *Soc Sci Medicine 1982* **71,** 187–95 (2010).

439. Pietras, S. A. & Goodman, E. Socioeconomic status gradients in inflammation in adolescence. *Psychosom Med* **75,** 442–8 (2013).

440. Steptoe, A., Owen, N., Kunz-Ebrecht, S. & Mohamed-Ali, V. Inflammatory cytokines, socioeconomic status, and acute stress responsivity. *Brain Behav Immun* **16,** 774–784 (2002).

441. Friedman, E. M. & Herd, P. Income, education, and inflammation: differential associations in a national probability sample (The MIDUS study). *Psychosom Med* **72,** 290–300 (2010).

442. Lockwood, K. G., John-Henderson, N. A. & Marsland, A. L. Early life socioeconomic status associates with interleukin-6 responses to acute laboratory stress in adulthood. *Physiol Behav* **188,** 212–220 (2018).

443. John-Henderson, N. A., Marsland, A. L., Kamarck, T. W., Muldoon, M. F. & Manuck, S. B. Childhood Socioeconomic Status and the Occurrence of Recent Negative Life Events as Predictors of Circulating and Stimulated Levels of Interleukin-6. *Psychosom Med* **78,** 91–101 (2016).

444. Ema, M. *et al.* cDNA cloning and structure of mouse putative Ah receptor. *Biochem Bioph Res Co* **184,** 246–253 (1992).

445. Burbach, K. M., Poland, A. & Bradfield, C. A. Cloning of the Ah-receptor cDNA reveals a distinctive ligand-activated transcription factor. *Proc National Acad Sci* **89,** 8185–8189 (1992).

446. Vicennati, V. Response of the Hypothalamic-Pituitary-Adrenocortical Axis to High-Protein/Fat and High-Carbohydrate Meals in Women with Different Obesity Phenotypes. *J Clin Endocr Metab* **87,** 3984–3988 (2002).

447. Vgontzas, A. N. *et al.* Sleep Apnea and Daytime Sleepiness and Fatigue: Relation to Visceral Obesity, Insulin Resistance, and Hypercytokinemia. *J Clin Endocrinol Metabolism* **85,** 1151–1158 (2000).

448. Mohamed-Ali, V. *et al.* Subcutaneous adipose tissue releases interleukin-6, but not tumor necrosis factor-alpha, in vivo. *J Clin Endocrinol Metabolism* **82,** 4196–200 (1997).

449. Shin, H. M. *et al.* Notch1 augments NF-κB activity by facilitating its nuclear retention. *Embo J* **25,** 129–138 (2005).

450. Fischer, A., Schumacher, N., Maier, M., Sendtner, M. & Gessler, M. The Notch target genes Hey1 and Hey2 are required for embryonic vascular development. *Gene Dev* **18,** 901–911 (2004).

451. Geiser, A. G. *et al.* Transforming growth factor beta 1 (TGF-beta 1) controls expression of major histocompatibility genes in the postnatal mouse: aberrant histocompatibility antigen expression in the pathogenesis of the TGF-beta 1 null mouse phenotype. *Proc National Acad Sci* **90,** 9944–9948 (1993).

452. Lienert, K. & Parham, P. Evolution of MHC class I genes in higher primates. *Immunol Cell Biol* **74,** 349–356 (1996).

453. Farah, M. J. The Neuroscience of Socioeconomic Status: Correlates, Causes, and Consequences. *Neuron* **96,** 56–71 (2017).

454. Wei, H. *et al.* Brain IL-6 elevation causes neuronal circuitry imbalances and mediates autism-like behaviors. *Biochim Biophys Acta* **1822,** 831–42 (2012).

455. Lutz, P.-E. *et al.* Association of a History of Child Abuse With Impaired Myelination in the Anterior Cingulate Cortex: Convergent Epigenetic, Transcriptional, and Morphological Evidence. *Am J Psychiat* **174,** 1185 1194 (2017).

456. Gauthier, J. *et al.* Truncating mutations in NRXN2 and NRXN1 in autism spectrum disorders and schizophrenia. *Hum Genet* **130,** 563–73 (2011).

457. Onay, H. *et al.* Mutation analysis of the NRXN1 gene in autism spectrum disorders. *Balk J Med Genet* **19,** 17–22 (2016).

458. Kim, H.-G. *et al.* Disruption of neurexin 1 associated with autism spectrum disorder. *Am J Hum Genet* **82,** 199–207 (2008).

459. Punzo, D. *et al.* Age-Related Changes in d-Aspartate Oxidase Promoter Methylation Control Extracellular d-Aspartate Levels and Prevent Precocious Cell Death during Brain Aging. *J Neurosci* **36,** 3064–3078 (2016).

460. Zafra, F. & Giménez, C. Glycine transporters and synaptic function. *Iubmb Life* **60,** 810–817 (2008).

461. Vandenberg, R. J., Ryan, R. M., Carland, J. E., Imlach, W. L. & Christie, M. J. Glycine transport inhibitors for the treatment of pain. *Trends Pharmacol Sci* **35,** 423–430 (2014).

462. Mostyn, S. N. *et al.* Identification of an allosteric binding site on the human glycine transporter, GlyT2, for bioactive lipid analgesics. *Elife* **8,** e47150 (2019).

463. Meur, A. M. *et al.* Reversible inhibition of the glycine transporter GlyT2 circumvents acute toxicity while preserving efficacy in the treatment of pain. *Brit J Pharmacol* **170,** 1053–1063 (2013).

464. Wood, P. L. The NMDA receptor complex: a long and winding road to therapeutics. *Idrugs Investigational Drugs J* **8,** 229–35 (2005).

465. Harvey, R. J. & Yee, B. K. Glycine transporters as novel therapeutic targets in schizophrenia, alcohol dependence and pain. *Nat Rev Drug Discov* **12,** 866–885 (2013).

466. Küpers, L. K. *et al.* Meta-analysis of epigenome-wide association studies in neonates reveals widespread differential DNA methylation associated with birthweight. *Nat Commun* **10,** 1893 (2019).

467. Gondalia, R. *et al.* Methylome-wide association study provides evidence of particulate matter air pollution-associated DNA methylation. *Environ Int* **132,** 104723 (2019).

468. Kennedy, E. M. *et al.* An integrated -omics analysis of the epigenetic landscape of gene expression in human blood cells. *Bmc Genomics* **19,** 476 (2018).

469. Liu, Y. *et al.* Methylomics of gene expression in human monocytes. *Hum Mol Genet* **22,** 5065 5074 (2013).

470. Mukherjee, R. *et al.* Non-Classical monocytes display inflammatory features: Validation in Sepsis and Systemic Lupus Erythematous. *Sci Rep-uk* **5,** 13886 (2015).

471. Ong, S.-M. *et al.* The pro-inflammatory phenotype of the human non-classical monocyte subset is attributed to senescence. *Cell Death Dis* **9,** 266 (2018).

472. Gainaru, G. *et al.* Increases in inflammatory and CD14dim/CD16pos/CD45pos patrolling monocytes in sepsis: correlation with final outcome. *Crit Care* **22,** 56 (2018).

473. Calle-Fabregat, C. de la, Morante-Palacios, O. & Ballestar, E. Understanding the Relevance of DNA Methylation Changes in Immune Differentiation and Disease. *Genes-basel* **11,** 110 (2020).

474. Walton, E. *et al.* Correspondence of DNA Methylation Between Blood and Brain Tissue and Its Application to Schizophrenia Research. *Schizophrenia Bull* **42,** 406–14 (2015).

475. Tesink, C. M. J. Y. *et al.* Neural correlates of pragmatic language comprehension in autism spectrum disorders. *Brain* **132,** 1941–1952 (2009).

476. Dube, S. R. *et al.* Long-Term Consequences of Childhood Sexual Abuse by Gender of Victim. *Am J Prev Med* **28,** 430–438 (2005).

477. Dube, S. R. *et al.* Cumulative childhood stress and autoimmune diseases in adults. *Psychosom Med* **71,** 243 250 (2009).

478. Szyf, M. The early life social environment and DNA methylation. *Epigenetics* **6,** 971–978 (2011).

479. Yang, B.-Z. *et al.* Child abuse and epigenetic mechanisms of disease risk. *Am J Prev Med* **44,** 101 107 (2013).

480. McGowan, P. O. *et al.* Epigenetic regulation of the glucocorticoid receptor in human brain associates with childhood abuse. *Nat Neurosci* **12,** 342 348 (2009).

481. Kobayashi, H. *et al.* Aberrant DNA methylation of imprinted loci in sperm from oligospermic patients. *Hum Mol Genet* **16,** 2542–2551 (2007).

482. Houshdaran, S. *et al.* Widespread Epigenetic Abnormalities Suggest a Broad DNA Methylation Erasure Defect in Abnormal Human Sperm. *Plos One* **2,** e1289 (2007).

483. Hammoud, S. S. *et al.* Distinctive chromatin in human sperm packages genes for embryo development. *Nature* **460,** 473–478 (2009).

484. Carrell, D. T. & Hammoud, S. S. The human sperm epigenome and its potential role in embryonic development. *Mhr Basic Sci Reproductive Medicine* **16,** 37–47 (2010).

485. Radford, E. J. *et al.* In utero undernourishment perturbs the adult sperm methylome and intergenerational metabolism. *Science* **345,** 1255903 (2014).

486. Manikkam, M., Tracey, R., Guerrero-Bosagna, C. & Skinner, M. K. Plastics Derived Endocrine Disruptors (BPA, DEHP and DBP) Induce Epigenetic Transgenerational Inheritance of Obesity, Reproductive Disease and Sperm Epimutations. *Plos One* **8,** e55387 (2013).

487. Wu, H., Hauser, R., Krawetz, S. A. & Pilsner, J. R. Environmental Susceptibility of the Sperm Epigenome During Windows of Male Germ Cell Development. *Curr Environ Heal Reports* **2,** 356–366 (2015).

488. Carone, B. R. *et al.* Paternally Induced Transgenerational Environmental Reprogramming of Metabolic Gene Expression in Mammals. *Cell* **143,** 1084–1096 (2010).

489. Rodgers, A. B., Morgan, C. P., Bronson, S. L., Revello, S. & Bale, T. L. Paternal stress exposure alters sperm microRNA content and reprograms offspring HPA stress axis regulation. *J Neurosci* **33,** 9003 9012 (2013).

490. Saavedra-Rodríguez, L. & Feig, L. A. Chronic social instability induces anxiety and defective social interactions across generations. *Biol Psychiat* **73,** 44 53 (2013).

491. Clarke, R. N., Klock, S. C., Geoghegan, A. & Travassos, D. E. Relationship between psychological stress and semen quality among in-vitro fertilization patients. *Hum Reprod* **14,** 753–758 (1999).

492. Fenster, L. *et al.* Effects of psychological stress on human semen quality. *J Androl* **18,** 194–202 (1997).

493. Marczylo, E. L., Amoako, A. A., Konje, J. C., Gant, T. W. & Marczylo, T. H. Smoking induces differential miRNA expression in human spermatozoa: A potential transgenerational epigenetic concern? *Epigenetics* **7,** 432–439 (2012).

494. Flanagan, J. M. *et al.* Intra- and Interindividual Epigenetic Variation in Human Germ Cells. *Am J Hum Genetics* **79,** 67–84 (2006).

495. Soubry, A. *et al.* Obesity-related DNA methylation at imprinted genes in human sperm: Results from the TIEGER study. *Clin Epigenetics* **8,** 51 (2016).

496. Donkin, I. *et al.* Obesity and Bariatric Surgery Drive Epigenetic Variation of Spermatozoa in Humans. *Cell Metab* **23,** 369–378 (2016).

497. Soubry, A., Hoyo, C., Jirtle, R. L. & Murphy, S. K. A paternal environmental legacy: Evidence for epigenetic inheritance through the male germ line. *Bioessays* **36,** 359–371 (2014).

498. Jones, P. A. & Baylin, S. B. The fundamental role of epigenetic events in cancer. *Nat Rev Genet* **3,** 415–428 (2002).

499. Anda, R. F. *et al.* Adverse Childhood Experiences and Smoking During Adolescence and Adulthood. *Jama* **282,** 1652–1658 (1999).

500. Roberts, A. L. *et al.* Women's Experience of Abuse in Childhood and Their Children's Smoking and Overweight. *Am J Prev Med* **46,** 249–258 (2014).

501. McLaughlin, K. A. *et al.* Childhood Adversities and Adult Psychiatric Disorders in the National Comorbidity Survey Replication II: Associations With Persistence of DSM-IV Disorders. *Arch Gen Psychiat* **67,** 124–132 (2010).

502. McLaughlin, K. A. *et al.* Childhood adversities and adult psychopathology in the National Comorbidity Survey Replication (NCS-R) III: Associations with functional impairment related to DSM-IV disorders. *Psychol Med* **40,** 847–859 (2010).

503. Widom, C. S., Czaja, S. J. & Dutton, M. A. Childhood victimization and lifetime revictimization. *Child Abuse Neglect* **32,** 785–796 (2008).

504. Desai, S., Arias, I., Thompson, M. P. & Basile, K. C. Childhood Victimization and Subsequent Adult Revictimization Assessed in a Nationally Representative Sample of Women and Men. *Violence Victims* **17,** 639–653 (2002).

505. Field, A. E. *et al.* Overweight, Weight Concerns, and Bulimic Behaviors Among Girls and Boys. *J Am Acad Child Adolesc Psychiatry* **38,** 754–760 (1999).

506. Bernstein, D. P. *et al.* Initial reliability and validity of a new retrospective measure of child abuse and neglect. *Am J Psychiat* **151,** 1132–1136 (1994).

507. Straus, M. A., Hamby, S. L., Finkelhor, D., Moore, D. W. & Runyan, D. Identification of Child Maltreatment With the Parent-Child Conflict Tactics Scales: Development and Psychometric Data for a National Sample of American Parents. *Child Abuse Neglect* **22,** 249–270 (1998).

508. Chartier, M. J., Walker, J. R. & Naimark, B. Health Risk Behaviors and Mental Health Problems as Mediators of the Relationship Between Childhood Abuse and Adult Health. *Am J Public Health* **99,** 847–854 (2009).

509. Dube, S. R., Felitti, V. J., Dong, M., Giles, W. H. & Anda, R. F. The impact of adverse childhood experiences on health problems: evidence from four birth cohorts dating back to 1900. *Prev Med* **37,** 268–277 (2003).

510. Radloff, L. S. The CES-D Scale. *Appl Psych Meas* **1,** 385–401 (1977).

511. Breslau, N., Peterson, E. L., Kessler, R. C. & Schultz, L. R. Short Screening Scale for DSM-IV Posttraumatic Stress Disorder. *Am J Psychiat* **156,** 908–911 (1999).

512. Schnurr, P., Vieilhauer, M., Weathers, F. & Findler, M. The Brief Trauma Questionnaire (BTQ). (1999).

513. Maksimovic, J., Gordon, L. & Oshlack, A. SWAN: Subset-quantile within array normalization for illumina infinium HumanMethylation450 BeadChips. *Genome Biol* **13,** R44 (2012).

514. Edgar, R., Domrachev, M. & Lash, A. E. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res* **30,** 207–210 (2002).

515. Peters, T. J. *et al.* De novo identification of differentially methylated regions in the human genome. *Epigenet Chromatin* **8,** 6 (2015).

516. Zou, H. & Hastie, T. Regularization and variable selection via the elastic net. *J Royal Statistical Soc Ser B Statistical Methodol* **67,** 301–320 (2005).

517. Wu, H. *et al.* Preconception urinary phthalate concentrations and sperm DNA methylation profiles among men undergoing IVF treatment: a cross-sectional study. *Hum Reprod* **32,** 2159–2169 (2017).

518. Jenkins, T. G. *et al.* Intra-sample heterogeneity of sperm DNA methylation. *Mhr Basic Sci Reproductive Medicine* **21,** 313 319 (2015).

519. WHO, W. H. O. WHO Laboratory Manual for the Examination and Processing of Human Semen. (2010).

520. Carver, J. A., Rekas, A., Thorn, D. C. & Wilson, M. R. Small Heat-shock Proteins and Clusterin: Intra- and Extracellular Molecular Chaperones with a Common Mechanism of Action and Function? *Iubmb Life* **55,** 661–668 (2003).

521. Ostermeier, G. C., Miller, D., Huntriss, J. D., Diamond, M. P. & Krawetz, S. A. Delivering spermatozoan RNA to the oocyte. *Nature* **429,** 154–154 (2004).

522. Stelzer, G. *et al.* The GeneCards Suite: From Gene Data Mining to Disease Genome Sequence Analyses. *Curr Protoc Bioinform* **54,** 1.30.1-1.30.33 (2016).

523. Smith, D. J. *et al.* Genome-wide analysis of over 106 000 individuals identifies 9 neuroticism-associated loci. *Mol Psychiatr* **21,** 749–757 (2016).

524. Seale, P. *et al.* PRDM16 controls a brown fat/skeletal muscle switch. *Nature* **454,** 961–967 (2008).

525. Cohen, P. *et al.* Ablation of PRDM16 and Beige Adipose Causes Metabolic Dysfunction and a Subcutaneous to Visceral Fat Switch. *Cell* **156,** 304–316 (2014).

526. Feinberg, J. I. *et al.* Paternal sperm DNA methylation associated with early signs of autism risk in an autism-enriched cohort. *Int J Epidemiol* **44,** 1199–1210 (2015).

527. Inslicht, S. S. *et al.* Increased cortisol in women with intimate partner violence-related posttraumatic stress disorder. *Psychoneuroendocrino* **31,** 825–838 (2006).

528. Griffin, M. G., Resick, P. A. & Yehuda, R. Enhanced Cortisol Suppression Following Dexamethasone Administration in Domestic Violence Survivors. *Am J Psychiat* **162,** 1192–1199 (2005).

529. Fernandez-Botran, R., Miller, J. J., Burns, V. E. & Newton, T. L. Correlations among inflammatory markers in plasma, saliva and oral mucosal transudate in post-menopausal women with past intimate partner violence. *Brain Behav Immun* **25,** 314–321 (2011).

530. Uddin, M. *et al.* Epigenetic and inflammatory marker profiles associated with depression in a community-based epidemiologic sample. *Psychol Med* **41,** 997–1007 (2011).

531. McGowan, P. O. & Kato, T. Epigenetics in mood disorders. *Environ Health Prev* **13,** 16–24 (2008).

532. Rodgers, A. B., Morgan, C. P., Leu, N. A. & Bale, T. L. Transgenerational epigenetic programming via sperm microRNA recapitulates effects of paternal stress. *Proc National Acad Sci* **112,** 13699–13704 (2015).

533. Pook, M., Krause, W. & Röhrle, B. Coping with infertility: distress and changes in sperm quality. *Hum Reprod* **14,** 1487–1492 (1999).

534. Zorn, B., Auger, J., Velikonja, V., Kolbezen, M. & Meden-Vrtovec, H. Psychological factors in male partners of infertile couples: relationship with semen quality and early miscarriage. *Int J Androl* **31,** 557–564 (2008).

535. Öst, A. *et al.* Paternal Diet Defines Offspring Chromatin State and Intergenerational Obesity. *Cell* **159,** 1352–1364 (2014).

536. Li, Y., Li, M., Liu, Y., Song, G. & Liu, N. A Microarray for MicroRNA Profiling in Spermatozoa from Adult Men Living in an Environmentally Polluted Site. *B Environ Contam Tox* **89,** 1111–1114 (2012).

537. Metzler-Guillemain, C. *et al.* Sperm mRNAs and microRNAs as candidate markers for the impact of toxicants on human spermatogenesis: an application to tobacco smoking. *Syst Biol Reprod Mec* **61,** 139–149 (2015).

538. Morgan, H. D., Santos, F., Green, K., Dean, W. & Reik, W. Epigenetic reprogramming in mammals. *Hum Mol Genet* **14,** R47–R58 (2005).

539. Sendler, E. *et al.* Stability, delivery and functions of human sperm RNAs at fertilization. *Nucleic Acids Res* **41,** 4104–4117 (2013).

540. Schagdarsurengin, U., Paradowska, A. & Steger, K. Analysing the sperm epigenome: roles in early embryogenesis and assisted reproduction. *Nat Rev Urol* **9,** 609–619 (2012).

541. Davis, T. L., Yang, G. J., McCarrey, J. R. & Bartolomei, M. S. The H19 methylation imprint is erased and re-established differentially on the parental alleles during male germ cell development. *Hum Mol Genet* **9,** 2885–2894 (2000).

542. Roberts, A. L., Lyall, K., Rich-Edwards, J. W., Ascherio, A. & Weisskopf, M. G. Association of Maternal Exposure to Childhood Abuse With Elevated Risk for Autism in Offspring. *Jama Psychiat* **70,** 508–515 (2013).

543. Dubowitz, H. *et al.* Type and Timing of Mothers' Victimization: Effects on Mothers and Children. *Pediatrics* **107,** 728–735 (2001).

544. Gavin, A. R., Hill, K. G., Hawkins, J. D. & Maas, C. The Role of Maternal Early-Life and Later-Life Risk Factors on Offspring Low Birth Weight: Findings From a Three-Generational Study. *J Adolescent Health* **49,** 166–171 (2011).

545. Bouvette-Turcot, A.-A., Bernier, A. & Meaney, M. Intergenerational Transmission of Psychosocial Risk: Maternal Childhood Adversity, Mother-Child Attachment, and Child Temperament. *Psychol Belg* **53,** 65–83 (2013).

546. Min, M. O., Singer, L. T., Minnes, S., Kim, H. & Short, E. Mediating Links Between Maternal Childhood Trauma and Preadolescent Behavioral Adjustment. *J Interpers Violence* **28,** 831–851 (2013).

547. Miranda, J. K., Osa, N. de la, Granero, R. & Ezpeleta, L. Multiple Mediators of the Relationships Among Maternal Childhood Abuse, Intimate Partner Violence, and Offspring Psychopathology. *J Interpers Violence* **28,** 2941–2965 (2013).

548. Cnattingius, S. The epidemiology of smoking during pregnancy: Smoking prevalence, maternal characteristics, and pregnancy outcomes. *Nicotine Tob Res* **6,** S125–S140 (2004).

549. Roberts, A. L. *et al.* Maternal experience of abuse in childhood and depressive symptoms in adolescent and adult offspring: A 21-year longitudinal study. *Depress Anxiety* **32,** 709–719 (2015).

550. Brand, S. R., Engel, S. M., Canfield, R. L. & Yehuda, R. The Effect of Maternal PTSD Following in Utero Trauma Exposure on Behavior and Temperament in the 9-Month-Old Infant. *Ann Ny Acad Sci* **1071,** 454–458 (2006).

551. Yehuda, R. & Bierer, L. M. The relevance of epigenetics to PTSD: Implications for the DSM-V. *J Trauma Stress* **22,** 427–434 (2009).

552. Non, A. L. Social epigenomics: are we at an impasse? *Epigenomics-uk* **0,** (2020).

553. Garagnani, P. *et al.* Methylation of ELOVL2 gene as a new epigenetic marker of age. *Aging Cell* **11,** 1132 1134 (2012).

554. Slieker, R. C., Relton, C. L., Gaunt, T. R., Slagboom, P. E. & Heijmans, B. T. Age-related DNA methylation changes are tissue-specific with ELOVL2 promoter methylation as exception. *Epigenet Chromatin* **11,** 25 (2018).

555. Bojesen, S. E., Timpson, N., Relton, C., Smith, G. D. & Nordestgaard, B. G. AHRR (cg05575921) hypomethylation marks smoking behaviour, morbidity and mortality. *Thorax* **72,** 646–653 (2017).

556. Miao, Z., Wang, Y. & Sun, Z. The Relationships Between Stress, Mental Disorders, and Epigenetic Regulation of BDNF. *Int J Mol Sci* **21,** 1375 (2020).

557. Tyrka, A. R. *et al.* Childhood maltreatment and methylation of FK506 binding protein 5 gene (FKBP5). *Dev Psychopathol* **27,** 1637–1645 (2015).

558. Provenzi, L., Giorda, R., Beri, S. & Montirosso, R. SLC6A4 methylation as an epigenetic marker of life adversity exposures in humans: A systematic review of literature. *Neurosci Biobehav Rev* **71,** 7–20 (2016).

559. Mansell, G. *et al.* Guidance for DNA methylation studies: statistical insights from the Illumina EPIC array. *Bmc Genomics* **20,** 366 (2019).

560. Saffari, A. *et al.* Estimation of a significance threshold for epigenome-wide association studies. *Genet Epidemiol* **42,** 20–33 (2018).

561. Benayoun, B. A., Pollina, E. A. & Brunet, A. Epigenetic regulation of ageing: linking environmental inputs to genomic stability. *Nat Rev Mol Cell Bio* **16,** 593–610 (2015).

562. Zhou, W. *et al.* DNA methylation loss in late-replicating domains is linked to mitotic cell division. *Nat Genet* **50,** 591–602 (2018).

563. Atlasi, Y. & Stunnenberg, H. G. The interplay of epigenetic marks during stem cell differentiation and development. *Nat Rev Genet* **18,** 643–658 (2017).

564. Alisch, R. S. *et al.* Age-associated DNA methylation in pediatric populations. *Genome Res* **22,** 623–632 (2012).

565. Martino, D. *et al.* Longitudinal, genome-scale analysis of DNA methylation in twins from birth to 18 months of age reveals rapid epigenetic change in early life and pair-specific effects of discordance. *Genome Biol* **14,** R42 (2013).

566. Pérez, R. F. *et al.* Longitudinal genome-wide DNA methylation analysis uncovers persistent early-life DNA methylation changes. *J Transl Med* **17,** 15 (2019).

567. Nagy, C. & Turecki, G. Sensitive periods in epigenetics: bringing us closer to complex behavioral phenotypes. *Epigenomics-uk* **4,** 445–457 (2012).

568. Yuan, T. *et al.* An Integrative Multi-scale Analysis of the Dynamic DNA Methylation Landscape in Aging. *Plos Genet* **11,** e1004996 (2015).

569. Simpkin, A. J. *et al.* Longitudinal analysis of DNA methylation associated with birth weight and gestational age. *Hum Mol Genet* **24,** 3752–3763 (2015).

570. Martino, D. J. *et al.* Evidence for age-related and individual-specific changes in DNA methylation profile of mononuclear cells during early immune development in humans. *Epigenetics* **6,** 1085–1094 (2011).

571. Acevedo, N. *et al.* Age-associated DNA methylation changes in immune genes, histone modifiers and chromatin remodeling factors within 5 years after birth in human blood leukocytes. *Clin Epigenetics* **7,** 34 (2015).

572. Allen, L. *et al.* Socioeconomic status and non-communicable disease behavioural risk factors in low-income and lower-middle-income countries: a systematic review. *Lancet Global Heal* **5,** e277–e289 (2017).

573. Giskes, K., Avendaňo, M., Brug, J. & Kunst, A. E. A systematic review of studies on socioeconomic inequalities in dietary intakes associated with weight gain and overweight/obesity conducted among European adults. *Obes Rev* **11,** 413–429 (2010).

574. Turrell, G. & Vandevijvere, S. Socio-economic inequalities in diet and body weight: evidence, causes and intervention options. *Public Health Nutr* **18,** 759–763 (2015).

575. Drenowatz, C. *et al.* Influence of socio-economic status on habitual physical activity and sedentary behavior in 8- to 11-year old children. *Bmc Public Health* **10,** 214 (2010).

576. Elhakeem, A., Cooper, R., Bann, D. & Hardy, R. Childhood socioeconomic position and adult leisure-time physical activity: a systematic review. *Int J Behav Nutr Phy* **12,** 92 (2015).

577. Stalsberg, R. & Pedersen, A. V. Effects of socioeconomic status on the physical activity in adolescents: a systematic review of the evidence. *Scand J Med Sci Spor* **20,** 368–383 (2010).

578. Jenkins, T. G., Liu, L., Aston, K. I. & Carrell, D. T. Pre-screening method for somatic cell contamination in human sperm epigenetic studies. *Syst Biol Reprod Mec* **64,** 146–155 (2018).

579. Rakyan, V. K., Down, T. A., Balding, D. J. & Beck, S. Epigenome-wide association studies for common human diseases. *Nat Rev Genet* **12,** 529 541 (2011).

580. Horvath, S. *et al.* Aging effects on DNA methylation modules in human brain and blood tissue. *Genome Biol* **13,** R97 (2012).

581. Fraser, H. B., Lam, L. L., Neumann, S. M. & Kobor, M. S. Population-specificity of human DNA methylation. *Genome Biol* **13,** R8 (2012).

582. Fagny, M. *et al.* The epigenomic landscape of African rainforest hunter-gatherers and farmers. *Nat Commun* **6,** 10047 (2015).

583. Galanter, J. M. *et al.* Differential methylation between ethnic sub-groups reflects the effect of genetic ancestry and environmental exposures. *Elife* **6,** e20532 (2017).

584. Moen, E. L. *et al.* Genome-Wide Variation of Cytosine Modifications Between European and African Populations and the Implications for Complex Traits. *Genetics* **194,** 987–996 (2013).

585. Rahmani, E. *et al.* Genome-wide methylation data mirror ancestry information. *Epigenet Chromatin* **10,** 1 (2017).

586. Gutierrez-Arcelus, M. *et al.* Passive and active DNA methylation and the interplay with genetic variation in gene regulation. *Elife* **2,** e00523 (2013).

587. Banovich, N. E. *et al.* Methylation QTLs Are Associated with Coordinated Changes in Transcription Factor Binding, Histone Modifications, and Gene Expression Levels. *Plos Genet* **10,** e1004663 (2014).

588. Hannon, E. *et al.* Methylation QTLs in the developing brain and their enrichment in schizophrenia risk loci. *Nat Neurosci* **19,** 48–54 (2016).

589. Dongen, J. van *et al.* Genetic and environmental influences interact with age and sex in shaping the human methylome. *Nat Commun* **7,** 11115 (2016).

590. Gaunt, T. R. *et al.* Systematic identification of genetic influences on methylation across the human life course. *Genome Biol* **17,** 61 (2016).

591. Kelsey, G., Stegle, O. & Reik, W. Single-cell epigenomics: Recording the past and predicting the future. *Science* **358,** 69–75 (2017).

592. Maurano, M. T. *et al.* Role of DNA Methylation in Modulating Transcription Factor Occupancy. *Cell Reports* **12,** 1184–1195 (2015).

593. Baubec, T. *et al.* Genomic profiling of DNA methyltransferases reveals a role for DNMT3B in genic methylation. *Nature* **520,** 243–247 (2015).

594. Pulecio, J., Verma, N., Mejía-Ramírez, E., Huangfu, D. & Raya, A. CRISPR/Cas9-Based Engineering of the Epigenome. *Cell Stem Cell* **21,** 431–447 (2017).

595. Liu, X. S. *et al.* Editing DNA Methylation in the Mammalian Genome. *Cell* **167,** 233-247.e17 (2016).

596. Tobi, E. W. *et al.* DNA methylation signatures link prenatal famine exposure to growth and metabolism. *Nat Commun* **5,** 5592 14 (2014).

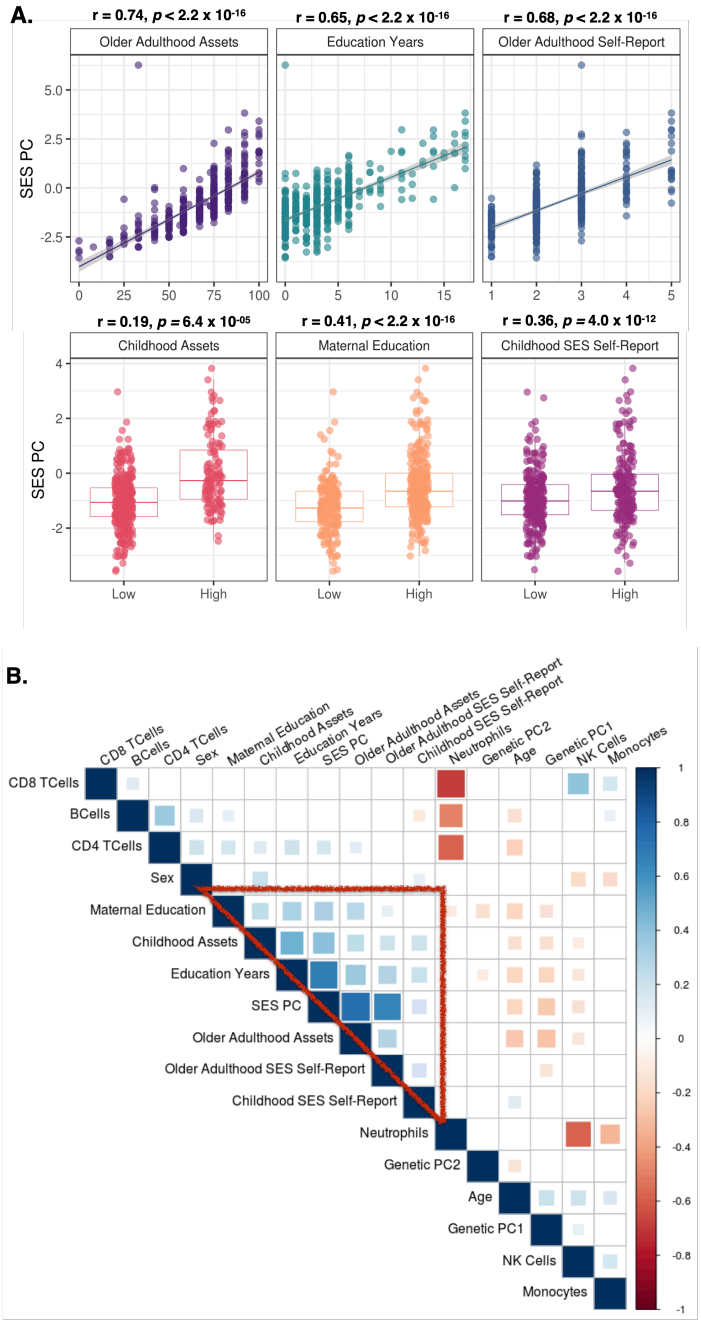
597. Belanger, C. F., Hennekens, C. H., Rosner, B. & Speizer, F. E. The Nurses' Health Study. *Ajn Am J Nurs* **78,** 1039–1040 (1978).

598. Setty, A. R., Curhan, G. & Choi, H. K. Smoking and the Risk of Psoriasis in Women: Nurses' Health Study II. *Am J Medicine* **120,** 953–959 (2007).

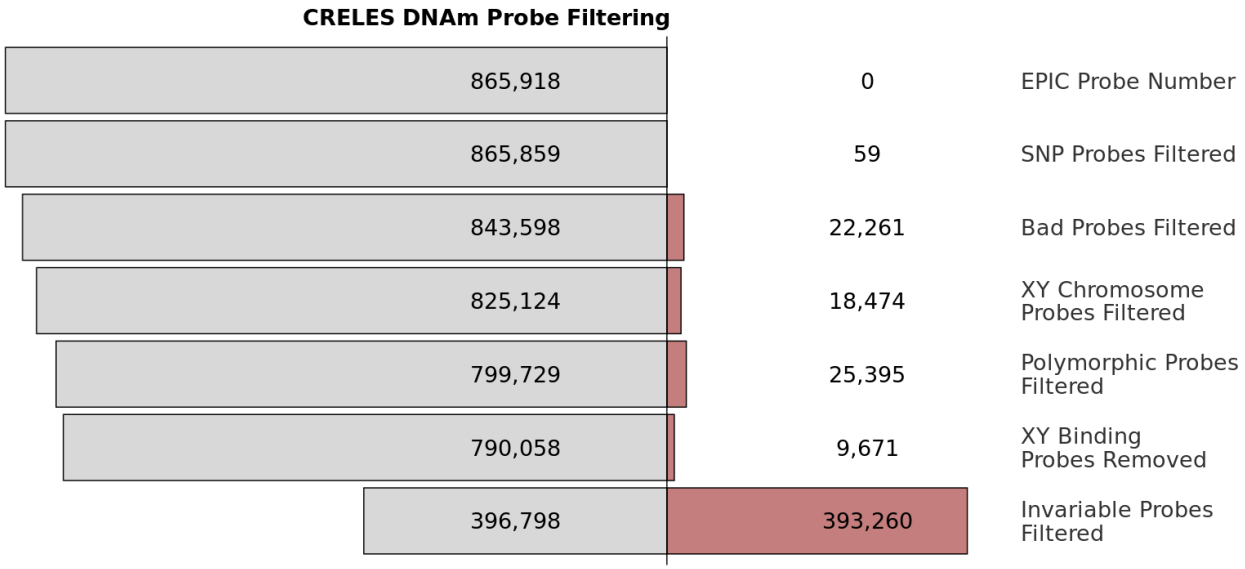
Appendices

Appendix A Supplementary Material for Chapter 2

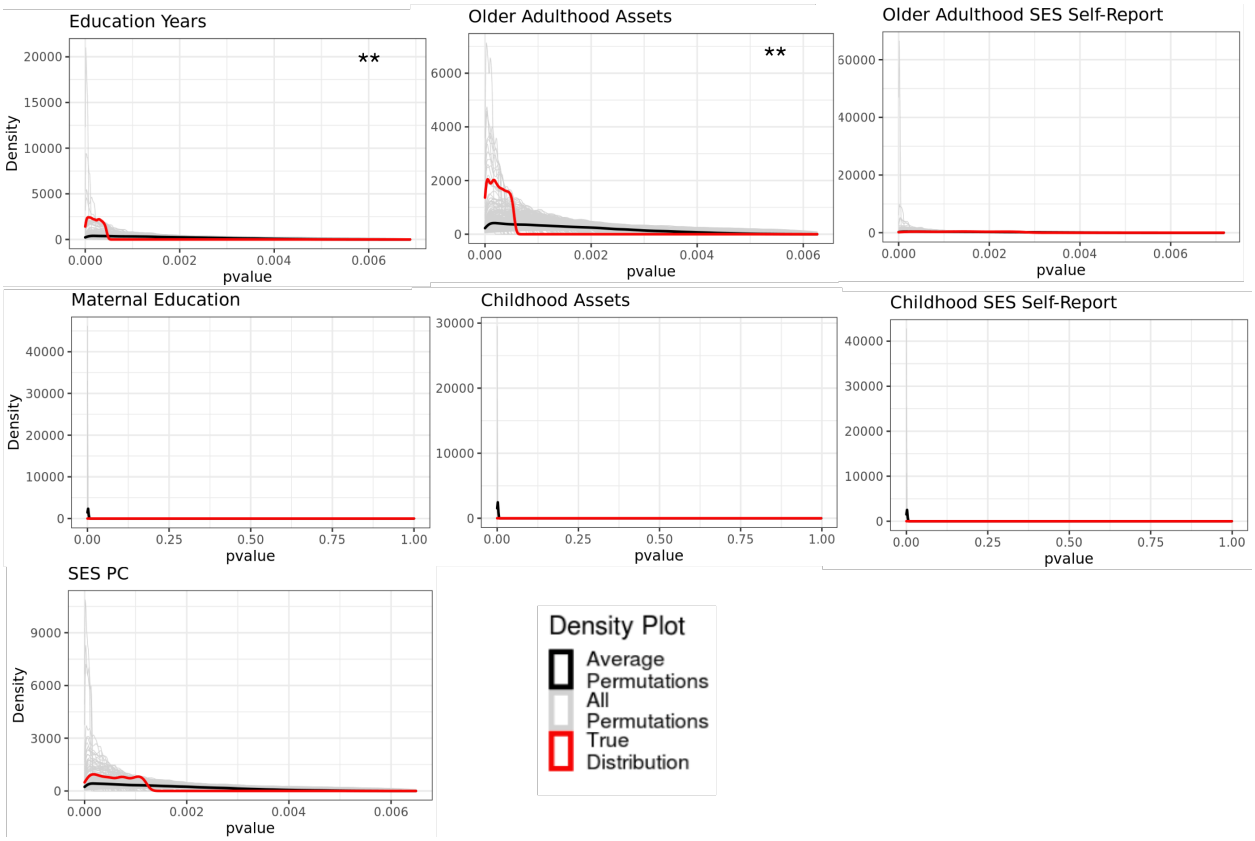
A.1 Supplementary Figures



Supplementary Figure 2.1 SES variables are not all correlated with one another but are all significantly correlated to the composite SES PC variable. (A) Associations between the SES PC composite variable and its association to each of the other six SES measures. Pearson's correlation was used for current life measures and ANOVA was used for the early life measures. (B) Pearson's correlation tests were performed on all variables where a heat map was produced with the color indicating the r-squared value and only associations with a p-value ≥ 0.05 are displayed.



Supplementary Figure 2.2 Probe summary of the pre-processing and normalization steps of the Illumina EPIC array.



Supplementary Figure 2.3 Randomized permutation analysis found more significant sites than expected by chance from the older adulthood assets and education years EWASs. These are *p*-value distributions for each SES variable summarizing 500 randomized permutations performed where the asterisk indicates EWASs with significantly more sites which had a *p*-value $\leq 1x10^{-6}$ than expected by chance. The grey lines represent each permutation, black lines represent the average of the permutations and the red lines represent the true distribution.

A.2 Supplementary Tables

Quality Control Step	Samples Removed	Probes Removed	Samples Remaining	Probes Remaining
Initial Numbers	-	-	511	618,540
Sample Outliers: Call Rate ≤ 0.97, p10 GC ≤ 0.5	9	-	502	618,540
EPIC and GSA Mismatched Samples ($r^2 \leq 0.8$)	6	-	496	618,540
Sex Mismatched Samples	4	-	492	618,540
Cluster Sep ≤ 0.45	-	6,636	492	611,904
AB R Mean ≤ 0.4	-	5,311	492	606,593
AB T Mean ≤ 0.2 or ≥ 0.8	-	398	492	606,195
Het Excess ≥ 0.2 or ≤ -0.3	-	12,070	492	594,125
HWE Chi-Square *p*-value ≤ 10^{-6}	-	81	492	594,044
AA Freq = 1 & AA T Mean ≥ 0.2	-	225	492	593,819
AA Freq = 1 & AA T Dev ≥ 0.04	-	18	492	593,801
BB Freq = 1 & BB T Mean ≤ 0.8	-	760	492	593,041
BB Freq = 1 & BB T Dev ≥ 0.04	-	68	492	592,973
AA or BB Freq = 0 & AB T Dev ≥ 0.5	-	0	492	592,973
AB Freq = 0 & MAF > 0	-	625	492	592,348
SNP Call Rate ≤ 0.97 or 10% GC	-	122	492	592,226
MAF ≤ 0.01	-	153,587	492	438,639
SNP LD Pruning (50kB, 5 SNPs, 2 VIF)	-	3,173	492	435,466
LD Pruning (Chr. 8, long LD, MHC)	-	13,183	492	422,283
Non-Autosomal Probes	-	1,800	492	420,483
Related Samples IBD PI_HAT ≥ 0.1875	10	-	482	420,483
MAF ≤ 0.05	-	280,554	482	139,929

Supplementary Table 2.1 Genotyping quality control and pre-processing summary of the Illumina GSA array.

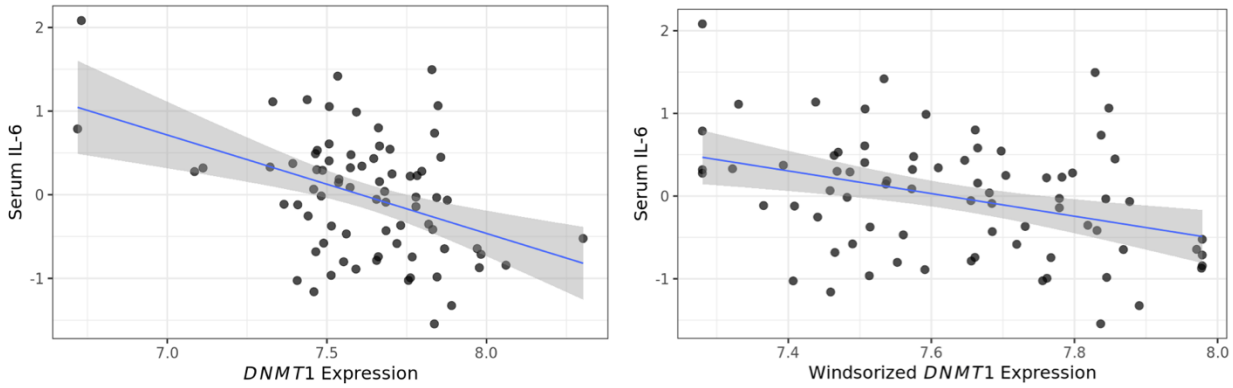
Supplementary Table 2.2 Associations of all SES measures with biomarkers and lifestyle behaviours. Pearson's correlations were performed on every of the SES variables to each of the biomarkers and lifestyle behaviours tested in the contribution analysis. CRP – C-reactive protein, DHEAS – dehydroepiandrosterone sulphate, Epi – epinephrine, Norepi – norepinephrine, BMI – body mass index.

	SES PC		**Older adulthood assets**		**Older adulthood SES self-report**		**Education years**		**Childhood SES self-report**		**Childhood assets**		**Maternal education**	
	p-value	r^2	*p*-value	r^2	*p*-value	r^2	*p*-value	r^2	*p*-value	r^2	*p*-value	r^2	*p*-value	r^2
CRP	$3.5x10^{-1}$	-0.04	$1.8x10^{-1}$	-0.06	$5.6x10^{-1}$	0.03	$2.5x10^{-1}$	-0.07	$2.3x10^{-1}$	0.06	$6.5x10^{-1}$	-0.02	$7.7x10^{-1}$	-0.01
Cortisol	$2.1x10^{-1}$	-0.08	$1.4x10^{-1}$	-0.09	$6.9x10^{-1}$	-0.02	$4.7x10^{-1}$	-0.04	$1.1x10^{-1}$	0.10	$4.2x10^{-1}$	-0.05	$7.4x10^{-1}$	0.02
DHEAS	$9.3x10^{-1}$	0.00	$5.4x10^{-1}$	-0.03	$2.5x10^{-1}$	-0.05	$3.4x10^{-1}$	0.04	$4.0x10^{-2}$*	-0.10*	$2.0x10^{-2}$*	-0.11*	$1.8x10^{-1}$	0.06
Epi	$1.2x10^{-1}$	-0.07	$4.4x10^{-2}$*	-0.09	$1.4x10^{-1}$	-0.07	$8.9x10^{-1}$	-0.01	$9.0x10^{-2}$	0.08	$5.4x10^{-1}$	-0.03	$9.0x10^{-1}$	0.01
Norepi	$4.6x10^{-1}$	-0.03	$4.0x10^{-1}$	-0.04	$3.1x10^{-1}$	0.05	$1.8x10^{-1}$	-0.04	$9.8x10^{-1}$	0.00	$4.6x10^{-2}$*	-0.09	$1.5x10^{-1}$	-0.07
Smoker	$9.7x10^{-2}$	-0.08	$3.0x10^{-1}$	-0.05	$4.0x10^{-2}$*	-0.10*	$1.8x10^{-1}$	-0.06	$1.8x10^{-1}$	-0.06	$2.0x10^{-3}$*	-0.14*	1	0
Drinker	$1.0x10^{-1}$	-0.07	$5.6x10^{-1}$	-0.03	$8.3x10^{-1}$	-0.01	$8.0x10^{-3}$*	-0.12*	$7.1x10^{-1}$	-0.02	$8.7x10^{-1}$	0.01	$2.0x10^{-3}$*	-0.15*
BMI	$6.0x10^{-4}$*	0.16*	$9.5x10^{-8}$*	0.24*	$3.8x10^{-1}$	0.04	$7.1x10^{-1}$	0.02	$8.1x10^{-1}$	-0.01	$3.0x10^{-3}$*	0.14*	$6.7x10^{-2}$	0.09

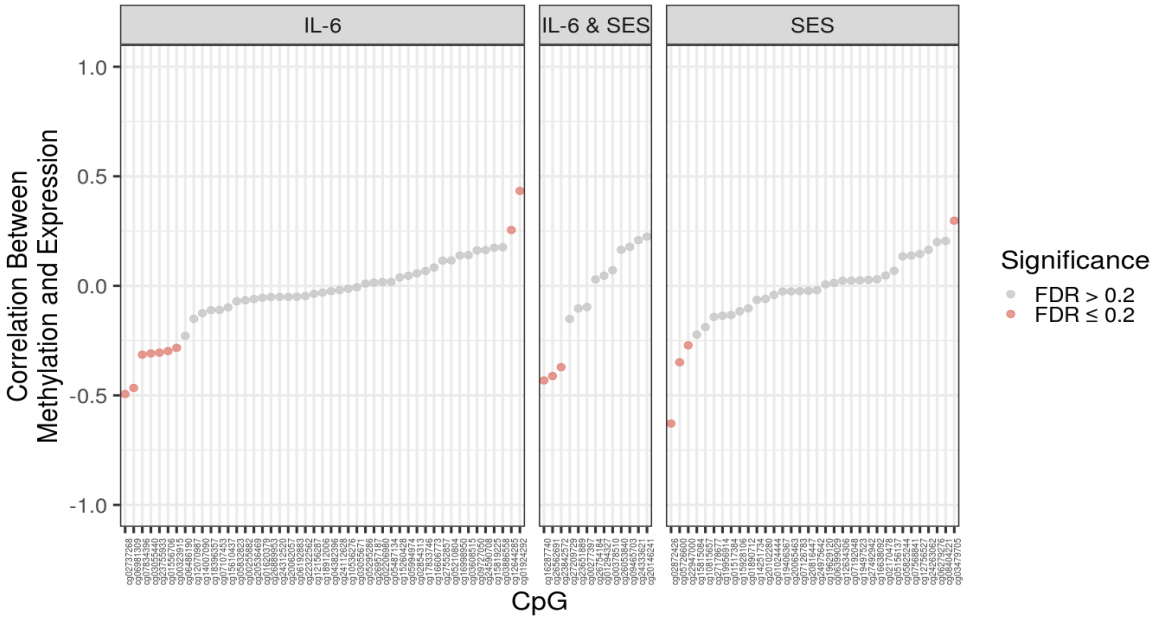
* *p*-value ≤ 0.05 or $| r^2 \geq 0.1 |$

Appendix B Supplementary Material for Chapter 3

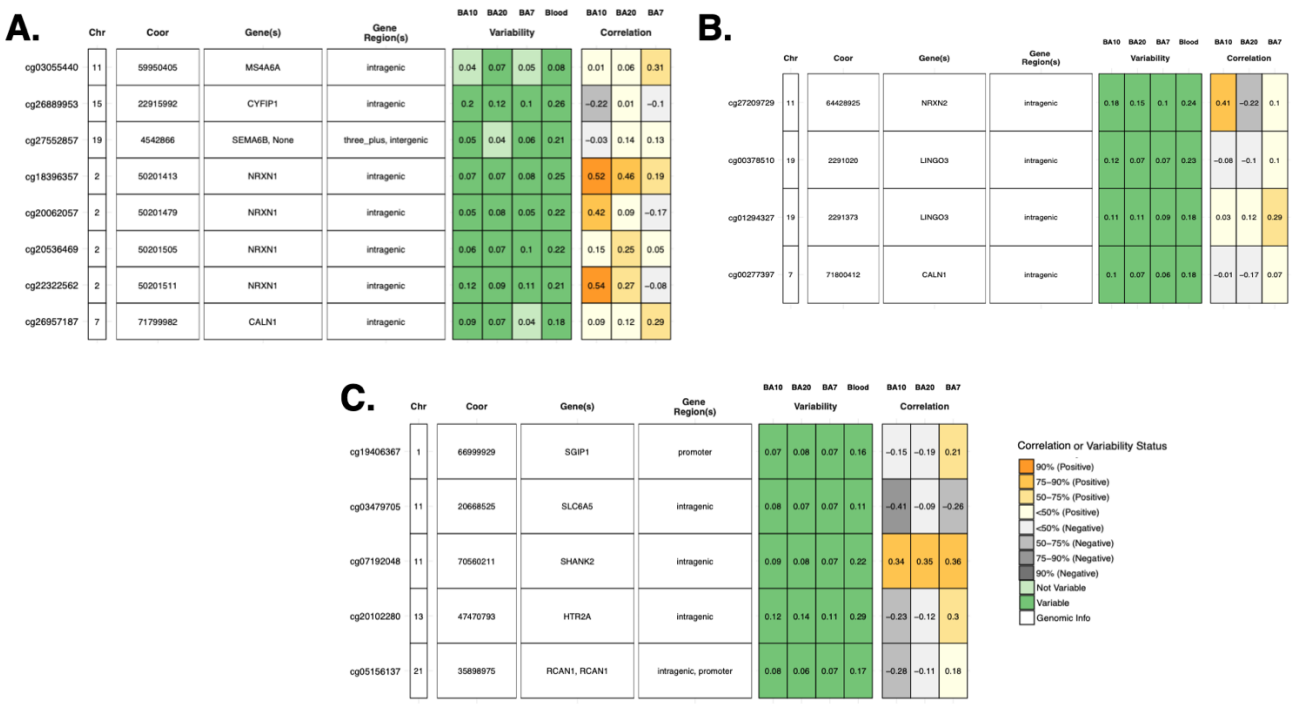
B.1 Supplementary Figures



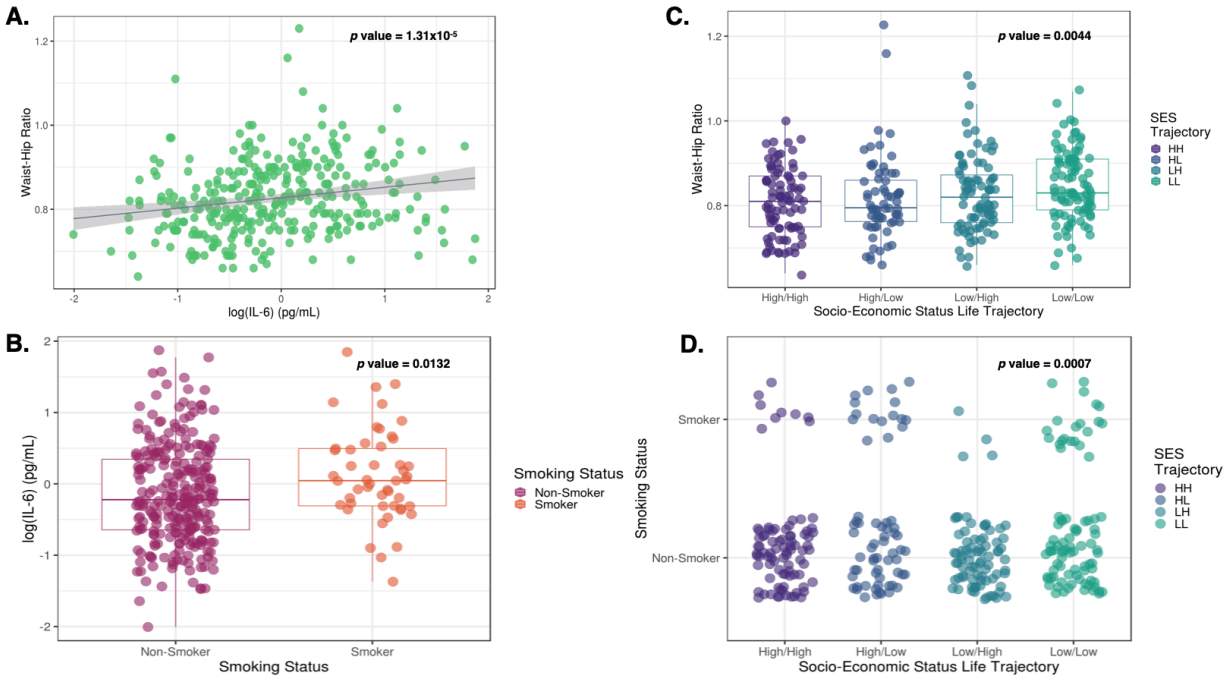
Supplementary Figure 3.1 *DNMT1* expression is negatively correlated with serum IL-6 levels in the CareGiver cohort. Scatter plot showing how increasing serum IL-6 levels were significantly associated with decreasing *DNMT1* expression in the CareGiver cohort (n = 78) unadjusted (left panel) and winsorized (right panel).



Supplementary Figure 3.2 Correlations between methylation of significant CpG sites and expression of genes CpG sites were located in using the CareGiver cohort data. UCSC was used to determine the gene name the CpG site was located in using human genome build GRCh37/hg19 which resulted in the investigation of 60/80 sites for the IL-6 analysis, 13/15 sites for the SES associated IL-6 hits, and 34/51 sites for the IL-6 independent SES analysis. Each panel represents the particular association the CpG site was found to be significant in with red color representing significant correlations (FDR ≤ 0.2).



Supplementary Figure 3.3 Methylation status of SES and/or IL-6 associated CpGs located in genes related to neurological function between blood and brain in BECon. Results from the BECon application where SES and/or IL-6 associated CpGs located in genes related to neurological functions were submitted and methylation levels from whole blood and three brain regions was correlated. These are tables breaking down the CpG annotation, variability measures in each tissue and correlation values between blood and each brain region (BA = Broadman Area). (A)Validated IL-6 associated CpGs. (B) IL-6 and SES associated CpGs. (C) IL-6 independent SES associated CpGs.



Supplementary Figure 3.4 Main effects were significantly associated with waist-to-hip ratio and smoking status. These figures display the associations with covariates which were significantly associated with my main effects of interest. Primarily waist-to-hip ratio and smoking status. (A) Scatter plot showing how serum IL-6 levels were significantly associated with WHR, the reported *p*-value resulted from a linear regression where age, sex and ethnicity were adjusted for. (B) Boxplot showing how serum IL-6 levels were significantly associated with smoking status, reported *p*-value is based on an ANCOVA adjusting for age, sex and ethnicity. (C) Boxplot showing how lifetime SES trajectory was significantly associated to WHR, reported *p*-value resulted from an ANCOVA where age, sex and ethnicity were adjusted for. (D) Scatter plot showing the distribution of individuals spanning the categories of lifetime SES trajectory and smoking status. Reported *p*-value resulted from a chi-squared test.

B.2 Supplementary Tables

Supplementary Table 3.1 Validation cohort demographic comparisons. Demographics of the Early Life and the validation cohort, CareGiver, after it was subset to match the Early Life demographics.

Demographic	Early Life	CareGiver
n	333	43
%Female	55	72
%Caucasian	73	70
Age Range (yrs)	15-55	22-54
Range log(IL-6 pg/ml)	-2.0 - 1.9	-1.5 - 1.5

Supplementary Table 3.2 Amount of DNAm sites which were significantly associated with serum IL-6 levels at various statistical thresholds. Statistical thresholds include FDR levels of 0.05, 0.10 and 0.20 with the respective nominal *p*-values. In addition, are thresholds of various effect size, delta beta (Δβ), cutoffs including |Δβ| at none, 0.03, 0.04 and 0.05.

FDR	≤ 0.05	≤ 0.10	≤ 0.20
Nominal *p* value	$\leq 1.9 \times 10^{-4}$	$\leq 8.2 \times 10^{-4}$	$\leq 4.1 \times 10^{-3}$
No Δβ	507	1,122	2,765
Δβ >= 3%	125	198	338
Δβ >= 4%	50	81	133
Δβ >= 5%	15	27	54

Supplementary Table 3.3 SES trajectory was associated with DNA methylation independent of serum IL-6 levels. Table summarizing the number of sites which passed various statistical thresholds with an effect size, delta beta (Δβ), | Δβ | ≥ 0.03. There are 51 unique sites which have an FDR ≤ 0.2 and an absolute Δβ ≥ 0.03 in any of the early/current life SES comparisons.

FDR		≤ 0.05	≤ 0.10	≤ 0.20
Nominal *p* value		≤4.3 x 10^{-6}	≤2.2 x 10^{-5}	≤2.9x 10^{-4}
No Δβ		15	31	267
abs(Δβ ≥ 3%)	Total	3	6	51
	HH–LL	0	0	5
	HH–HL	0	0	9
	HH - LH	2	5	28
	HL–LL	0	0	5
	HL-LH	0	2	30
	LL-LH	1	2	19

Supplementary Table 3.4 The results for DNAm sites found to be associated with serum IL-6 and SES trajectories and were also significantly correlated to the expression levels of the genes the sites were located in. UCSC was used to determine the gene name the CpG site was located in using human genome build GRCh37/hg19 which resulted in the investigation of 60/80 sites for the IL-6 analysis, 13/15 sites for the SES associated IL-6 hits, and 34/51 sites for the IL-6 independent SES analysis. This is a list of the DNAm sites which were significantly correlated to expression of the associated gene (FDR ≤ 0.2).

CpG	Previously Associated Variable	Associated Gene	Genomic Location	*p* value	FDR	Beta Coefficient
cg00323915	IL-6	GIMAP4	5'UTR	1.26e-02	6.86e-02	-1.05e-05
cg01556706	IL-6	FHIT	Body	8.56e-03	5.14e-02	-1.05e-03
cg01924292	IL-6	ADIPOR2	5'UTR	8.24e-05	1.32e-03	1.95e-04
cg02737268	IL-6	CDC25B	Body	5.07e-06	3.04e-04	-1.06e-03
cg03055440	IL-6	MS4A6A	5'UTR	6.26e-03	4.64e-02	-9.30e-06
cg06981309	IL-6	PLSCR1	5'UTR	1.93e-05	5.79e-04	-1.77e-04
cg07834396	IL-6	MSRB2	Body	5.40e-03	4.63e-02	-3.05e-05
cg12644285	IL-6	CHD2	3'UTR	2.53e-02	1.27e-01	3.50e-04
cg23755933	IL-6	TNPO3	3'UTR	6.96e-03	4.64e-02	-5.06e-04
cg16287740	IL-6 and SES	GOLGA3	3'UTR	8.78e-05	1.32e-03	-2.08e-04
cg23842572	IL-6 and SES	MPRIP	Body	8.87e-04	8.87e-03	-2.83e-04
cg26562691	IL-6 and SES	PRKCB	Body	1.99e-04	2.39e-03	-1.14e-04
cg02872426	SES	DDO	TSS200	9.27e-10	3.15e-08	-7.27e-03
cg03479705	SES	SLC6A5	Body	8.65e-03	9.81e-02	2.39e-03
cg05726600	SES	TMEM176B	TSS200	1.88e-03	3.19e-02	-7.57e-05
cg22947000	SES	BCMO1	TSS200	1.69e-02	1.43e-01	-1.70e-03

Supplementary Table 3.5 Contribution of smoking status and waist-to-hip ratio (WHR) to the DNAm associations observed with main effects. (A) The number of sites where at least 10% of the change in DNAm was associated with serum IL-6 levels were found to be contributed by WHR or smoking status. (B) The number of sites where at least 10% of the change in DNAm was associated with lifetime SES status is contributed by WHR or smoking status.

A.

		Contribution ≥ 10%	
IL-6	Base Model	WHR	Smoking
# of sites	80	17	15

B.

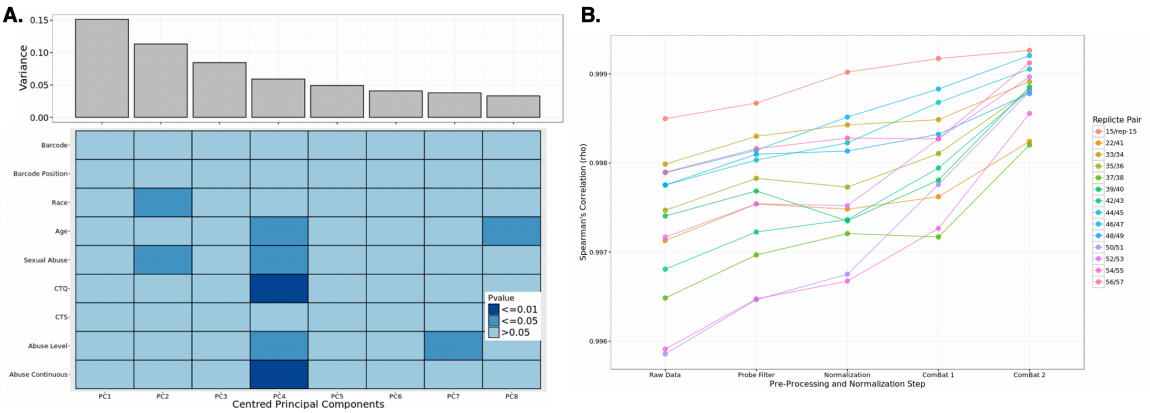
		Contribution ≥ 10%	
Early/Current Life SES	Base Model	WHR	Smoking
Unique	51	0	3
HH-LL	5	0	1
HH-HL	9	0	1
HH-LH	28	0	1
HL-LL	5	0	0
HL-LH	30	0	1
LL-LH	19	0	2

Supplementary Table 3.6 The change in methylation across serum IL-6 levels at 15 CpGs was contributed in part by lifetime SES trajectory status, smoking and/or WHR. Medium-confidence IL-6 associated CpGs were characterized by their significance level, effect size, genomic association, SES and lifestyle behaviour contributions. UCSC was used to determine the gene name the CpG site was located in using human genome build GRCh37/hg19. The max SES trajectory delta beta value reports the highest effect size of the six SES comparisons.

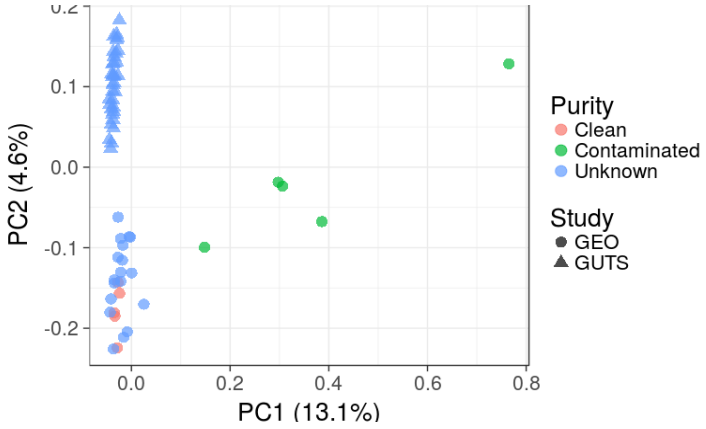
CpG	*p* value	FDR	IL-6 Δβ	Max SES Trajectory Δβ	Gene Name	SES Trajectory Adjusted *p* value	SES Trajectory % Contribution	Smoking Adjusted *p* value	Smoking % Contribution	WHR Adjusted *p* value	WHR % Contribution
cg00277397	9.55e-04	1.05e-01	-3.30%	2.1% (HH-LH)	CALN1	9.86e-02	3.84	8.80e-02	7.22	7.31e-02	14.44
cg00378510	1.36e-03	1.25e-01	-4.60%	-4.2% (LL-LH)	LINGO3	9.86e-02	7.60	2.02e-03	17.72	2.39e-01	7.76
cg01294327	2.59e-03	1.66e-01	-4.20%	-3.6% (LL-LH)		1.69e-01	6.53	3.31e-04	14.29	3.25e-01	6.43
cg06766960	3.93e-03	1.97e-01	6.61%	-5.0% (HH-LL)		1.69e-01	15.87	4.11e-01	7.89	2.47e-01	8.18
cg09465703	1.41e-03	1.26e-01	-3.50%	-2.1% (HL-LH)	JMJD8	1.69e-01	1.01	4.62e-01	7.46	2.70e-01	3.06
cg12073466	4.99e-04	8.09e-02	-4.85%	-3.2% (HL-LH)		1.81e-01	-0.74	9.85e-01	17.64	9.87e-02	8.59
cg16287740	2.37e-04	5.77e-02	4.70%	3.3% (HL-LH)	TNPO3	1.81e-01	-1.97	4.06e-01	12.47	4.44e-01	2.60
cg20146241	1.00e-03	1.08e-01	3.14%	-1.7% (HH-HL)	RCAN3	1.81e-01	12.08	3.30e-05	29.21	7.31e-02	9.54
cg23651889	5.10e-04	8.17e-02	3.90%	-3.5% (HH-HL)	LPP/miR-28-5p	9.86e-02	14.75	8.93e-05	32.36	1.58e-01	4.08
cg23842572	1.91e-06	3.62e-03	3.20%	-2.0% (HH-HL)	MPRIP	9.86e-02	6.72	1.34e-06	14.17	4.31e-02	6.17
cg24333621	6.92e-04	9.28e-02	-5.60%	2.9% (HH-LH)	SLC35B4	1.69e-01	1.52	4.62e-01	-10.32	6.21e-01	-2.86
cg26053840	6.88e-04	9.25e-02	3.20%	2.1% (HH-HL)	GLO1	1.75e-01	11.77	7.04e-02	15.81	2.33e-02	14.49
cg26562691	5.27e-04	8.29e-02	4.80%	3.1% (HH-HL)	PRKCB	1.81e-01	7.05	1.48e-01	6.69	8.11e-02	7.16
cg26754184	6.18e-04	8.88e-02	3.30%	1.9% (HH-HL)	OR3A2	1.69e-01	12.96	2.83e-01	6.22	2.65e-01	4.52
cg27209729	4.03e-04	7.27e-02	-5.20%	2.5% (HH-LH)	NRXN2	1.69e-01	-0.17	2.98e-03	17.26	1.72e-01	2.10

Appendix C Supplementary Material for Chapter 4

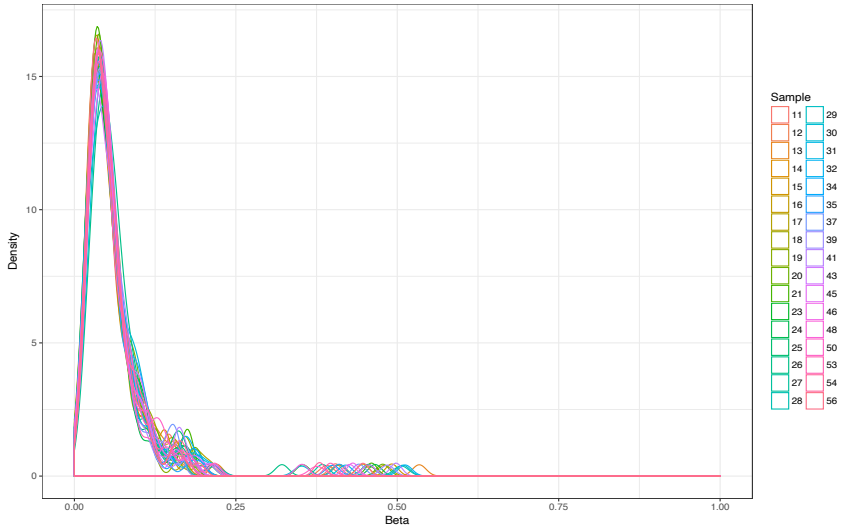
C.1 Supplementary Figures



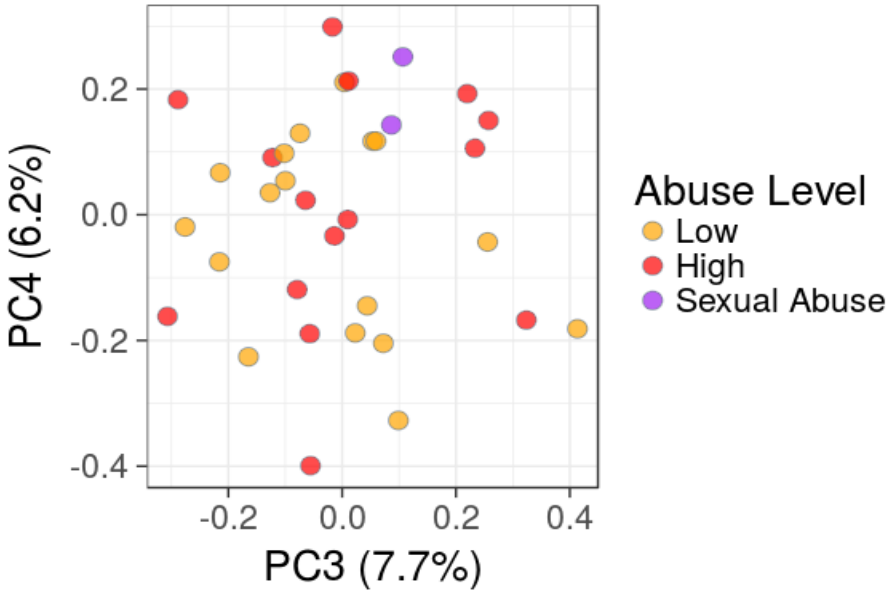
Supplementary Figure 4.1 Batch effects removed and correlations between replicate samples increased after normalization and pre-processing of 450K array DNA methylation data. (A) After probe filtering, quantile inter-sample normalization, SWAN intra-sample normalization and ComBat correction for batch effects there were no significant correlations between batch variables (Barcode and Barcode Position) and variation present in DNAm data as found using PCA. (B) Correlation measurements between replicate samples throughout normalization. All replicates display the highest correlations by the end of the pipeline. Abbreviations: CTQ: Childhood Trauma Questionnaire; CTS: Conflict Tactic Scales.



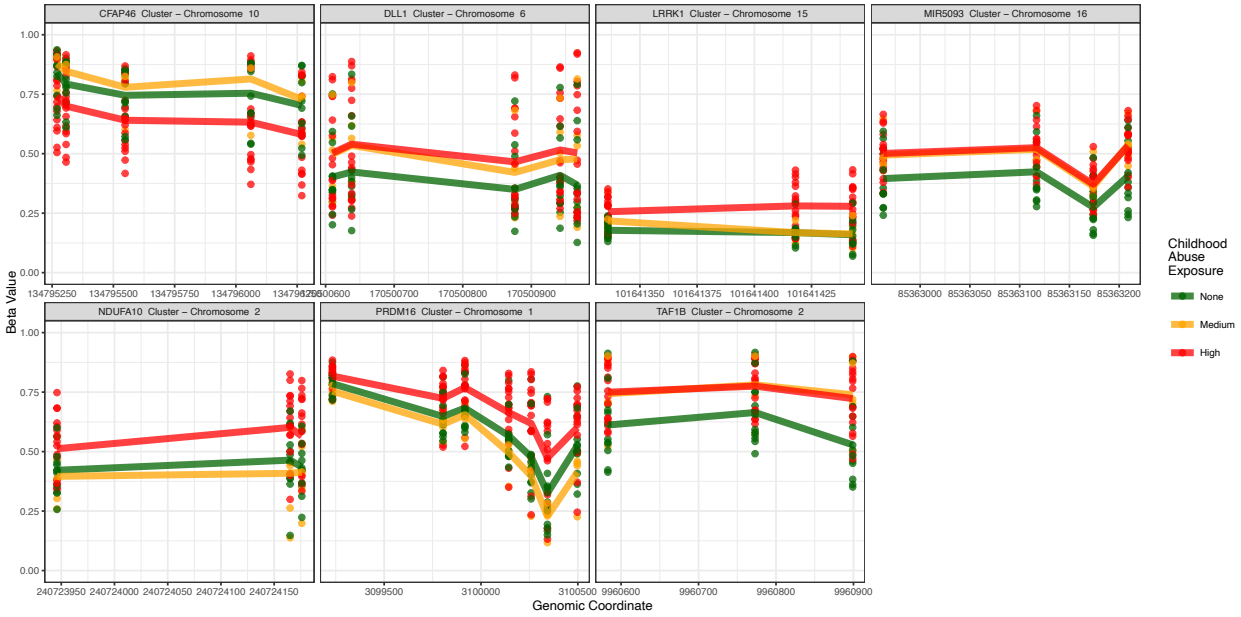
Supplementary Figure 4.2 The DNA methylation data of GUTS samples clustered with publicly available, known purified sperm Illumina 450K DNA methylation data (GEO accession GSE108058(514)**).** The GSE108058 dataset contained five contaminated semen samples, five purified semen samples and 20 semen samples of unknown purity. We merged this dataset with our own data and performed principal component analysis. The vast majority of variation in methylation is associated with tissue heterogeneity, therefore the first few PCs should be correlated with the purity of the semen samples. Plotting PC1 against PC2 (for visualization purposes), our samples cluster with the pure semen samples, providing evidence that we successfully purified the semen samples.



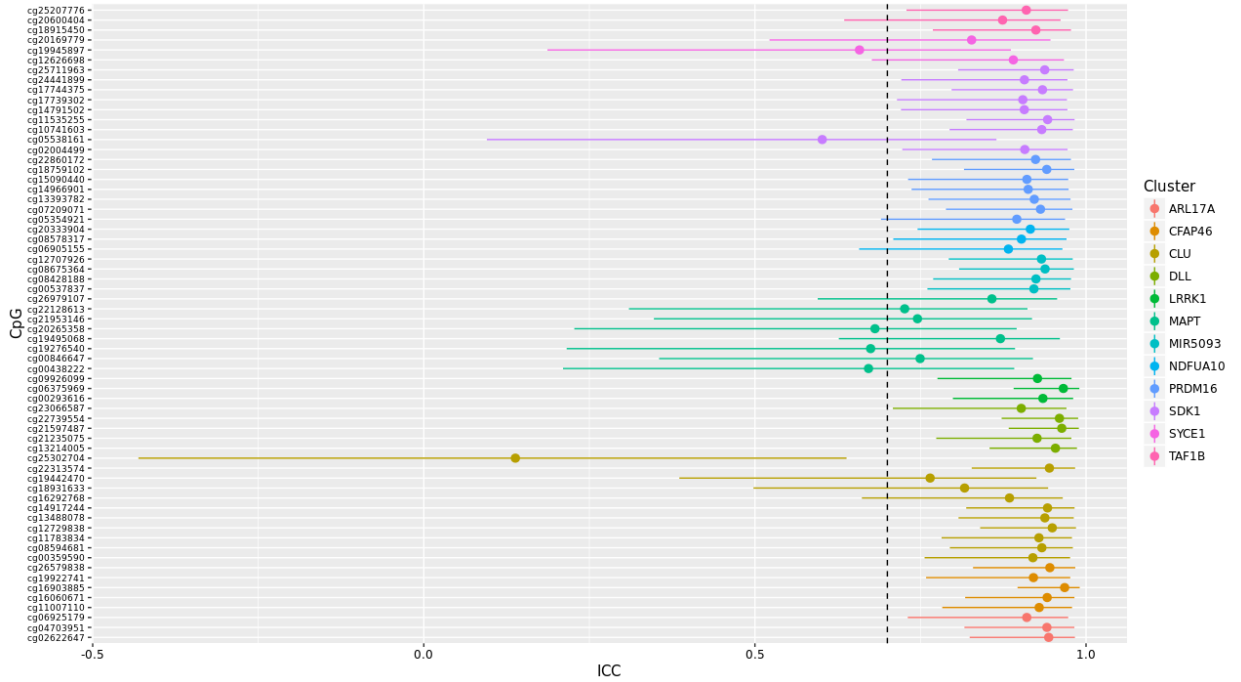
Supplementary Figure 4.3 Beta value distribution of probes within two paternally expressed imprinting control regions (ICR) (*HYMAI* - ICR: chr6: 144326000 – 144330000; *GNAS-AS* - ICR: chr20: 57425000 – 57429000) shows dominant peak at 0. Within the listed genomic regions, we plotted the beta values of all probes underlying the *HYMAI* ICR (61 probes) and the *GNAS-AS* ICR (69 probes) for each individual in our study. Both of these regions are paternally expressed, and therefore if our samples contain purified haploid male gametes, we would anticipate these regions would be fully unmethylated (as opposed to hemi-methylated in somatic tissue).



Supplementary Figure 4.4 Scatterplot of the loading on PC4 (y-axis) versus PC3 (x-axis, for visualization) for each individual show associations with PC4 are not driven by a small subset of samples. The two men who were exposed to sexual abuse are indicated in purple. The scatterplot suggests that these men are not outliers amongst men who experienced abuse (shown in red).



Supplementary Figure 4.5 Remaining seven of 12 statistically significant childhood abuse DMRs. Differentially methylated regions (DMRs) were defined as regions that differed statistically by abuse exposure at an FDR ≤ 0.05, had a mean $\Delta\beta \geq 5\%$ across probes, and were confirmed using replicates. Other than the *PRDM16* DMR, which is located in the body of that gene, the remaining DMRs are intergenic in location and are labelled with the name of the gene whose transcription start site is closest.

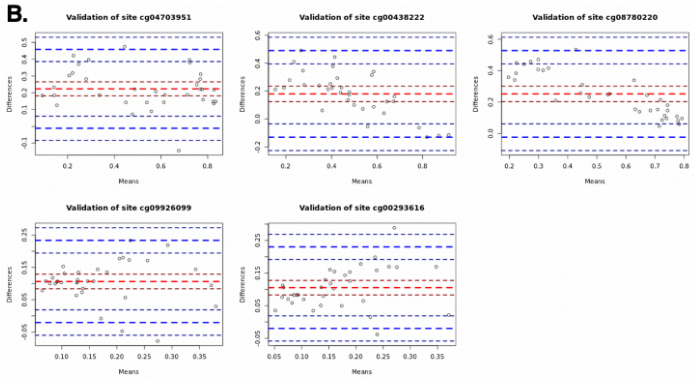


Supplementary Figure 4.6 Probes located in DMRs have high intraclass correlation (ICC) when comparing biological replicate sample DNA methylation values. 90% of probes in significant DMRs have ICC ≥ 0.7 (dashed line). A single ICC score was obtained for each probe using 24 samples (2 biological replicates per participant). A value of 1 would indicate the biological replicates are in perfect concordance with one another. Whiskers represent 95% confidence intervals.

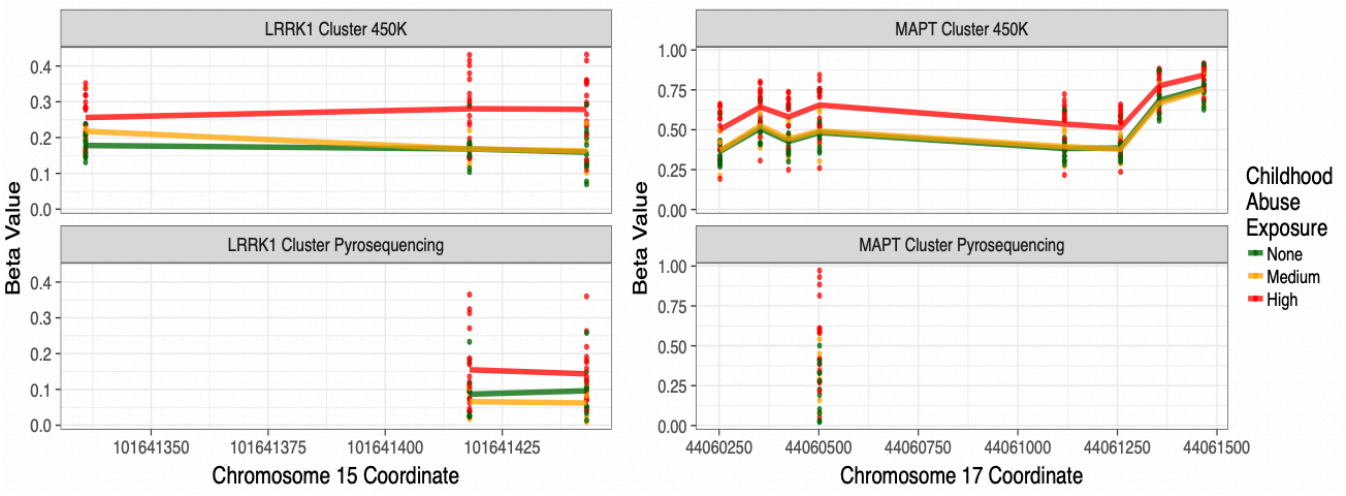
A.

CpG Site	Correlation Value	P-Value	FDR
cg04703951	0.893	1.2×10^{-12}	4.7×10^{-12}
cg00438222	0.830	1.3×10^{-09}	3.8×10^{-09}
cg08780220	0.935	4.4×10^{-16}	2.2×10^{-15}
cg09926099	0.749	3.5×10^{-07}	2.2×10^{-15}
cg00293616	0.748	3.6×10^{-07}	7.0×10^{-07}

B.



Supplementary Figure 4.7 Measurements of DNAm using the 450K array and the pyrosequencing platform were highly correlated. For a certain subset of sites, pyrosequencing assays were designed to confirm DNA methylation measurements to ensure findings were not specific to the 450K array platform. (A) Measurement correlations and *p*-values obtained using Spearman's rank method. (B) Bland-Altman plots of DNA methylation sites validated in pyrosequencing made by comparing the difference in measurements by the average value between platforms.

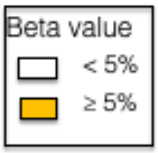


Supplementary Figure 4.8 DNA methylation measurements were confirmed at additional CpG sites measured with pyrosequencing within DMRs identified with the Illumina Infinium HumanMethylation450 BeadChip. Of the sites selected for confirmation with pyrosequencing, these are the subset of sites which were found in a subset of significant DMRs.

C.2 Supplementary Tables

Sample	Median Beta Value	
	GNAS-AS ICR	*HYMAI* ICR
11	0.047	0.039
12	0.047	0.037
13	0.047	0.033
14	0.049	0.037
15	0.045	0.035
16	0.047	0.035
17	0.049	0.038
18	0.049	0.038
19	0.050	0.039
20	0.047	0.039
21	0.046	0.037
23	0.051	0.032
24	0.049	0.040
25	0.047	0.038
26	0.056	0.046
27	0.049	0.040
28	0.050	0.048

Sample	Median Beta Value	
	GNAS-AS ICR	*HYMAI* ICR
29	0.051	0.040
30	0.050	0.045
31	0.050	0.039
32	0.047	0.040
34	0.049	0.034
35	0.051	0.034
37	0.052	0.042
39	0.051	0.041
41	0.048	0.038
43	0.053	0.038
45	0.050	0.038
46	0.047	0.037
48	0.050	0.038
50	0.048	0.039
53	0.046	0.038
54	0.051	0.043
56	0.048	0.035



Supplementary Table 4.1 Methylation status of two paternally expressed imprinting control regions (ICR) (*HYMAI* - ICR: chr6: 144326000 – 144330000, 61 probes; *GNAS-AS* - ICR: chr20: 57425000 – 57429000, 69 probes) displayed values close to 0 suggesting a highly purified sperm sample. Within the listed genomic regions, we examined all probes underlying the *HYMAI* ICR (61 probes) and the *GNAS-AS* ICR (69 probes) which were measured on the Illumina 450K array and calculated the median beta value for each region for each individual in our study. Both of these regions are paternally expressed, and therefore if our samples contain purified haploid male gametes, we would anticipate these regions would be fully unmethylated (as opposed to hemi-methylated in somatic tissue).

	Probes with largest PC4 scores				
	1%	2%	3%	4%	5%
	N				
All probes	8,794	17,590	26,384	35,180	43,974
Restricted to probes with Δβ≥5%	1,137	1,443	1,617	1,718	1,793
Restricted to probes with Δβ≥10%	97	110	119	122	126

Supplementary Table 4.2 Probes with the highest loading on PC4 and meeting minimum values for differences in DNAm β values (Δβ) by childhood abuse exposure. Probes with top PC4 scores as defined by percentiles and Δβ levels. Δβ values were calculated by taking the absolute difference between the mean β for high versus no childhood abuse. Probes in the 1st percentile of positive or negative loading on PC4 and with Δβ ≥ 5% are highlighted in blue.

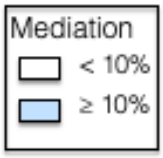
	Genomic Information							
Cluster Name	**Location**	**Width**	**Gene Association**	**Gene Location**	**Nearest TSS Gene**	**Max Distance to Nearest TSS (bp)**	**Enhancer**	**CpG Island**
ARL17A	chr17:43578568-43578911	344			AF119889	-2056		
MAPT	chr17:44060252-44061467	1216	MAPT	Body	MAPT	1236	Yes	Yes
CLU	chr8:27467783-27470629	2847	CLU	Body,1stExon,5'UTR,TSS200,TSS1500	CLU	-1361	Yes	Yes
LRRK1	chr15:101641336-101641443	108			LRRK1	48016		
PRDM16	chr1:3099233-3100497	1265	PRDM16	Body	MIR4251	55959		Yes
CFAP46	chr10:134795270-134796269	1000			AK125849	17232		Yes
MIR5093	chr16:85362963-85363209	247			LOC727710	-26207	Yes	Yes
TAF1B	chr2:9960583-9960899	317			TAF1B	-22987		
DLL1	chr6:170500610-170500967	358			LOC154449	71047		
SYCE1	chr10:135381727-135381914	188	SYCE1	5'UTR	SYCE1	1149		Yes
NDFUA10	chr2:240723946-240724176	231			LOC150935	39623		
SDK1	chr7:4242866-4244372	1507	SDK1	Body	SDK1	-24249		

Supplementary Table 4.3 Genomic annotation of regions significantly differentially methylated by childhood abuse (DMRs). DMR information was obtained using annotation provided from Illumina for the Infinium 450K array.

	p value	FDR
cg04703951	4.67E-03	8.38E-03
Pos2	2.87E-03	6.45E-03
Pos3	1.02E-02	1.31E-02
Pos4	5.59E-03	8.38E-03
Pos5	1.66E-03	6.45E-03
cg08780220	1.95E-03	6.45E-03
cg00438222	2.26E-03	6.45E-03
cg09926099	1.78E-02	2.00E-02
cg00293616	5.77E-02	5.77E-02

Supplementary Table 4.4 Four additional sites detected by pyrosequencing but not measured by the 450K array within the *ARL17A* DMR were significantly associated with childhood abuse exposure in the same direction as original DMR sites. Linear regression for novel sites identified by pyrosequencing within the *ARL17A* DMR was performed and found to be significantly associated.

Cluster Name	Number of probes	Model 1: BMI and smoking	Model 2: Depressive and PTSD symptoms	Model 3: Lifetime Trauma
ARL17A	3	5.67%	11.23%	16.65%
CFAP46	5	-7.74%	-8.97%	-20.44%
CLU	11	5.76%	13.63%	5.28%
DLL1	5	-3.09%	9.34%	23.70%
LRRK1	3	0.12%	0.79%	2.62%
MAPT	8	1.92%	-1.13%	3.28%
MIR5093	4	1.18%	3.72%	5.14%
NDFUA10	3	-2.35%	5.68%	15.80%
PRDM16	7	-8.22%	-12.94%	-25.86%
SDK1	8	-4.45%	-15.01%	2.87%
SYCE1	3	-2.90%	-9.23%	-14.90%
TAF1B	3	-8.93%	5.29%	14.04%



Supplementary Table 4.5 Mean percent mediation across probes within childhood abuse differentially methylated regions (DMR) by adulthood BMI and smoking, mental health symptoms, and lifetime trauma exposure. Percent mediation = ([beta estimate $_{\text{base model}}$-beta estimate $_{\text{adjusted model}}$]/beta estimate $_{\text{base model}}$)*100. Negative percentages indicate a stronger association of childhood abuse with DNAm in adjusted models compared to the base model.

Probe Name	DMR Name	Model 1: BMI and smoking	Model 2: Depressive and PTSD symptoms	Model 3: Lifetime Trauma
cg02622647	ARL17A	5.57%	11.10%	16.07%
cg04703951	ARL17A	4.45%	10.62%	15.68%
cg06925179	ARL17A	6.99%	11.97%	18.20%
cg11007110	CFAP46	-6.19%	-6.01%	-14.57%
cg16060671	CFAP46	-8.50%	-9.08%	-20.00%
cg16903885	CFAP46	-7.56%	-9.03%	-19.21%
cg19922741	CFAP46	-8.45%	-11.76%	-26.80%
cg26579838	CFAP46	-8.00%	-8.98%	-21.62%
cg00359590	CLU	4.34%	10.94%	4.06%
cg08594681	CLU	6.00%	14.37%	6.00%
cg11783834	CLU	4.92%	13.46%	5.62%
cg12729838	CLU	5.84%	14.03%	5.48%
cg13488078	CLU	5.68%	13.68%	4.99%
cg14917244	CLU	3.69%	12.05%	3.35%
cg16292768	CLU	3.55%	13.88%	-0.24%
cg18931633	CLU	-2.63%	15.48%	14.68%
cg19442470	CLU	5.59%	18.00%	4.74%
cg22313574	CLU	4.82%	14.86%	5.24%
cg25302704	CLU	21.63%	9.15%	4.17%
cg13214005	DLL	-1.83%	8.65%	23.49%
cg21235075	DLL	-4.30%	11.68%	26.98%
cg21597487	DLL	-4.12%	11.34%	25.93%
cg22739554	DLL	-3.23%	4.47%	20.08%
cg23066587	DLL	-1.96%	10.56%	22.05%
cg00293616	LRRK1	0.08%	2.11%	3.30%
cg06375969	LRRK1	0.00%	-1.32%	2.70%
cg09926099	LRRK1	0.30%	1.57%	1.85%
cg00438222	MAPT	1.66%	1.33%	5.44%
cg00846647	MAPT	2.82%	1.05%	2.05%
cg19276540	MAPT	2.06%	1.23%	8.14%
cg19495068	MAPT	1.08%	-6.12%	-2.56%
cg20265358	MAPT	1.82%	-2.35%	1.84%
cg21953146	MAPT	1.74%	-0.26%	3.79%
cg22128613	MAPT	1.54%	0.38%	5.99%
cg26979107	MAPT	2.61%	-4.30%	1.55%
cg00537837	MIR5093	1.45%	4.88%	7.08%
cg08428188	MIR5093	1.06%	3.82%	4.36%
cg08675364	MIR5093	1.09%	2.48%	4.11%
cg12707926	MIR5093	1.10%	3.69%	5.01%
cg06905155	NDFUA10	-3.85%	8.05%	19.81%
cg08578317	NDFUA10	-2.48%	4.59%	16.26%
cg20333904	NDFUA10	-0.73%	4.41%	11.33%
cg05354921	PRDM16	-14.54%	-29.66%	-50.00%
cg07209071	PRDM16	-8.57%	-14.82%	-26.59%
cg13393782	PRDM16	-6.78%	-7.46%	-19.39%
cg14966901	PRDM16	-8.14%	-14.04%	-28.10%
cg15090440	PRDM16	-5.30%	-3.58%	-15.26%
cg18759102	PRDM16	-8.22%	-13.18%	-25.14%
cg22860172	PRDM16	-5.97%	-8.86%	-16.58%
cg02004499	SDK1	-5.18%	-16.56%	2.23%
cg10741603	SDK1	-3.97%	-12.82%	1.41%
cg11535255	SDK1	-4.58%	-13.38%	8.50%
cg14791502	SDK1	-6.10%	-17.97%	0.05%
cg17739302	SDK1	-4.42%	-15.93%	-0.63%
cg17744375	SDK1	-4.25%	-14.76%	4.60%
cg24441899	SDK1	-3.11%	-13.42%	1.27%
cg25711963	SDK1	-4.03%	-15.23%	5.52%
cg12626698	SYCE1	-3.41%	-8.22%	-14.91%
cg19945897	SYCE1	-2.01%	-6.76%	-12.45%
cg20169779	SYCE1	-3.27%	-12.69%	-17.32%
cg18915450	TAF1B	-9.91%	4.93%	15.21%
cg20600404	TAF1B	-9.26%	6.55%	14.59%
cg25207776	TAF1B	-7.64%	4.39%	12.32%

Mediation
<10%
≥10%

Supplementary Table 4.6 Percent change in association of childhood abuse with DNAm at probes within childhood abuse differentially methylated regions (DMR) further adjusted for: adulthood BMI and smoking; mental health symptoms; and lifetime trauma exposure. Percent mediation = [beta estimate $_{\text{base model}}$-beta estimate $_{\text{adjusted model}}$]/beta estimate $_{\text{base model}}$. Negative percentages indicate a stronger association of childhood abuse with DNAm compared to the base model.

Probe Name	DMR Name	Base Model	Model 1: BMI and smoking	Model 2: Depressive and PTSD symptoms	Model 3: Lifetime Trauma
cg02622647	ARL17A	-1.329	-1.255	-1.181	-1.115
cg04703951	ARL17A	-1.042	-0.995	-0.931	-0.878
cg06925179	ARL17A	-0.560	-0.521	-0.493	-0.458
cg11007110	CFAP46	-0.441	-0.468	-0.467	-0.505
cg16060671	CFAP46	-0.510	-0.554	-0.557	-0.612
cg16903885	CFAP46	-0.421	-0.453	-0.459	-0.502
cg19922741	CFAP46	-0.403	-0.437	-0.450	-0.511
cg26579838	CFAP46	-0.461	-0.497	-0.502	-0.560
cg00359590	CLU	0.319	0.305	0.284	0.306
cg08594681	CLU	0.488	0.459	0.418	0.459
cg11783834	CLU	0.471	0.447	0.407	0.444
cg12729838	CLU	0.324	0.305	0.279	0.306
cg13488078	CLU	0.375	0.354	0.324	0.357
cg14917244	CLU	0.639	0.615	0.562	0.618
cg16292768	CLU	0.293	0.283	0.252	0.294
cg18931633	CLU	0.194	0.199	0.164	0.166
cg19442470	CLU	0.154	0.145	0.126	0.146
cg22313574	CLU	0.528	0.503	0.450	0.501
cg25302704	CLU	0.042	0.033	0.038	0.040
cg13214005	DLL	0.690	0.703	0.631	0.528
cg21235075	DLL	0.468	0.488	0.414	0.342
cg21597487	DLL	0.559	0.582	0.496	0.414
cg22739554	DLL	0.529	0.546	0.506	0.423
cg23066587	DLL	0.513	0.523	0.459	0.400
cg00293616	LRRK1	0.636	0.635	0.622	0.615
cg06375969	LRRK1	0.371	0.371	0.376	0.361
cg09926099	LRRK1	0.564	0.562	0.555	0.553
cg00438222	MAPT	0.584	0.574	0.576	0.552
cg00846647	MAPT	0.458	0.445	0.453	0.449
cg19276540	MAPT	0.449	0.439	0.443	0.412
cg19495068	MAPT	0.396	0.392	0.420	0.406
cg20265358	MAPT	0.495	0.486	0.507	0.486
cg21953146	MAPT	0.399	0.392	0.400	0.384
cg22128613	MAPT	0.503	0.495	0.501	0.473
cg26979107	MAPT	0.393	0.383	0.410	0.387
cg00537837	MIR5093	0.413	0.407	0.393	0.384
cg08428188	MIR5093	0.373	0.369	0.359	0.357
cg08675364	MIR5093	0.350	0.346	0.341	0.336
cg12707926	MIR5093	0.331	0.327	0.319	0.314
cg06905155	NDFUA10	0.381	0.396	0.350	0.306
cg08578317	NDFUA10	0.542	0.555	0.517	0.454
cg20333904	NDFUA10	0.581	0.585	0.555	0.515
cg05354921	PRDM16	0.251	0.288	0.323	0.377
cg07209071	PRDM16	0.458	0.497	0.525	0.579
cg13393782	PRDM16	0.288	0.308	0.310	0.344
cg14966901	PRDM16	0.336	0.363	0.383	0.430
cg15090440	PRDM16	0.190	0.201	0.197	0.219
cg18759102	PRDM16	0.528	0.571	0.597	0.660
cg22860172	PRDM16	0.363	0.385	0.396	0.424
cg02004499	SDK1	-0.809	-0.851	-0.943	-0.791
cg10741603	SDK1	-0.516	-0.536	-0.582	-0.509
cg11535255	SDK1	-0.466	-0.487	-0.528	-0.426
cg14791502	SDK1	-0.644	-0.684	-0.760	-0.644
cg17739302	SDK1	-0.484	-0.505	-0.561	-0.487
cg17744375	SDK1	-0.444	-0.463	-0.509	-0.423
cg24441899	SDK1	-0.433	-0.447	-0.491	-0.428
cg25711963	SDK1	-0.368	-0.383	-0.424	-0.348
cg12626698	SYCE1	0.472	0.488	0.511	0.542
cg19945897	SYCE1	0.532	0.543	0.568	0.598
cg20169779	SYCE1	0.526	0.543	0.592	0.617
cg18915450	TAF1B	0.437	0.481	0.416	0.371
cg20600404	TAF1B	0.527	0.576	0.493	0.450
cg25207776	TAF1B	0.686	0.738	0.656	0.602

Supplementary Table 4.7 Beta estimates from associations of childhood abuse with DNAm at probes in differentially methylated regions (DMR), base model and further adjusted for: 1) adulthood BMI and smoking; 2) adulthood mental health symptoms; and 3) lifetime trauma exposure.