

Epigenome-wide association analysis of daytime sleepiness in the Multi-Ethnic Study of Atherosclerosis reveals African-American specific associations

Richard Barfield^{1,2}, Heming Wang^{3,4,5}, Yongmei Liu⁶, Jennifer A Brody⁷, Brenton Swenson^{7,8},
Ruitong Li^{4,10}, Traci M. Bartz^{7,8}, Nona Sotoodehnia⁷, Yii-der I. Chen⁹, Brian E. Cade^{3,4,5}, Han
Chen^{10,11,12}, Sanjay R. Patel¹³, Xiaofeng Zhu¹⁴, Sina A. Gharib¹⁵, W. Craig Johnson¹⁶, Jerome I.
Rotter⁹, Richa Saxena^{3,4,5,17}, Shaun Purcell^{3,18}, Xihong Lin^{10,19}, Susan Redline^{4,5,20}, Tamar
Sofer^{4,5,10}

1. Department of Epidemiology, University of Washington, Seattle, WA
2. Public Health Sciences Division, Fred Hutchinson Cancer Research Center, Seattle, WA
3. Program in Medical and Population Genetics, Broad Institute, Cambridge, MA
4. Division of Sleep and Circadian Disorders, Department of Medicine, Brigham and Women's Hospital, Boston, MA
5. Division of Sleep Medicine, Harvard Medical School, Boston, MA
6. Department of Epidemiology and Prevention, Division of Public Health Sciences, Wake Forest University School of Medicine, Winston-Salem, NC
7. Cardiovascular Health Research Unit, Department of Medicine, University of Washington, Seattle, WA
8. Institute for Public Health Genetics, University of Washington, Seattle, WA.
9. The Institute for Translational Genomics and Population Sciences, Departments of Pediatrics and Medicine, Los Angeles Biomedical Research Institute at Harbor-UCLA Medical Center, Torrance, CA
10. Department of Biostatistics, Harvard T. H. Chan School of Public Health, Boston, MA
11. Human Genetics Center, Department of Epidemiology, Human Genetics and Environmental Sciences, School of Public Health, The University of Texas Health Science Center at Houston, Houston, TX.
12. Center for Precision Health, School of Public Health & School of Biomedical Informatics, The University of Texas Health Science Center at Houston, Houston, TX.
13. Division of Pulmonary, Allergy and Critical Care Medicine, University of Pittsburgh Medical Center, Pittsburgh, PA.
14. Department of Population and Quantitative Health Sciences, School of Medicine, Case Western Reserve University, Cleveland, OH.
15. Computational Medicine Core, Center for Lung Biology, University of Washington Medicine Sleep Center, Division of Pulmonary, Critical Care, and Sleep Medicine, University of Washington, Seattle, WA.
16. Department of Biostatistics, University of Washington, Seattle, WA.

17. Center for Genomic Medicine and Department of Anesthesia, Pain, and Critical Care Medicine, Massachusetts General Hospital, Boston, MA.
18. Department of Psychiatry, Brigham & Women's Hospital, Harvard Medical School, Boston, MA
19. Department of Statistics, Harvard University, Cambridge, MA
20. Division of Pulmonary, Critical Care, and Sleep Medicine, Beth Israel Deaconess Medical Center, Boston, MA

Corresponding Author:

Tamar Sofer

Brigham and Women's Hospital

Sleep and Circadian Disorders, room 225C

221 Longwood Ave

Boston MA 02115

tsofer@bwh.harvard.edu

Abstract

Study Objectives: Excessive daytime sleepiness (EDS) is a consequence of inadequate sleep, or of a primary disorder of sleep-wake control. Population variability in prevalence of EDS and susceptibility to EDS are likely due to genetic and biological factors as well as social and environmental influences. Epigenetic modifications (such as DNA methylation-DNAM) are potential influences on a range of health outcomes. Here, we explored the association between DNAM and daytime sleepiness quantified by the Epworth Sleepiness Scale (ESS).

Methods: We performed multi-ethnic and ethnic-specific epigenome-wide association studies for DNAM and ESS in 619 individuals from the Multi-Ethnic Study of Atherosclerosis. Replication was assessed in the Cardiovascular Health Study (CHS). Genetic variants in genes proximal to ESS-associated DNAM were analyzed to identify methylation quantitative trait loci and followed with replication of genotype-sleepiness associations in the UK Biobank.

Results: 61 methylation sites were associated with ESS ($FDR \leq 0.1$) in African Americans only, including an association in *KCTD5*, a gene strongly implicated in sleep. One association (cg26130090) replicated in CHS African Americans (p-value 0.0004). We identified a sleepiness-associated methylation site in the gene *RAII*, a gene associated with sleep and circadian phenotypes. In a follow-up analysis, a genetic variant within *RAII* associated with both DNAM and sleepiness score. The variant's association with sleepiness was replicated in the UK Biobank.

Conclusions: Our analysis identified methylation sites in multiple genes that may be implicated in EDS. These sleepiness-methylation associations were specific to African Americans. Future work is needed to identify mechanisms driving ancestry-specific methylation effects.

Keywords: Diversity; Epigenetics; Race/ethnic heterogeneity; excessive daytime sleepiness; sleep-wake; methylation; genomics.

Statement of Significance

Excessive daytime sleepiness is associated with negative health outcomes such as reduction in quality of life, increased workplace accidents, and cardiovascular mortality. There are race/ethnic disparities in excessive daytime sleepiness, however, the environmental and biological mechanisms for these differences are not yet understood. We performed an association analysis of DNA methylation, measured in monocytes, and daytime sleepiness within a racially diverse study population. We detected numerous DNA methylation markers associated with daytime sleepiness in African Americans, but not in European and Hispanic Americans. Future work is required to elucidate the pathways between DNA methylation, sleepiness, and related behavioral/environmental exposures.

Introduction

Excessive daytime sleepiness (EDS), is estimated to affect between 10 to 20% of the population^{1,2}. EDS is associated with numerous adverse clinical, behavioral and public health outcomes, including work and vehicular accidents³⁻¹⁰; reduced health-related quality of life¹¹⁻¹³; cognitive and performance deficits^{14,15}; and increased rates of stroke and total and cardiovascular mortality^{16,17}. There are multiple mechanisms for EDS, including insufficient sleep duration due to behavioral, social or work-related factors; sleep disruption due to a sleep disorder (sleep

apnea, periodic limb movement disorder); circadian misalignment; or the presence of a primary disorder of hypersomnia that affects the central sleep-wake control processes (e.g., narcolepsy). The prevalence of EDS has increased over the past decades, likely due to an increase in working hours and availability of electronic devices³. A rise in EDS prevalence may also reflect the increased prevalence of obesity, which is associated with increased sleepiness^{18,19}. This association is hypothesized to be due to obesity-related co-morbidities that reduce sleep quality, including obstructive sleep apnea, as well as to metabolic and neuroendocrine effects of adipokines on wake promoting neurons²⁰.

Although EDS is prevalent in the population, there appears to be large inter-individual differences in propensity for sleepiness following sleep deprivation, sleep fragmentation, or sleep apnea^{21,22}. Similarly, there is significant population variability in prevalence of EDS²³. Differences in sleepiness, including vulnerability or resilience to sleep disrupting influences, have been suggested to be due to genetic and other biological differences, although social and environmental influences also likely play a role^{24,25}. Generally, the bases for differences in sleepiness between populations are not well understood.

Epigenetic modifications are increasingly recognized to mediate the impact of environmental influences on gene expression and on prevalence and severity of a wide range of health outcomes, including neuropsychiatric, metabolic and cardiovascular diseases. The most studied epigenetic marker, DNA methylation (DNAm), occurs when a methyl group is added to a cytosine that is followed by a guanine on the genome (a “CpG” site). Changes in DNA methylation occur in response to a wide range of exposures, many of which are associated with sleep and sleepiness, such as obesity, diet, and stress²⁶⁻³². Therefore, we postulated that changes in DNAm would associate with variation in sleepiness and these changes may be population-

specific, providing insights into underlying susceptibility to sleepiness that may be partially attributed to environmental factors. To the best of our knowledge, there has been no reported study on DNA methylation and sleepiness.

In this paper, we performed an Epigenome Wide Association Study (EWAS) of EDS quantified by the Epworth Sleepiness Scale (ESS) in the Multi Ethnic Study of Atherosclerosis. We leveraged the racial/ethnic diversity of the sample to explore potential differences in EWAS associations by background groups. We performed replication analysis of methylation sites-ESS associations in the Cardiovascular Health Study (CHS) and identified cis-meQTLs in genes harboring associated methylation sites for replication analysis in the UK Biobank.

Methods

Study Sample

The study population consisted of participants from the Multi-Ethnic Study of Atherosclerosis (MESA), a prospective, longitudinal cohort study established to study factors associated with the development of cardiovascular disease. MESA clinic visits were first performed between 2000 and 2002 in six field centers across the United States when participants were free of known cardiovascular disease³³. The subset of MESA individuals in this study is composed of those who participated in a sleep examination conducted in conjunction with MESA Exam 5 (described in detail previously^{23,33}), and who also participated in the MESA DNAm study³⁴. The blood draws for the methylation study were obtained during MESA Exam 5 (2010-2012)³⁵ on a random subset of MESA participants at four of the six field centers: John Hopkins University, University of Minnesota, Columbia University, and Wake Forest University. A total of 623

individuals were available with both sleep data and DNAm data with four removed due to having missing ESS. The final study sample included 619 individuals: 132 African Americans (AAs), 202 Hispanic Americans (HAs), and 285 European Americans (EAs). The study was approved by the Institutional Review Boards of each participating site, and participants provided written informed consent, including use of genetic data.

DNAm Collection and Processing

Methods for collection and assays for DNAm have been described previously³⁵. In brief, peripheral blood was separated into mononuclear cells (CD14+) within two hours of collection. DNAm in monocytes was measured using the Illumina HumanMethylation450 BeadChip. Residual cell contamination in the monocyte data was assessed using Gene Set Enrichment Analysis, providing enrichment scores for neutrophils, B cells, T cells, and Natural Killer cells from Gene Expression data collected using the Illumina HumanHT-12 v4 Expression BeadChip and Illumina Bead Array Reader^{35,36}. The DNAm data underwent quality control tests prior to analysis using the lumi Bioconductor package³⁷. These included color bias adjustments via smooth quantile normalization, median background adjustment, standard quantile adjustment, checks for potential sex or race mismatches, and outlier detection via multidimensional plots. Additional details on pre-processing of the DNAm data can be found in Liu et al³⁴. Following DNAm preprocessing, each of 484,882 methylation probes for each person has a Beta-value, representing the proportion of methylated monocytes at that site and person.

We further prepared the methylation data for analysis based on meta information and sample-specific characteristics, as follows. First, we excluded 61,219 probes that were within 10bp of a SNP with a minor allele frequency greater than 1% in European, African, or American

populations from the 1000 genomes reference data³⁸. This was done to exclude loci where variation is solely due to genetic polymorphisms. We also removed any non-cg methylation probes and probes on sex chromosomes. Next, we focused on 73,082 CpG sites that had an inter-decile range greater than 0.1. Methylation values in these sites are unlikely driven by technical variability and focusing on them will reduce the multiple testing burden³⁹. Then, we used the software program ComBat⁴⁰ on these 73,082 CpG sites to remove any signal due to technical artifacts of chip and position on chip effects, while maintaining correlation with our primary set of covariates (self-reported ancestry, recruitment site, sex, age, and residual cell type enrichment). ComBat was run on the M-values (logit transformed Beta-values)⁴¹. We then removed any cross-reactive probes as defined by Chen et al⁴², leaving us with 66,028 DNAm probes to analyze.

Single Nucleotide Polymorphism (SNP) Data

For consenting individuals, DNA was extracted from whole blood and genotyped on Affymetrix 6.0 GWAS array. Standard quality control methods for SNP- and sample-level quality were applied, including the exclusion of participants and SNPs with over 5% missing call rates. This resulted in 895,289 genotyped variants in 615 individuals with DNAm data. For downstream analysis, we excluded SNPs with minor allele frequency (MAF) less than 1%. Further details on the genotype and quality control can be found in Vargas et al⁴³. We calculated the top 5 principal components (PCs) from the Linkage-Disequilibrium pruned set SNPs with $MAF \geq 5\%$. PCs were calculated in the combined sample and also in each ethnic-specific group. There were four AA individuals that had missing SNP information; therefore, the PCs were imputed with the mean value of the respective PC in the AA subset that had SNP information.

Sleep Assessments

As part of the MESA Sleep Examination (2010-2013), participants completed standardized questionnaires and underwent a single night in-home polysomnography (Compumedics Somte Systems, Abbotsville, Australia, AU0) and 7 day wrist actigraphy (Philips-Respironics Spectrum, Murrysville, PA), as described before²³. The primary sleep measure was daytime sleepiness as quantified by the Epworth Sleepiness Scale (ESS), an 8-item validated instrument that asks the individual to assess likelihood of dozing off in a variety of daily activities using a 4 point (0-3) scale⁴⁴. ESS scores vary from 0 to 24, with higher scores denoting more sleepiness⁴⁴. ESS was assessed within one year of blood draw for DNA methylation. We identified a set of additional sleep phenotypes/exposures that might be associated with both DNAm and ESS, and thus are potential mediators or confounders of any DNAm and ESS association. These measures were: insomnia (report of doctor diagnosed insomnia or the Women's Health Initiative (WHI) Insomnia Rating Scale⁴⁵, a validated scale varying from 0 to 20); Apnea-Hypopnea Index [AHI] (assessed by polysomnography as the sum of all apneas and hypopneas associated with 3% or more oxygen desaturation divided by total sleep time); overnight hypoxemia (derived by polysomnography as the percentage of sleep time the participant spent with oxyhemoglobin saturation less than 90% [Per90]); and sleep duration (actigraphy derived average sleep duration over the 7 monitoring nights as described before²³).

Covariates

Age, sex and race/ethnicity were self-reported. Other behavioral, socioeconomic and lifestyle exposures that may confound sleepiness-methylation associations were also assessed. Alcohol use was based on self-reported (yes/no); smoking status was classified by ever/former/never;

depressive symptoms were based on responses to the Center for Epidemiological Studies (CES) depression scale⁴⁶; having less than a college education; mother having less than a college education, the latter two variables representing socio-economic status. Body mass index (BMI) was calculated from weight and height measured at MESA Exam 5. Dietary variables were derived from a food frequency questionnaire: long chain score (sum of Omega fatty acids mg/d), total fats, carbohydrates, and saturated fatty acids (rams) and Total Alternative Healthy Eating Index - 2010 (Total AHEI-2010)⁴⁷.

Statistical Analysis

Our complete analysis approach is depicted in Figure 1. We first describe our discovery analyses, followed by sensitivity, gene expression, local methylation expressive quantitative trait loci (cis-meQTL), and replication analyses.

Discovery Analysis for Methylation-ESS Association

The primary analysis outcome was ESS score, square-root transformed to achieve approximate normality. It was analyzed in a linear regression on DNAm M-values generated following application of ComBat (as described earlier). The primary analyses were adjusted for sex, age, residual cell type enrichment, recruitment site, and the top 5 genetic PCs. We first analyzed the transformed ESS on the combined multi-ethnic sample with an additional covariate for ethnicity (EA, AA, or HA) and PCs calculated in the overall group. We then performed race/ethnic group-specific analyses adjusting for the top five PCs calculated exclusively for each respective race/ethnicity group. To account for multiple testing, we calculated false discovery rate (FDR) controlling q-values in each of the analyses, i.e. the multi-ethnic and each ethnic-group specific

analyses⁴⁸. An association was deemed as significant if the q-value was less than 0.1⁴⁸, thus controlling the FDR at the 10% level in each of the analyses. We evaluated heterogeneity in any race/ethnic specific results by fitting a linear model using the overall sample with self-reported race/ethnicity interaction terms with DNAm and testing these interaction terms using an F-test via analysis of variance (ANOVA).

Sensitivity Analysis for methylation-ESS associations

We next identified behavioral, socioeconomic and lifestyle exposures, including variation in sleep and sleep disturbances, that may influence the sleepiness-methylation associations and further adjusted for these factors in sensitivity analyses conducted in significant probe-trait primary associations. Covariates considered for the sensitivity analysis were: BMI, sleep duration (actigraphy-based), insomnia, AHI, nocturnal hypoxemia [Per90], alcohol use, smoking status, CES depression scale, having less than a college education, mother having less than a college education, and dietary variables, as defined earlier in the manuscript. We also examined moderate sleep apnea defined as an AHI greater than 15, less than five hours of sleep, and overnight hypoxemia (dichotomized as Per90 greater than 5%). We only assessed variables that had a p-value less than 0.1 for their association with ESS after adjusting for age and sex. These were included together as covariates in the regression model for CpG sites with a q-value less than 0.1. We handled missingness of continuous variables (up to five missing values per variable) by imputing the missing value to the mean of the variable calculated in the rest of the study participants of the same self-reported race, and removed individuals with missing categorical values. Missing information by variable is provided in Supplementary Table 1. Finally, we evaluated whether an observed change in effect size estimate between a methylation

site and ESS after including each or sets of covariates was higher than by chance via permutation. We permuted the additional exposure/sensitivity variables 10,000 times and calculated the proportion of permutations in which the observed change in DNAm-ESS effect sizes was higher compared to the DNAm-ESS effect size in the un-permuted sensitivity variables.

DNAm-ESS replication analysis in the Cardiovascular Health Study (CHS)

The CHS is a population-based cohort study of risk factors for coronary heart disease and stroke in adults ≥ 65 years conducted across four field centers⁴⁹. The original predominantly European ancestry cohort of 5,201 persons was recruited in 1989-1990 from random samples of the Medicare eligibility lists; subsequently, an additional predominantly African-American cohort of 687 persons was enrolled for a total sample of 5,888. CHS was approved by institutional review committees at each field center and individuals in the present analysis had available DNA and gave appropriate informed consent for use of genetic information. DNA methylation was measured on 336 European ancestry and 329 African-American ancestry participants at study year 5. The samples were randomly selected among participants without presence of coronary heart disease, congestive heart failure, peripheral vascular disease, valvular heart disease, stroke or transient ischemic attack at study baseline or lack of available DNA at study year 5. Sleepiness was assessed at year 6 using the ESS.

Methylation measurements were performed at the Institute for Translational Genomics and Population Sciences at the Harbor-UCLA Medical Center--Los Angeles Biomedical Research Institute using the Infinium HumanMethylation450 BeadChip (Illumina Inc, San Diego, CA). Quality control was performed in the minfi R package⁵⁰⁻⁵² (version 1.12.0,

<http://www.bioconductor.org/packages/release/bioc/html/minfi.html>). Samples with low median intensities of below 10.5 (log₂) across the methylated and unmethylated channels, samples with a proportion of probes falling detection of greater than 0.5%, samples with QC probes falling greater than 3 standard deviation from the mean, sex-check mismatches, failed concordance with prior genotyping or > 0.5% of probes with a detection p-value > 0.01 were removed. In total, 11 samples were removed for sample QC resulting in a sample of 323 European-ancestry and 326 African-American samples. Methylation values were normalized using the SWAN quantile normalization method⁵². Since white blood cell proportions were not directly measured in CHS they were estimated from the methylation data using the Houseman method⁵³.

We performed replication look-up of the significant (q-value<0.1) associations in CHS AA and EA. We adjusted for age, sex, recruitment site, current smoking status, white blood cell counts, and top five genetic PCs. For this, we examined results in both individual methylation sites, and used a binomial test⁵⁴ for the overall agreement in the direction of associations between discovery and replication studies. A methylation site was considered to be statistically associated with ESS if its one-sided p-values (with direction of association guided by the direction of DNAm-sleepiness association in MESA discovery) was less than 0.05 divided by number of FDR significant sites in MESA, a Bonferroni threshold controlling the Family-Wise Error Rate of replication⁵⁵. We analyzed both CHS AA and EA samples separately.

Testing for cis-meQTLs in MESA

We hypothesized that a combination of DNA methylation and genetic association plays a role in sleepiness. To study this hypothesis, we attempted to identify variants that are associated with both methylation and sleepiness. From a causal pathway perspective, this analysis considers the

setting in which a genetic variant predicts both DNAm and sleepiness, where the association with sleepiness is either direct (pleiotropic) or mediated through DNAm. An example of the latter mechanism, a genetic variant could affect the methylation status of a CpG site by disrupting binding sites for regulatory proteins, which in turn influences sleepiness via a molecular mechanism not directly related to the product of the gene of interest. Due to our limited sample size, we implemented a screening process on genetic variants in the CpG region, followed by a replication analysis of the SNP-sleepiness association in an independent cohort (UK Biobank).

First, we mapped the top ten methylation sites to their respective genes to assess if genetic variants in that gene were associated with both ESS and DNAm. Specifically, for each gene that a top probe mapped to, we first performed association analysis between genotyped SNPs within 100kb of the gene and the associated DNAm site adjusting for age, sex, top five PCs, study site, and residual cell type. This identified local (cis) meQTLs (screening step 1). SNPs with associations significant at $P < 0.005$ (accounting for ten comparisons) were next assessed for their association with sleepiness, with and without additional adjustment for the DNAm (screening step 2). SNPs with $P < 0.05$ in screening step 2 were carried forward to independent replication analysis (see below) of the SNP-sleepiness association in the UK Biobank.

Replication analysis of cis-meQTL-sleepiness association in the UK Biobank

We further looked up the SNP-sleepiness associations using individuals of European ancestries from a GWAS of sleepiness in UK Biobank. The UK Biobank is a large prospective study for a wide range of genetics and health outcomes in over 500,000 participants aged 40-69 years

recruited from 2006 to 2010 in the United Kingdom⁵⁶. Genotyping and imputation data are described in Bycroft et al. (2017)⁵⁷. Sleepiness was assessed by self-reported responses to the question “How likely are you to dose off or fall asleep during the daytime when you don’t mean to? (e.g. when working, reading or driving)” with the options of “Never/rarely”, “sometimes”, “often”, “all of the time”, “do not know”, and “prefer not to answer.” Association analysis was performed in 452,071 individuals adjusted for age, sex, genotyping array, 10 PCs and genetic relatedness matrix.

Gene expression analysis

For genes corresponding to the top 10 DNAm sites, we used gene expression data in MESA to study the evidence for association between DNAm and gene expression (of the same gene), and gene expression association with ESS. Gene expression profiling and processing are described in detail in Liu et al.³⁴. In brief, gene expression was measured using Illumina HumanHT-12 v4 Expression BeadChip array in mononuclear cells, on the same participants with DNA methylation measures. We performed (1) association analysis of DNAm and gene expression, using all individuals with available data (234 AA, 386 HA, and 582 EA individuals), with ESS as the outcome and DNAm as exposure; and (2) association analysis of gene expression (as exposure) and squared-root ESS (as the outcome) in the set of individuals that had both gene expression and ESS measured. Complete details are provided in the Supplementary Material.

Results

Study Characteristics

The MESA sample consisted of 619 individuals (53% female) with a mean age of 68 years, with 132 AA, 202 HA, and 285 EA individuals (Table 1). The mean ESS was 6 and 19% were

classified as having EDS (ESS>10), which did not vary significantly by ethnicity in this analytical sample ($p=0.20$) (although ESS was higher in AA than whites in the larger MESA cohort²³). Sleep duration was significantly shorter in the AA group compared to the other groups (5.02×10^{-7}). Median AHI was 14.1, and 46% of the sample had moderate or more severe sleep apnea (AHI>15). Across groups, sleep apnea was more prevalent and severe in the HA group (Table 1). Significant differences were observed for BMI, having a college education, smoking status, sleep apnea, sleep duration, dietary long chain score, and Total AHEI and alcohol consumption by ancestry group (Table 1).

Associations Between Sleepiness and DNAm

In the combined analysis across race/ethnic groups, no significant associations were observed between DNAm M-values and ESS. In race/ethnicity stratified analyses, significant associations were only observed in the AA group, where 61 DNAm probes were significantly associated with ESS at $q\text{-value} \leq 0.1$ (Table 3). The top probe was in *KCTD5* ($p=3.96 \times 10^{-7}$), a gene with known sleep-wake effects⁵⁸. Two significant probes were within *RAI1*, a gene associated with several sleep and circadian phenotypes^{59,60}, and 5 probes were within *KIF3*, involved in cytokinesis with effects on photoreceptors⁶¹. We display the top three associations in Figure 2. In 56 of the 61 probes, decreased methylation was associated with more sleepiness (i.e., higher ESS scores) (Supplement Table 2). Testing for any interaction effects between these 61 probes and self-reported ethnicity on the transformed ESS scores in the overall group identified 58 of the 61 associations with a p-value for interaction less than 0.05. Table 3 provides all associations passing the significance threshold, along with their corresponding genes (from Illumina annotation), and replication p-values in CHS. Supplemental Table 2 includes additional

information on these associations, including interaction p-values, and effects and standard errors in the overall group, HA, AA and EA, as well as estimated effect sizes, SDs, replication p-values (one-sided) and usual two-sided p-values in the CHS replication samples. Histogram of p-values in each analysis are provided in the Supplementary Material.

Sensitivity Analysis

We next evaluated potential explanatory factors or confounders between the ESS-associated DNAm probes in the AA sample. Supplement Table 3 shows p-value between all considered sensitivity parameters and partial R^2 with ESS after adjusting for age and sex. WHI insomnia score, alcohol use, mother having less than a college education, and sleep duration were all associated with sleepiness (square-root transformed ESS) at a p-value of 0.1 adjusting for age and sex in the AA group (Supplementary Table 3). We included these four variables as additional covariates in the primary model to assess if they modified the association between DNAm and ESS. In our 61 probes, we see a change in the effect size over 10% for 45 of the probes (Supplement Table 4). However, the change for any individual variable was less than 10% (Supplement Table 4). We additionally performed a permutation study to compare the observed changes to what would be expected by chance. The change in association of 51 DNAm sites was significantly different from zero with permutation p-value < 0.05, i.e., with less than 5% of permutations demonstrating the same or higher change in estimated association (Supplement Table 4 last column). Supplement Table 4 summarizes the results for the top ten probes. This suggests that the selected exposure variables confound/modify the DNAm-ESS association. Supplemental Table 4 also shows how much the association between DNAm and transformed ESS change after the inclusion of each individual sensitivity parameter by itself. There is some

evidence of confounding (effect change over 10% for 45 of the 61 probes). The majority of adjusted associations are still nominally significant (range 2.33×10^{-3} to 4.25×10^{-6}) suggesting that these variables do not explain away the association between DNAm and ESS.

Replication of DNAm-sleepiness association in CHS

Of the 61 probes found significant in our discovery analysis, 60 were assayed in the CHS AA dataset (n=238). Replication analysis results are provided in Table 3. Globally, the direction of DNAm-sleepiness associations largely agreed between MESA and CHS AA for these 61 probes (binomial p-value=0.005, Supplement Table 2). Forty of the sixty (67% agreement) probes had the same effect sign. Only one probe of the 60 was significant at a Bonferroni level of 0.05 in the CHS AA analysis, cg26130090 (32945 base pairs from *SMUG1* chromosome 12 (MESA p-value = 1.92×10^{-5} , CHS p-value = 3.95×10^{-4})). The CHS EA analysis (n=245) agreed with the MESA EA results, which showed no evidence of association in these methylation sites. Consistent with null results in MESA EA, no individual associations replicated in CHS EA, and globally, there was no agreement in the directions of associations between MESA AA and CHS EA (binomial p-value = 0.9393). There was no agreement in direction of effect comparing MESA AA vs MESA EA or MESA HA.

Cis-meQTL analysis and UK Biobank replication

We performed cis-meQTL analysis for the top ten probes which mapped to seven genes that harbored significant DNAm-sleepiness associations in our AA sample. Thirty SNPs in four genes (*KIFC3*, *KDM2B*, *TGFB1L1*, and *RAI1*) passed screening step 1 ($P < 0.005$; Supplementary Table 5). Of these, 15 SNPs passed screening step 2 and were carried forward for testing the

SNP-sleepiness association in the UK Biobank data set of European Ancestry individuals. We determined significance in this replication by a p-value threshold of $0.05/15 = 0.003$. The results from this analysis are reported in Supplement Table 6. A single SNP passed this threshold, rs9896285 in *RAII* ($p=0.0013$).

Gene expression analysis

Comprehensive results are provided in Supplementary Tables 7 and 8. Briefly, we considered three genes corresponding to the top 10 DNAm sites: *KCTD5* (two transcripts, one DNAm site), *KDM2B* (one transcript and one DNAm site), and *RAII* (two transcripts and two DNAm sites). For *KDM2B*, cg09411874 methylation was associated with increased expression in the HA group ($p=0.04$), but there was no strong evidence of *KDM2B* expression association with ESS ($p=0.18$ in the HA group). The two DNAm sites associated with *RAII* had $p=0.08$ in the combined cohort in association with expression of one of the *RAII* transcripts. However, the estimated association of this transcript with ESS was low (and with $p>0.6$ for all analyzed groups). Interestingly, larger expression of the other transcript of *RAII* was significantly associated with decreased ESS ($p=0.009$ in the combined cohort). We did not detect any associations in the analysis of *KCTD5*.

Discussion

This study reports for the first time a large-scale DNA methylation association analysis in a multi-ethnic population designed to detect DNAm markers associated with excessive daytime sleepiness. We analyzed the association both in the pooled MESA cohort and separately in each race/ethnicity group (AA, EA, and HA). In our initial discovery stage, there were 61 DNAm markers associated with sleepiness at the FDR rate of < 0.1 , solely in the AA group. Of these, 60 were studied in association with the ESS score in the CHS cohort. One of the associations, that of cg26130090 (intergenic), replicated in CHS AAs ($p\text{-value} = 0.0004$), and there was an overall

directional agreement in the significant associations in MESA AA and CHS AA (Binomial p-value=0.005). Additional top DNAm associations were observed in the *RAI1* and *KCTD5* genes, which have strong biological relevance and were previously reported as associated with several sleep traits^{58-60,62-64}.

Our study focused on the multi-ethnic epigenetic bases of daytime sleepiness. Several studies have shown that even after accounting for sleep duration and sleep disorders, sleepiness is less in EA compared to other ethnic groups^{23,65-68}. Compared to EA groups, AAs also have a higher prevalence of both short and long sleep durations and sleep apnea syndrome (defined by the presence of an elevated apnea hypopnea index plus daytime sleepiness²³) and in the larger MESA cohort, AA had higher ESS than EA²³. Race/ethnicity differences in sleep patterns and susceptibility to EDS appear early in life. AA children have been shown to have shorter sleep duration than white children as early as one year of age⁶⁹. In children with sleep apnea treated with adenotonsillectomy, AA children have higher sleepiness scores than EA children both at baseline and following treatment⁷⁰. Here, in our study of a middle-aged to older community multi-racial sample, we found significant epigenetic findings only in the AA sample. The specificity for the findings in the AA sample could reflect differences in underlying genetic architecture that influence susceptibility to methylation or differences in environmental exposures, such as air pollution, or stress that differentially affect different groups⁷¹⁻⁷⁴. We explored this latter hypothesis by a series of sensitivity analyses that explored potential explanatory effects due to sleep variation and other lifestyle exposures. These analyses showed that adjusting for mother having less than a college education, alcohol use, sleep duration, and insomnia modestly attenuated the effect estimates, suggesting potential mechanisms for epigenetic findings. However, these measures do not characterize lifetime exposures nor

included precise measurements of the social or physical environments, emphasizing the need for additional research that links key aspects of an individual's environmental and sociological exposures with epigenetic markers and health conditions such as sleepiness as important routes for understanding health disparities.

A few of our sleepiness associated DNAm sites are within genes that have strong evidence for associations with sleep related processes. Our most significant DNAm association with excessive sleepiness (MESA AA p-value 3.69×10^{-7}), was for cg26609398 in the *KCTD5* gene. While this probe did not replicate in CHS, it was suggestive (CHS AA p-value 0.028) with the association in the same direction. *KCTD5* is a mammalian ortholog of the *Drosophila* gene *Inc*, a gene expressed in multiple sleep associated regions in the central nervous system and shown to play a fundamental role in sleep in *Drosophila* by interacting with Cullin-3 ubiquitin ligase (*Cul3*). When mutated, both genes (*KCTD5* and *Cul3*) produce similar phenotypes characterized by markedly shortened and fragmented sleep together with altered synaptic structure and impaired synaptic transmission in relevant neuronal circuits⁵⁸. This suggests that alterations in protein dynamics (e.g. ubiquitination) might disrupt synaptic function and impact sleep. Future research is needed to examine how epigenetic changes in *KCTD5* and similar genes influence gene expression and core elements of the sleep homeostat, with initial work indicating potential protein interactions with the circadian genes *FBXW11* and *BTRC*⁷⁵.

Two of our top ten associated methylation sites were within *RAII* (retinoic acid induced-1). *RAII* encodes a protein that is highly expressed in neuronal tissues and is involved in early neural differentiation and transcriptional regulation of circadian clock components^{60,62-64}. Variants in this gene have been associated with circadian rhythm disruption, several abnormal sleep traits, neurobehavioral problems and obesity^{62,64}. In addition, SNPs in *RAII* have been

found to be associated with sleep apnea⁵⁹. Moreover, our cis-meQTL analysis detected a SNP in *RAII* influencing sleepiness in an EA population from the UK Biobank. Our analysis in MESA suggests that the SNP rs9896285 has both direct and indirect (through DNAm) effect on excessive sleepiness. However, the methylation association with *RAII* did not replicate in CHS. In another secondary analysis, we found that the expression of *RAII* in mononuclear cells was associated with ESS in MESA. More work is needed to further dissect the association of *RAII* markers with ESS.

Other findings in our study, while not as strong as *KCTD5* and *RAII*, have some evidence supporting the epigenetic association between sleepiness and DNAm. A significantly associated DNAm site mapped to *KDM2B*, a histone demethylase gene. DNAm at this gene has been found to be associated with BMI^{28,76}. Obesity may promote sleepiness through effects on sleep disruption or the effects of inflammatory cytokines (that are augmented by obesity-associated inflammation) on central processes regulating sleep and sleepiness^{77,78}. However, our sensitivity analyses did not support a significant influence of obesity on our observed associations. This could be because BMI does not optimally capture the underlying obesity-related phenotypes, or because methylation of *KDM2B* regulates both BMI and sleepiness via a shared process. The methylation site, cg17585031, was near *KSRI* (Kinase Suppressor of Ras 1), a protein suppressor of RAS signaling, a pathway that controls cellular growth and differentiation⁷⁹. *KSRI* has been found to be up-regulated after sleep-restriction⁸⁰ in 26 participants studied in a laboratory-based assessment. DNAm at this gene has been associated with sleep apnea as well, though in a small study at a high significance level (FDR<0.5)⁸¹. This *KSRI* probe (cg17585031) did not replicate in CHS AA and had opposite sign of association compared to the probe reported in this gene from the sleep apnea study⁸¹. A significant association was observed for the CpG site near a

master cytokine *TGFBI*. Experimental sleep fragmentations in mice has shown rapid increases in expression of *TGFBI* in the hypothalamus and hippocampus⁸².

A unique feature of this study is study of a multi-ethnic population. We identified DNAm-EDS associations that are only present in AAs. DNAm differences in these sites may explain some of the differences in EDS between AAs and EAs, highlighting potential mechanisms to explore in studies of health disparities. The main limitation of our association study is that we had small sample sizes in the group-specific analyses. Despite the limited sample size, we could directly replicate one association (after accounting for multiple testing) and observed a global consistency of the significant association between the discovery cohort MESA AA, and the replication study of CHS AAs. Such global consistency was not observed between MESA AA and CHS EA, strengthening the race/ethnic specific association evidence. Lack of further replication may relate to many factors, including the older sample of CHS participants (who are 5 years older than MESA participants, on average), biases and misclassification in self-reported sleepiness assessment⁸³, and potential difference in their exposure history compared to MESA participants. Another limitation was our pruning of DNA methylation sites was stringent though similar has been done prior³⁹. Given our small sample size, this was a good way to reduce the multiple testing burden by excluding probes that were likely not informative. Finally, a limitation which is shared by many epigenome-wide association studies, is the lack of identified exposure explaining the change in methylation and the group-specificity of the associations. We did study potential mechanism by considering cis-meQTLs analysis and a sensitivity analysis adjusting for various, plausible, exposure variable (sleep, lifestyle, and nutrition variables) which may be independently associated with either DNAm, ESS, or both, but these analyses did not lead to strong hypotheses about causal pathways.

In conclusion, we identified multiple DNAm probes associated with sleepiness. The majority of these probes indicated that hypomethylation was positively associated with sleepiness, with some of our top probes being near genes known to be associated with critical neuronal processes that influence sleep. Taken together, our findings suggest that these sleepiness-associated DNAm sites are related to genes that have a biologically meaningful function with regards to sleep. The significance of our findings in AAs identifies biological areas that may contribute to the observed differences in sleepiness between AAs and EAs, providing an avenue for further investigation, both genetic and environmental, of health disparities. For example, within MESA, the *RAI1* SNP (that associates with cg27208169) had MAF of 0.28 in AA, 0.04 in EA, and 0.11 in HA. In the UK biobank, the MAF was 0.06. As mentioned in the introduction, poor sleep outcomes are associated with negative life outcomes and a recent review presented extensive evidence that obstructive sleep apnea and aging share various biological pathways⁸⁴. Future studies with larger sample sizes may elucidate these genomic pathways and how they relate to potential biological or sociological patterns that are being reflected in excessive sleepiness.

Acknowledgments

We thank the participants of the UK Biobank; replication of the SNP sleepiness association was performed under application 6818. RTB was funded by NCI grant T32CA094880 and NHLBI grant R01HL113338. SR and TS were partially funded by R35 HL135818. HW was supported by NIH NHLBI R01HL113338 (to SR) and Sleep Research Society Foundation Career Development Award 018-JP-18 (to HW). RS was supported by NIH R01DK107859, R01DK102696 and R01DK105072.

MESA:

This research was supported by the Multi-Ethnic Study of Atherosclerosis (MESA).

MESA and the MESA SHARe project are conducted and supported by the National Heart, Lung, and Blood Institute (NHLBI) in collaboration with MESA investigators. Support for MESA is provided by contracts HHSN268201500003I, N01-HC-95159, N01-HC-95160, N01-HC-95161, N01-HC-95162, N01-HC-95163, N01-HC-95164, N01-HC-95165, N01-HC-95166, N01-HC-95167, N01-HC-95168, N01-HC-95169, UL1-TR-000040, UL1-TR-001079, UL1-TR-001420, UL1-TR-001881, and DK063491.

MESA SLEEP was funded by R01HL098433. Funding support for the Sleep Polysomnography dataset was provided by grant HL56984.

The MESA Epigenomics Studies were funded by R01HL101250, R01 DK103531-01, R01DK103531, R01 AG054474, and R01 HL135009-01 to Wake Forest University Health Sciences.

The provision of genotyping data was supported in part by the National Center for Advancing Translational Sciences, TSCI grant UL1TR001881, and the National Institute of Diabetes and Digestive and Kidney Disease Diabetes Research (DRC) grant DK063491. Funding for MESA SNP Health Association Resource (SHARe) genotyping was provided by NHLBI Contract N02-HL-64278. Genotyping was performed at Affymetrix (Santa Clara, California, USA) and the Broad Institute of Harvard and MIT (Boston, Massachusetts, USA) using the Affymetrix Genome-Wide Human SNP Array 6.0.

The authors thank the other investigators, the staff, and the participants of the MESA study for their valuable contributions. A full list of participating MESA investigators and institutions can be found at <http://www.mesa-nhlbi.org>.

CHS: Infrastructure for the CHARGE Consortium is supported in part by the National Heart, Lung, and Blood Institute grant R01HL105756. The CHS research was supported by NHLBI contracts HHSN268201200036C, HHSN268200800007C, HHSN268201800001C, N01HC55222, N01HC85079, N01HC85080, N01HC85081, N01HC85082, N01HC85083, N01HC85086; and NHLBI grants U01HL080295, U01HL130114, K08HL116640, R01HL087652, R01HL092111, R01HL103612, R01HL105756, R01HL103612, R01HL111089, R01HL116747 and R01HL120393 with additional contribution from the National Institute of Neurological Disorders and Stroke (NINDS). Additional support was provided through R01AG023629 from the National Institute on Aging (NIA), Merck Foundation / Society of Epidemiologic Research as well as Laughlin Family, Alpha Phi Foundation, and Locke Charitable Foundation. A full list of principal CHS investigators and institutions can be found at CHS-NHLBI.org. The provision of genotyping data was supported in part by the National Center for Advancing Translational Sciences, CTSI grant UL1TR001881, and the National Institute of Diabetes and Digestive and Kidney Disease Diabetes Research Center (DRC) grant DK063491 to the Southern California Diabetes Endocrinology Research Center.

The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Disclosure Statement

Financial Disclosure: SR receives grants from the NIH, American Sleep Medicine Foundation and Jazz Pharmaceuticals. She received consulting fees from Jazz Pharmaceuticals.

Non-Financial Disclosure:

Abbreviations

- AAs: African American
- AHI: Apnea-Hypopnea Index
- ANOVA: Analysis of Variance
- BMI: Body mass index
- CES: Center for Epidemiological Studies
- CHS: Cardiovascular Health Study
- DNAm: DNA methylation
- EAs: European Americans
- EDS: Excessive daytime sleepiness
- ESS: Epworth Sleepiness Scale
- FDR: False Discovery Rate
- GWAS: Genome-Wide association study
- HAs: Hispanic Americans
- meQTLs: methylation expressive quantitative trait loci
- MESA: Multi-ethnic Study of Atherosclerosis

- PC: Principal Component
- Per90: nocturnal hypoxemia
- SNP: Single Nucleotide Polymorphism
- WHI: Women's Health Initiative

Table legends

Table 1: Characteristics of 619 MESA participants. Columns show phenotype characteristics in overall group and each reported ethnicity along with a test for difference in the phenotype between the reported ethnicity.

Table 2: Characteristics of 483 CHS participants. Columns show phenotype characteristics as well as separately in European and African Americans.

Table 3: Association analysis results in the MESA AA discovery sample and in the CHS AA replication sample. Table rows correspond to methylation sites associated with the Epworth Sleepiness Scale (square root transformed) with $FDR < 10\%$ in the discovery sample.

Figure Legends

Figure 1: Analysis workflow for assessing associations between DNAm and ESS.

Figure 2: Top three association in AA. Left column shows the probe vs squared-root transformed ESS in the overall sample, right column shows the probe vs transformed ESS in AA only.

1. Hublin C, Kaprio J, Partinen M, Heikkila K, Koskenvuo M. Daytime sleepiness in an adult, Finnish population. *J Intern Med.* 1996; 239 (5): 417-423.
2. Ohayon MM. From wakefulness to excessive sleepiness: what we know and still need to know. *Sleep Med Rev.* 2008; 12 (2): 129-141.
3. Tsai SC. Excessive sleepiness. *Clin Chest Med.* 2010; 31 (2): 341-351.
4. Mullins HM, Cortina JM, Drake CL, Dalal RS. Sleepiness at work: a review and framework of how the physiology of sleepiness impacts the workplace. *J Appl Psychol.* 2014; 99 (6): 1096-1112.
5. Cohen DA, Wang W, Wyatt JK, et al. Uncovering residual effects of chronic sleep loss on human performance. *Sci Transl Med.* 2010; 2 (14): 14ra13.
6. Connor J, Whitlock G, Norton R, Jackson R. The role of driver sleepiness in car crashes: a systematic review of epidemiological studies. *Accid Anal Prev.* 2001; 33 (1): 31-41.
7. Ellen RL, Marshall SC, Palayew M, Molnar FJ, Wilson KG, Man-Son-Hing M. Systematic review of motor vehicle crash risk in persons with sleep apnea. *J Clin Sleep Med.* 2006; 2 (2): 193-200.
8. Herman J, Kafoa B, Wainiqolo I, et al. Driver sleepiness and risk of motor vehicle crash injuries: a population-based case control study in Fiji (TRIP 12). *Injury.* 2014; 45 (3): 586-591.
9. Philip P, Sagaspe P, Lagarde E, et al. Sleep disorders and accidental risk in a large group of regular registered highway drivers. *Sleep Med.* 2010; 11 (10): 973-979.
10. Teran-Santos J, Jimenez-Gomez A, Cordero-Guevara J. The association between sleep apnea and the risk of traffic accidents. Cooperative Group Burgos-Santander. *N Engl J Med.* 1999; 340 (11): 847-851.
11. Briones B AN, Whalen C, et al. Relationship between general health status, sleepiness, and sleep apnea. *Am J Respir Crit Care.* 1995; 151: A104.
12. Gooneratne NS, Weaver TE, Cater JR, et al. Functional outcomes of excessive daytime sleepiness in older adults. *Journal of the American Geriatrics Society.* 2003; 51 (5): 642-649.
13. Vinnikov D, Blanc PD, Alilin A, Zutler M, Holty JC. Fatigue and sleepiness determine respiratory quality of life among veterans evaluated for sleep apnea. *Health Qual Life Outcomes.* 2017; 15 (1): 48.
14. Ohayon MM, Vecchierini MF. Daytime sleepiness and cognitive impairment in the elderly population. *Archives of internal medicine.* 2002; 162 (2): 201-208.

15. Merlino G, Piani A, Gigli GL, et al. Daytime sleepiness is associated with dementia and cognitive decline in older Italian adults: a population-based study. *Sleep Med.* 2010; 11 (4): 372-377.
16. Newman AB, Spiekerman CF, Enright P, et al. Daytime sleepiness predicts mortality and cardiovascular disease in older adults. The Cardiovascular Health Study Research Group. *J Am Geriatr Soc.* 2000; 48 (2): 115-123.
17. Qureshi AI, Giles WH, Croft JB, Bliwise DL. Habitual sleep patterns and risk for stroke and coronary heart disease: a 10-year follow-up from NHANES I. *Neurology.* 1997; 48 (4): 904-911.
18. Ng WL, Orellana L, Shaw JE, Wong E, Peeters A. The relationship between weight change and daytime sleepiness: the Sleep Heart Health Study. *Sleep Med.* 2017; 36: 109-118.
19. Vgontzas AN, Bixler EO, Chrousos GP. Obesity-related sleepiness and fatigue: the role of the stress system and cytokines. *Ann N Y Acad Sci.* 2006; 1083: 329-344.
20. Panossian LA, Veasey SC. Daytime sleepiness in obesity: mechanisms beyond obstructive sleep apnea--a review. *Sleep.* 2012; 35 (5): 605-615.
21. Van Dongen HP, Baynard MD, Maislin G, Dinges DF. Systematic interindividual differences in neurobehavioral impairment from sleep loss: evidence of trait-like differential vulnerability. *Sleep.* 2004; 27 (3): 423-433.
22. Kapur VK, Baldwin CM, Resnick HE, Gottlieb DJ, Nieto FJ. Sleepiness in patients with moderate to severe sleep-disordered breathing. *Sleep.* 2005; 28 (4): 472-477.
23. Chen X, Wang R, Zee P, et al. Racial/Ethnic Differences in Sleep Disturbances: The Multi-Ethnic Study of Atherosclerosis (MESA). *Sleep.* 2015; 38 (6): 877-888.
24. Cacioppo JT, Hawkley LC, Berntson GG, et al. Do lonely days invade the nights? Potential social modulation of sleep efficiency. *Psychol Sci.* 2002; 13 (4): 384-387.
25. Dinges DF. The state of sleep deprivation: From functional biology to functional consequences. *Sleep Med Rev.* 2006; 10 (5): 303-305.
26. Dick KJ, Nelson CP, Tsaprouni L, et al. DNA methylation and body-mass index: a genome-wide analysis. *Lancet.* 2014; 383 (9933): 1990-1998.
27. Klengel T, Pape J, Binder EB, Mehta D. The role of DNA methylation in stress-related psychiatric disorders. *Neuropharmacology.* 2014; 80: 115-132.
28. Mendelson MM, Marioni RE, Joehanes R, et al. Association of Body Mass Index with DNA Methylation and Gene Expression in Blood Cells and Relations to Cardiometabolic Disease: A Mendelian Randomization Approach. *PLoS Med.* 2017; 14 (1): e1002215.
29. Saunderson EA, Spiers H, Mifsud KR, et al. Stress-induced gene expression and behavior are controlled by DNA methylation and methyl donor availability in the dentate gyrus. *Proc Natl Acad Sci U S A.* 2016; 113 (17): 4830-4835.
30. Quach A, Levine ME, Tanaka T, et al. Epigenetic clock analysis of diet, exercise, education, and lifestyle factors. *Aging (Albany NY).* 2017; 9 (2): 419-446.
31. Pauwels S, Ghosh M, Duca RC, et al. Dietary and supplemental maternal methyl-group donor intake and cord blood DNA methylation. *Epigenetics.* 2017; 12 (1): 1-10.
32. Sharp GC, Arathimos R, Reese SE, et al. Maternal alcohol consumption and offspring DNA methylation: findings from six general population-based birth cohorts. *Epigenomics.* 2018; 10 (1): 27-42.
33. Bild DE, Bluemke DA, Burke GL, et al. Multi-ethnic study of atherosclerosis: objectives and design. *American journal of epidemiology.* 2002; 156 (9): 871-881.

34. Liu Y, Ding J, Reynolds LM, et al. Methyloomics of gene expression in human monocytes. *Hum Mol Genet.* 2013; 22 (24): 5065-5074.
35. Reynolds LM, Taylor JR, Ding J, et al. Age-related variations in the methylome associated with gene expression in human monocytes and T cells. *Nat Commun.* 2014; 5: 5366.
36. Subramanian A, Tamayo P, Mootha VK, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A.* 2005; 102 (43): 15545-15550.
37. Du P, Kibbe WA, Lin SM. lumi: a pipeline for processing Illumina microarray. *Bioinformatics.* 2008; 24 (13): 1547-1548.
38. *Illumina450ProbeVariants.db: Annotation Package combining variant data from 1000 Genomes Project for Illumina HumanMethylation450 Bead Chip probes* [computer program]. 2013.
39. Meng H, Joyce AR, Adkins DE, et al. A statistical method for excluding non-variable CpG sites in high-throughput DNA methylation profiling. *BMC Bioinformatics.* 2010; 11: 227.
40. Johnson WE, Li C, Rabinovic A. Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics.* 2007; 8 (1): 118-127.
41. Du P, Zhang X, Huang CC, et al. Comparison of Beta-value and M-value methods for quantifying methylation levels by microarray analysis. *BMC Bioinformatics.* 2010; 11: 587.
42. Chen YA, Lemire M, Choufani S, et al. Discovery of cross-reactive probes and polymorphic CpGs in the Illumina Infinium HumanMethylation450 microarray. *Epigenetics.* 2013; 8 (2): 203-209.
43. Vargas JD, Manichaikul A, Wang XQ, et al. Common genetic variants and subclinical atherosclerosis: The Multi-Ethnic Study of Atherosclerosis (MESA). *Atherosclerosis.* 2016; 245: 230-236.
44. Johns MW. A new method for measuring daytime sleepiness: the Epworth sleepiness scale. *Sleep.* 1991; 14 (6): 540-545.
45. Levine DW, Kripke DF, Kaplan RM, et al. Reliability and validity of the Women's Health Initiative Insomnia Rating Scale. *Psychol Assess.* 2003; 15 (2): 137-148.
46. Radloff LS. The CES-D Scale: A Self-Report Depression Scale for Research in the General Population. *Applied Psychological Measurement.* 1977; 1 (3): 385-401.
47. Chiuve SE, Fung TT, Rimm EB, et al. Alternative dietary indices both strongly predict risk of chronic disease. *J Nutr.* 2012; 142 (6): 1009-1018.
48. Storey JD. A direct approach to false discovery rates. *Journal of the Royal Statistical Society: Series B (Statistical Methodology).* 2002; 64 (3): 479-498.
49. Fried LP, Borhani NO, Enright P, et al. The Cardiovascular Health Study: design and rationale. *Ann Epidemiol.* 1991; 1 (3): 263-276.
50. Aryee MJ, Jaffe AE, Corrada-Bravo H, et al. Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. *Bioinformatics.* 2014; 30 (10): 1363-1369.
51. Maksimovic J, Gordon L, Oshlack A. SWAN: Subset-quantile within array normalization for illumina infinium HumanMethylation450 BeadChips. *Genome Biol.* 2012; 13 (6): R44.
52. Fortin JP, Labbe A, Lemire M, et al. Functional normalization of 450k methylation array data improves replication in large cancer studies. *Genome Biol.* 2014; 15 (12): 503.

53. Houseman EA, Accomando WP, Koestler DC, et al. DNA methylation arrays as surrogate measures of cell mixture distribution. *BMC Bioinformatics*. 2012; 13: 86.
54. Carlson CS, Matise TC, North KE, et al. Generalization and dilution of association results from European GWAS in populations of non-European ancestry: the PAGE study. *PLoS Biol*. 2013; 11 (9): e1001661.
55. Sofer T, Heller R, Bogomolov M, et al. A powerful statistical framework for generalization testing in GWAS, with application to the HCHS/SOL. *Genet Epidemiol*. 2017; 41 (3): 251-258.
56. Sudlow C, Gallacher J, Allen N, et al. UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Med*. 2015; 12 (3): e1001779.
57. Bycroft C, Freeman C, Petkova D, et al. Genome-wide genetic data on ~500,000 UK Biobank participants. *bioRxiv*. 2017.
58. Li Q, Kellner DA, Hatch HAM, et al. Conserved properties of *Drosophila* Insomniac link sleep regulation and synaptic function. *PLoS Genet*. 2017; 13 (5): e1006815.
59. Chen H, Cade BE, Gleason KJ, et al. Multiethnic Meta-Analysis Identifies RAI1 as a Possible Obstructive Sleep Apnea-related Quantitative Trait Locus in Men. *Am J Respir Cell Mol Biol*. 2018; 58 (3): 391-401.
60. Smith AC, Dykens E, Greenberg F. Sleep disturbance in Smith-Magenis syndrome (del 17 p11.2). *Am J Med Genet*. 1998; 81 (2): 186-191.
61. Jiang L, Wei Y, Ronquillo CC, et al. Heterotrimeric kinesin-2 (KIF3) mediates transition zone and axoneme formation of mouse photoreceptors. *J Biol Chem*. 2015; 290 (20): 12765-12778.
62. Boone PM, Reiter RJ, Glaze DG, Tan DX, Lupski JR, Potocki L. Abnormal circadian rhythm of melatonin in Smith-Magenis syndrome patients with RAI1 point mutations. *Am J Med Genet A*. 2011; 155A (8): 2024-2027.
63. Cheng HY, Dziema H, Papp J, et al. The molecular gatekeeper *Dexas1* sculpts the photic responsiveness of the mammalian circadian clock. *J Neurosci*. 2006; 26 (50): 12984-12995.
64. Williams SR, Zies D, Mullegama SV, Grotewiel MS, Elsea SH. Smith-Magenis syndrome results in disruption of *CLOCK* gene transcription and reveals an integral role for RAI1 in the maintenance of circadian rhythmicity. *Am J Hum Genet*. 2012; 90 (6): 941-949.
65. Durrence HH, Lichstein KL. The sleep of African Americans: a comparative review. *Behavioral sleep medicine*. 2006; 4 (1): 29-44.
66. Jean-Louis G, von Gizycki H, Zizi F, Dharawat A, Lazar JM, Brown CD. Evaluation of sleep apnea in a sample of black patients. *J Clin Sleep Med*. 2008; 4 (5): 421-425.
67. Redline S, Tishler PV, Hans MG, Tosteson TD, Strohl KP, Spry K. Racial differences in sleep-disordered breathing in African-Americans and Caucasians. *Am J Respir Crit Care Med*. 1997; 155 (1): 186-192.
68. Whitney CW, Enright PL, Newman AB, Bonekat W, Foley D, Quan SF. Correlates of daytime sleepiness in 4578 elderly persons: the Cardiovascular Health Study. *Sleep*. 1998; 21 (1): 27-36.
69. Cespedes EM, Gillman MW, Kleinman K, Rifas-Shiman SL, Redline S, Taveras EM. Television viewing, bedroom television, and sleep duration from infancy to mid-childhood. *Pediatrics*. 2014; 133 (5): e1163-1171.

70. Paruthi S, Buchanan P, Weng J, et al. Effect of Adenotonsillectomy on Parent-Reported Sleepiness in Children with Obstructive Sleep Apnea. *Sleep*. 2016; 39 (11): 2005-2012.
71. Cunliffe VT. The epigenetic impacts of social stress: how does social adversity become biologically embedded? *Epigenomics*. 2016; 8 (12): 1653-1669.
72. Alasaari JS, Lagus M, Ollila HM, et al. Environmental stress affects DNA methylation of a CpG rich promoter region of serotonin transporter gene in a nurse cohort. *PLoS One*. 2012; 7 (9): e45813.
73. Baccarelli A, Wright RO, Bollati V, et al. Rapid DNA methylation changes after exposure to traffic particles. *Am J Respir Crit Care Med*. 2009; 179 (7): 572-578.
74. Nwanaji-Enwerem JC, Dai L, Colicino E, et al. Associations between long-term exposure to PM2.5 component species and blood DNA methylation age in the elderly: The VA normative aging study. *Environ Int*. 2017; 102: 57-65.
75. Kim TY, Siesser PF, Rossman KL, et al. Substrate trapping proteomics reveals targets of the betaTrCP2/FBXW11 ubiquitin ligase. *Mol Cell Biol*. 2015; 35 (1): 167-181.
76. Demerath EW, Guan W, Grove ML, et al. Epigenome-wide association study (EWAS) of BMI, BMI change and waist circumference in African American adults identifies multiple replicated loci. *Hum Mol Genet*. 2015; 24 (15): 4464-4479.
77. Jewett KA, Krueger JM. Humoral sleep regulation; interleukin-1 and tumor necrosis factor. *Vitam Horm*. 2012; 89: 241-257.
78. Zielinski MR, Gerashchenko D, Karpova SA, et al. The NLRP3 inflammasome modulates sleep and NREM sleep delta power induced by spontaneous wakefulness, sleep deprivation and lipopolysaccharide. *Brain Behav Immun*. 2017; 62: 137-150.
79. Muller J, Cacace AM, Lyons WE, McGill CB, Morrison DK. Identification of B-KSR1, a novel brain-specific isoform of KSR1 that functions in neuronal signaling. *Mol Cell Biol*. 2000; 20 (15): 5529-5539.
80. Moller-Levet CS, Archer SN, Bucca G, et al. Effects of insufficient sleep on circadian rhythmicity and expression amplitude of the human blood transcriptome. *Proc Natl Acad Sci U S A*. 2013; 110 (12): E1132-1141.
81. Chen YC, Chen TW, Su MC, et al. Whole Genome DNA Methylation Analysis of Obstructive Sleep Apnea: IL1R2, NPR2, AR, SP140 Methylation and Clinical Phenotype. *Sleep*. 2016; 39 (4): 743-755.
82. Dumaine JE, Ashley NT. Acute sleep fragmentation induces tissue-specific changes in cytokine gene expression and increases serum corticosterone concentration. *Am J Physiol Regul Integr Comp Physiol*. 2015; 308 (12): R1062-1069.
83. Beaudreau SA, Spira AP, Stewart A, et al. Validation of the Pittsburgh Sleep Quality Index and the Epworth Sleepiness Scale in older black and white women. *Sleep Med*. 2012; 13 (1): 36-42.
84. Gaspar LS, Alvaro AR, Moita J, Cavadas C. Obstructive Sleep Apnea and Hallmarks of Aging. *Trends Mol Med*. 2017; 23 (8): 675-692.

Table 1: MESA Study Participants Characteristics

	<u>Overall</u>	<u>AA</u>	<u>HA</u>	<u>EA</u>	<u>P-value</u>
Total	619	132	202	285	
Age (years)	68.7 (9.24)	69.1 (9)	67.3(8.88)	69.6(9.6)	0.024
Sex, males (%)	290 (47%)	53 (40%)	99 (49%)	138 (48%)	0.22
Study Site					<0.001
John Hopkins University	165 (27%)	79 (60%)	0 (0%)	86 (30%)	
Columbia University	193 (31%)	51 (39%)	104 (51%)	38 (13%)	
Wake Forest University	25 (4%)	2 (1%)	0 (0%)	23 (8%)	
University of Minnesota	236 (38%)	0 (0%)	98 (49%)	138 (49%)	
Epworth Sleepiness Scale (ESS) Score	6.0 (4.04)	6.7 (4.5)	5.8 (4.1)	5.9 (3.8)	0.08
Excessive Daytime Sleepiness (ESS>10)	85 (14%)	24 (18%)	23 (11%)	38 (13%)	0.2036
BMI (kg/m ²)	29.8 (5.6)	30.7 (5.5)	30.3 (5.4)	29 (5.6)	0.002
AHI (events/hr)	19.6 (18.8)	17.7 (17.8)	21.4 (18.3)	19.1 (19.6)	0.18
Moderate to Severe Apnea (AHI>15)	267 (43%)	53 (40%)	105 (52%)	109 (38%)	0.03
Nocturnal Hypoxemia (%time < 90% saturation)	4.2 (10.3)	3.3 (9.4)	3.4 (7)	5.2 (12.5)	0.097
Nocturnal Hypoxemia >=5%	108 (17%)	18 (14%)	34 (17%)	56 (20%)	0.17
Sleep Duration (minutes)	391.42 (79.8)	360.48 (87.1)	385.13 (78.8)	410.2 (71.8)	< 0.001
Less than 5 hours of sleep a night	73 (12%)	32 (24%)	22 (11%)	19 (6%)	< 0.001
Less than a college education (%)	192 (32%)	40 (31%)	98 (52%)	54 (19%)	< 0.001
Mother having less than a college education (%)	485 (81%)	107 (84%)	181 (96%)	197 (70%)	<0.001
Doctor Diagnosed Insomnia	34 (6%)	9 (7%)	14 (7%)	11 (4%)	0.26
WHI Insomnia Rating Scale	7.7 (4.68)	7.7 (4.51)	8.0 (4.86)	7.4 (4.63)	0.49
CES Depression Scale	8.3 (7.3)	8.2 (7.2)	9.2 (7.9)	7.8 (6.9)	0.14
Currently Drink Alcohol	286 (46%)	53 (40%)	66 (33%)	167 (59%)	<0.001
Smoking Status					0.005
Never	242	50	90	102	

	Former	325	62	100	163	
	Current	51	20	11	20	
Nutrients: Total Fat (g)		73 (44.1)	79.1 (53.5)	68.7 (44.1)	73.06 (38.9)	0.1126
Nutrients: Total Carbohydrates (g)		215.1 (115.4)	226.9 (144.9)	218.1 (125.3)	207.7 (90.1)	0.2636
Nutrients: Total Saturated Fatty Acids (g)		22.2 (14.7)	22.2 (17.4)	20.5 (14)	23.3 (13.7)	0.1189
Long chain fatty acid score (mg/d)		2.6 (2.36)	3.3 (2.8)	2.5 (2.4)	2.2 (2.03)	<0.001
Total AHEI-2010		48.6(12.5)	48.3(12.2)	46.7 (13.7)	50 (11.6)	0.0175

Data presented as mean (SD) or n (%). CES depression scale originated from the Center for Epidemiological Studies. AHEI is the Alternative Healthy Eating Index. WHI is the Women's Health Initiative ranges from 0 to 20. Test of difference either done via regression (continuous variables) or chi-squared test (categorical variable).

Table 2: CHS Study Participants Characteristics

	<u>Overall</u>	<u>African Americans</u>	<u>European Americans</u>	<u>P-value</u>	
Total	483	238	245		
Age, years	74.03 (5.3)	72.79 (5.4)	75.23(4.9)	<0.001	
Sex, males (%)	177 (37%)	82 (34%)	95 (38%)	0.3729	
Epworth Sleepiness Scale Score	5.98 (4.1)	6.38 (4.4)	5.58 (3.8)	0.037	
Excessive Daytime Sleepiness (ESS>10)	72 (15%)	43 (18%)	29 (12%)	0.07077	
BMI (kg/m ²)	27.8 (5)	29.0 (4.9)	26.7 (4.8)	<0.001	
Currently Drink Alcohol	209 (43%)	97 (41%)	112 (46%)	0.3136	
Smoking Status				0.07077	
	Never	216 (45%)	108 (45%)	108 (44%)	
	Former	207 (43%)	93 (39%)	114 (47%)	
	Current	60 (12%)	37 (16%)	23 (9%)	

Data is presented as mean (SD) or n (%).

Table 3: Methylation sites associated (FDR q-value \leq 0.1) with the Epworth Sleepiness Scale (square root transformed) in African Americans, MESA (n= 132) along with their association in CHS AA (n=238)

Probe	CHR.	Location	Gene	Effect			Effect	
				Estimate in MESA AA ^a	MESA P-value ^a	MESA P-value ^b	Estimate in CHS AA ^c	P-value in CHS AA ^c
cg26609398	16	2752235	KCTD5	-1.98	3.69E-07	1.06E-03	-3.3	2.77E-02
cg26927427	16	57832230	KIFC3	-1.27	7.71E-07	2.48E-04	-0.11	4.60E-01
cg09411874	12	121905955	KDM2B	-2.04	5.62E-06	1.18E-03	-3.27	3.30E-02
cg16591159	16	31487813	TGFB11	-0.77	6.60E-06	7.56E-04	0.28	6.62E-01
cg00917156	16	57831841	KIFC3	-1.07	9.14E-06	2.20E-03	0.4	5.84E-01
cg27208169	17	17603584	RAI1	-1.75	9.50E-06	4.59E-04	0.99	7.68E-01
cg20359888	15	85362883	ALPK3	-2.36	1.07E-05	3.55E-03	-2.7	6.09E-02
cg17585031	17	25798942	KSR1	-2.42	1.25E-05	1.46E-02	0.16	5.19E-01
cg03661817	16	57831745	KIFC3	-0.66	1.47E-05	8.80E-04	0.37	6.68E-01
cg14720773	17	17603531	RAI1	-1.46	1.80E-05	2.13E-03	0.71	7.17E-01
cg16111231	6	160399700	IGF2R	-1.62	1.80E-05	2.24E-01	-3.4	5.75E-02
cg16523643	15	37395171		1.21	1.90E-05	1.61E-03	NA	NA
cg26130090	12	54615724		-1.74	1.92E-05	1.91E-03	-6.16	3.95E-04
cg23781467	6	31760233	VAR5	-1.96	2.22E-05	1.12E-03	-0.36	4.13E-01
cg25782445	7	99006594	BUD31, PDAP1	2.52	2.35E-05	4.77E-04	0.34	4.46E-01
cg18115658	5	138720490	SLC23A1	-2.1	2.36E-05	9.12E-03	-2.55	9.40E-02
cg20676580	19	44115063	SRRM5, ZNF428	-2.37	2.84E-05	3.32E-03	-1.92	2.45E-01
cg07026490	16	57837114	KIFC3	-2.48	3.15E-05	3.29E-03	-1.91	2.22E-01
cg12473797	1	3036358	PRDM16	-1.46	3.24E-05	8.56E-04	-1.86	9.01E-02

cg27312338	16	57832198	KIFC3	-0.91	3.31E-05	1.27E-03	-0.25	4.20E-01
cg24643922	10	134236975		-1.39	3.36E-05	5.49E-04	-0.57	3.80E-01
cg24338748	17	34947916	DHRS11	-1.48	3.47E-05	4.90E-04	0.61	6.53E-01
cg25722983	1	36840028	STK40	-2.5	3.85E-05	4.38E-04	-0.16	4.63E-01
cg00898920	2	241504756		-2.04	4.13E-05	1.09E-02	-2.98	9.86E-02
cg08462988	3	147087225		-1.2	4.70E-05	2.56E-04	-0.07	4.84E-01
cg16218721	2	177017352	HOXD4	1.97	4.81E-05	9.68E-03	1.53	2.51E-01
cg21186098	8	108208344		-1.87	4.84E-05	8.15E-04	1.15	7.83E-01
cg10127660	8	21996234	REEP4	-1.75	5.14E-05	2.93E-04	0.2	5.46E-01
cg18328933	3	52008538	ABHD14B,ABHD14A	-1.99	5.18E-05	9.30E-03	-4.02	1.80E-02
cg02391615	11	73061896	ARHGEF17	-1.96	5.95E-05	4.93E-04	-1.99	1.47E-01
cg13253439	6	30614397	C6orf136	-1.8	6.04E-05	1.63E-04	0.31	5.75E-01
cg24751129	6	42928920	GNMT	-1.46	6.08E-05	9.93E-04	-2.61	3.67E-02
cg04610235	17	56592440	MTMR4	-2.68	6.19E-05	6.98E-04	0.39	5.61E-01
cg20161984	4	84205961	COQ2	3.31	6.26E-05	2.15E-04	4.07	8.54E-02
cg09284397	7	128312406	FAM71F2	-1.55	6.41E-05	2.90E-03	-3.32	8.02E-02
cg20492868	1	42202048	HIVEP3	-2.07	6.42E-05	2.22E-04	-3.94	1.65E-02
cg27045264	16	31441954		-1.3	6.62E-05	2.95E-02	-1.72	1.23E-01
cg10603290	5	14154788	TRIO	-1.17	6.65E-05	1.47E-02	-0.22	4.11E-01
cg19628469	19	856107	ELANE	-0.92	6.71E-05	5.96E-04	0.78	7.12E-01
cg20003638	20	44687730	SLC12A5	-0.9	6.85E-05	8.60E-04	0.04	5.12E-01
cg22798214	1	212692190		-1.95	6.89E-05	8.53E-03	-2.91	5.74E-02
cg09304397	2	235580461		-1.66	7.30E-05	6.90E-03	-0.11	4.69E-01
cg09076650	7	73424891		-1.57	7.39E-05	2.99E-02	0.73	6.79E-01
cg00549412	21	43933482	SLC37A1	-1.15	7.48E-05	5.79E-03	-1.46	1.33E-01
cg27646484	16	31210736		-2.3	7.62E-05	8.82E-04	-0.81	3.44E-01

cg25588826	17	7755583	KDM6B	-1.9	7.67E-05	2.32E-03	-0.89	3.30E-01
cg01771651	6	33162444	RXRB	-2.14	7.94E-05	2.05E-02	-4.43	1.40E-02
cg20652042	12	6883826	LAG3	-1.98	8.08E-05	1.22E-04	0.65	6.23E-01
cg04673737	1	41830161		-2.86	8.13E-05	2.05E-02	-4.8	1.79E-02
cg14020176	17	72764985	SLC9A3R1	-1.78	8.35E-05	2.19E-03	0.58	6.07E-01
cg01373896	19	1854724	KLF16	-1.34	8.38E-05	3.79E-03	-2.08	1.66E-01
cg04172049	17	47931091		-2.09	8.44E-05	2.57E-03	-1.83	2.15E-01
cg02625481	2	183667124		-1.97	8.52E-05	8.57E-04	0.43	5.84E-01
cg27144592	16	783916	NARFL	-0.96	8.55E-05	4.63E-04	-1.09	1.59E-01
cg17076667	10	121418050	BAG3	-1.06	8.65E-05	1.61E-01	0.28	6.02E-01
cg11069824	1	6205325	CHD5	-1.92	8.83E-05	4.98E-03	-1.1	2.47E-01
cg25371385	2	157177665		1.29	9.83E-05	6.55E-02	1.09	2.51E-01
cg10857774	19	46286683	DMPK,DMWD	-1.95	9.98E-05	2.34E-03	-1.67	2.09E-01
cg00947878	16	57832005	KIFC3	-0.95	9.99E-05	3.83E-03	0.45	7.08E-01
cg14780449	11	44578801		-2.37	1.05E-04	5.73E-04	0.47	5.97E-01
cg14204600	14	104000460	TRMT61A	-1.99	1.05E-04	6.16E-03	-2.51	9.14E-02

^a Models adjusted for age, sex, residual cell type enrichment, site of recruitment, and top five principal components calculated in AA.

^b Models adjusted for age, sex, residual cell type enrichment, recruitment site, top five principal components in the overall sample, and self-reported ancestry when testing for interaction between DNAm M-value and self-reported ancestry in overall sample. ^c Association results in CHS AAs. P-value is based on one-sided p-values informed by the direction of effect. Genomic location is provided in genome build 37.



