

## **Analysis of C-reactive protein omics-measures associates methylation risk score with sleep health and related health outcomes**

Ziqing Wang<sup>1,2</sup>, Danielle A Wallace<sup>1,2,3</sup>, Brian W Spitzer<sup>1</sup>, Tianyi Huang<sup>4</sup>, Kent Taylor<sup>5</sup>, Jerome I Rotter<sup>5</sup>, Stephen S Rich<sup>6</sup>, Peter Y Liu<sup>7</sup>, Martha L. Daviglus<sup>8</sup>, Lifang Hou<sup>8</sup>, Alberto R Ramos<sup>9</sup>, Sonya Kaur<sup>9</sup>, J Peter Durda<sup>10</sup>, Hector M González<sup>11</sup>, Myriam Fornage<sup>12,13</sup>, Susan Redline<sup>2,3</sup>, Carmen R Isasi<sup>14</sup>, Tamar Sofer<sup>1,2,3,15</sup>

<sup>1</sup>Cardiovascular Institute, Beth Israel Deaconess Medical Center, Boston, MA, USA

<sup>2</sup>Department of Medicine, Harvard Medical School, Boston, MA, USA

<sup>3</sup>Division of Sleep Medicine and Circadian Disorders, Department of Medicine, Brigham and Women's Hospital, Boston, MA, USA

<sup>4</sup>Laboratory of Epidemiology and Population Sciences, Intramural Research Program, National Institute on Aging, Baltimore, MD, USA

<sup>5</sup>The Institute for Translational Genomics and Population Sciences, Department of Pediatrics, The Lundquist Institute for Biomedical Innovation at Harbor-UCLA Medical Center, Torrance, CA, USA

<sup>6</sup>Center for Public Health Genomics, University of Virginia School of Medicine, Charlottesville, VA, USA

<sup>7</sup>Division of Genetics, Lundquist Institute at Harbor-UCLA Medical Center, Torrance, CA, USA

<sup>8</sup>Department of Preventive Medicine, Northwestern University Feinberg School of Medicine, Chicago, IL, USA

<sup>9</sup>Department of Neurology, University of Miami Miller School of Medicine, Miami, FL, USA

<sup>10</sup>Department of Pathology and Laboratory Medicine, University of Vermont, Burlington, Vermont, USA

<sup>11</sup>Department of Neurosciences and Shiley-Marcos Alzheimer's Disease Center, University of California, San Diego, La Jolla, CA, USA

<sup>12</sup>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, USA

<sup>13</sup>Brown Foundation Institute of Molecular Medicine, McGovern Medical School, University of Texas Health Science Center at Houston, Houston, TX, USA

<sup>14</sup>Department of Epidemiology & Population Health, Department of Pediatrics, Albert Einstein College of Medicine, Bronx, NY, USA

<sup>15</sup>Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, MA, USA

## Abstract

### Introduction

DNA methylation (DNAm) predictors of high sensitivity C-reactive protein (CRP) offer a stable and accurate means of assessing chronic inflammation, bypassing the CRP protein fluctuations secondary to acute illness. Poor sleep health is associated with elevated inflammation (including elevated blood CRP levels) which may explain associations of sleep insufficiency with metabolic, cardiovascular and neurological diseases. Our study aims to characterize the relationships among sleep health phenotypes and CRP markers —blood, genetic, and epigenetic indicators— within the Hispanic Community Health Study/Study of Latinos (HCHS/SOL).

### Methods

In HCHS/SOL, methylation risk scores (MRS)-CRP and polygenetic risk score (PRS)-CRP were constructed separately as weighted sums of methylation beta values or allele counts, respectively, for each individual. Sleep health phenotypes were measured using self-reported questionnaires and objective measurements. Survey-weighted linear regression established the association between the multiple sleep phenotypes (obstructive sleep apnea (OSA), sleep duration, insomnia and excessive sleepiness symptom), cognitive assessments, diabetes and hypertension with CRP markers while adjusting for age, sex, BMI, study center, and the first five principal components of genetic ancestry in HCHS/SOL.

### Results

We included 2221 HCHS/SOL participants (age range 37-76 yrs, 65.7% female) in the analysis. Both the MRS-CRP (95% confidence interval (CI): 0.32-0.42,  $p = 3.3 \times 10^{-38}$ ) and the PRS-CRP (95% CI: 0.15-0.25,  $p = 1 \times 10^{-14}$ ) were associated with blood CRP level. Moreover, MRS-CRP was associated with sleep health phenotypes (OSA, long sleep duration) and related conditions (diabetes and hypertension), while PRS-CRP markers were not associated with these traits. Circulating CRP level was associated with sleep duration and diabetes. Associations between OSA traits and metabolic comorbidities weakened after adjusting for MRS-CRP, most strongly for diabetes, and least for hypertension.

### Conclusions

MRS-CRP is a promising estimate for systemic and chronic inflammation as reflected by circulating CRP levels, which either mediates or serves as a common cause of the association between sleep phenotypes and related comorbidities, especially in the presence of diabetes.

## Introduction

Sleep disorders are risk factors for impaired neurological function, metabolic and cardiovascular diseases (CVD)<sup>1</sup>. For instance, obstructive sleep apnea (OSA), characterized by episodic upper airflow narrowing or collapse<sup>2</sup>, causes intermittent hypoxia, systemic inflammation and endothelial dysfunction. These physiological disturbances are considered key pathways linking OSA to increased risk for hypertension, cardiovascular and cerebrovascular diseases<sup>3,4</sup>. Sleep fragmentation and restriction-characteristics of many sleep disorders, including insomnia, have also been shown to elevate inflammation level and cardiovascular morbidity<sup>5-7</sup>. Furthermore, excessive daytime sleepiness (EDS) is a symptom of insufficient sleep or heightened sleep propensity and is also found associated with inflammation<sup>8</sup>.

C-reactive protein (CRP), as a blood marker of systemic inflammation, is commonly used as a risk indicator for cardiometabolic diseases, including coronary heart disease, diabetes, and hypertension<sup>9-11</sup>. However, circulating CRP levels fluctuate following acute diseases, diminishing its stability and accuracy in estimating chronic inflammation. While sleep disorders, including OSA, have been associated with systemic inflammation<sup>12</sup>, early studies reported conflicting results on associations of OSA with blood CRP levels, likely due to limitations such as small sample size, study designs overlooking confounding factors including body mass index (BMI), detection threshold of CRP and so on<sup>13,14</sup>.

DNA methylation (DNAm) is an epigenetic modification essential for transcriptional regulation and various fundamental biological processes<sup>15</sup>, impacted by both genetics and environment<sup>16</sup>. While DNAm has site-specific effects, overall DNAm variability over time in adults is relatively stable<sup>17</sup>. Methylation risk scores (MRS) have been developed as weighted sums of DNAm at CpG sites and tested for their associations with health outcomes or biomarkers (such as aging or CRP) to infer exposures and disease states, similar to polygenic risk scores (PRS). Recently, an MRS for CRP (MRS-CRP) has emerged as a more accurate marker for chronic inflammation than circulating CRP, exhibiting enhanced test-retest reliability and stronger associations with long term health outcomes<sup>18-21</sup>. CRP levels are also partly determined by genetics, with genetic loci associated with CRP explain approximately 10% of variation in blood CRP level<sup>21,22</sup>.

The current study aims to investigate the relationships between circulating, epigenetic and genetic CRP patterns and common sleep problems (OSA; insomnia and EDS; sleep duration); as well as sleep-related comorbidities (hypertension, diabetes and cognitive decline); in the Hispanic Community Health Study/Study of Latinos (HCHS/SOL) cohort. For secondary analysis, we estimated the associations of sleep phenotypes with sleep-associated comorbidities, with and without adjusting for MRS-CRP to assess the extent to which MRS-CRP potentially explains the associations. Additionally, lasso penalized regression was performed to select key CRP-associated CpG sites associated with sleep traits and their comorbidities. The contribution of this paper is two-fold: 1) addressing the knowledge gap regarding the associations between

MRS-CRP and sleep traits; 2) exploring associations between CRP omics measures with sleep, metabolic syndrome and cognitive function in a Hispanic population.

## Methods

Figure 1 illustrates the overview of the study design. Since our primary interest is modeling associations in the HCHS/SOL cohort, we first constructed various PRS-CRPs in an independent cohort, the Multi-Ethnic Study of Atherosclerosis (MESA) using multiple methods and summary statistics. The best-performing PRS-CRP, selected based on its association with circulating CRP in MESA, and MRS-CRP were then computed using data from the Hispanic Community Health Study/Study of Latinos (HCHS/SOL) cohort, followed by association tests with blood CRP levels. We evaluated associations of these CRP omics-measures with sleep, cognitive and metabolic comorbidities via survey linear regression, and conducted secondary analyses to further explore the relationships between systemic inflammation and these health conditions from epigenetic and genetic perspectives.

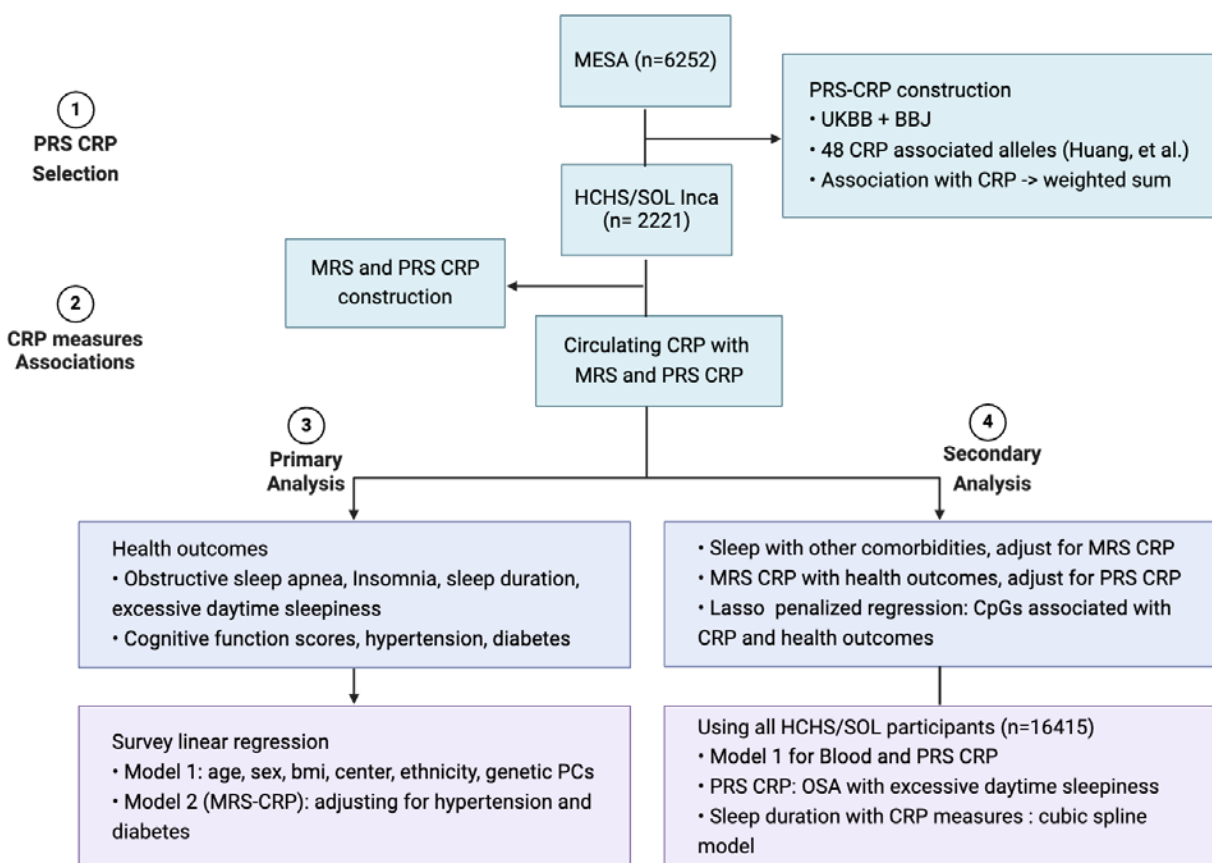


Figure 1. Study design. Best-performing polygenic risk score (PRS) of C-reactive protein (CRP) was first selected in MESA as the one explaining the highest variance in circulating CRP. Then association of circulating CRP with Methylation risk score (MRS) and PRS constructed in SOL-INCA was confirmed before evaluating the association of these CRP measures with sleep and other health outcomes.

## The MESA Dataset

The MESA is a multi-center longitudinal cohort study comprising 6,814 men and women aged 45–84 years, self-identified from four race/ethnic groups without overt clinical CVD at the time of enrollment (Exam 1; 2000-2002)<sup>23</sup>. The current study included all participants with CRP measured at Exam 1 (N=6,415) with appropriate consent and excluded N=163 individuals without genetic data in the current dataset.

**Assessment of CRP and Demographic Characteristics.** High sensitivity CRP was measured during MESA Exam 1 using a BNII nephelometer with particle-enhanced immunonephelometric assays (Dade Behring Inc, Deerfield, IL). Sociodemographic characteristics, including age, sex and medical history, were collected using standard self-reported questionnaires. To report demographic characteristics, participants were stratified by clinical CRP risk groups for CVD (< 1mg/L [low risk], 1–3 mg/L [Borderline], and  $\geq$  3mg/L [elevated risk])<sup>24</sup>.

**Whole-genome sequencing (WGS) and PRS construction.** MESA WGS data was obtained from the Trans-Omics in Precision Medicine (TOPMed) program Freeze 10 release, with reads aligned to the human-genome build GRCh38 by a common pipeline across all sequencing centers<sup>25</sup>. Variant and genotype calling was performed jointly on all samples in a given TOPMed freeze<sup>25</sup>. Minor allele frequency (MAF) threshold for variant inclusion was set to  $MAF > 0.01$  (computed over MESA participants) and genetic relatedness was computed centrally by TOPMed using the PC-Relate and algorithm<sup>26</sup>.

Multiple methods were employed to construct genome-wide association analysis (GWAS)-derived PRS-CRP using summary statistics from UK Biobank (UKBB)<sup>27</sup> and Biobank Japan (BBJ)<sup>28</sup> for MESA and HCHS/SOL participants. UKBB data were divided into sets representing different genetic backgrounds: AFR (African, n = 6203), AMR (Amerindian, n = 937), EAS (East Asian, n = 2564), EUR (European, n = 400,094), and SAS (South Asian, n = 8397)<sup>27</sup>.

First, UKBB-EAS and BBJ data were meta-analyzed using GWAMA (Genome-Wide Association Meta-Analysis) software<sup>29</sup> with default parameters allowing indel alleles. The GWAMA output, together with other UKBB ancestry-specific GWAS summary statistics, was then submitted to PRS-CSx<sup>30</sup> for development of variant weights. In PRS-CSx, variants present in at least one of the UKBB linkage disequilibrium reference panels were used to develop posterior effect estimates for each of the five genetic backgrounds<sup>30</sup>, based on which PRSs were constructed using PRSice2<sup>31</sup> without clumping or thresholding. We also developed PRSs using PRSice2, based on UKBB+BBJ meta-analyzed GWAS summary statistics, with clumping and thresholding set to  $R^2 = 0.1$ , distance = 1000 Kb, and p-value =  $5 \times 10^{-8}$ ,  $1 \times 10^{-7}$ ,  $1 \times 10^{-6}$ ,  $1 \times 10^{-5}$ ,  $1 \times 10^{-4}$ ,  $1 \times 10^{-3}$ ,  $1 \times 10^{-2}$ , 0.1.

A simple PRS summing the effect estimates for 48 genome-wide significant alleles associated with CRP<sup>8</sup> (referred to as Huang et al. PRS) was also constructed using PLINK<sup>32</sup>. For comparison purposes, each calculated PRS was standardized by subtracting its mean and divided by its standard deviation.

**PRS association analysis with CRP levels in MESA.** Association between each PRS-CRP and blood CRP level was assessed using simple linear regression in MESA, while adjusting for age, sex, body mass index (BMI), race, study site and first 5 genetic PCs. The best-performing PRS-CRP was determined as the one explaining highest variance in blood CRP levels and used in subsequent analysis in HCHS/SOL. We also carried forward a PRS developed as a weighted summation of PRS-CSx PRSs with weights determined by the regression coefficients from CRP regressed over all ancestry-specific PRS-CSx PRSs in MESA, where we further used either all MESA individuals, or only MESA Hispanic individuals. We do not report the results from association analyses of this PRS in MESA due to overfitting (as weights are developed based on analysis in MESA), but it was used in HCHS/SOL based on past work that demonstrated that such a weighted sum tends to outperform other PRSs<sup>33</sup>.

**The HCHS/SOL dataset.** The HCHS/SOL is a population-based cohort study with participants of participants who were recruited from four metropolitan areas across the United States (Bronx NY, Chicago IL, Miami FL, and San Diego CA)<sup>34</sup>. By design, participants represented diverse Hispanic/Latino origins (and genetic admixture) and had a wide age range (18-74 years old at baseline). Among the total 16,415 participants, 2,221 from the Study of Latinos-Investigation of Neurocognitive Aging (SOL-INCA) ancillary study were considered for inclusion due to available genetic data and data on methylation, sleep phenotypes and circulating CRP. Sociodemographic and lifestyle variables were self-reported, including age, sex, household income level, smoking status, cigarette and alcohol consumption, as well as self-identified Hispanic background. Obesity parameters BMI and waist-hip ratio were measured following standard procedure at baseline. BMI was calculated as body weight (kg) divided by height squared (m<sup>2</sup>), and waist-hip ratio as waist circumference divided by hip circumference.

**Sleep assessment.** Multiple sleep phenotypes were assessed using self-reported questionnaires from the baseline visit<sup>35</sup>. Sleep duration was derived as weighted average of weekday and weekend sleep. Long and short sleep variables were derived from sleep duration as binary outcomes, with thresholds set at more than 9 hours and less than 6 hours. Scores indicative of insomnia and excessive daytime sleepiness were also measured using the 5-item Women's Health Initiative Insomnia Rating Scale (WHIIRS)<sup>36</sup> and the 8-item Epworth Sleepiness Scale (ESS)<sup>37</sup>, respectively. OSA was assessed using a self-applied sleep apnea monitor (ARES Unicorder 5.2; B-Alert, Carlsbad, CA) that measured nasal airflow, position, snoring, actigraphy,

and hemoglobin oxygen saturation (SpO<sub>2</sub>)<sup>38</sup> during overnight sleep. Obstructive respiratory events were defined as a 50% or greater reduction in airflow with associated desaturations of 3% or larger, lasting for at least 10 seconds. Apnea–hypopnea index (AHI) was subsequently computed as the sum of such events per estimated sleep hour<sup>35</sup>. Given the importance of measures of hypoxia in characterizing OSA severity<sup>39</sup>, we also studied several measures of overnight desaturation (minimum SpO<sub>2</sub>, average SpO<sub>2</sub> and % Time SpO<sub>2</sub><90 - the percentage of cumulative time with oxygen saturation below 90% in total sleep time).

**Assessment of C-reactive Protein and Relevant Comorbidities.** In HCHS/SOL, high sensitivity CRP was assayed with a Roche Modular P Chemistry Analyzer (Roche Diagnostics Corporation, IN) using the immunoturbidimetric method (Roche Diagnostics Corporation, IN) during the baseline visit<sup>40</sup> (Visit 1). Demographic characteristics of the HCHS/SOL target population were summarized by clinical CRP risk groups for CVD (< 1mg/L [low risk], 1–<3 mg/L [Borderline], and ≥ 3mg/L [elevated risk])<sup>24</sup>.

Diabetes status was identified (Visit 1) based on the American Diabetes Association definition, taking into account fasting serum glucose level (≥126 mg/dL), glucose tolerance test (≥200 mg/dL), medication use and hemoglobin A1c levels (HbA1c, ≥ 6.5%), whichever was available<sup>41</sup>. Hypertension was identified as the presence of either systolic (≥140 mmHg) or diastolic blood pressure (≥90 mmHg) exceeding the threshold, or by the use of antihypertensive medications<sup>41</sup>. Cognitive function score at baseline (Visit 1) was computed as the mean of the standardized scores of three cognitive tests<sup>42</sup>: 1) Brief-Spanish English Verbal Learning Test (B-SEVLT; verbal episodic learning and memory); 2) Word Fluency test (WF; verbal fluency); 3) Digit Symbol Substitution test (DSS; processing speed and executive function)<sup>42</sup>. SOL-INCA ancillary study, conducted concurrently or within a year of Visit 2 (7 years after Visit 1, on average), repeated the same neuropsychological tests with additional two Trail Making Tests (TMT-A and B; executive function), which were also included in the calculation of the cognitive function score<sup>42</sup>. Cognitive change was defined as the difference between the cognitive function score at SOL-INCA and Visit 1.

**Genotype imputation and PRS construction.** HCHS/SOL participants were genotyped on a custom Illumina HumanOmni2.5-8v1-1array with 150,000 added custom variants, including ancestry informative single-nucleotide polymorphism (SNPs)<sup>43</sup>. Over 12,000 samples were imputed using the TOPMed imputation server<sup>43</sup>. As in MESA, MAF threshold was applied for variant exclusion in the SOL-INCA dataset, requiring MAF≥0.01, where MAF was computed in HCHS/SOL. PRS-CRP for the HCHS/SOL dataset used in the study was computed using the same method as the best performing PRS-CRP selected in MESA, and PRS summation of PRS-CSx was computed with weights obtained from regression of CRP on ancestry-specific PRS-CSx PRSs using all MESA individuals and only MESA Hispanic individuals, as described above.

**Epigenome-wide methylation quantification and MRS construction.** Methylation profiling of whole blood samples from SOL-INCA participants across the genome was performed using the Infinium MethylationEPIC array (Illumina Inc, San Diego, CA). The quality control process involved filtering for sex and SNPs mismatches, as well as removal of control and blind duplicate samples. Subsequently, methylation data were normalized across all samples using the R package SeSAMe<sup>44</sup> to mask 105,454 probes, followed by dye bias correction, and normal-exponential deconvolution using out-of-band probes (Noob) background subtraction<sup>45</sup>. The final methylation values were calculated as the beta-values, defined as the percentage of methylation signal for each individual CpG site. The obtained beta-values underwent further correction for type-2 probe bias using beta mixture quantile normalization (BMIQ) from the WateRmelon package<sup>46</sup>, followed by ComBat batch correction to account for position on array, slide and plate<sup>47</sup>. Cell type proportions were estimated using a reference-based method based on the genome-wide methylation data.

Methylation risk scores (MRS) of CRP were constructed as weighted sums of each individual's DNAm values of pre-developed CpG sites using *MethParquet* package (v0.1.0)<sup>48</sup> in R (v4.3.1)<sup>49</sup>, selected via an Elastic Net regression, associated blood CRP level<sup>21</sup>, and standardized across the sample by subtracting the mean before dividing it by its standard deviation. Of the 1468 CpGs previously identified for constructing the MRS for CRP<sup>21</sup>, 1041 were available for MRS calculation.

**CRP association analysis.** Association analyses were performed using survey linear regression models (*survey* package v4.2.1)<sup>50</sup> in R (v4.3.1)<sup>49</sup> to account for the HCHS/SOL study design and generate inference applicable to the HCHS/SOL target population. Study design features accounted for stratification, sampling weights, and non-response. Blood CRP level was modeled as a continuous variable in all analyses and log-transformed to account for the right skewed distribution. OSA related traits and cognitive phenotypes mentioned above were modeled as continuous variables, while excessive daytime sleepiness (EDS, ESS >10), insomnia (WHIIRS ≥10), hypertension and diabetes were binary indicators.

The associations of PRS- and MRS-CRP with blood CRP level were first validated, followed by estimating the association between sleep traits and related comorbidities with each CRP measure. The primary model 1 adjusted for age, sex, BMI, study center, Hispanic/Latino background, and first 5 genetic PCs in case of PRS and MRS-CRP. Model 2 was fitted further adjusting for hypertension and diabetes status. Statistical significance was based on multiple comparison adjustment with the Benjamini-Hochberg FDR correction on model 1 (FDR q-value < 0.05), grouped by each CRP variables. The primary model 1 adjusted for age, sex, BMI, study center, and first 5 genetic ancestry principal components (PCs) in case of PRS and MRS-CRP. Model 2 was fitted further adjusting for hypertension and diabetes status. Statistical



significance was based on multiple comparison adjustment with the Benjamini-Hochberg FDR correction on model 1 (FDR  $q$ -value $<0.05$ ), grouped by each CRP variables.

### **Secondary and sensitivity analyses.**

We estimated the associations of sleep phenotypes that were found to be associated with MRS-CRP in model 1, with diabetes, hypertension and cognitive scores with and without adjusting for MRS-CRP. In addition, lasso penalized regression was applied to select, from the CRP-associated CpG sites used to construct MRS-CRP, those also important to sleep and health outcomes identified in previous association analyses, adjusting for age, sex, BMI, study center, first 5 genetic PCs and six cell-type proportions in blood estimated based on DNA methylation data. Particularly, AHI was modeled as binary variable based on the threshold of 15 to categorize OSA groups into moderate to severe and mild to normal cases.

We also compared MRS-CRP associations with health outcomes from model 1 with and without adjusting for PRS-CRP. To account for potential U-shaped association of sleep duration with MRS- and blood CRP, we fitted penalized cubic spline models using *mgcv* package (v1.9.1)<sup>51</sup> in R (v4.3.1)<sup>49</sup>, adjusting for the same covariates as in model 1. The basis dimension for the spline term ( $k$ ) was selected by minimizing the Akaike Information Criterion (AIC).

Using all HCHS/SOL participants ( $n=16,415$ ) and those with genotype data ( $n=12,343$ ), we repeated the primary association analysis (model 1) for blood-CRP and PRS-CRP respectively, to increase power for detecting associations. Furthermore, we classified OSA cases (AHI $>5$ ) according to the absence or presence of EDS cases and tested the association of OSA cases with and without EDS with PRSs, as described in the previous paper<sup>8</sup>.

Given that DNAm profile derived from whole blood represents an average across all present cell types, and the association of systemic inflammation with central obesity and smoking, we performed sensitivity analyses for MRS-CRP adjusting for cell-type proportions, smoking status and waist-hip ratio, respectively.

## **Results**

**Demographics and outcome characteristics of study samples.** Sample characteristics for MESA are summarized in Table S1. The number of participants in the three CVD risk groups is evenly distributed. There are race/ethnic differences in CRP levels, with the proportion of Chinese individuals decreasing by 18% while the proportion of Black and Hispanic individuals increased by 10% between the low risk and elevated risk groups.

Descriptive statistics of demographic characteristics in the primary analytical sample are provided in Table 1. Overall, women constitute 65.8% of this sample, with higher female percentage, 75.7%, among participants with blood CRP level larger than 3 mg/L. Note that the high proportion of women is due to prioritization of individuals with Mild Cognitive Impairment (MCI) into the SOL-INCA methylation study, combined with higher MCI prevalence in women

compared to men. Most OSA related measurements, along with hypertension and diabetes, exhibit greater severity and incidence rate in parallel with higher CRP blood level, whereas sleep duration, insomnia, daytime sleepiness and cognitive scores show little difference across the three risk groups (Table 1).

Table 1. Demographic characteristics of HCHS/SOL stratified by cardiovascular risk groups based on C-reactive protein levels (mg/L)

	Overall (n = 2221)	Low risk (<1) (n = 456)	Borderline (1-3) (n = 849)	Elevated risk (>3) (n = 916)
Age (Mean (SD))	56.66 (8.22)	56.77 (8.33)	56.57 (8.10)	56.69 (8.26)
BMI (Mean (SD))	30.23 (5.51)	26.86 (3.79)	29.38 (4.57)	32.70 (5.87)
Sex (%) - Female	1435 (58.4)	227 (44.9)	520 (53.6)	688 (69.8)
CRP (mg/L) (Mean (SD))	4.32 (10.30)	0.60 (0.21)	1.87 (0.58)	8.44 (15.07)
<b>Background (%)</b>				
Central American	214 (6.1)	44 (5.4)	78 (6.9)	92 (5.6)
Cuban	380 (28.4)	70 (25.0)	126 (25.0)	184 (33.2)
Dominican	241 (10.2)	49 (10.6)	93 (10.1)	99 (10.0)
Mexican	754 (26.9)	178 (34.4)	313 (29.2)	263 (20.9)
Puerto Rican	432 (20.6)	77 (17.4)	156 (19.8)	199 (23.1)
South American	159 (6.1)	32 (5.4)	64 (6.9)	63 (5.7)
<b>Sleep traits (mean (SD))</b>				
AHI	8.16 (12.46)	6.11 (10.37)	7.41 (10.80)	9.94 (14.48)
Min SpO2	85.75 (6.49)	86.67 (6.34)	86.26 (6.14)	84.78 (6.77)
Average SpO2	96.29 (0.89)	96.46 (0.81)	96.34 (0.80)	96.15 (1.0)
% Time SpO2<90	1.03 (2.99)	0.69 (2.34)	0.88 (2.67)	1.36 (3.53)
ESS <sup>a)</sup>	6.07 (4.96)	5.95 (4.86)	5.94 (4.94)	6.25 (5.04)
WHIIRS <sup>b)</sup>	7.96 (5.57)	7.52 (5.37)	7.78 (5.42)	8.35 (5.78)
Sleep Duration (Mean (SD))	7.80 (1.43)	7.70 (1.30)	7.79 (1.36)	7.85 (1.56)
<b>Comorbidities (%)</b>				
Diabetes	678 (30.8)	101 (22.0)	235 (28.6)	342 (37.3)
Hypertension	988 (46.1)	163 (37.2)	367 (42.8)	458 (53.7)
<b>Cognitive traits (Mean (SD))</b>				
Global Cognitive Function				
Baseline	0.06 (0.76)	0.06 (0.76)	0.07 (0.77)	0.06 (0.75)
Global Cognitive Change	-0.18 (0.56)	-0.14 (0.53)	-0.19 (0.58)	-0.18 (0.57)

a) ESS: Epworth Sleepiness Scale, b) WHIIRS: Women's Health Initiative Insomnia Rating

**Performance-based selection of PRS for CRP.** Table S2 provides the number of SNPs used in different PRSs. The best performing PRS-CRP was selected as the one with the highest CRP blood level variance explained in MESA (Table S3). European ancestry-specific PRS-CSx explained

more than 9% of the CRP level. Consequently, the PRS-CSx of European ancestry (PRS-EUR) was selected for subsequent association analysis in HCHS/SOL dataset. However, previous multi-cohort studies have shown that a weighted sum trained over multiple PRSs yields stronger association with related phenotypes<sup>33</sup>. Also, PRS weighted summation is the intended, developed use of PRS-CSx PRSs. Therefore, in HCHS/SOL, we also constructed a weighted sum of the five ancestry-specific PRS-CSx with adaptive weights (PRS-wsum), obtained using all MESA participants (PRS-wsum-All) and only MESA Hispanic participants (summation weights are provided in Table S4). Because PRS-wsum-all and PRS-wsum-Hispanic had comparable associations with blood CRP level in HCHS/SOL (Figure S1), we focus PRS-wsum-all in subsequent association analyses. Table S5 provides the number of SNPs used for PRS-CSx construction in HCHS/SOL.

**Associations between PRS, MRS-CRP and blood CRP level.** Both PRS and MRS-CRP are associated with blood CRP level (Figure 2). MRS-CRP showed an almost two-fold higher association than the two PRSs, yielding a coefficient of 1.43 mg/L per one standard deviation increase in MRS-CRP. On the other hand, one standard deviation increase in weighted sum of PRS-CRP amounts to 1.23 mg/L increase in blood CRP level.

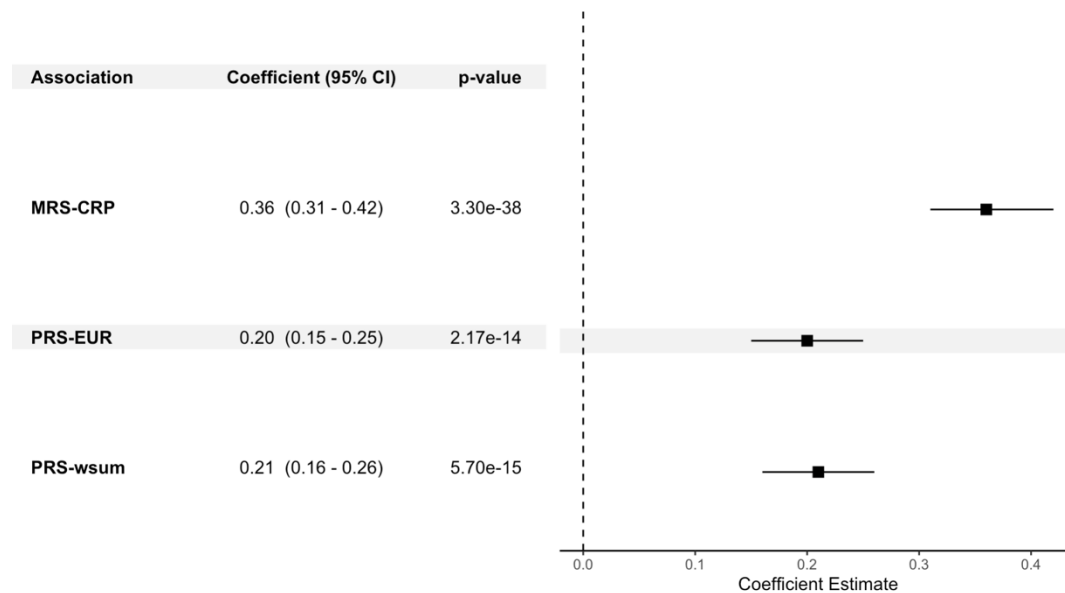


Figure 2. Association of polygenic risk score (PRS) and methylation risk score (MRS) for C-reactive protein (CRP) with blood CRP level.

**CRP associations with sleep and other comorbidities.** Figure 3 summarizes the associations between the CRP measures and phenotypic outcomes with corresponding 95% confidence intervals. Interestingly, MRS-CRP is the only metric showing statistically significant associations with AHI (1.11 [0.36, 1.86]), minimum (-0.56 [-1.05, -0.08]) and average (-0.10 [-0.17, -0.04])

SpO<sub>2</sub> (Figure 3A), implying that MRS-CRP correlates positively with OSA severity. Although not exhibiting significant associations with OSA traits, the coefficient of blood CRP level displays the same direction as that of MRS-CRP.

As shown in Figure 3B, MRS-CRP is also associated with insomnia (odds ratio (OR) = 1.19 [1.02, 1.38]) and long sleep duration (OR = 1.29 [1.08, 1.55]), while blood CRP level is associated with EDS (OR = 0.81 [0.68, 0.96]) and overall sleep duration (0.13 [0.04, 0.21]), and long sleep duration (OR = 1.31, [1.13, 1.53]). Meanwhile, short sleep duration is not significantly associated with any of the CRP measures (Figure 3B). On the other hand, PRS-CRPs are not associated with any sleep traits (Figure 3).

Regarding other comorbidities, the relationship between individuals with greater genetic susceptibility to CRP, indicated by higher standardized weighted sum of PRS-CRP, and lower cognitive function at baseline (-0.05 [-0.1, -0.01]) fails to pass the multiple testing correction (Figure 3A). Surprisingly, MRS-CRP remains the only CRP measure to have strong associations ( $q$ -value < 0.001) with both cardio-metabolic diseases, diabetes (OR = 2.44, [2.06, 2.90]) and hypertension (OR = 1.29, [1.10, 1.50]) (Figure 3B).

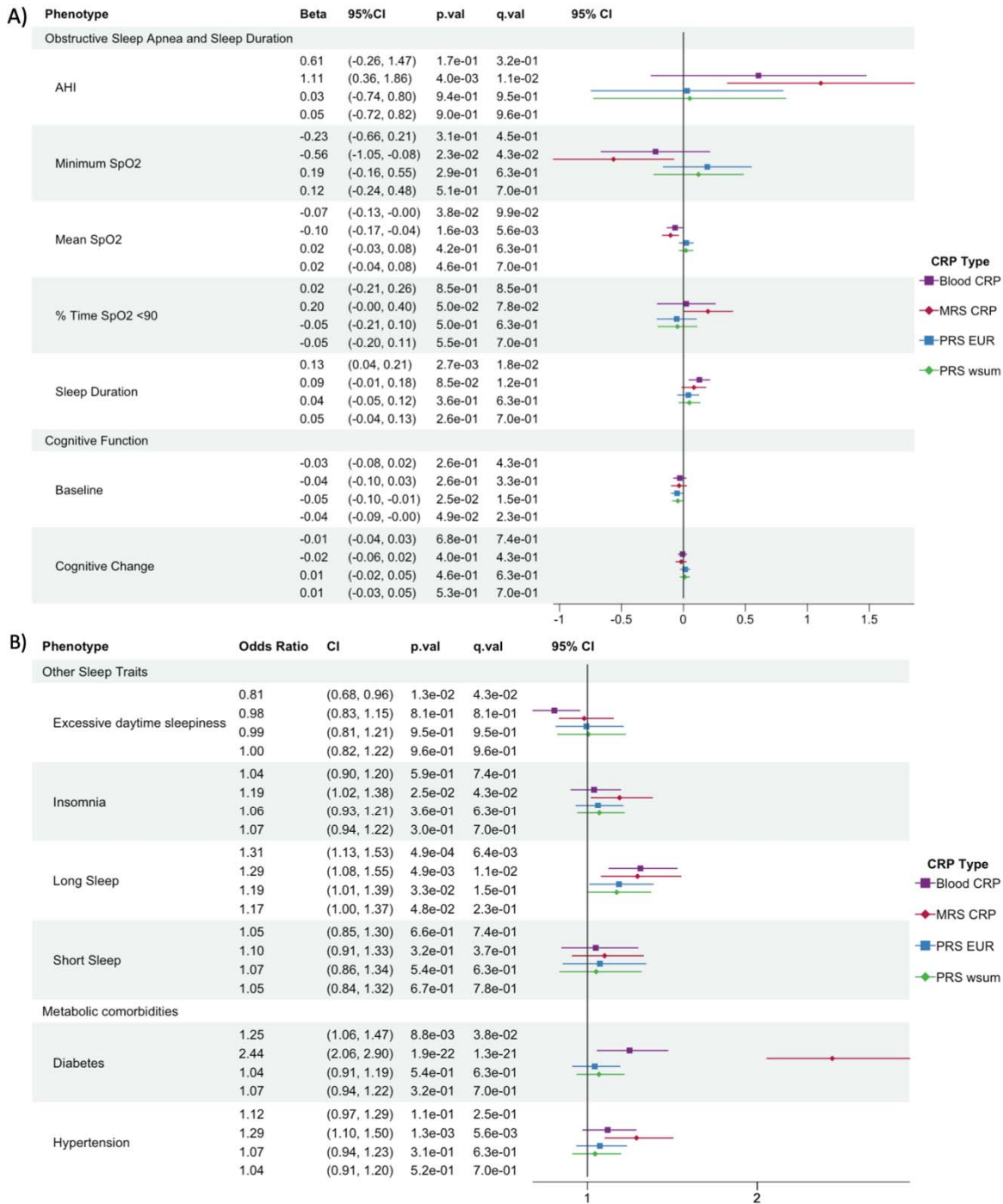


Figure 3. Forest plot showing associations between C-reactive protein (CRP) measures and measured health outcomes. A) obstructive sleep apnea (OSA), sleep duration and cognitive traits; B) odds ratio for binary cardio-metabolic and other sleep outcomes. From left to right model coefficients or odds ratio in case of binary outcomes, 95% confidence interval (95% CI), p value (p.val) and FDR corrected q value (q.val). MRS: methylation risk score; PRS: polygenic risk score

The effect of MRS-CRP on the health outcomes decreased slightly in model 2 (Figure S2). More particularly, its associations with OSA and diabetes remain statistically significant after adjusting for cardio-metabolic diseases, except for minimum SpO<sub>2</sub>, insomnia, long sleep duration and hypertension (Figure S2).

### Secondary analyses.

**Association between sleep and related comorbidities.** AHI and minimum SpO<sub>2</sub> were associated with diabetes, cognitive function and hypertension, and average SpO<sub>2</sub> was associated with the two cardio-metabolic diseases (Figure S3A). Notably, cognitive function score at baseline was positively associated with OSA severity, yet inversely associated with insomnia and long sleep duration (Figure S3A). Adjusting for MRS-CRP reduced the effect size of the associations with the two cardio-metabolic disorders (Figure S3B), suggesting that MRS-CRP captures the inflammatory component shared by sleep phenotypes and these comorbidities. Interestingly, the strongest change was observed for diabetes, whereas adjusting for MRS-CRP had a much smaller impact on associations with hypertension and cognitive scores (Figure S3). Adjusting for PRS-CRP essentially had no effect on the associations of MRS-CRP with OSA phenotypes, long sleep and metabolic comorbidities (Figure S4).

**Cubic spline model for sleep duration.** Our cubic spline model results supported the previously reported U-shaped (or J-shaped) association between sleep duration and CRP markers<sup>52,53</sup>, where optimal sleep time (around 7 hours) corresponds to lowest level of both MRS- and blood CRP level (Figure S5). The more robust relationship between long sleep and CRP markers (Figure S5) agrees with a previous study supporting the J-shaped association<sup>52</sup>. The protective effect of short sleep on blood CRP persisted in analyses that utilized data from the larger sample, while the rate of increase in blood CRP level associated with longer sleep becomes less pronounced (Figure S5).

**Association analysis using all HCHS/SOL participants.** Interestingly, associations of blood CRP level with AHI (0.41 [0.16, 0.66]) and cognitive function at baseline (-0.04 [-0.06, -0.01]) become statistically significant after increasing the sample size to all HCHS/SOL participants, yet its association with EDS become statistically insignificant (Figure S6). When it comes to PRS-CRPs, their association p-values remain unchanged across these health outcomes (Figure S6). Table S6 provides the results of association of PRS-CRPs with OSA risk, with and without EDS, in comparison with no OSA. In contrast to the previous study<sup>8</sup>, our computed PRS-CRP was negatively associated with OSA with EDS (Table S6).

**Lasso penalized regression.** Out of the list of the 1,041 CpG sites used in the construction of the MRS-CRP score, lasso penalized regression identified 66 and 12 CpG sites as pertinent to AHI and minimum SpO<sub>2</sub>, respectively, yet 0 CpG sites for average SpO<sub>2</sub> and insomnia (Table S7-S8). In terms of diabetes and hypertension, 55 and 6 CRP associated CpG sites were selected by lasso penalized regression (results not shown). Among them, 6 CpG sites are shared between OSA related measures, and 5 shared across OSA and one of the two metabolic disorders (Table 2).

Table 2. CpGs selected by Lasso regression across obstructive sleep apnea and metabolic comorbidities

CpG	Chromosome	Gene Name	Phenotypes
cg00572560	10		AHI, Diabetes
cg00574958	11	CPT1A	AHI, Min SpO <sub>2</sub> , Diabetes
cg03246954	19	MKMK2	AHI, Min SpO <sub>2</sub>
cg06690548	4	SLC7A11	AHI, Hypertension
cg14476101	1	PHGDH	AHI, Min SpO <sub>2</sub> , Hypertension
cg14656297	9	FXN	AHI, Min SpO <sub>2</sub>
cg19693031	1	TXNIP	AHI, Diabetes, Hypertension
cg23281327	10		AHI, Min SpO <sub>2</sub>
cg23440058	3	KALRN	AHI, Min SpO <sub>2</sub>

**Sensitivity analysis.** Adjusting for cell-type proportions, smoking status and waist-hip ratio yielded comparable results in association significance level to that of primary analysis for MRS-CRP (Figure S7). Surprisingly, using waist-hip ratio as parameter for obesity instead of BMI augmented considerably the association between MRS-CRP and these health outcomes except for diabetes (Figure S7).

## Discussion

MRS-CRP stands out as the only CRP measure associated with insomnia and OSA severity markers including AHI, minimum and average SpO<sub>2</sub>. Blood CRP level, on the other hand, is only associated with AHI after increasing the sample pool to all HCHS/SOL participants and not associated with average SpO<sub>2</sub> in either participants with OSA<sup>54</sup> or the general Hispanic population. That said, blood CRP is likely a less specific and accurate marker for OSA compared to MRS-CRP. Our results supported the superior performance of MRS-CRP to circulating CRP in measuring chronic inflammation and comorbidities<sup>18-21</sup>.

Several factors could impact the performance of circulating CRP. First of all, fluctuations in circulating CRP levels are highly susceptible to circadian misalignment, acute medical condition and genetic predisposition<sup>55,56</sup>. Furthermore, CRP molecules exist in two structural forms,

pentameric and monomeric<sup>57</sup>. The former is synthesized in the liver and regulated by proinflammatory cytokines. While the latter is dissociated from the pentameric CRP at local inflammation sites such as atherosclerotic plaque, further escalating inflammatory activities<sup>58</sup>. However, current high sensitivity assays measure only the pentameric form<sup>57</sup>, thus limiting the representation of overall inflammatory status. In contrast, MRS-CRP, derived from a mixture of cellular populations in the blood, reflects epigenetic changes resulting from cellular responses to inflammation. This makes MRS-CRP a promising marker for stratifying population risk across numerous outcomes, including chronic inflammation, sleep and metabolic health, thereby enhancing accuracy and clinical inference.

CRP has emerged as a useful clinical marker for OSA and cardiovascular comorbidities, due to its independent association with apnea severity and hypertension<sup>59</sup>. Given that OSA is recognized as an antecedent risk factor for secondary hypertension<sup>60</sup>, the modest change in association between OSA traits and hypertension after adjusting for MRS-CRP suggests a relatively small contribution of inflammation to the elevated blood pressure. Nevertheless, the association between sleep outcomes and diabetes were greatly attenuated and even no longer statistically significant after adjusting for MRS-CRP, which is in alignment with the mitigating effect of CRP over the association of OSA severity and incident diabetes<sup>61</sup>. These findings indicate that systemic inflammation may function as a mediator or a common factor underlying the relationship between sleep phenotypes and diabetes.

Both MRS- and blood CRP markers manifest a J-shaped association with sleep duration. Previous research has linked diabetes, hypertension, depression and low-grade inflammation<sup>62</sup> to prolonged sleep time, which is generally considered as an indicator of poor health condition<sup>63</sup>. Congruent with our findings, the relationship between short sleep duration and CRP levels is no longer statistically significant after adjusting for confounding variables<sup>52,53</sup>, probably due to the strong associations of sex, race/ethnicity, and age with both extreme sleep durations and blood CRP level<sup>52</sup>. Interestingly, the slightly protective effect of short sleep on CRP levels in Hispanic (Figure S5) and Asian population<sup>52</sup> warrants future studies to replicate these results using larger sample sizes among different races/ethnic groups, and to investigate how inflammation interplays with sleep duration and cardiovascular comorbidities.

The non-significant associations observed for PRS-CRP with sleep and health outcomes indicate a weak relationship between genetic component of CRP with related comorbidities. Circulating CRP level is thus likely the consequence of these health conditions<sup>9,64</sup>, not the other way around. One possible reason for the inconsistent association results of PRS-CRP with OSA cases with EDS from the previous paper<sup>8</sup> is the distinct assessment criteria for OSA and EDS employed by the two studies, aside from intrinsic differences across sampled populations. In the study conducted by Huang and colleagues<sup>8</sup>, OSA cases were based on clinical diagnosis and their EDS assessment showed only a moderate correlation with ESS. Nevertheless, PRS-CRP showed a significant association with blood pressure, albeit not with diabetes, in European population<sup>65</sup>. These



conflicting results emphasize the need for more research to develop PRS-CRPs and estimate their associations across populations of different ancestries and extra caution in its application for clinical assessment and inference.

Given that sleep questionnaires, OSA assessments, blood sampling for CRP DNAm assays were all conducted during the baseline visit, we were limited to performing association analysis. Another limitation of current study stems from the potential bias in self-reported sleep measurements, which often fall short of capturing context-specific sleep patterns influenced by environmental and lifestyle factors. Overall, it remains to be determined whether the inflammation-associated DNAm causes or results from sleep and metabolic disorders, or if a third scenario exists where both are affected by a shared underlying causal factor. In light of the established associations, the underlying epigenetic mechanisms behind the observed inflammatory processes and how they relate to sleep and its co-morbidities still need to be elucidated.

To address this question, we applied lasso regression to select key CRP-associated CpGs related to health outcomes. Consequently, 6 CpGs were found to be shared between OSA traits and metabolic disorders. Specifically, cg19693031 methylation in the thioredoxin interacting protein (TXNIP) gene is inversely associated with HbA1c<sup>66</sup>, fasting blood glucose<sup>67</sup>, triglycerides levels<sup>66</sup> and blood pressure<sup>68</sup>. Various studies confirmed its strong association with incident and prevalent type 2 diabetes<sup>69-71</sup>. Hypomethylation of cg19693031 enhances TXNIP expression, which in turn promotes inflammation and activates monocytes towards an inflammatory state<sup>71,72</sup>. The negative coefficient of this locus obtained by lasso regression supports its protective role against the disease onset (Table S7-S9). This suggests a possible connection between OSA and diabetes through hyperglycemia-induced inflammation<sup>73</sup>.

Hypermethylation at cg00574958 of carnitine palmitoyltransferase-1A (CPT1A) also showed a protective association with AHI, minimum SpO<sub>2</sub>, and diabetes (Table S7-S9), consistent with previous studies of diabetes<sup>74,75</sup>. CPT1A mediates energy homeostasis by regulating the rate of mitochondrial fatty acid oxidation<sup>76</sup>. A multi-cohort study reported associations between carbohydrate and fat intake with changes at CPT1A-cg00574958 locus, with implication in mediating risks of metabolic traits including BMI, diabetes, hypertension, as well as triglyceride and glucose levels<sup>74</sup>.

Solute carrier family 7 member 11 (SLC7A11)-cg06690548 and phosphoglycerate dehydrogenase (PHGDH)-cg14476101 have been frequently reported together as showing protective association against blood pressure<sup>77,78</sup>. Such protective effect was also observed in HCHS/SOL, with hypermethylation at both sites associated with improved AHI and SpO<sub>2</sub> levels, as well as lower hypertension risk (Table S7-S9). PHGDH-cg14476101 is associated with triglyceride level<sup>79</sup> and urinary albumin-to-creatinine ratio<sup>78</sup> by regulating serine biosynthesis and lipid homeostasis<sup>80</sup>. DNAm change in cg06690548 affects the expression of SLC7A11 to modulate metabolic reprogramming and the defense mechanism against oxidative stress<sup>77,81</sup>.

Overall, these lasso-selected CpG sites reiterate the important role of DNAm in the pathogenesis underlying complex diseases, and may serve as potential biomarkers for inflammatory stress associated with metabolic syndrome and OSA traits, bridged by fat, glucose and energy homeostasis. The large effect sizes of some CpGs for AHI and minimum SpO<sub>2</sub>, such as MKNK2-cg03246954, PFDN2-cg00816397, ATF2-cg02622866 and ARSA-cg07453718, indicate strong associations with OSA, and potential applicability as disease biomarkers (Table S7-S8). Future research could benefit from confirming causality of these associations as well as their directionality, and expanding to populations with other ancestries to test efficacy of the epigenetic predictor of CRP.

### **Conclusion**

A CRP measure based on DNA methylation status at select CpG sites outperforms the genetic and blood measure of circulating CRP in association with several OSA-related sleep traits and metabolic comorbidities commonly linked to chronic inflammation. However, associations with other sleep and cognitive measures were mixed. Our results also suggest a probable shared mechanism between sleep phenotypes and metabolic syndrome, related to systemic inflammation and modeled by MRS-CRP, particularly diabetes. Consequently, the incorporation of MRS-CRP, versus circulating CRP, in clinical assessment has the potential to improve population risk stratification for adverse health outcomes

### **Data Availability**

HCHS/SOL data are available through application to the data base of genotypes and phenotypes (dbGaP) accession phs000810, or via data use agreement with the HCHS/SOL Data Coordinating Center (DCC) at the University of North Carolina at Chapel Hill, see collaborators website: <https://sites.csc.unc.edu/hchs/>.

TOPMed-MESA freeze 10 WGS data are available by application to dbGaP according to the study specific accession: MESA: “phs001211”. MESA phenotype data are available from dbGaP according to the MESA accession: MESA: “phs000209”.

Summary statistics from GWAS of CRP were downloaded from the PAN-UKBB project (<https://pan.ukbb.broadinstitute.org/>; phenocode 30710) and BioBank Japan PheWeb at <https://pheweb.jp/pheno/CRP>.

PRS-CRP alleles and weights will be provided at the Github repository: [https://github.com/ZWangTen/CRP\\_markers\\_association](https://github.com/ZWangTen/CRP_markers_association).

### **Acknowledgement**

The authors thank the staff and participants of MESA and HCHS/SOL for their important contributions.

HCHS/SOL investigators website - <http://www.csc.unc.edu/hchs/>. The work was supported by National Heart Lung and Blood Institute (NHLBI) grants R01HL161012 to T.S., National Institute on Aging (NIA) grants R01AG080598 to T.S., RF1AG061022 to MF and HMG, and R01AG048642, R56AG048642 and R01AG075758 to HMG. The Hispanic Community Health Study/Study of Latinos is a collaborative study supported by contracts from the National Heart, Lung, and Blood Institute (NHLBI) to the University of North Carolina (HHSN268201300001I / N01-HC-65233), University of Miami (HHSN268201300004I / N01-HC-65234), Albert Einstein College of Medicine (HHSN268201300002I / N01-HC-65235), University of Illinois at Chicago (HHSN268201300003I / N01-HC-65236 Northwestern Univ), and San Diego State University (HHSN268201300005I / N01-HC-65237). The following Institutes/Centers/Offices have contributed to the HCHS/SOL through a transfer of funds to the NHLBI: National Institute on Minority Health and Health Disparities, National Institute on Deafness and Other Communication Disorders, National Institute of Dental and Craniofacial Research, National Institute of Diabetes and Digestive and Kidney Diseases, National Institute of Neurological Disorders and Stroke, NIH Institution-Office of Dietary Supplements.

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. MESA Family is conducted and supported by the National Heart, Lung, and Blood Institute (NHLBI) in collaboration with MESA investigators. Support is provided by grants and contracts R01HL071051, R01HL071205, R01HL071250, R01HL071251, R01HL071258, R01HL071259, and by the National Center for Research Resources, Grant UL1RR033176. 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. Molecular data for the Trans-Omics in Precision Medicine (TOPMed) program was supported by the National Heart, Lung and Blood Institute (NHLBI). Genome sequencing for “NHLBI TOPMed: Multi-Ethnic Study of Atherosclerosis” (phs001416.v2.p1) was performed at Broad Institute Genomics Platform (HHSN268201500014C, 3U54HG003067-13S1. Core support including centralized genomic read mapping and genotype calling, along with variant quality metrics and filtering were provided by the TOPMed Informatics Research Center (3R01HL-117626-02S1; contract HHSN268201800002I). Core support including phenotype harmonization, data management, sample-identity QC, and general program coordination were provided by the TOPMed Data Coordinating Center (R01HL-

120393; U01HL-120393; contract HHSN268201800001I). We gratefully acknowledge the studies and participants who provided biological samples and data for TOPMed.

### **Ethics Statement**

HCHS/SOL: This study was approved by the institutional review boards (IRBs) at each field center, where all participants gave written informed consent, and by the Non-Biomedical IRB at the University of North Carolina at Chapel Hill, to the HCHS/SOL Data Coordinating Center. All IRBs approving the study are: Non-Biomedical IRB at the University of North Carolina at Chapel Hill, Chapel Hill, NC; Einstein IRB at the Albert Einstein College of Medicine of Yeshiva University, Bronx, NY; IRB at Office for the Protection of Research Subjects (OPRS), University of Illinois at Chicago, Chicago, IL; Human Subject Research Office, University of Miami, Miami, FL; Institutional Review Board of San Diego State University, San Diego, CA.

MESA: All MESA participants provided written informed consent, and the study was approved by the Institutional Review Boards at The Lundquist Institute (formerly Los Angeles BioMedical Research Institute) at Harbor-UCLA Medical Center, University of Washington, Wake Forest School of Medicine, Northwestern University, University of Minnesota, Columbia University, and Johns Hopkins University.

This work was approved by the Beth Israel Deaconess Medical Center Committee on Clinical Investigators (protocol #2023P000538).

## Reference

1. Jackson, C. L., Redline, S. & Emmons, K. M. Sleep as a potential fundamental contributor to disparities in cardiovascular health. *Annual Review of Public Health* vol. 36 417–440 Preprint at <https://doi.org/10.1146/annurev-publhealth-031914-122838> (2015).
2. Pataka, A. & Riha, R. L. The obstructive sleep apnoea/hypopnoea syndrome - An overview. *Respir Med CME* **2**, 111–117 (2009).
3. Maeder, M., Schoch, O. & Rickli, H. A clinical approach to obstructive sleep apnea as a risk factor for cardiovascular disease. *Vasc Health Risk Manag* **85** (2016) doi:10.2147/VHRM.S74703.
4. Colla-Machado, P. E. *et al.* [Prevalence of silent cerebrovascular lesions in patients with obstructive sleep apnea syndrome]. *Rev Neurol* **62**, 113–7 (2016).
5. Meier-Ewert, H. K. *et al.* Effect of sleep loss on C-Reactive protein, an inflammatory marker of cardiovascular risk. *J Am Coll Cardiol* **43**, 678–683 (2004).
6. Grandner, M. A. *et al.* Extreme Sleep Durations and Increased C-Reactive Protein: Effects of Sex and Ethnoracial Group. *Sleep* **36**, 769–779 (2013).
7. Ayas, N. T. *et al.* A Prospective Study of Sleep Duration and Coronary Heart Disease in Women. *Arch Intern Med* **163**, 205 (2003).
8. Huang, T. *et al.* Genetic Predisposition to Elevated C-Reactive Protein and Risk of Obstructive Sleep Apnea. *Am J Respir Crit Care Med* **209**, 329–331 (2024).
9. Mirhafez, S. R. *et al.* Serum high-sensitivity C-reactive protein as a biomarker in patients with metabolic syndrome: evidence-based study with 7284 subjects. *Eur J Clin Nutr* **70**, 1298–1304 (2016).
10. Onat, A., Can, G. & Hergenc, G. Serum C-reactive protein is an independent risk factor predicting cardiometabolic risk. *Metabolism* **57**, 207–214 (2008).
11. Megawati, G. *et al.* Comparison of hs-CRP in Adult Obesity and Central Obesity in Indonesia Based on Omega-3 Fatty Acids Intake: Indonesian Family Life Survey 5 (IFLS 5) Study. *Int J Environ Res Public Health* **20**, 6734 (2023).
12. Kheirandish-Gozal, L. & Gozal, D. Obstructive Sleep Apnea and Inflammation: Proof of Concept Based on Two Illustrative Cytokines. *Int J Mol Sci* **20**, 459 (2019).
13. Imani, M. M. *et al.* Evaluation of blood levels of C-reactive protein marker in obstructive sleep apnea: A systematic review, meta-analysis and meta-regression. *Life* vol. 11 Preprint at <https://doi.org/10.3390/life11040362> (2021).
14. Li, K., Wei, P., Qin, Y. & Wei, Y. Is C-reactive protein a marker of obstructive sleep apnea? *Medicine (United States)* vol. 96 Preprint at <https://doi.org/10.1097/MD.0000000000006850> (2017).
15. Yong, W.-S., Hsu, F.-M. & Chen, P.-Y. Profiling genome-wide DNA methylation. *Epigenetics Chromatin* **9**, 26 (2016).
16. Jaenisch, R. & Bird, A. Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. *Nat Genet* **33**, 245–254 (2003).
17. Byun, H.-M. *et al.* Temporal Stability of Epigenetic Markers: Sequence Characteristics and Predictors of Short-Term DNA Methylation Variations. *PLoS One* **7**, e39220 (2012).
18. Wielscher, M. *et al.* DNA methylation signature of chronic low-grade inflammation and its role in cardio-respiratory diseases. *Nat Commun* **13**, (2022).

19. Stevenson, A. J. *et al.* Characterisation of an inflammation-related epigenetic score and its association with cognitive ability. *Clin Epigenetics* **12**, (2020).
20. Verschoor, C. P., Vlasschaert, C., Rauh, M. J. & Paré, G. A DNA methylation based measure outperforms circulating CRP as a marker of chronic inflammation and partly reflects the monocytic response to long-term inflammatory exposure: A Canadian Longitudinal Study on Aging analysis. *Aging Cell* **22**, (2023).
21. Hillary, R. F. *et al.* Blood-based epigenome-wide analyses of chronic low-grade inflammation across diverse population cohorts. *Cell Genomics* **4**, 100544 (2024).
22. Ligthart, S. *et al.* Genome Analyses of >200,000 Individuals Identify 58 Loci for Chronic Inflammation and Highlight Pathways that Link Inflammation and Complex Disorders. *The American Journal of Human Genetics* **103**, 691–706 (2018).
23. Bild, D. E. *et al.* Multi-Ethnic Study of Atherosclerosis: Objectives and design. *Am J Epidemiol* **156**, 871–881 (2002).
24. Pearson, T. A. *et al.* Markers of Inflammation and Cardiovascular Disease. *Circulation* **107**, 499–511 (2003).
25. Taliun, D. *et al.* Sequencing of 53,831 diverse genomes from the NHLBI TOPMed Program. *Nature* **590**, 290–299 (2021).
26. Conomos, M. P., Reiner, A. P., Weir, B. S. & Thornton, T. A. Model-free Estimation of Recent Genetic Relatedness. *The American Journal of Human Genetics* **98**, 127–148 (2016).
27. Bycroft, C. *et al.* The UK Biobank resource with deep phenotyping and genomic data. *Nature* **562**, 203–209 (2018).
28. Nagai, A. *et al.* Overview of the BioBank Japan Project: Study design and profile. *J Epidemiol* **27**, S2–S8 (2017).
29. Mägi, R. & Morris, A. P. GWAMA: software for genome-wide association meta-analysis. *BMC Bioinformatics* **11**, 288 (2010).
30. Ruan, Y. *et al.* Improving polygenic prediction in ancestrally diverse populations. *Nat Genet* **54**, 573–580 (2022).
31. Euesden, J., Lewis, C. M. & O’Reilly, P. F. PRSice: Polygenic Risk Score software. *Bioinformatics* **31**, 1466–1468 (2015).
32. Chang, C. C. *et al.* Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* **4**, 7 (2015).
33. Sofer, T. *et al.* A polygenic risk score for Alzheimer’s disease constructed using APOE-region variants has stronger association than APOE alleles with mild cognitive impairment in Hispanic/Latino adults in the U.S. *Alzheimers Res Ther* **15**, 146 (2023).
34. Sorlie, P. D. *et al.* Design and Implementation of the Hispanic Community Health Study/Study of Latinos. *Ann Epidemiol* **20**, 629–641 (2010).
35. Redline, S. *et al.* Sleep-disordered Breathing in Hispanic/Latino Individuals of Diverse Backgrounds. The Hispanic Community Health Study/Study of Latinos. *Am J Respir Crit Care Med* **189**, 335–344 (2014).
36. Cespedes, E. M. *et al.* Joint associations of insomnia and sleep duration with prevalent diabetes: The Hispanic Community Health Study/Study of Latinos (HCHS/SOL). *J Diabetes* **8**, 387–397 (2016).

37. Johns, M. W. A New Method for Measuring Daytime Sleepiness: The Epworth Sleepiness Scale. *Sleep* **14**, 540–545 (1991).
38. Westbrook, P. R. *et al.* Description and Validation of the Apnea Risk Evaluation System. *Chest* **128**, 2166–2175 (2005).
39. Wali, S. *et al.* The correlation between oxygen saturation indices and the standard obstructive sleep apnea severity. *Ann Thorac Med* **15**, 70 (2020).
40. McCurley, J. L. *et al.* Chronic stress, inflammation, and glucose regulation in U.S. Hispanics from the HCHS/SOL Sociocultural Ancillary Study. *Psychophysiology* **52**, 1071–1079 (2015).
41. Vásquez, P. M. *et al.* Physical activity, stress, and cardiovascular disease risk: HCHS/SOL Sociocultural Ancillary Study. *Prev Med Rep* **20**, 101190 (2020).
42. Zhang, Y. *et al.* Genetic associations between sleep traits and cognitive ageing outcomes in the Hispanic Community Health Study/Study of Latinos. *EBioMedicine* **87**, 104393 (2023).
43. Conomos, M. P. *et al.* Genetic Diversity and Association Studies in US Hispanic/Latino Populations: Applications in the Hispanic Community Health Study/Study of Latinos. *The American Journal of Human Genetics* **98**, 165–184 (2016).
44. Ding, W., Kaur, D., Horvath, S. & Zhou, W. Comparative epigenome analysis using Infinium DNA methylation BeadChips. *Brief Bioinform* **24**, (2023).
45. Triche, T. J., Weisenberger, D. J., Van Den Berg, D., Laird, P. W. & Siegmund, K. D. Low-level processing of Illumina Infinium DNA Methylation BeadArrays. *Nucleic Acids Res* **41**, e90–e90 (2013).
46. Pidsley, R. *et al.* A data-driven approach to preprocessing Illumina 450K methylation array data. *BMC Genomics* **14**, 293 (2013).
47. 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. *Bioinformatics* **28**, 882–883 (2012).
48. Wang, Z., Cassidy, M., Wallace, D. A. & Sofer, T. MethParquet: an R package for rapid and efficient DNA methylation association analysis adopting Apache Parquet. *Bioinformatics* **40**, (2024).
49. R Core Team. R: A Language and Environment for Statistical Computing. <https://www.R-project.org> (2023).
50. Lumley, T. *Complex Surveys: A Guide to Analysis Using R: A Guide to Analysis Using R.* (John Wiley and Sons, 2010).
51. Wood, S. N. Stable and Efficient Multiple Smoothing Parameter Estimation for Generalized Additive Models. *J Am Stat Assoc* **99**, 673–686 (2004).
52. Grandner, M. A. *et al.* Extreme Sleep Durations and Increased C-Reactive Protein: Effects of Sex and Ethnoracial Group. *Sleep* **36**, 769–779 (2013).
53. Gupta, K. *et al.* Sleep duration, baseline cardiovascular risk, inflammation and incident cardiovascular mortality in ambulatory U.S. Adults: National health and nutrition examination survey. *Am J Prev Cardiol* **8**, 100246 (2021).
54. Arnardottir, E. S. *et al.* The interaction of obstructive sleep apnea and obesity on the inflammatory markers C-reactive protein and interleukin-6: The Icelandic Sleep Apnea Cohort. *Sleep* **35**, 921–932 (2012).

55. Nathan, P., Gibbs, J. E., Rainger, G. E. & Chimen, M. Changes in Circadian Rhythms Dysregulate Inflammation in Ageing: Focus on Leukocyte Trafficking. *Front Immunol* **12**, (2021).
56. Freese, J., Klement, R. J. & Lötzerich, H. The inflammation paradox: Why are Tsimane protected against Western diseases while Westerners are not? *F1000Res* **7**, 252 (2018).
57. Melnikov, I. *et al.* The monomeric C-reactive protein level is associated with the increase in carotid plaque number in patients with subclinical carotid atherosclerosis. *Front Cardiovasc Med* **9**, (2022).
58. Rajab, I. M., Hart, P. C. & Potempa, L. A. How C-Reactive Protein Structural Isoforms With Distinctive Bioactivities Affect Disease Progression. *Front Immunol* **11**, (2020).
59. Gaines, J. *et al.* Gender differences in the association of sleep apnea and inflammation. *Brain Behav Immun* **47**, 211–217 (2015).
60. Dopp, J. M., Reichmuth, K. J. & Morgan, B. J. Obstructive sleep apnea and hypertension: Mechanisms, evaluation, and management. *Curr Hypertens Rep* **9**, 529–534 (2007).
61. Nagayoshi, M. *et al.* Obstructive sleep apnea and incident type 2 diabetes. *Sleep Med* **25**, 156–161 (2016).
62. Tan, X., Chapman, C. D., Cedernaes, J. & Benedict, C. Association between long sleep duration and increased risk of obesity and type 2 diabetes: A review of possible mechanisms. *Sleep Med Rev* **40**, 127–134 (2018).
63. Stamatakis, K. A. & Punjabi, N. M. Long sleep duration: A risk to health or a marker of risk? *Sleep Med Rev* **11**, 337–339 (2007).
64. Tabassum, R., Mia, A. R., Reza-Ul-Haq, K. M., Yesmin, M. & Faruqui, J. M. C-reactive Protein Level in Type-2 Diabetic Patients Attending Mymensingh Medical College Hospital, Mymensingh. *Mymensingh Med J* **26**, 56–60 (2017).
65. Prins, Bram. P. *et al.* Investigating the Causal Relationship of C-Reactive Protein with 32 Complex Somatic and Psychiatric Outcomes: A Large-Scale Cross-Consortium Mendelian Randomization Study. *PLoS Med* **13**, e1001976 (2016).
66. Miller, R. G., Mychaleckyj, J. C., Onengut-Gumuscu, S., Orchard, T. J. & Costacou, T. TXNIP DNA methylation is associated with glycemic control over 28 years in type 1 diabetes: findings from the Pittsburgh Epidemiology of Diabetes Complications (EDC) study. *BMJ Open Diabetes Res Care* **11**, (2023).
67. Tsai, H. H. *et al.* Interaction between a diabetes-related methylation site (TXNIP cg19693031) and variant (GLUT1 rs841853) on fasting blood glucose levels among non-diabetics. *J Transl Med* **20**, (2022).
68. Richard, M. A. *et al.* DNA Methylation Analysis Identifies Loci for Blood Pressure Regulation. *The American Journal of Human Genetics* **101**, 888–902 (2017).
69. Cardona, A. *et al.* Epigenome-Wide Association Study of Incident Type 2 Diabetes in a British Population: EPIC-Norfolk Study. *Diabetes* **68**, 2315–2326 (2019).
70. Chambers, J. C. *et al.* Epigenome-wide association of DNA methylation markers in peripheral blood from Indian Asians and Europeans with incident type 2 diabetes: a nested case-control study. *Lancet Diabetes Endocrinol* **3**, 526–534 (2015).
71. Domingues, A., Jolibois, J., Marquet de Rougé, P. & Nivet-Antoine, V. The Emerging Role of TXNIP in Ischemic and Cardiovascular Diseases; A Novel Marker and Therapeutic Target. *Int J Mol Sci* **22**, 1693 (2021).



72. Rong, J. *et al.* Thioredoxin-interacting protein promotes activation and inflammation of monocytes with DNA demethylation in coronary artery disease. *J Cell Mol Med* **24**, 3560–3571 (2020).
73. Tseng, H. H. L., Vong, C. T., Kwan, Y. W., Lee, S. M.-Y. & Hoi, M. P. M. TRPM2 regulates TXNIP-mediated NLRP3 inflammasome activation via interaction with p47 phox under high glucose in human monocytic cells. *Sci Rep* **6**, 35016 (2016).
74. Lai, C. Q. *et al.* Carbohydrate and fat intake associated with risk of metabolic diseases through epigenetics of CPT1A. *American Journal of Clinical Nutrition* **112**, 1200–1211 (2020).
75. Das, M. *et al.* Association of DNA methylation at CPT1A locus with metabolic syndrome in the genetics of lipid lowering drugs and diet network (GOLDN) study. *PLoS One* **11**, (2016).
76. Schlaepfer, I. R. & Joshi, M. CPT1A-mediated Fat Oxidation, Mechanisms, and Therapeutic Potential. *Endocrinology* **161**, (2020).
77. Zhang, X. *et al.* The Interplay of Epigenetic, Genetic, and Traditional Risk Factors on Blood Pressure: Findings from the Health and Retirement Study. *Genes (Basel)* **13**, (2022).
78. Kho, M. *et al.* Epigenetic loci for blood pressure are associated with hypertensive target organ damage in older African Americans from the genetic epidemiology network of Arteriopathy (GENOA) study. *BMC Med Genomics* **13**, (2020).
79. Truong, V. *et al.* Blood triglyceride levels are associated with DNA methylation at the serine metabolism gene PHGDH. *Sci Rep* **7**, (2017).
80. Kang, Y. P. *et al.* PHGDH supports liver ceramide synthesis and sustains lipid homeostasis. *Cancer Metab* **8**, 6 (2020).
81. Lewerenz, J. *et al.* The Cystine/Glutamate Antiporter System  $x_c^-$  in Health and Disease: From Molecular Mechanisms to Novel Therapeutic Opportunities. *Antioxid Redox Signal* **18**, 522–555 (2013).