# Influence of Genetic Ancestry on Gene-Environment Interactions of Polygenic Risk and Sociocultural Factors: Results from the Hispanic Community Health Study/Study of Latinos

**Jayati Sharma**<sup>1</sup>, Cristin E. McArdle<sup>2</sup>, Mariaelisa Graff<sup>3</sup>, Christina Cordero<sup>4</sup>, Martha Daviglus<sup>5</sup>, Linda C. Gallo<sup>6</sup>, Carmen R. Isasi<sup>7</sup>, Tanika N. Kelly<sup>8</sup>, Krista M. Perreira<sup>9</sup>, Gregory A. Talavera<sup>6</sup>, Jianwen Cai<sup>10</sup>, Kari E. North<sup>3</sup>, Lindsay Fernández-Rhodes<sup>2,3</sup>, Genevieve L. Wojcik<sup>1</sup>

- 1. Department of Epidemiology, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, MD, United States
- 2. Department of Biobehavioral Health, College of Health and Human Development, The Pennsylvania State University, University Park, PA, United States
- 3. Department of Epidemiology, Gillings School of Global Public Health, University of North Carolina at Chapel Hill, Chapel Hill, NC, United States
- 4. Department of Psychology, University of Miami, Miami, FL, United States
- 5. Institute for Minority Health Research, University of Illinois at Chicago, Chicago, IL, United States
- 6. Department of Psychology, San Diego State University, San Diego, CA, United States
- 7. Department of Epidemiology and Population Health, Albert Einstein College of Medicine, Bronx, NY, United States
- 8. Department of Medicine, University of Illinois at Chicago, Chicago, IL, United States
- 9. Department of Social Medicine, University of North Carolina School of Medicine, Chapel Hill, NC, United States
- 10. Department of Biostatistics, Gillings School of Global Public Health, University of North Carolina at Chapel Hill, Chapel Hill, NC, United States

It is made available under a CC-BY-NC-ND 4.0 International license .

#### 1

## 2 ABSTRACT

Background: Many present analyses of Hispanic/Latino populations in epidemiologic research aggregate all members of this ethnic group, despite immense diversity in genetic backgrounds, environment, and culture between and across Hispanic/Latino background groups. Using the Hispanic Community Health Study/Study of Latinos (HCHS/SOL), we examined the role of selfidentified background group and genetic ancestry proportions in gene-environment interactions influencing the relationship between body mass index (BMI) and a polygenic score for BMI (PGS<sub>BMI</sub>).

Methods: Weighted univariate and multivariable generalized linear models were executed to compare the effects of environmental variables identified *a priori* by McArdle et al. 2021. Both Amerindigenous (AME) ancestry proportion and background group identity were statistically modeled as confounders both through stratified and joint analyses to understand their influence on the relationship between BMI and PGS<sub>BMI</sub>, while incorporating gene-environment interactions of PCS v dist and PCS v applied and pCS v dist and PCS v dist and PCS v dist.

15 of PGS x diet and PGS x age-at-immigration.

16 **Results**: After complex survey weighting, 7,075 participants remained in the analytic sample, 17 representing individuals of six background groups: Central American, Cuban, Dominican, 18 Mexican, Puerto Rican, and South American. The distributions of key environmental and 19 sociocultural variables were heterogeneous between Hispanic/Latino background groups. 20 Associations of these variables with AME ancestry were similarly heterogeneous upon 21 stratification, indicating confounding by background group. In a predictive model for BMI 22 incorporating health, immigration, and environmental variables, PGS<sub>BM</sub> performance decreased 23 with increasing AME ancestry proportion. In this model, most statistically significant GxE 24 interactions lost significance after ancestry and background stratification, except for PGS x age-25 at-immigration interactions in some strata: Mexican background individuals born in the US 26 compared to those >=21 years old at migration ( $\beta$ =1.33, p<0.01), Dominican background 27 individuals 6-12 years old at migration compared to those >=21 years old at migration ( $\beta$ =4.38, 28 p<0.001), and Cuban background individuals 0-5 years old at migration compared to those >=21 29 years old at migration ( $\beta$ =2.20, p=0.015), where US-born includes individuals born in the US 50 30 states/DC.

31 **Conclusions**: Controlling for self-identified background group identity and genetic ancestry did 32 not eliminate statistically significant differences in interactions between AME ancestry and 33 environmental variables in certain strata of AME ancestry among some Hispanic/Latino 34 background groups in HCHS/SOL.

It is made available under a CC-BY-NC-ND 4.0 International license .

### 35 INTRODUCTION

36 Hispanic/Latino groups in the United States (US) comprise an inherently diverse ethnicity. 37 and yet are often modeled as a monolithic group in epidemiologic analyses, despite their unique sociocultural and genetic composition. This issue is especially prevalent in presentation of race-38 39 and ethnicity-stratified health and socioeconomic statistics in the US, where all members of 40 Hispanic/Latino ethnicity are aggregated into a singular category regardless of their specific 41 Hispanic/Latino background or birthplace.<sup>1,2</sup> In a systematic review and meta-analysis of cardiovascular mortality in Hispanic/Latino populations in the US that identified lower mortality 42 43 among Hispanic/Latino individuals than their non-Hispanic white counterparts, authors identified only one of 17 included studies that stratified participants by place of birth.<sup>3</sup> Recent literature 44 45 examining the role of acculturation and heterogeneous sociocultural landscapes within Hispanic/Latino backgrounds has started to address this scarcity, through the study of topics 46 47 ranging from substance-use treatment outcomes to food insecurity and obesity research.<sup>4,5</sup> These important distinctions are also reflected in the genetic diversity within the Hispanic/Latino ethnicity. 48 evidenced in various analyses of population structure and genetic ancestry proportions of 49 Hispanic/Latino populations.<sup>6,7</sup> Diverse Hispanic/Latino experiences and histories, including 50 51 complex geographic histories shaped through colonization and immigration, have and continue 52 to shape the cultures, behaviors, and health of Hispanic/Latino groups throughout the US.

53 As the second-largest ethnic or racial group in the US and as a historically marginalized 54 population, Hispanic/Latino groups are the subject of many public health studies focused on 55 health inequities.<sup>8</sup> One facet of these inequities is in the burden of chronic diseases, exacerbated by obesity, which 44.8% of adult Hispanic/Latino individuals in the US face.<sup>9</sup> While lifestyle factors 56 57 and health behaviors have been studied as predictors of obesity, the incorporation of genetics and gene-environment interactions are another promising avenue through which to understand 58 the individual-level impact of these factors.<sup>10</sup> Gene-environment (GxE) interactions, which 59 60 characterize the joint influence of genetic and environmental variables (such as health behaviors), 61 are an important area of study that may inform future precision health applications of screenings 62 and therapeutics designed to prevent and treat chronic diseases caused by pre-existing 63 conditions such as obesity. In Hispanic/Latino populations, disproportionate levels of exposure to 64 obesogenic environments via poor diet quality, low socioeconomic status, poor education, and healthcare bias may interact with obesity-associated genetic polymorphisms and may contribute 65 to group-level disparities by Hispanic/Latino background.<sup>11</sup> 66

67 Studying disease disparities and GxE interactions across populations, particularly of 68 diverse genetic ancestries, can advance our understanding of the complex relationships between genetic and lifestyle or behavioral factors that may have been otherwise unobservable in non-69 70 genetic analyses. One approach to examining genetic ancestry is by estimating admixture proportions, in which each individual's genome is apportioned based on its similarity to either a 71 72 reference population, or other genomic segments in the sample.<sup>12,13</sup> Proportions of inferred 73 genetic ancestry, or admixture proportions, have been shown to vary widely by Hispanic/Latino 74 background, typically modeled with European, African, and Native American (AME: referred to in 75 this manuscript as Amerindigenous) ancestry components showing higher African ancestry in Caribbean populations (Dominican, Puerto Rican) and higher Amerindigenous ancestry in 76 Mexican participants.<sup>14,15</sup> Reflecting differences in migration patterns, differences in ancestry 77

proportions have also been seen to persist based on geographic position in the United States
 when examined across Hispanic/Latino participants.<sup>16</sup>

80 Recently, an effective way to model the genetic liability of complex traits and diseases is 81 through polygenic scores (PGS), a relatively comprehensive metric comprising a weighted sum 82 of many genetic variants (single nucleotide polymorphisms, or SNPs) associated with a given 83 particular trait. However, there are noted challenges with the use of PGS across diverse 84 populations. Recent research has identified the decreased PGS<sub>BMI</sub> performance in non-European 85 participants, in particular those from African, South Asian, East Asian, and Hispanic/Latino backgrounds, when samples of European ancestries are used to train such PGSs.<sup>17</sup> This may be 86 87 due to differences in the underlying genetic architecture, including linkage disequilibrium patterns and allele frequencies, but could also be due to differing environmental influences.<sup>18</sup> Decreased 88 89 PGS performance has also been observed when examining non-genetic factors that differ between training and test sets, such as age and sex.<sup>19,20</sup> However, there is an open question as 90 91 to the degree to which ancestry and environmental differences jointly contribute to PGS 92 performance.

93 The Hispanic Community Health Study/Study of Latinos (HCHS/SOL) is an ideal cohort to 94 examine these relationships as participants provide background information during study 95 enrollment as well as extensive well-characterized longitudinal data. A previous analysis by 96 McArdle et al. examined the influence of PGS-by-diet and PGS-by-acculturation interactions on 97 the genetic risk of obesity via a PGS<sub>BMI</sub> trained on European ancestry samples from UK Biobank and GIANT and applied in HCHS/SOL.<sup>21</sup> The authors found dietary patterns and age at 98 99 immigration to be significant modifiers of the effect of individuals' genetic risk on obesity; the effect 100 of the PGS<sub>BMI</sub> on BMI was different based on different levels of acculturation and healthy diet. 101 Specifically, in their full model, the authors identified that a one-standard deviation (SD) increase in the PGS<sub>BMI</sub> was associated with a 1.10 kg/m<sup>2</sup> increase in BMI ( $\beta$ =1.10,p<0.001), adjusted for 102 103 various demographic, sociocultural, and environmental variables, which differed substantially 104 when stratified by sex (males:  $\beta$ =0.79, p<0.001; females:  $\beta$ =1.45, p<0.001). A separate 105 exploratory analysis stratified by self-reported Hispanic/Latino background observed significant 106 heterogeneity for PGS<sub>BMI</sub>, with weaker effects in individuals of South American background 107 ( $\beta$ =0.91, p<0.001) than of Mexican background ( $\beta$ =1.73, p<0.001). Complicating this, the 108 modifying effects of age-at-immigration and healthy diet on PGS<sub>BMI</sub> showed different direction of 109 effects across background groups, such as when comparing individuals were born in the US 50 110 states/DC to those having migrated to the US after the age of 20 years old (South American 111 background:  $\beta$ =-1.74,p<0.05; Mexican background:  $\beta$ =1.09,p<0.05).

112 It remains unclear, however, the extent to which background within Hispanic/Latino 113 ethnicity influences these relationships and what factors would influence these interactions. Given 114 this discrepancy in predictive performance, we hypothesize that there exist ancestry-driven 115 differences in the performance of this PGS<sub>BMI</sub> in Hispanic/Latino populations. In addition, it is 116 unclear how the intersection of ancestry and environmental differences may influence the 117 performance of a PGS. Building on prior work, we hypothesize that the GxE interactions between 118 a  $PGS_{BMI}$  and immigration history and diet variables varied between background groups of 119 Hispanic/Latino ethnicity as a function of both group-specific ancestry differences as well as 120 differences in environment. There is an urgent need to disentangle these complex factors to better 121 characterize the joint roles of genetics and environmental influences on human health, as well as

It is made available under a CC-BY-NC-ND 4.0 International license .

122 demonstrate the need to model Hispanic/Latino populations appropriately in genetic research. As

such, this paper broadly seeks to (1) better understand the role of inferred genetic ancestry with

BMI, both between and within Hispanic/Latino groups, and (2) expand the analysis of McArdle *et* 

125 *al.* to incorporate AME ancestry proportion and examine the influence of group heterogeneity on

- 126 interactions between the afore-mentioned sociocultural variables and genetic risk for obesity.
- 127

## 128 METHODS

**Study Design.** The HCHS/SOL study design and methodology have been described in detail elsewhere.<sup>22</sup> From four sites (Bronx, Miami, Chicago, San Diego) in the US, 16,415 self-identified Hispanic/Latino participants aged 18-74 were recruited and had physical, behavioral, sociocultural, and biometric measurements collected at a baseline examination between 2008-2011 and second clinic visit between 2014-2017. Participants were recruited to HCHS/SOL through a two-stage area household probability design and therefore some participants are related.<sup>22</sup>

136 In the present study, we examined data from an analytic subset of HCHS/SOL participants 137 who were included in the analysis conducted by McArdle et al.<sup>21</sup>: those who had consented for 138 genetic data collection (at visit 1) and analysis (at both visits), and whose information was 139 successfully linked to the PGS<sub>BMI</sub> data constructed in HCHS/SOL. This analytic subset was then 140 further restricted to participants with visit 1 data, with estimated admixture proportions for genetic 141 ancestry, and no missing covariate data (n=7,282). The inclusion/exclusion criteria and the 142 resulting sample size of this study are described in Supplementary Figure 1.

143 Variable Definitions. Immigration-related variables used in this analysis were immigrant 144 generation (first or second) and age at immigration to the US. Immigrant generation was defined 145 as "first generation" as being foreign-born with foreign-born parents, including those born in a US 146 territory. "Second generation" was defined as those who were born in the US or those who were 147 foreign-born with at least one US-born parent. Age at immigration to the US was defined among 148 the foreign/territory-born participants based on their current age and years residing in the US. 149 Following McArdle et al., five categories were defined: Born in US 50 states/DC (hereafter, US 150 born), 0-5 years old at migration, 6-12 years old at migration, 13-20 years old at migration, and 151  $\geq$ 21 years old at migration.

152 Several other health and environmental variables were also included as covariates in the 153 full model, derived from the analysis conducted by McArdle et al. These consisted of sex (binary: male or female); age; age<sup>2</sup>; self-reported Hispanic/Latino background identity and descent 154 155 (Central American, Cuban, Dominican, Mexican, Puerto Rican, and South American): marital 156 status (currently married, yes/no); educational status (less than high school vs. high school and 157 beyond); employment status (retired, not retired or employed, employed  $\leq$  35 h/week, employed 158 > 35 h/week); type 2 diabetes status (yes/no); cardiovascular disease (CVD) status, e.g. CHD or 159 Stroke, but not counting angina or transient ischemic attacks (yes/no); sleep duration (h/day); 160 sweetened beverages consumption (servings/day); whether the WHO's 2008 Global Physical Activity guidelines for Americans criteria<sup>23</sup> were met (yes/no); alcohol use level (no current use, 161 162 low-level use, and high-level use); cigarette use (never, current, and former); Center for

163 Epidemiological Studies Depression (CES-D) 10-item Summary Score<sup>24</sup>; and a single item ethnic 164 identity score (5-level ordinal variable).

Healthy diet was defined using in part the JAMA Healthy Diet Score, which used data from the baseline 24-hour dietary recall to assign individuals a value ranging from 1 to 5 based on meeting sex-specific quintiles of predicted daily intake of saturated fatty acids, fiber, calcium, and potassium. Scoring above the 60th percentile of this score was defined as meeting the criteria for having a "healthy diet," consistent with other definitions of healthy diet in prior literature.<sup>25</sup>

BMI was calculated at the baseline visit for all participants, as individuals' measured weight divided by their height, squared (kg/m<sup>2</sup>). Participants with BMI <18.5 or >70.0 were excluded from the analysis, as were those <20 years of age at the visit, consistent with quality control measures employed by other literature examining the genetic variation of BMI and obesity.<sup>26</sup>

**Genetic Data.** Genotyping, quality control, and imputation methods employed in the HCHS/SOL cohort have been described elsewhere previously.<sup>27,28</sup> Briefly, DNA was extracted from individual blood samples and genotyped on the SOL HCHS Custom 15041502 B3 array (i.e., Illumina Omni2.5M array + 150,000 custom informative SNPs).<sup>27</sup> Standard quality control filters were enacted and then imputed using 1000 Genomes Project phase 3 reference populations. Principal components (PCs) of genetic ancestry, with 1000 Genomes reference populations, were constructed to account for confounding by population stratification.

181 The PGS<sub>BMI</sub> employed in this analysis was derived from effect sizes published in a 182 European-ancestry genome-wide association study (GWAS) of BMI in almost 700,000 participants enrolled in the UK Biobank and in the GIANT consortium, and applied to genome-183 wide data from HCHS/SOL.<sup>29</sup> Using SNPs that passed a minor allele frequency cutoff of 5% in 184 185 HCHS/SOL, the PGS<sub>BMI</sub> was calculated from effect sizes in this GWAS via the LDpred method and an infinitesimal model.<sup>30</sup> Several PGSs were compared to identify the best-fitting PGS<sub>BMI</sub> 186 187 which explained 7.4% of the variance in inverse normalized BMI in the HCHS/SOL cohort, 188 adjusted for various relevant covariates. For consistency of interpretation, this PGS<sub>BMI</sub> was then 189 standardized to a mean of 0 and a standard deviation of 1.

190 Genetic ancestry estimation. Genetic ancestry was estimated using genotyped data that 191 passed the above quality control filters using a larger sample as detailed in a previous publication 192 including other studies participating in the Population Architecture using Genomics and 193 Epidemiology (PAGE) Study Genotype data were phased with SHAPEIT2 and imputed into 1000 194 Genomes phase 3 reference data using IMPUTE version 2.3.2.<sup>31</sup> Unrelated individuals were 195 subset from the data up to 2<sup>nd</sup> degree relatives (N=45,255) and genotyped sites were pruned using r<sup>2</sup><0.1 (M=132,591) in PLINK.<sup>32</sup> Ancestry proportions were estimated using ADMIXTURE 196 with 10 unsupervised runs assuming k=5.13 These five clusters were inferred to represent 197 198 European (EUR), African (AFR), Amerindigenous (AME), East Asian (EAS), and Pacific Islander 199 (PI) genetic ancestries given their distributions within self-identified racial and/or ethnic 200 categories. Across all ten runs, there was only one mode as determined by pong (https://github.com/ramachandran-lab/pong), indicating stable estimates.<sup>13,33</sup> Proportion of AME 201 202 ancestry was expressed as a continuous value assigned to all participants within this HCHS/SOL 203 sample (N=7,282).

204 **Statistical analysis.** An exploratory analysis was conducted to characterize the variables that 205 were most conceptually relevant to the model constructed herein, based on those chosen by

It is made available under a CC-BY-NC-ND 4.0 International license .

McArdle *et al.* and described above. Table 1 p-values were calculated through the Wilcoxon ranksum test for complex survey samples, which tests whether the group values are different from one another and via Pearson's Chi-squared test with Rao & Scott's second-order correction for categorical variables, which tests differences between observed and expected frequencies.

Models and observations used in summary measures were weighted using complex survey weights specified by HCHS/SOL protocols and all analyses were conducted in R version 4.1.1. Associations with proportion of AME ancestry were calculated through univariate surveyweighted generalized linear models of AME ancestry proportion on each individual explanatory variable included in Table 1, using the {gtsummary} package in R.<sup>34</sup>

To assess the contribution of AME ancestry proportion to the relationship between immigration-related variables, diet, and other measures as predictors of BMI, a generalized linear model was fit to the data comprising immigration-related and environmental variables as specified a priori by McArdle et al. The original model creation strategy used augmented backwards elimination and tests of GxE interaction that contributed to the development of a full model to predict BMI, and excluded variables such as SASH Language Scale, marital status, CESD-10 Depression Scale. The full model implemented in this analysis is shown below:

222 $BMI \sim PGS_{BMI} + 5$  Principal Components (PCs) + Center + Age + Age<sup>2</sup> + Sex + Diabetes223+ Sleep Duration + Cigarette Use + Physical Activity + CVD + Alcohol Use + Sweetened224Beverage Consumption + Immigrant Generation + Employment Status + Education +225Income + 5-category Age at Immigration + Healthy Diet Score + PGS\_{BMI} \*Healthy Diet226Score + PGS\_{BMI} \*Age at Immigration

To understand the influence of both ancestry and Hispanic/Latino background group, three modeling comparisons were employed: 1) implementing the full model with and without an AME ancestry proportion term, 2) stratifying the full model by quartile of AME ancestry, and 3) stratifying the full model by both group and by tertile of AME ancestry, since quartile stratification yielded too few observations in each stratification bin.

All models and observations used in summary measures were weighted using complex survey weights specified by HCHS/SOL protocols and all analyses were conducted in R version 4.1.1. Survey-weighted models were executed through the {survey} package in R.

235

## 236 **RESULTS**

## 237 Observed heterogeneity in key variables between Hispanic/Latino background groups

238 A total of 7,282 participants were included in the full analysis. After weighting to 239 account for HCHS/SOL's complex survey design and restricting to those with complete data for 240 all included covariates, 7,075 weighted observations remained in the sample (representing 7,282 241 individuals). Detailed participant characteristics are described in Table 1. Within the weighted 242 sample, individuals comprised six background groups: Central American (N=602), Cuban 243 (N=1,717), Dominican (N=636), Mexican (N=2,560), Puerto Rican (N=1,157), and South 244 American (N=403). We identified significant heterogeneity in population characteristics by 245 background group, including BMI, immigrant generation, prevalent cardiovascular disease (CVD), 246 and the PGS<sub>BMI</sub>. (Table 1) This difference is especially apparent between Caribbean and 247 Central/South American geographical representations of Hispanic/Latino groups. Since some of

It is made available under a CC-BY-NC-ND 4.0 International license .

the largest sample sizes are found in Mexican (N=2,560) and Puerto Rican (N=1,157) groups,
and these two groups adequately demonstrate differences in geography by recruitment site and
immigration-related histories, we will use these strata to illustrate these trends.

251 Overall, the average BMI across all Hispanic/Latino participants was 28.9 (IQR: 25.9, 252 32.7). However, the distributions of BMI were significantly different between Hispanic/Latino 253 groups (p<0.001), with the mean BMI was slightly lower at 28.7 (IQR: 26.1, 32.3) in Mexican 254 background individuals, while slightly higher in Puerto Rican background individuals at 29.6 (IQR: 255 26.1, 34.3). Significant heterogeneity was also observed when comparing genetic risk for BMI as 256 estimated by the standardized PGS<sub>BMI</sub> (p<0.001; Table 1). The weighted mean PGS<sub>BMI</sub> in the 257 pooled sample was 0.01 (IQR: -0.065, 0.067). However, the mean PGS<sub>BMI</sub> was higher in the Mexican background group at 0.07 (IQR: -0.61, 0.72), and lower in the Puerto Rican background 258 259 group at -0.03 (IQR: -0.69, 0.64), an inverse trend to BMI which showed lower values in Mexican background individuals and higher values in Puerto Rican background individuals. 260

261 Considering the main variables thought to interact with PGS<sub>BMI</sub> and influence prediction of 262 BMI (Figure 1), we note differences in proxy measures for acculturation used in this analysis. Overall, 80.8% of Hispanic/Latino individuals in this study identified as 1<sup>st</sup> generation. However, 263 when stratified, 79% of Mexican background participants identified as 1<sup>st</sup> generation, compared 264 265 to only 53% of Puerto Rican background individuals (Table 1). Additionally, there was significant heterogeneity in age-at-immigration between groups, with 50% of Mexican background 266 267 participants immigrating after the age of 21 years old yet this proportion was only 21% of Puerto 268 Rican background participants. Diet is another culture- and region-specific variable that 269 demonstrates sociocultural differences. Only 29% of Mexican background individuals were below 270 the 60th sex-specific percentile of their JAMA Healthy Diet score (less healthy diet), compared to 271 80% of Puerto Rican background individuals. This difference is not adequately depicted in the 272 overall distribution of diet scores, where 52% of members of Hispanic/Latino ethnicity score below 273 their sex-specific 60th percentile. Overall, this demonstrates marked heterogeneity in these key 274 variables between Hispanic/Latino backgrounds which would be typically modeled as a single 275 homogenous groups in genetic studies.

Figure 1. Conceptual Framework of Full Analytic Model



276

# 277 <u>Relationships between inferred ancestry components with risk factors and traits differ by</u> 278 <u>Hispanic/Latino background groups</u>

279 In line with our conceptual framework (Figure 1), genetic ancestry, especially inferred AME 280 ancestry in Hispanic/Latino ethnic groups, can capture finer genetic background composition in 281 addition to self-identified background group that may confound the association of interest between PGS<sub>BMI</sub> and BMI.<sup>38</sup> We estimate differences in AME ancestry proportion distributions by 282 283 Hispanic/Latino background group, and their associations with other risk factors (Table 2, Figure 284 2). While the pooled sample has on average 29% AME ancestry with noted variance (IQR: 9%, 285 48%), when stratified by background group reflecting geography and historical patterns of 286 migration and colonialism, stark differences arise. Central American, Mexican, and South 287 American background groups have higher mean AME ancestry of 45%, 44%, and 47%, 288 respectively, while Cuban, Dominican, and Puerto Rican background groups, have lower mean AME ancestry proportions of 4%, 6%, and 13%, respectively (Supp Fig 1). 289

290 To assess the relationship between these key variables and global ancestry proportions 291 both in aggregate and between Hispanic/Latino groups, we conducted a univariate analysis 292 between each variable and estimated proportion of AME ancestry. We observed similar 293 heterogeneity to what was described above, largely driven by substantial AME ancestry 294 distribution differences between background groups (Table 2). Specifically, we observed the 295 association of AME ancestry to be significantly negatively associated with BMI among all 296 participants ( $\beta$ =-0.001, p=0.044), but that effect is null when examined stratified by group (Figure 297 2A). When taken in conjunction with group-level differences in BMI as detailed above, this lack of 298 statistically significant relationship after stratification indicates possible confounding of the 299 relationship between ancestry and BMI by group membership.

300 Associations between the PGS<sub>BMI</sub> and AME ancestry are similarly heterogeneous. Overall, 301 there is a statistically significant positive association ( $\beta$ =0.030, p<0.001) among HCHS/SOL 302 Hispanic/Latino participants. However, this association differs by background group with a larger 303 effect size among Mexican background individuals ( $\beta$ =0.040, p<0.001), but attenuated signals in 304 Puerto Rican background individuals ( $\beta$ =-0.001, p=0.39) (Figure 2B). In contrast to BMI, this 305 indicates that the association between ancestry and PGS distribution is not spurious, but stronger 306 in specific subgroups with higher AME ancestry proportions.

307 We observe significant heterogeneity in the association of AME ancestry proportions with 308 other relevant variables, including prevalent cardiovascular disease (CVD) (Figure 2). The 309 relationship between AME ancestry and CVD across the entire cohort appears statistically 310 significant ( $\beta$ =-0.054 (95% CI: -0.076,-0.031), p<0.001). Importantly, when stratified by 311 background group, this statistical significance disappears completely in all groups (p>0.05). This 312 is consistent with our observation that certain groups have higher prevalence of CVD than others 313 (e.g., Puerto Rican individuals at 7.9% vs. Mexican individuals at 4.8%), and that AME proportions 314 vary widely by background group as well, indicating confounding by background group. Similar 315 inferences are made regarding the Employment Status, Physical Activity, Marital Status, and 316 CESD-10 Item Depression Score variables (Table 2). Taken together, these results caution the 317 use of a pooled Hispanic/Latino participant sample to characterize the role of ancestry proportions 318 on human health or relevant trait distributions.

319



# **Figure 2.** Distributions of Selected Variables by AME Ancestry Proportion Aggregated and by Background



differences (p<0.05) between categories.

Prevalent CVD

Prevalent CVD

Given high heterogeneity in the univariate analyses, we sought to understand the influence of this heterogeneity in a model incorporating the above variables into a genetic risk prediction model for BMI. Specifically, to further investigate if differential performance of the PGS<sub>BMI</sub> is reflective of potential population stratification not captured by principal components accounting for global genetic ancestry, we compared the performance of the extended model for BMI previously employed by McArdle *et al.* with and without an independent AME ancestry

It is made available under a CC-BY-NC-ND 4.0 International license .

proportion term. In the pooled sample of all Hispanic/Latino participants, the addition of AME ancestry proportion to the full model did not meaningfully change the main effect of PGS<sub>BMI</sub> on BMI ( $\beta$ =2.47, p<0.001 vs.  $\beta$ =2.46, p<0.001) or model fit with an R<sup>2</sup> of 0.170 in the original full model from McArdle et al. and an R<sup>2</sup> of 0.171 with the addition of AME ancestry proportions (Supp Table 1). Coefficients for other terms were consistent between models, including interaction terms with PGS<sub>BMI</sub> with healthy diet and with age at immigration (Supp Table 1).

335 We then explored how proportion of AME ancestry may affect both the individual 336 coefficients and overall model fit by stratifying the full model by inferred AME ancestry into 337 quartiles in the pooled sample (Q1: 0-8.9%, Q2: 8.9-28.9%, Q3: 28.9-48.2%, and Q4: 48.2-99.9% 338 AME ancestry). In contrast to the previous analysis which seeks to control for confounding, this 339 analysis assesses the role of overall genetic background (AME ancestry proportion) on the ability 340 for the PGS to capture the genetic liability for BMI. Overall, model fit (R<sup>2</sup>) generally decreased 341 with increasing quartiles of AME ancestry (0.222, 0.202, 0.213, and 0.164, respectively; Table 3). 342 Due to the collinearity of increasing inferred AME with decreasing EUR, this finding is consistent 343 with prior literature showing generally decreased performance with increasing genetic distance from the training sample, which was of European ancestries.<sup>40</sup> The adjusted effect size of the 344 PGS<sub>BMI</sub> on BMI remained generally consistent with increasing guartile of AME ancestry (guartile 345 346 1:  $\beta$ =2.23 (1.32,3.14), p<0.001; quartile 4:  $\beta$ =2.40 (1.27,3.52), p<0.001) (Supp Table 2).

347 Further, we examined the role of ancestry proportion on the effect modification of PGS<sub>BMI</sub> 348 by age-at-immigration and diet, which revealed heterogeneity by AME ancestry proportion 349 guartile. The estimates of PGS<sub>BMI</sub> x age-at-immigration interaction comparing those born in the 350 US to immigrating as adults (>21 years) generally became stronger with increasing quartile of 351 AME ancestry for the lowest three quartiles, though only statistically significant in the third quartile 352 of AME ancestry (quartile 3:  $\beta$ =0.788 (95% CI: 0.183,1.29), p=0.01). The estimates of the PGS<sub>BM</sub>-353 diet interaction were variable and only significant (p<0.05) in the highest quartile of AME ancestry 354 (quartile 4: β=-0.536 (95% CI: -1.05, -0.03), p=0.04) (Supp Table 2). These interactions are 355 directionally consistent with their unstratified complete model estimates for the PGS<sub>BMI</sub>-diet 356 interaction ( $\beta$ =-0.398 (95% CI: -0.725, -0.07), p=0.017) and the PGS<sub>BMI</sub>-immigration interaction 357 for adults >21 years compared to those born in the US ( $\beta$ =0.514 (95% CI: -0.126,1.15), p=0.12) 358 (Supp Table 1).

# Interaction of PGS<sub>BMI</sub> with non-genetic factors and admixture proportions differs by Hispanic/Latino background group

361 To evaluate the potential modification of both AME ancestry and background group on the 362 performance of PGS<sub>BMI</sub> and its interaction with sociocultural factors: we partitioned AME ancestry 363 into tertiles in the pooled sample due to limited sample size when data is parsed by both variables. 364 The ancestry tertiles corresponded to 0-12.5%, 12.5-42.8%, and 42.8-99% inferred AME 365 ancestry. When stratified by both tertiles of AME ancestry and by background, model  $R^2$ 366 decreased with increasing tertiles in all background groups with appreciable AME ancestry 367 (Mexican, Central American, and South American). In Caribbean (Cuban, Dominican, Puerto 368 Rican) groups with lower AME ancestry proportions on average, trends were inconsistent with 369 insufficient data in some cells (Table 4). The adjusted effect of the PGS<sub>BM</sub> on BMI generally 370 became stronger with increasing AME ancestry tertile in each background group, contrary to the 371 trend observed without background stratification (Supp Tables 3.1-3.6). This trend was observed

in all groups except in the Mexican background group, where the PGS<sub>BMI</sub> coefficient became attenuated in the higher AME ancestry tertile ( $\beta$ =1.21 (95% CI: 0.190, 2.22), p=0.02) compared to the lower tertile ( $\beta$ =3.25 (95% CI: 1.72, 4.77), p<0.001).

375 The estimates of the PGS<sub>BMI</sub>-immigration interaction (reference group are participants who 376 immigrated at 21 years old or older) were only statistically significant in three strata; 1) in the 1st 377 tertile of AME ancestry in the Cuban background group, comparing those aged 0-5 years at 378 immigration ( $\beta$ =2.20 (0.43, 3.98), p=0.02), 2) in the 1st AME ancestry tertile of the Dominican 379 background group, comparing those aged 6-12 years at immigration ( $\beta$ =4.38 (2.14, 6.62), 380 p<0.001), and 3) in the 2nd AME ancestry tertile of the Mexican background group, comparing 381 those born in the US ( $\beta$ =1.33 (0.59, 2.07), p<0.001). (Supplementary Figure 2) Expanding on the 382 last strata for Mexican background individuals, participants who immigrated at  $\geq 21$  years old 383 had on average an increase of 3.25 units of BMI (95% CI: 1.72, 4.77) for every standard deviation 384 increase of PGS<sub>BMI</sub>. However, for participants born in the US, every one unit increase in PGS<sub>BMI</sub> 385 is associated with a 4.58 unit increase in BMI (1.33 units higher than the reference group). As 386 such, the effect of the PRS is statistically higher in this contrast. The PGS<sub>BMI</sub>-immigration 387 interaction remained statistically insignificant and was directionally heterogeneous in all other 388 categories of background groups and subsequent ancestry tertiles, despite larger sample sizes 389 in those bins. This indicates that even when explicitly controlling for ancestry, through both AME 390 ancestry proportions and principal components, there remain differences between 391 Hispanic/Latino groups for this GxE interaction (PGS<sub>BMI</sub>-immigration).

392 Beyond the inherent heterogeneity in model fit demonstrated here between background 393 groups of a Hispanic/Latino cohort, we observe that AME ancestry proportion thereby differentially 394 affects the GxE interactions underlying BMI-diet-age at immigration relationships. As AME 395 ancestry proportion increases, the model fit worsens. Interestingly, the statistically significant GxE 396 interactions of PGS-by-diet and PGS-by-age at immigration generally did persist upon 397 stratification by ancestry quartile and by both ancestry tertile and background, except in the 1st 398 tertiles of Mexican and Dominican groups and in the 2nd tertile of the Central American 399 background group, where estimates were directionally heterogeneous between groups. This 400 effect size heterogeneity across ancestry tertiles within some groups may indicate the 401 intersectional role of genetic ancestry with sociocultural influences which could affect the 402 performance of the PGS.

## 403 **DISCUSSION**

404 Through the present analyses, we sought to characterize these complex relationships 405 between a polygenic score, genetic ancestry, background group, and various lifestyle, 406 sociodemographic, and immigration-related variables in BMI prediction in a Hispanic/Latino 407 population (HCHS/SOL). Our inquiry was four-fold: (1) what are the individual distributions of 408 these variables across Hispanic/Latino backgrounds?; (2) what is these variables' relationship to 409 genetic ancestry?; (3) can inferred genetic ancestry explain previously identified heterogeneity 410 between Hispanic/Latino backgrounds in an integrated BMI risk model?; and (4) does additional 411 self-identified background group-level stratification offer additional insights to comprehensive 412 inferred ancestry modeling? Across all analyses, we identified persistent heterogeneity in 413 univariate and multivariate associations of these variables with PGS<sub>BMI</sub>, particularly due to

It is made available under a CC-BY-NC-ND 4.0 International license .

414 confounding by Hispanic/Latino background group that indicates the importance of background-415 level stratification in avoiding spurious findings in this population.

416 First, we identified high levels of heterogeneity among the various genetic and 417 environmental variables in this study population, underscoring the important distinctions to be made between groups of self-identified Hispanic/Latino ethnicity. In fact, aggregating groups of 418 419 this ethnicity obscures meaningful heterogeneity between geographical and cultural groups of 420 Hispanic/Latino ethnicity and may lead to false conclusions. This is consistent with other studies 421 of acculturation and BMI in Hispanic/Latino populations. For example, Khan et al. found that the 422 effect of acculturation on BMI among Mexican Americans in the Hispanic Health and Nutrition 423 Examination Survey was stronger than among Puerto Ricans and Cubans, which the authors 424 hypothesized as being at least partly a function of longer duration of "exposure to (mainland) US 425 culture" among Mexican Americans, and may have also been driven by a higher sample size in 426 this group.<sup>10</sup>

427 Second, we observed that when stratified by background group, the relationships of these 428 variables with genetic ancestry, modeled here as inferred AME ancestry admixture proportion, 429 were heterogeneous. Genetic analyses of Hispanic/Latino participants may be inclined to 430 incorporate Amerindigenous or other ancestry proportions into the model, as an important 431 predictor of environmental variables of interest. It has been reported that those of Hispanic/Latino ethnicity in the US have, on average, 18.0% AME ancestry.<sup>35</sup> The analysis of AME ancestry in 432 433 this cohort, however, demonstrates significant variation ranging from an average of 4% to 47% 434 AME ancestry in different Hispanic/Latino background groups. In our analysis, we identify 435 heterogeneity in associations between various health-related variables and ancestry proportions, 436 both between pooled and stratified results, as well as between the groups themselves. Therefore, 437 in studies involving Hispanic/Latino participants, background group stratification is important to 438 prevent inferences of spurious associations of such variables with AME ancestry, which are 439 instead driven primarily by group heterogeneity and differences in ancestry proportions between 440 background groups.

441 Third, we found that the fit of our predictive model for BMI as explained by a PGS<sub>BMI</sub> along 442 with various health, immigration, and environmental variables, and genetic ancestry worsened with increasing proportions of AME ancestry. This observation is consistent with previous 443 444 demonstrations of poor transferability of PGS trained in European-ancestries to other 445 populations.<sup>36,37</sup> This is not surprising given of the PGS<sub>BMI</sub> here was trained on a largely European 446 ancestry sample, which inherently does not reflect the genetic composition of HCHS/SOL 447 participants.<sup>29</sup> However, it is notable that there is significant heterogeneity within the pooled 448 Hispanic/Latino participants. Longstanding patterns of migration, conflict, and other sociopolitical 449 factors influenced the current structure of genetic ancestry of Hispanic/Latino groups being a 450 combination of European, African, and Amerindigenous ancestry, giving rise to different allele frequencies and linkage disequilibrium (LD) correlation structures.<sup>35</sup> This poor performance by 451 452 ancestry has consequences in the ability to detect GxE interactions, as demonstrated by the high 453 heterogeneity in interaction effect estimates of PGS with immigration and diet-related variables, 454 particularly related to the confounding effect of background group on the relationship of these 455 variables and BMI.

It is made available under a CC-BY-NC-ND 4.0 International license .

456 Fourth, when we subsequently stratified this model by both AME ancestry and background 457 group, we identified heterogeneous effects in the specific contrasts in which GxE interactions 458 persisted. McArdle et al. previously identified few statistically significant GxE interactions in an 459 analysis of the full model stratified by background group: those who immigrated to the US had a higher BMI than those born in the US, comparing those with a higher to lower PGS<sub>BMI</sub>.<sup>21</sup> In our 460 analysis that was stratified by both background group and AME ancestry, our results also 461 462 indicated a modest effect of younger age at immigration to the US on the PGS<sub>BMI</sub>- BMI relationship 463 in Cuban, Dominican, and Mexican background groups. Those who were brought to the US at 464 young ages from different countries may also reflect different demographic backgrounds and 465 sociocultural environments that represent varying risk-increasing or risk-decreasing relationships 466 of health- or environment-related risk factors with BMI and obesity. Group-specific histories of 467 colonization, along with geopolitical and temporal patterning of how specific Hispanic/Latino 468 groups became part of or relocated to the US, play a major role in these sociocultural and demographic differences, with immigration and land loss beginning in the mid-19<sup>th</sup> century and 469 470 continuing to the present day through many immigration barriers and policies.<sup>38</sup> In recent history. 471 migration from Mexico to the US peaked between 2000-2010, before the implementation of 472 stricter border policies that curtailed the routine back and forth or "circular" migration patterns 473 between Mexico and the US.<sup>39,40</sup> The Caribbean islands of Puerto Rico, the Dominican Republic. 474 and Cuba, despite their geographical proximity, have had more nuanced US migration and 475 immigration patterns. While the former was annexed by the US in 1898 and saw migration 476 numbers peak after World War II, Cubans were mostly isolated from the US during this time.<sup>39</sup> This long standing relationship changed, however, with the 1959 communist revolution, which 477 drove the immigration of hundreds of thousands of Cuban migrants to the US.<sup>39</sup> In recent years. 478 479 the immigration of Hispanic/Latino people to the US has continued, including many from Central 480 and South American countries, albeit at a slower pace in the face of stricter immigration policy.<sup>41</sup> 481 Our results are at least partially consistent with other non-genetic analyses of obesity in 482 Hispanic/Latino populations in the US that reflect the important role of US residency and 483 immigration in obesity and that age at immigration is not as significant an obesity risk factor in 484 these populations.<sup>42</sup>

485 This study's limitations lie primarily in small sample size and potential for information bias 486 and misclassification. Loss of statistical significance in the GxE interactions is likely driven at least 487 in part by small sample sizes in the increasingly minute sub-classifications of the analytic sample. 488 However, the maintained statistical significance of these interactions and their significant effect 489 size change in two of the background groups suggests that group analyses could identify 490 differential modifying effects of diet and age-at-immigration on the effect on BMI of PGS<sub>BMI</sub>. Our 491 analyses were also restricted to incorporation of AME ancestry, despite the notably important 492 contribution of African (AFR) ancestry, particularly in Dominican background participants. High 493 AFR ancestry likely also contributes differentially to poor model fit observed for this group in Table 494 5.<sup>43</sup> We did not explore associations of immigration-related, environmental, and diet variables with 495 African (AFR) ancestry, due to issues arising from collinearity of AFR ancestry with AME ancestry 496 and extremely limited sample sizes poorly powered to detect effects. Further studies to examine 497 the role of AFR ancestry in these GxE frameworks are needed.

498 An additional limitation is that the measurement of immigration-related variables (i.e. age-499 at-immigration, nativity, and years lived in the US) requires the simplification of innately complex

It is made available under a CC-BY-NC-ND 4.0 International license .

500 concepts that reflect years and often generations of migratory and environmental changes. 501 Importantly, these analyses cannot capture the circular nature of migration of Hispanic/Latino 502 groups, for example among individuals of Mexican and Puerto Rican backgrounds who often 503 migrate between the US 50 states/DC and Mexico or Puerto Rico within and across generations.<sup>39,40,44,45</sup> However, current analyses of health and health-related behaviors in US 504 505 Hispanic/Latino populations have consistently used acculturation and immigration-related 506 variables as proxies for complex migratory and environmental patterns and have robustly 507 identified associations with health-related variables, including BMI.<sup>46–48</sup>

508 Despite controlling for self-identified background, AME ancestry proportion, and principal 509 components, our analyses still identified significant heterogeneity in GxE interactions, indicating 510 the influence of non-genetic factors and complex social environments in the poor performance of 511 a model exploring a European-ancestry-derived PGS<sub>BMI</sub> in a non-European sample. Taken 512 together, we demonstrate that environmental variables play an important role in the effect of 513 genetics on obesity risk in Hispanic/Latino groups. However, the consideration of group identity 514 within Hispanic/Latino ethnicity and other large ethnic groups requires increased urgency, 515 especially as the incorporation of precision medicine into clinical practice and preventative care 516 looms closer. Heeding calls for movement towards an integrated risk score requires 517 acknowledgement of the complex genetic and environmental profiles of this population, 518 specifically the explicit modeling of their multi-layered sociodemographic and relationally complex 519 migration histories of which may confound relationships of genetic and environmental variables.

520 To this aim, we recommend that cohorts must collect finer data on Hispanic/Latino 521 participants that will continue to be used to study this population. This recommendation falls in 522 line with recent consensus around for the appropriate use of race and ethnicity as population 523 descriptors in both genetic and biomedical research: to use more granular levels of participants' self-identity rather than ancestry-driven or broadly defined categorization.<sup>49,50</sup> Additionally, clinical 524 525 applications of genetic association analyses and of analyses incorporating gene-environment 526 interactions should pay attention to different modifying effects of environmental variables between 527 Hispanic/Latino background groups. More broadly, epidemiologic and other analyses should 528 stratify Hispanic/Latino participant populations by their background group identity to understand 529 the true nature of public health relationships in these groups, and to prevent the publication of 530 sweeping generalizations that are not reflective of true associations between important variables.

It is made available under a CC-BY-NC-ND 4.0 International license .

### 531 Acknowledgements

532 The authors thank the staff and participants of HCHS/SOL for their important contributions. 533 (Investigators website - http://www.cscc.unc.edu/hchs/) The Hispanic Community Health 534 Study/Study of Latinos is a collaborative study supported by contracts from the National Heart, 535 Lung, and Blood Institute (NHLBI) to the University of North Carolina (HHSN2682013000011 / 536 N01-HC-65233), University of Miami (HHSN268201300004I / N01-HC-65234), Albert Einstein 537 College of Medicine Rev. December 4, 2023 (HHSN268201300002I / N01-HC-65235), University 538 of Illinois at Chicago (HHSN268201300003I / N01- HC-65236 Northwestern Univ), and San Diego 539 State University (HHSN268201300005I 1 N01-HC-65237). The following 540 Institutes/Centers/Offices have contributed to the HCHS/SOL through a transfer of funds to the 541 NHLBI: National Institute on Minority Health and Health Disparities, National Institute on Deafness 542 and Other Communication Disorders, National Institute of Dental and Craniofacial Research, 543 National Institute of Diabetes and Digestive and Kidney Diseases, National Institute of 544 Neurological Disorders and Stroke, NIH Institution-Office of Dietary Supplements. Additional 545 funding for JS and GLW comes from the National Human Genome Research Institute 546 (R35HG011944).

It is made available under a CC-BY-NC-ND 4.0 International license .

#### 547 **REFERENCES**

- Ogden CL, Carroll MD, Fryar CD, Flegal KM. Prevalence of Obesity Among Adults and Youth:
   United States, 2011-2014. NCHS Data Brief. 2015;(219):1-8.
   https://www.ncbi.nlm.nih.gov/pubmed/26633046
- 551 2. Krogstad JM. Hispanics only group to see its poverty rate decline and incomes rise. 552 Published 2014. Accessed September 19, August 7, 2024. 553 https://www.pewresearch.org/short-reads/2014/09/19/hispanics-only-group-to-see-its-554 poverty-rate-decline-and-incomes-rise/
- 555 3. Cortes-Bergoderi M, Goel K, Murad MH, et al. Cardiovascular mortality in Hispanics 556 compared to non-Hispanic whites: a systematic review and meta-analysis of the Hispanic 557 paradox. *Eur J Intern Med*. 2013;24(8):791-799. doi:10.1016/j.ejim.2013.09.003
- Chartier KG, Carmody T, Akhtar M, Stebbins MB, Walters ST, Warden D. Hispanic
   Subgroups, Acculturation, and Substance Abuse Treatment Outcomes. *J Subst Abuse Treat*.
   2015;59:74-82. doi:10.1016/j.jsat.2015.07.008
- 561 Smith TM, Colón-Ramos U, Pinard CA, Yaroch AL. Household food insecurity as a 5. 562 determinant of overweight and obesity among low-income Hispanic subgroups: Data from 563 2011-2012 California Health Interview Survey. Appetite. 2016;97:37-42. the 564 doi:10.1016/j.appet.2015.11.009
- Manichaikul A, Palmas W, Rodriguez CJ, et al. Population structure of Hispanics in the United
   States: the multi-ethnic study of atherosclerosis. *PLoS Genet*. 2012;8(4):e1002640.
   doi:10.1371/journal.pgen.1002640
- 568 7. Gravel S, Zakharia F, Moreno-Estrada A, et al. Reconstructing Native American migrations
  569 from whole-genome and whole-exome data. *PLoS Genet*. 2013;9(12):e1004023.
  570 doi:10.1371/journal.pgen.1004023
- 5718.US Census Bureau. Measuring Racial and Ethnic Diversity for the 2020 Census. Accessed572March17,2022.https://www.census.gov/newsroom/blogs/random-573samplings/2021/08/measuring-racial-ethnic-diversity-2020-census.html
- Hales CM, Carroll MD, Fryar CD, Ogden CL. Prevalence of obesity and severe obesity among adults: United States, 2017-2018. *NCHS Data Brief*. 2020;(360):1-8. Accessed February 27, 2022. https://www.cdc.gov/nchs/data/databriefs/db360-h.pdf
- 577 10. Khan LK, Sobal J, Martorell R. Acculturation, socioeconomic status, and obesity in Mexican
  578 Americans, Cuban Americans, and Puerto Ricans. *Int J Obes Relat Metab Disord*.
  579 1997;21(2):91-96. doi:10.1038/sj.ijo.0800367
- Yracheta JM, Alfonso J, Lanaspa MA, et al. Hispanic Americans living in the United States
   and their risk for obesity, diabetes and kidney disease: Genetic and environmental
   considerations. *Postgrad Med.* 2015;127(5):503-510. doi:10.1080/00325481.2015.1021234
- Tang H, Peng J, Wang P, Risch NJ. Estimation of individual admixture: analytical and study
   design considerations. *Genet Epidemiol*. 2005;28(4):289-301. doi:10.1002/gepi.20064
- Alexander DH, Novembre J, Lange K. Fast model-based estimation of ancestry in unrelated
   individuals. *Genome Res.* 2009;19(9):1655-1664. doi:10.1101/GR.094052.109

- 14. Homburger JR, Moreno-Estrada A, Gignoux CR, et al. Genomic Insights into the Ancestry
  and Demographic History of South America. *PLoS Genet*. 2015;11(12):e1005602.
  doi:10.1371/journal.pgen.1005602
- 590 15. Moreno-Estrada A, Gravel S, Zakharia F, et al. Reconstructing the population genetic history 591 of the Caribbean. *PLoS Genet*. 2013;9(11):e1003925. doi:10.1371/journal.pgen.1003925
- 592 16. Bryc K, Velez C, Karafet T, et al. Colloquium paper: genome-wide patterns of population
  593 structure and admixture among Hispanic/Latino populations. *Proc Natl Acad Sci U S A*.
  594 2010;107 Suppl 2(Supplement 2):8954-8961. doi:10.1073/pnas.0914618107
- 595 17. Martin AR, Kanai M, Kamatani Y, Okada Y, Neale BM, Daly MJ. Clinical use of current polygenic risk scores may exacerbate health disparities. *Nat Genet*. 2019;51(4):584-591.
  597 doi:10.1038/s41588-019-0379-x
- 598 18. Bitarello BD, Mathieson I. Polygenic scores for height in admixed populations. *G3* (*Bethesda*). 2020;10(11):4027-4036. doi:10.1534/g3.120.401658
- Mostafavi H, Harpak A, Agarwal I, Conley D, Pritchard JK, Przeworski M. Variable prediction
   accuracy of polygenic scores within an ancestry group. *Elife*. 2020;9.
   doi:10.7554/eLife.48376
- Components of UK Biobank. *Nat Commun.* 2021;12(1):1-13. doi:10.1038/s41467-021-21073y
- McArdle CE, Bokhari H, Rodell CC, et al. Findings from the Hispanic Community Health
  Study/Study of Latinos on the Importance of Sociocultural Environmental Interactors:
  Polygenic Risk Score-by-Immigration and Dietary Interactions. *Front Genet*. 2021;12:1853.
  doi:10.3389/FGENE.2021.720750/BIBTEX
- LaVange LM, Kalsbeek WD, Sorlie PD, et al. Sample Design and Cohort Selection in the
  Hispanic Community Health Study/Study of Latinos. *Ann Epidemiol*. 2010;20(8):642-649.
  doi:10.1016/J.ANNEPIDEM.2010.05.006
- 613 23. World Health Organization 2020 Guidelines on Physical Activity and Sedentary Behaviour.
- Andresen EM, Malmgren JA, Carter WB, Patrick DL. Screening for depression in well older
  adults: evaluation of a short form of the CES-D (Center for Epidemiologic Studies Depression
  Scale). *Am J Prev Med.* 1994;10(2):77-84. https://www.ncbi.nlm.nih.gov/pubmed/8037935
- Liu K, Daviglus ML, Loria CM, et al. Healthy lifestyle through young adulthood and the
  presence of low cardiovascular disease risk profile in middle age: the Coronary Artery Risk
  Development in (Young) Adults (CARDIA) study. *Circulation*. 2012;125(8):996-1004.
  doi:10.1161/CIRCULATIONAHA.111.060681
- 621 26. Gong J, Nishimura KK, Fernandez-Rhodes L, et al. Trans-ethnic analysis of metabochip data
  622 identifies two new loci associated with BMI. *Int J Obes.* 2017;42(3):384-390.
  623 doi:10.1038/ijo.2017.304

- Conomos MP, Laurie CA, Stilp AM, et al. Genetic Diversity and Association Studies in US
   Hispanic/Latino Populations: Applications in the Hispanic Community Health Study/Study of
   Latinos. *Am J Hum Genet*. 2016;98(1):165-184. doi:10.1016/j.ajhg.2015.12.001
- Sofer T, Moon JY, Isasi CR, et al. Relationship of genetic determinants of height with
  cardiometabolic and pulmonary traits in the Hispanic Community Health Study/Study of
  Latinos. *Int J Epidemiol.* 2018;47(6):2059-2069. doi:10.1093/ije/dyy177
- Sidorenko J, Kemper KE, et al. Meta-analysis of genome-wide association studies
   for height and body mass index in ~700000 individuals of European ancestry. *Hum Mol Genet.* 2018;27(20):3641-3649. doi:10.1093/hmg/ddy271
- 30. Vilhjálmsson BJ, Yang J, Finucane HK, et al. Modeling Linkage Disequilibrium Increases
  Accuracy of Polygenic Risk Scores. *Am J Hum Genet*. 2015;97(4):576-592.
  doi:10.1016/j.ajhg.2015.09.001
- 31. Wojcik GL, Graff M, Nishimura KK, et al. Genetic analyses of diverse populations improves
  discovery for complex traits. *Nature*. 2019;570(7762):514-518. doi:10.1038/s41586-0191310-4
- 639 32. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK:
  640 rising to the challenge of larger and richer datasets. *Gigascience*. 2015;4(1):7.
  641 doi:10.1186/s13742-015-0047-8
- 33. Behr AA, Liu KZ, Liu-Fang G, Nakka P, Ramachandran S. pong: fast analysis and
  visualization of latent clusters in population genetic data. *Bioinformatics*. 2016;32(18):28172823. doi:10.1093/BIOINFORMATICS/BTW327
- Sjoberg DD. Gtsummary: Presentation-Ready Data Summary and Analytic Result Tables.;
   2021. https://CRAN.R-project.org/package=MetBrewer
- Bryc K, Durand EY, Macpherson JM, Reich D, Mountain JL. The genetic ancestry of African Americans, Latinos, and European Americans across the United States. *Am J Hum Genet*. 2015;96(1):37-53. doi:10.1016/j.ajhg.2014.11.010
- 36. Duncan L, Shen H, Gelaye B, et al. Analysis of polygenic risk score usage and performance
  in diverse human populations. *Nat Commun.* 2019;10(1):1-9. doi:10.1038/s41467-01911112-0
- 653 37. Grinde KE, Qi Q, Thornton TA, et al. Generalizing polygenic risk scores from Europeans to 654 Hispanics/Latinos. *Genet Epidemiol*. 2019;43(1):50-62. doi:10.1002/gepi.22166
- 655 38. Alba FAF. Mexico: The New Migration Narrative. Accessed March 31, 2022.
   656 https://www.migrationpolicy.org/article/mexico-new-migration-narrative
- Blizzard JBB, Batalova J. Cuban Immigrants in the United States. Accessed February 27,
   2022. https://www.migrationpolicy.org/article/cuban-immigrants-united-states-2018
- 40. Rosenblum MR, Kandel WA, Seelke CR, Wasem RE. Mexican migration to the United States:
  Policy and trends. Published 2012. Accessed March 18, 2022.
  https://sgp.fas.org/crs/row/R42560.pdf

It is made available under a CC-BY-NC-ND 4.0 International license .

- 41. Moslimani M. Facts on Latinos in the U.S. Published August 16, 2023. Accessed August 6,
  2024. https://www.pewresearch.org/race-and-ethnicity/fact-sheet/latinos-in-the-us-factsheet/
- 42. Isasi CR, Ayala GX, Sotres-Alvarez D, et al. Is acculturation related to obesity in
  Hispanic/Latino adults? Results from the Hispanic Community Health Study/Study of Latinos. *J Obes.* 2015;2015. doi:10.1155/2015/186276
- 43. Tajima A, Hamaguchi K, Terao H, et al. Genetic background of people in the Dominican
  Republic with or without obese type 2 diabetes revealed by mitochondrial DNA
  polymorphism. *J Hum Genet*. 2004;49(9):495-499. doi:10.1007/s10038-004-0179-7
- 44. Duany J. Mobile livelihoods: The sociocultural practices of circular migrants between Puerto
  Rico and the United States. *Int Migr Rev.* 2002;36(2):355-388. doi:10.1111/j.17477379.2002.tb00085.x
- 45. Durand J, Massey DS. Evolution of the Mexico-U.s. migration system: Insights from the
  Mexican Migration Project. Ann Am Acad Pol Soc Sci. 2019;684(1):21-42.
  doi:10.1177/0002716219857667
- 46. Lara M, Gamboa C, Kahramanian MI, Morales LS, Bautista DEH. Acculturation and Latino
  health in the United States: a review of the literature and its sociopolitical context. *Annu Rev Public Health*. 2005;26:367-397. doi:10.1146/annurev.publhealth.26.021304.144615
- 47. Fernández-Rhodes L, Butera NM, Lodge EK, et al. Demographic and sociocultural risk
  factors for adulthood weight gain in Hispanic/Latinos: results from the Hispanic Community
  Health Study / Study of Latinos (HCHS/SOL). *BMC Public Health*. 2021;21(1):2064.
  doi:10.1186/s12889-021-11848-9
- 684 48. Guadamuz JS, Durazo-Arvizu RA, Daviglus ML, et al. Immigration Status and Disparities in
  685 the Treatment of Cardiovascular Disease Risk Factors in the Hispanic Community Health
  686 Study/Study of Latinos (Visit 2, 2014-2017). *Am J Public Health*. 2020;110(9):1397-1404.
  687 doi:10.2105/AJPH.2020.305745
- 49. National Academies of Sciences, Engineering, and Medicine. 2023. Using Population
  Descriptors in Genetics and Genomics Research: A New Framework for an Evolving Field.
  Washington, DC: The National Academies Press. https://doi.org/10.17226/26902.
- 50. National Academies of Sciences, Engineering, and Medicine. 2024. Rethinking Race and
  Ethnicity in Biomedical Research. Washington, DC: The National Academies Press.
  https://doi.org/10.17226/27913.

694

## TABLES

Table 1. Selected Population Characteristics of the HCHS/SOL Analytic Subsample Overall (n=7,282) and Stratified by Background

Characteristic <sup>1</sup>	Overall	Central American	Cuban	Dominican	Mexican	Puerto Rican	South American	p-value
Total	7282	846	1331	659	2701	1211	534	
Weighted	7075	602	1717	636	2560	1157	403	
AGE (years)	43 (33, 54)	40 (32, 51)	48 (39, 59)	43 (33, 53)	40 (31, 51)	45 (34, 56)	44 (36, 53)	<0.001
STUDY CENTER								<0.001
Brooklyn	1,965 (28%)	81 (13%)	117 (6.8%)	594 (93%)	252 (9.8%)	825 (71%)	96 (24%)	
Chicago	1,033 (15%)	111 (19%)	21 (1.2%)	6 (0.9%)	613 (24%)	219 (19%)	63 (16%)	
Miami	2,287 (32%)	358 (59%)	1,567 (91%)	30 (4.7%)	43 (1.7%)	76 (6.6%)	214 (53%)	
San Diego	1,790 (25%)	52 (8.6%)	12 (0.7%)	6 (0.9%)	1,653 (65%)	38 (3.3%)	30 (7.4%)	
AGE AT MIGRATION								<0.001
Born in the US	1,255 (18%)	36 (5.9%)	144 (8.4%)	48 (7.5%)	478 (19%)	514 (44%)	36 (9.0%)	
0 - <6 yrs	337 (4.8%)	8 (1.3%)	35 (2.0%)	19 (3.0%)	114 (4.5%)	153 (13%)	8 (1.9%)	
6 - 12 yrs	326 (4.6%)	25 (4.1%)	67 (3.9%)	33 (5.1%)	111 (4.3%)	79 (6.8%)	12 (3.0%)	
13 - 20 yrs	1,173 (17%)	121 (20%)	141 (8.2%)	119 (19%)	582 (23%)	167 (14%)	42 (10%)	
>=21 yrs	3,985 (56%)	413 (69%)	1,330 (77%)	417 (66%)	1,275 (50%)	245 (21%)	306 (76%)	
IMMIGRANT GENERATION								<0.001
1st Generation	5,714 (81%)	567 (94%)	1,570 (91%)	587 (92%)	2,012 (79%)	612 (53%)	367 (91%)	
EDUCATION								<0.001
Less than HS	2,160 (31%)	227 (38%)	339 (20%)	230 (36%)	903 (35%)	390 (34%)	72 (18%)	
INCOME								<0.001
<\$30,000 USD	4,626 (65%)	454 (75%)	1,235 (72%)	462 (73%)	1,481 (58%)	739 (64%)	257 (64%)	
EMPLOYMENT STATUS								<0.001
Retired and not currently working	649 (9.2%)	25 (4.1%)	202 (12%)	65 (10%)	122 (4.8%)	209 (18%)	27 (6.6%)	
Not retired and not currently working	2,581 (36%)	193 (32%)	735 (43%)	232 (37%)	883 (34%)	422 (36%)	116 (29%)	
Employed <= 35h/wk	1,190 (17%)	141 (23%)	194 (11%)	115 (18%)	501 (20%)	152 (13%)	86 (21%)	
Employed > 35h/wk	2,656 (38%)	244 (40%)	586 (34%)	224 (35%)	1,054 (41%)	374 (32%)	174 (43%)	

Sweetened Beverage Consumption (Servings/day)	1.58 2.34)	(0.97,	1.73 (1.06, 2.61)	1.48 2.11)	(0.95,	1.31 1.85)	(0.74,	1.59 2.38)	(0.97,	1.77 2.66)	(1.12,	1.64 2.37)	(1.02,	<0.001
Meets 2008 Physical Activity Guidelines	4,703 (6	6%)	421 (70%)	970 (569	%)	424 (67	%)	1,807 (	71%)	795 (69	9%)	287 (719	%)	<0.001
JAMA Healthy Diet Score														<0.001
>60th sex-specific percentile	3,393 (4	8%)	264 (44%)	719 (429	%)	143 (22	%)	1,827 (	71%)	231 (20	0%)	210 (529	%)	
Sleep Duration (h/day)	7.93 8.71)	(7.00,	7.79 (7.00, 8.57)	8.00 8.93)	(7.21,	7.86 8.93)	(7.00,	8.00 8.71)	(7.29,	7.64 8.79)	(6.71,	7.71 8.50)	(6.79,	<0.001
CIGARETTE USE														<0.001
Never	4,223 (6	0%)	390 (65%)	938 (559	%)	501 (79	%)	1,586 (	62%)	570 (49	9%)	236 (59	%)	
Former	1,365 (1	9%)	119 (20%)	342 (209	%)	80 (13%	<b>6</b> )	524 (20	0%)	191 (17	7%)	109 (27	%)	
Current	1,489 (2	1%)	93 (15%)	437 (259	%)	55 (8.6%	%)	450 (18	3%)	396 (34	1%)	58 (14%	<b>)</b>	
ALCOHOL USE LEVEL														0.004
Non-drinker	3,382 (4	8%)	336 (56%)	857 (50%	%)	290 (46	%)	1,157 (	45%)	567 (49	9%)	175 (439	%)	
Low-risk drinker	3,252 (4	6%)	228 (38%)	745 (439	%)	312 (49	%)	1,222 (	48%)	527 (46	S%)	218 (54	%)	
At-risk drinker	442 (6.2	%)	38 (6.3%)	115 (6.7	%)	34 (5.4%	%)	181 (7.	1%)	63 (5.5	%)	10 (2.6%	6)	
Diabetes History	1,154 (1	6%)	87 (14%)	298 (179	%)	96 (15%	<b>6</b> )	417 (10	5%)	216 (19	9%)	41 (10%	) )	0.025
Prevalent CVD	462 (6.5	%)	37 (6.1%)	151 (8.8	%)	40 (6.2%	%)	123 (4.	8%)	92 (7.9	%)	20 (4.8%	6)	0.055
BMI	28.9 (25.9, 32	2.7)	28.8 (26.2, 32.4)	29.0 (25.7, 32	2.6)	29.3 (26.0, 3	2.7)	28.7 (26.1, 3	32.3)	29.6 (26.1, 3	34.3)	27.9 (24.9, 30	0.8)	<0.001
BMI PGS	0.01 (-0.65, 0	.67)	0.07 (-0.54, 0.67)	-0.14 (-0.88,0.	.55)	0.09 (-0.55, (	).69)	0.07 (-0.61,	0.72)	-0.03 (-0.69,	0.64)	-0.02 (-0.64, 0	).62)	<0.001
AME Ancestry Proportion	0.29 (0.09, 0.4	48)	0.45 (0.37, 0.54)	0.04 (0.02, 0.	08)	0.06 (0.05, 0	.08)	0.47 (0.37, 0	0.60)	0.13 (0.11, 0	).15)	0.44 (0.29, 0.	61)	<0.001

<sup>1</sup>Median (IQR); n (%)

	Overall (n=7075)			Mexican (n=	2560)		Puerto Rican (n=1157)			
Characteristic	Beta	95% CI	p-value	Beta	95% CI	p-value	Beta	95% CI	p-value	
Age	-0.003	-0.004, - 0.003	<0.001	-0.003	-0.003, - 0.002	<0.001	0.000	0.000, 0.000	0.024	
Study Center										
Bronx	REF	REF	REF	REF	REF	REF	REF	REF	REF	
Chicago	0.300	0.286, 0.313	<0.001	-0.207	-0.238, -0.177	<0.001	0.011	0.007, 0.016	<0.001	
Miami	0.017	0.004, 0.030	0.009	-0.333	-0.399, -0.266	<0.001	-0.002	-0.011, 0.008	0.73	
San Diego	0.248	0.235, 0.262	<0.001	-0.351	-0.381, -0.321	<0.001	0.007	-0.008, 0.022	0.38	
Age at Immigration										
>=21 years	REF	REF	REF	REF	REF	REF	REF	REF	REF	
US BORN	-0.079	-0.095, -0.062	<0.001	-0.094	-0.114, -0.074	<0.001	-0.003	-0.008, 0.002	0.26	
0-5 years	-0.082	-0.112, -0.052	<0.001	-0.054	-0.093, -0.0015	0.006	0.002	-0.005, 0.009	0.64	
6-12 years	-0.052	-0.079, -0.025	<0.001	-0.020	-0.054, 0.014	0.24	0.006	-0.002, 0.013	0.13	
13-20 years	0.059	0.043, 0.074	<0.001	0.002	-0.015, 0.019	0.81	0.002	-0.004, 0.008	0.48	
Immigrant Generation	-0.075	-0.091, -0.060	<0.001	-0.098	-0.115, -0.080	<0.001	-0.005	-0.009, -0.001	0.018	
Education	-0.078	-0.090, -0.066	<0.001	-0.064	-0.077, -0.051	<0.001	-0.004	-0.008, 0.000	0.078	
Income	0.004	-0.008, 0.016	0.55	-0.060	-0.074, -0.046	<0.001	0.001	-0.003, 0.006	0.49	
Employment Status										
Retired and not currently working	REF	REF	REF	REF	REF	REF	REF	REF	REF	
Not retired and not currently working	0.062	0.042, 0.082	<0.001	0.034	0.005, 0.062	0.020	0.001	-0.004, 0.006	0.65	

Table 2. AME Ancestry Proportion Univariate Association with Selected Variables and Background Groups (Individuals with missing data were dropped)

Employed <= 35h/wk	0.133	0.110, 0.155	<0.001	0.044	0.014, 0.075	0.004	-0.006	-0.014, 0.001	0.11
Employed > 35h/wk	0.107	0.087, 0.126	<0.001	0.046	0.018, 0.074	0.002	0.002	-0.004, 0.007	0.55
Sweetened Beverage Consumption (svgs/day)	0.016	0.011, 0.022	<0.001	0.018	0.011, 0.024	<0.001	0.001	-0.001, 0.003	0.22
Meets2008PhysicalActivityGuidelines	0.037	0.025, 0.049	<0.001	0.008	-0.007, 0.022	0.29	-0.001	-0.005, 0.003	0.64
Top 40th Percentile Diet Score	0.135	0.124, 0.146	<0.001	-0.030	-0.046, -0.014	<0.001	0.001	-0.004, 0.005	0.82
Sleep Duration (h/day)	-0.001	-0.005, 0.003	0.59	-0.002	-0.007, 0.004	0.53	0.000	-0.001, 0.001	0.85
Cigarette Use									
Never	REF	REF	REF	REF	REF	REF	REF	REF	REF
Former	-0.025	-0.039, -0.011	<0.001	-0.037	-0.054, -0.021	<0.001	-0.001	-0.006, 0.004	0.74
Current	-0.090	-0.104, -0.075	<0.001	-0.048	-0.067, -0.029	<0.001	-0.002	-0.006, 0.003	0.49
Alcohol Use Level									
Non-drinker	REF	REF	REF	REF	REF	REF	REF	REF	REF
Low-risk drinker	-0.013	-0.025, -0.002	0.026	-0.023	-0.037, -0.009	0.001	0.000	-0.004, 0.003	0.81
At-risk drinker	-0.031	-0.057, -0.005	0.020	-0.041	-0.071, -0.010	0.009	0.007	-0.002, 0.016	0.11
Diabetes History	0.000	-0.014, 0.014	0.97	0.025	0.008, 0.042	0.003	0.003	-0.001, 0.007	0.19
Prevalent CVD	-0.054	-0.076, -0.031	<0.001	0.001	-0.030, 0.031	0.95	0.001	-0.006, 0.008	0.76
BMI	-0.001	-0.002, 0.000	0.044	0.000	-0.001, 0.001	0.97	0.000	0.000, 0.000	0.20
BMI PGS	0.030	0.024, 0.036	<0.001	0.040	0.033, 0.047	<0.001	-0.001	-0.003, 0.001	0.39

# Table 3. Comparing R<sup>2</sup> values of multivariable regression results, stratified by AME ancestry proportion quartile

AME Ar	<b>D</b> <sup>2</sup>	
Quartile	Range	ĸ
Quartile 1	(0.000, 0.089)	0.222
Quartile 2	(0.089, 0.289)	0.202
Quartile 3	(0.289, 0.482)	0.213
Quartile 4	(0.482, 0.999)	0.164

## Table 4. Comparing R<sup>2</sup> values of multivariable regression results, stratified by AME ancestry proportion tertile & background group

AME Ance	estry Proportion	Background Group (R <sup>2</sup> (n))								
Tertile	Range	Central American	Cuban	Dominican	Mexican	Puerto Rican	South American			
1	(0.000, 0.125)	N/A (14)	0.238 (1225)	0.206 (658)	N/A (17)	0.298 (476)	N/A (38)			
2	(0.125, 0.428)	0.295 (360)	0.420 (106)	N/A (1)	0.252 (1012)	0.277 (735)	0.351 (213)			
3	(0.428, 0.999)	0.220 (472)	N/A (0)	N/A (0)	0.156 (1672)	N/A (0)	0.333 (283)			