The American Journal of Human Genetics, Volume *107*

Supplemental Data

A Genome-wide Association Study Discovers 46 Loci of the Human Metabolome in the Hispanic Community Health Study/Study of Latinos

Elena V. Feofanova, Han Chen, Yulin Dai, Peilin Jia, Megan L. Grove, Alanna C. Morrison, Qibin Qi, Martha Daviglus, Jianwen Cai, Kari E. North, Cathy C. Laurie, Robert C. Kaplan, Eric Boerwinkle, and Bing Yu

Supplemental Figures

Figure S1. Manhattan plots for known and previously unreported genetic loci affecting the metabolites levels for A. Lipid-related metabolites; B. Amino-acid-related metabolites; C. Other metabolites. For each of the corresponding super-pathway, known signals are shown in gray; color scheme for previously unreported loci for other metabolites is represented in the legend.

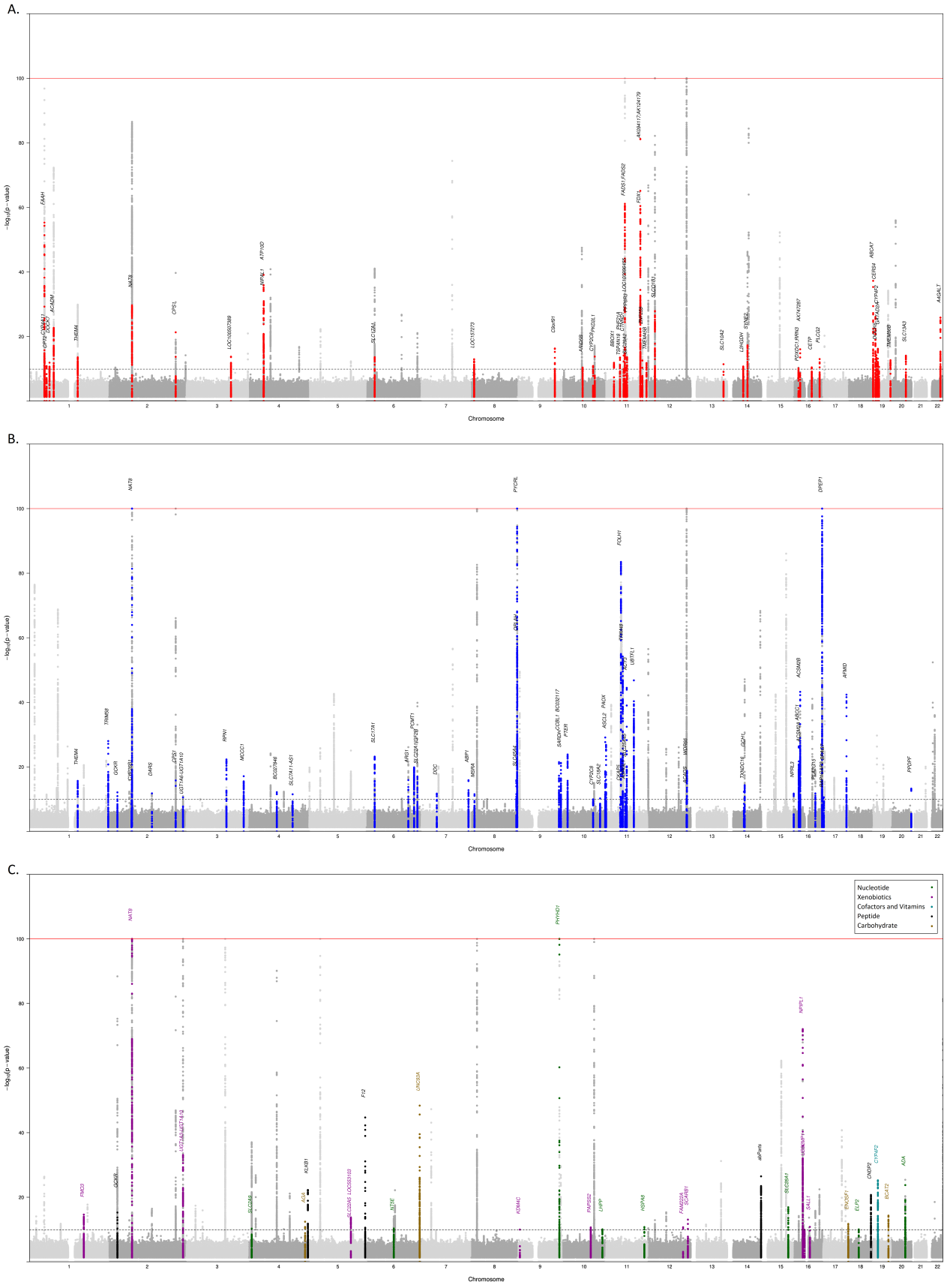
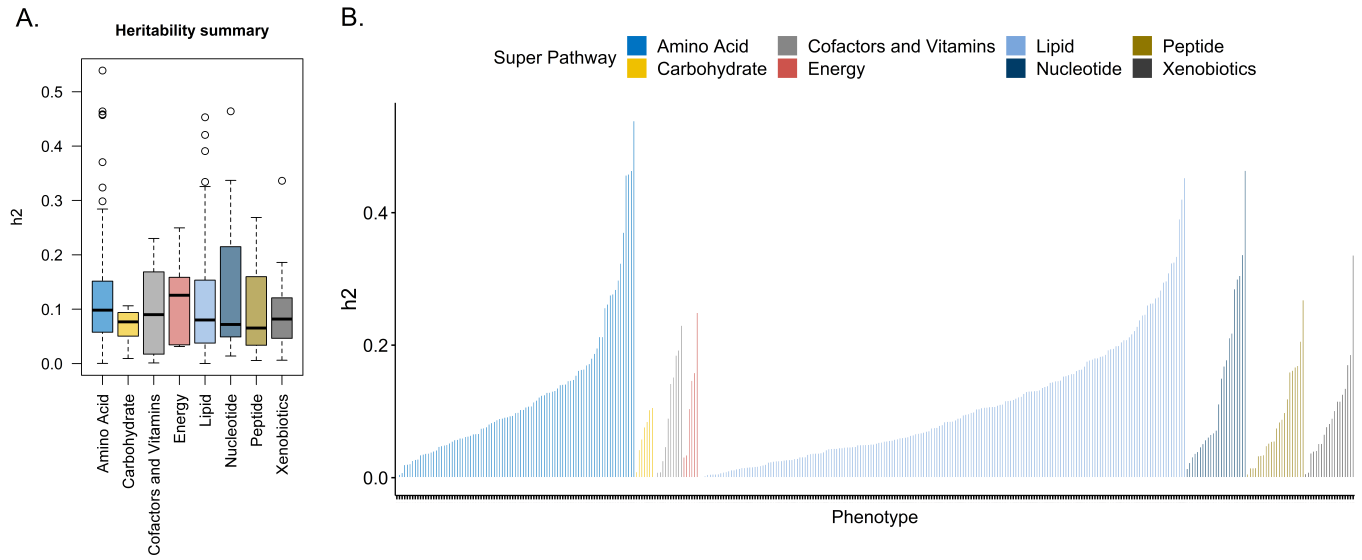


Figure S2. Heritability estimates for 366 metabolites*. A. Heritability estimates summary by Super Pathway. B. Heritability estimates by Super Pathway.



* 366 out of 640 metabolites had positive heritability estimate in ldsc.

Figure S3. Previously unreported variants with globalized effect associated with various phenotypes in previously published GWAS (PhenoScanner). Direction of effect is shown for the minor allele.

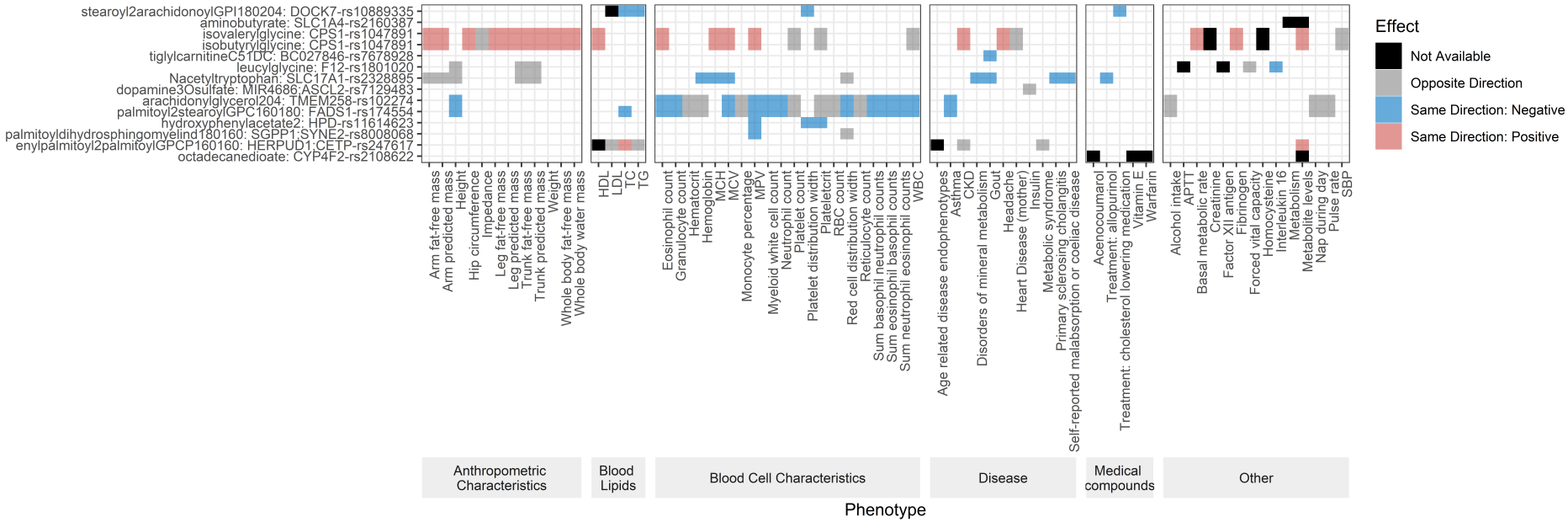


Figure S4. Tissue-specific enrichment analysis for 39 metabolites in 47 GTEx tissues. Color of each cell is proportional to $-\log_{10}$ (BH adjusted *P*-value) from Fisher's exact test [89]. The tissues labeled with '1' indicate those top 1 ranked tissues from the enrichment test.

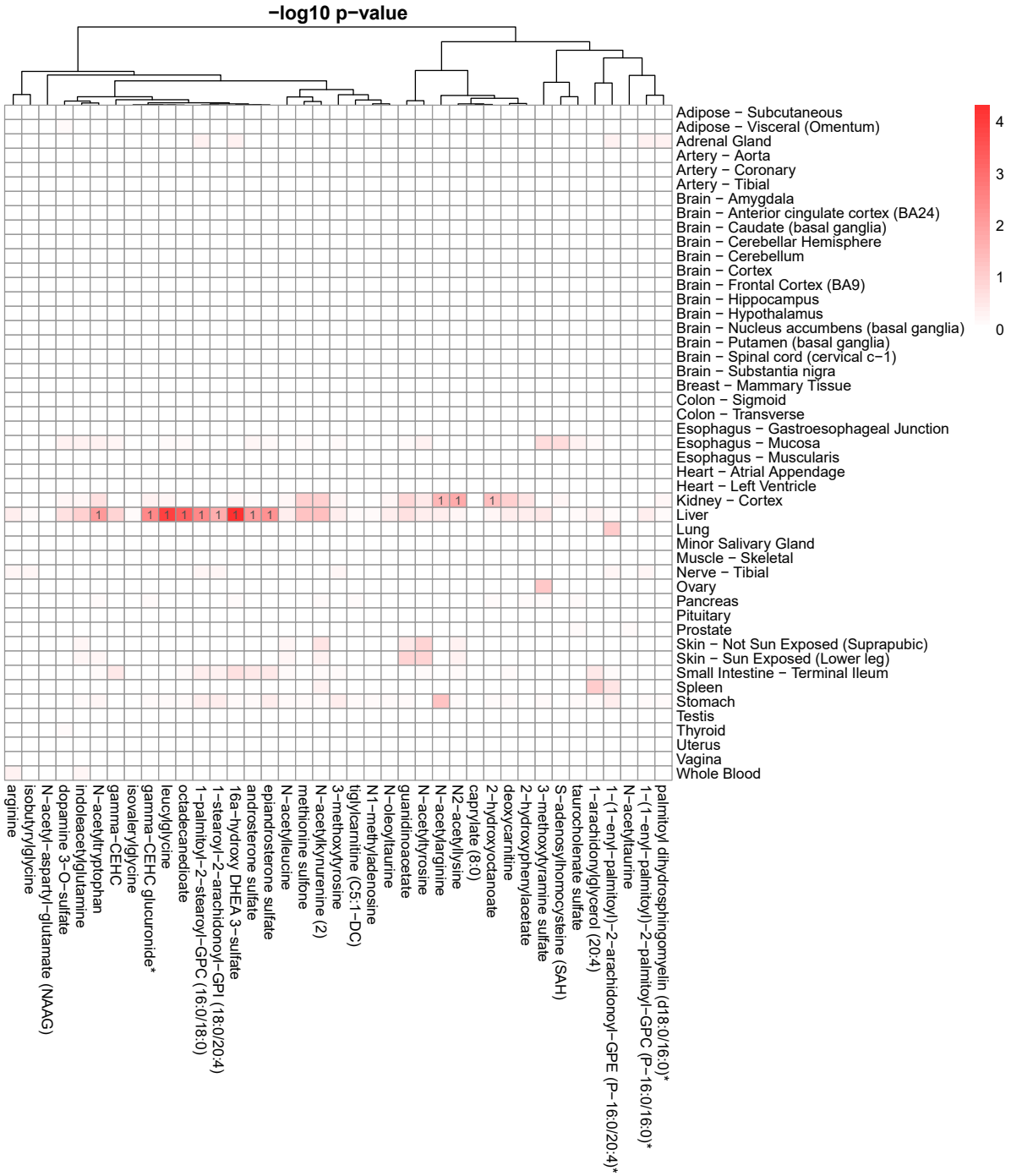
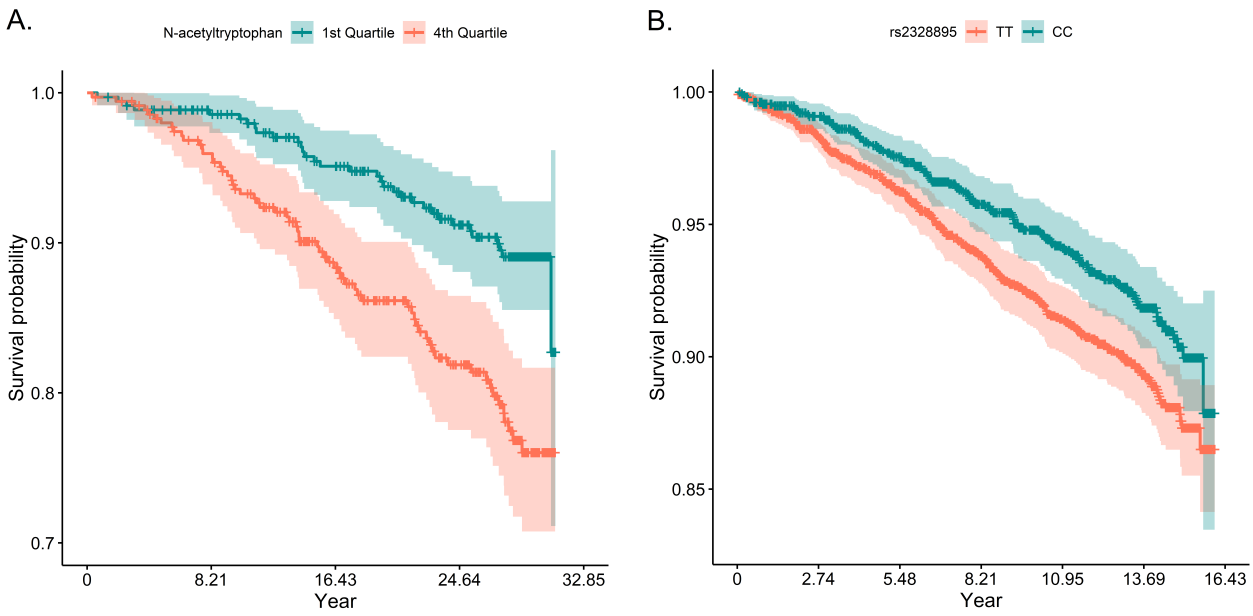


Figure S5. Survival analysis, ARIC, incident CHD: 2002-2017. A. For participants with the levels of N-acetyltryptophan in the 1st and the 4th quartiles; B. For participants >60 years of age homozygous for rs2328895.



Supplemental Methods

Conditional analysis

For metabolites and metabolite sets that reached genome-wide significance, we performed conditional analysis to identify the driving variants in the associated regions. In each set of the correlated metabolites, we defined metabolite associated genetic regions as containing all statistically significant variants within 500kb from each other. Additionally, 500kb was added to each side of the region to account for linkage disequilibrium, and for each metabolite set overlapping regions were merged. We identified 390 region-metabolite set pairs containing statistically significant variants (**Table S3**). We further merged all the overlapping region-metabolite associations to identify 158 non-overlapping genetic loci, including 497 genetic locus-metabolite pairs containing statistically significant variants (**Table S3**).

In 35 genetic locus-metabolite pairs, only one statistically significant variant (P-value$\leq 1.23 \times 10^{-10}$) was identified; therefore, such variants were considered driving variants and no conditional analysis was performed.

For each of 608 statistically significant independent variant-metabolite association, proportion of variance in corresponding metabolite explained by the variant was calculated using R [1].

Locus-specific Investigations

In HCHS/SOL, prevalent coronary heart disease (CHD) was identified using self-reported medical history and electrocardiogram reports of possible old myocardial infarction (MI), angina, MI, or procedure (angioplasty, stent, and bypass) [2]. Alcohol and tobacco use were assessed using a questionnaire, and participants were categorized as never, former, and current alcohol/tobacco users [3, 4].

In ARIC, information on heart failure (HF) and CHD was obtained at the baseline, and then every year using telephone interviews and hospital medical record review [5]. Individuals were followed up for events from baseline to 31 December 2017, and those who were lost to follow-up were censored at the date of last contact. The diagnosis of HF was based on *International Classification of Diseases, Ninth Revision* (ICD-9) code 428, or ICD-10, code I50 [6]. A CHD event was defined as a validated definite or probable hospitalized MI, a definite CHD death, an unrecognized MI defined by ARIC electrocardiogram readings, or coronary revascularization [7, 8]. Cigarette smoking and alcohol use were self-reported at the baseline, classified as current, never, and former use. Alcohol use was obtained using dietary intake questionnaire [9], while cigarette use was assessed during an interview [10].

For the analysis of N-oleoyl-taurine and rs324420 with smoking and drinking status in HCHS/SOL and ARIC, we considered significant associations reaching the Bonferroni-adjusted P-value<0.006, accounting for 2 outcomes (smoking and drinking), 2 traits tested (N-oleoyl-taurine and rs324420), and 2 cohorts.

Supplemental References

1. Shim H, Chasman DI, Smith JD, Mora S, Ridker PM, Nickerson DA, Krauss RM and Stephens M. A multivariate genome-wide association analysis of 10 LDL subfractions, and their response to statin treatment, in 1868 Caucasians. *PLoS One*. 2015;10:e0120758.
2. Gonzalez HM, Tarraf W, Rodriguez CJ, Gallo LC, Sacco RL, Talavera GA, Heiss G, Kizer JR, Hernandez R, Davis S, Schneiderman N, Daviglus ML and Kaplan RC. Cardiovascular health among diverse Hispanics/Latinos: Hispanic Community Health Study/Study of Latinos (HCHS/SOL) results. *Am Heart J*. 2016;176:134-44.
3. HCHS/SOL V2-Alcohol Use QUESTIONNAIRE. https://sites.cscc.unc.edu/hchs/system/files/forms/ALE_QXQ.pdf. Accessed September 26, 2019. 2014.
4. HCHS/SOL V2-Tobacco Use. https://sites.cscc.unc.edu/hchs/system/files/forms/TBE_QXQ.pdf. Accessed September 26, 2019. 2014.
5. The Atherosclerosis Risk in Communities (ARIC) Study: design and objectives. The ARIC investigators. *Am J Epidemiol*. 1989;129:687-702.
6. Yamagishi K, Folsom AR, Rosamond WD, Boerwinkle E and Aric Investigators. A genetic variant on chromosome 9p21 and incident heart failure in the ARIC study. *Eur Heart J*. 2009;30:1222-8.
7. Barbalic M, Reiner AP, Wu C, Hixson JE, Franceschini N, Eaton CB, Heiss G, Couper D, Mosley T and Boerwinkle E. Genome-wide association analysis of incident coronary heart disease (CHD) in African Americans: a short report. *PLoS Genet*. 2011;7:e1002199.
8. White AD, Folsom AR, Chambless LE, Sharret AR, Yang K, Conwill D, Higgins M, Williams OD and Tyroler HA. Community surveillance of coronary heart disease in the Atherosclerosis Risk in Communities (ARIC) Study: methods and initial two years' experience. *J Clin Epidemiol*. 1996;49:223-33.
9. Williams OD, Stinnett S, Chambless LE, Boyle KE, Bachorik PS, Albers JJ and Lippel K. Populations and methods for assessing dyslipoproteinemia and its correlates. The Lipid Research Clinics Program Prevalence Study. *Circulation*. 1986;73:I4-11.
10. Home Interview Atherosclerosis Risk In Communities Study. https://sites.cscc.unc.edu/aric/sites/default/files/public/forms/HOMA.pdf. Accessed April 4, 2019. 1987.

Supplemental Acknowledgments

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.