The American Journal of Human Genetics, Volume *109*

Supplemental information

Predicted gene expression in ancestrally diverse populations leads to discovery of susceptibility loci for lifestyle and cardiometabolic traits

Heather M. Highland, Genevieve L. Wojcik, Mariaelisa Graff, Katherine K. Nishimura, Chani J. Hodonsky, Antoine R. Baldassari, Alanna C. Cote, Iona Cheng, Christopher R. Gignoux, Ran Tao, Yuqing Li, Eric Boerwinkle, Myriam Fornage, Jeffrey Haessler, Lucia A. Hindorff, Yao Hu, Anne E. Justice, Bridget M. Lin, Danyu Lin, Daniel O. Stram, Christopher A. Haiman, Charles Kooperberg, Loic Le Marchand, Tara C. Matise, Eimear E. Kenny, Christopher S. Carlson, Eli A. Stahl, Christy L. Avery, Kari E. North, Jose Luis Ambite, Steven Buyske, Ruth J. Loos, Ulrike Peters, Kristin L. Young, Stephanie A. Bien, and Laura M. Huckins

Figure S1 A

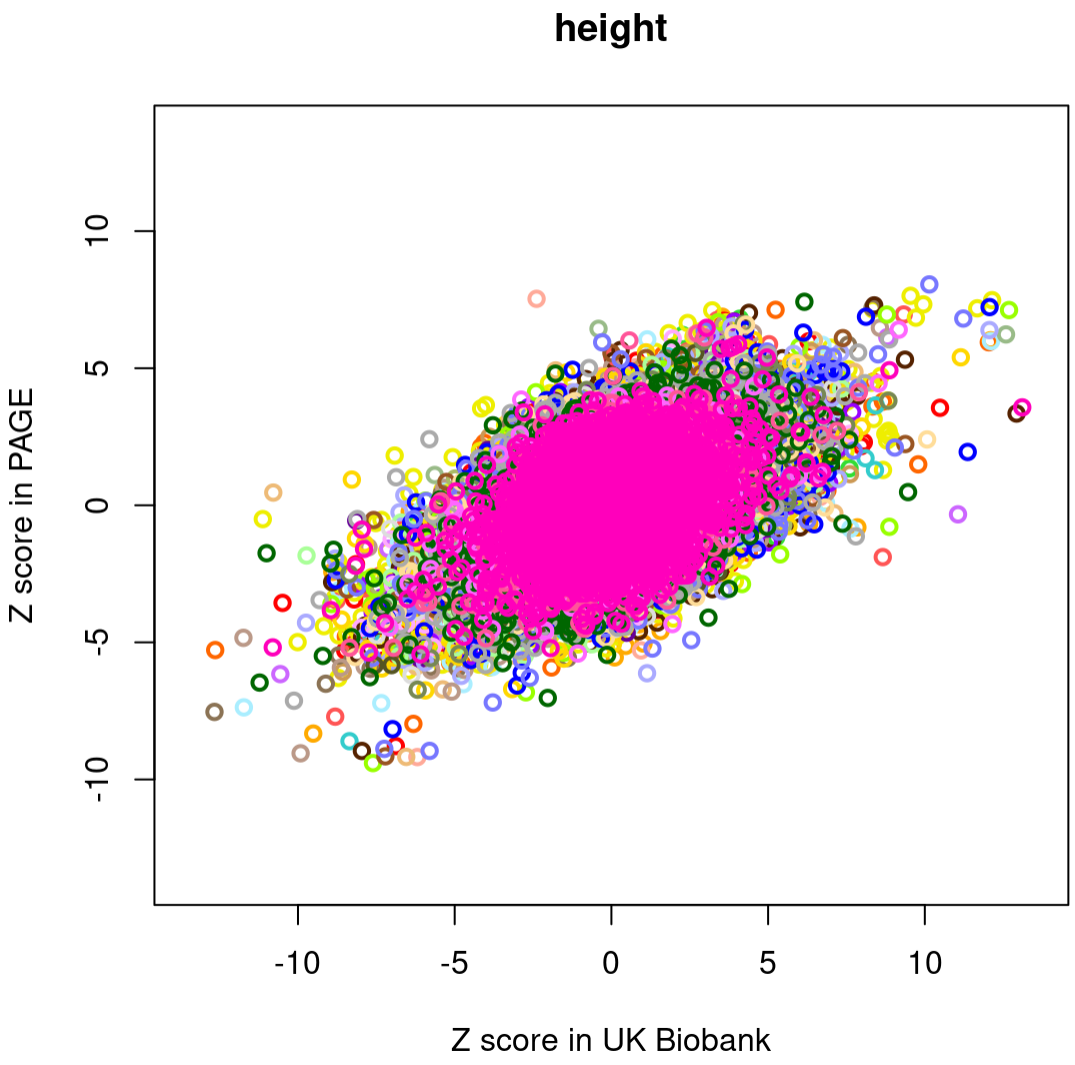


Figure S1 B

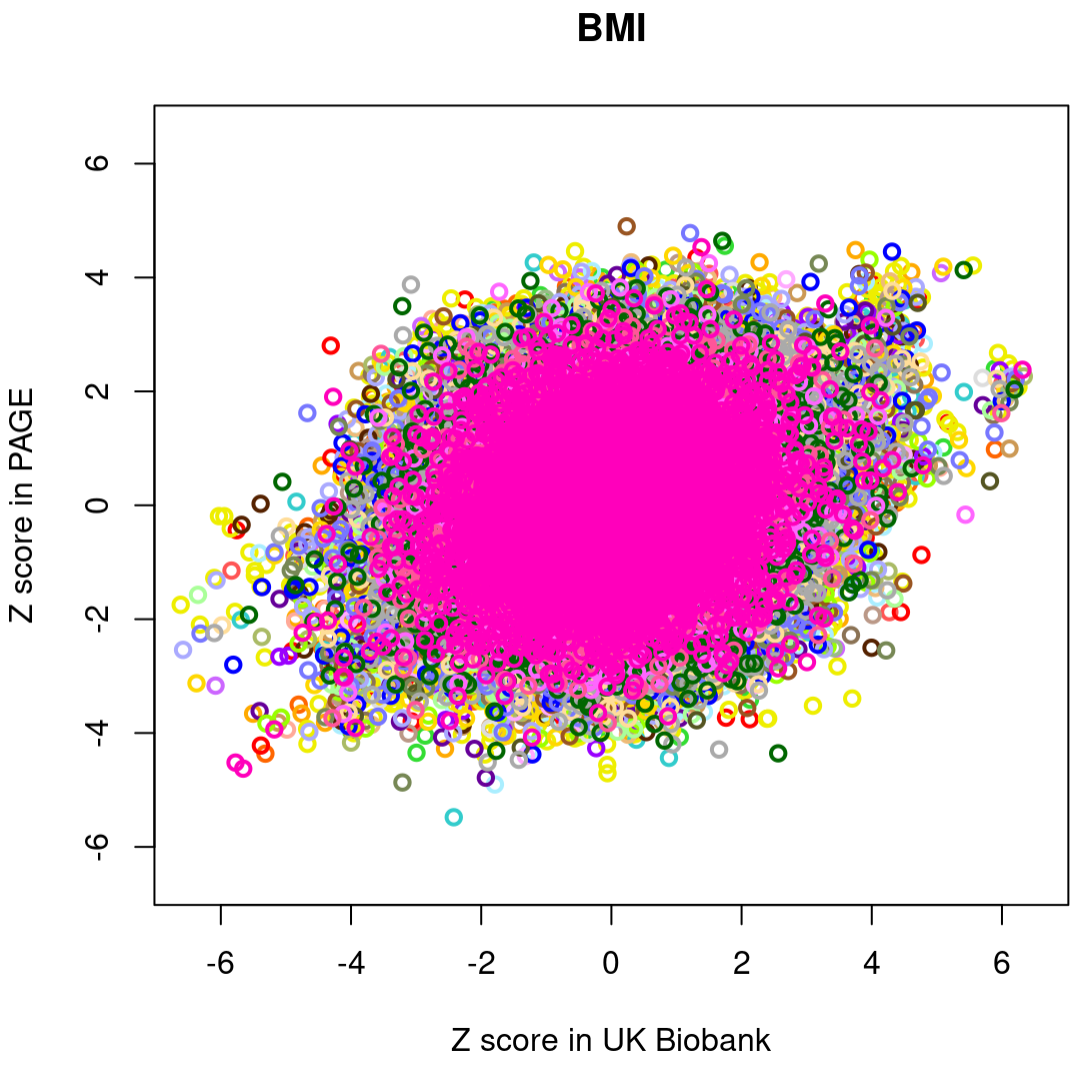
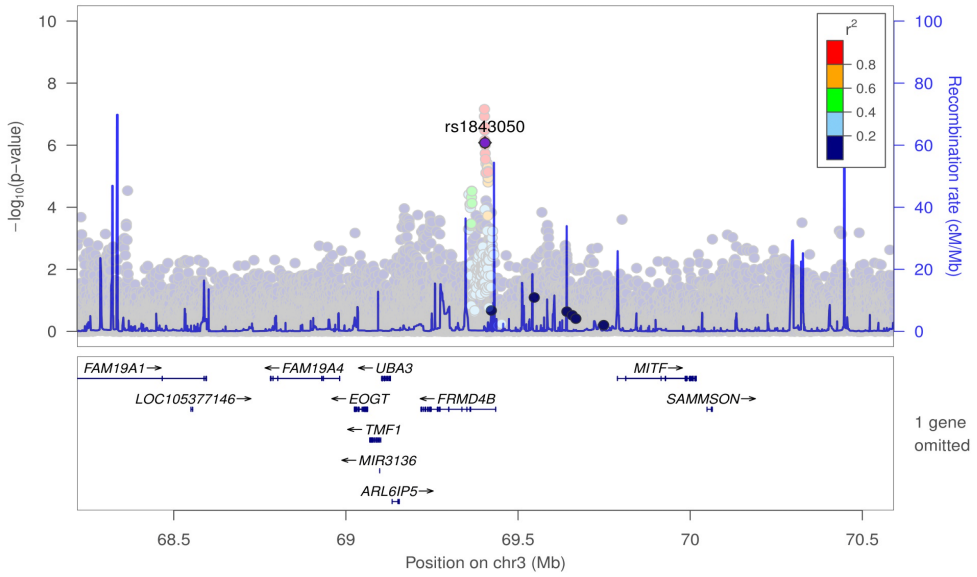
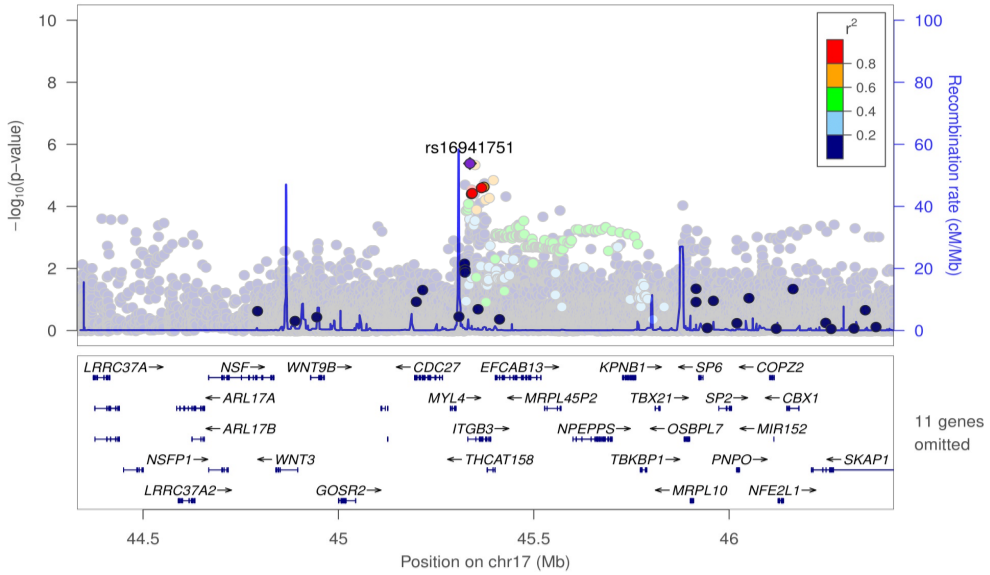


Figure S1 Comparison of Effects between PAGE and $UKBB_{50k}$ results. Z-scores for each tissue specific GReX association in S-PrediXcan for $UKBB_{50k}$ (x-axis) and PrediXcan association in PAGE (y-axis) for height (A) and BMI (B). Dot colors represent standard GTEx tissues colors.

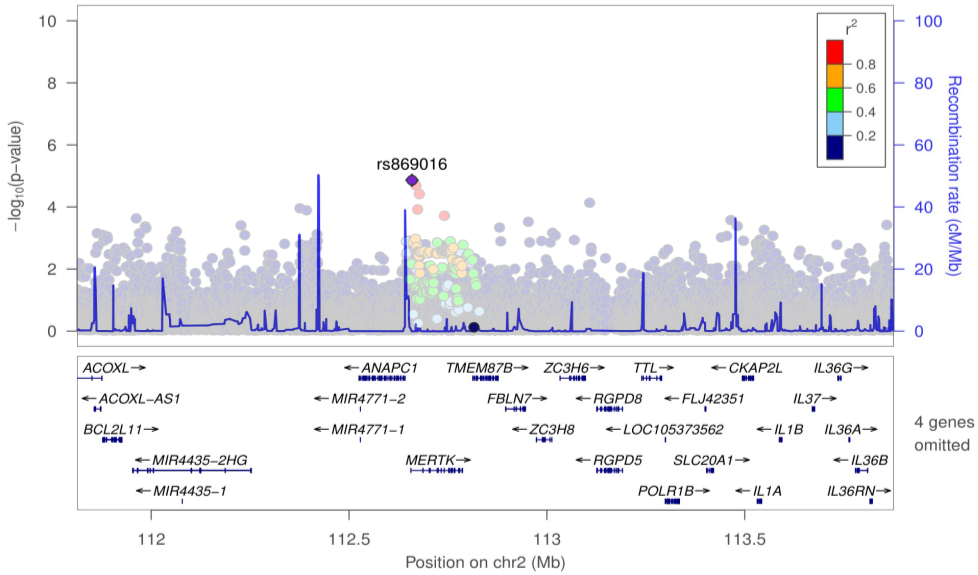
A) PR interval *FRMD4B* – Whole blood



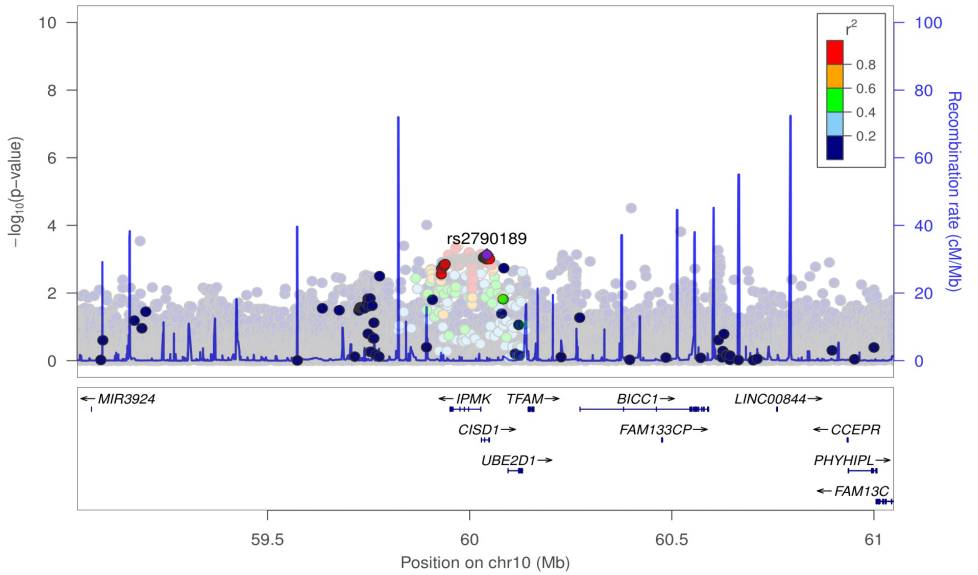
B) PR interval *ITGB3* – Heart left ventricle



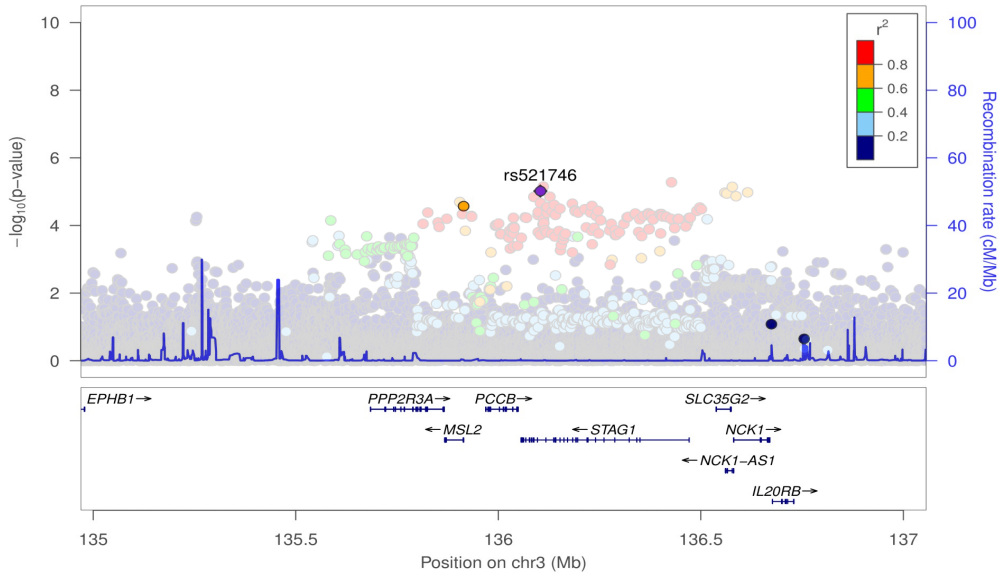
C) PR interval *TMEM87B* – Adrenal gland



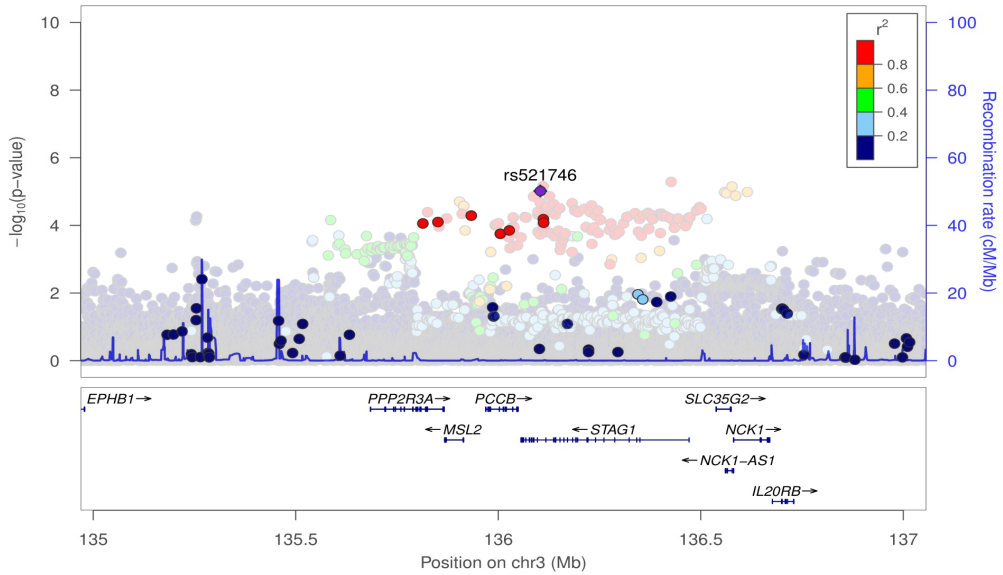
D) QRS interval *CISD1* – Whole blood



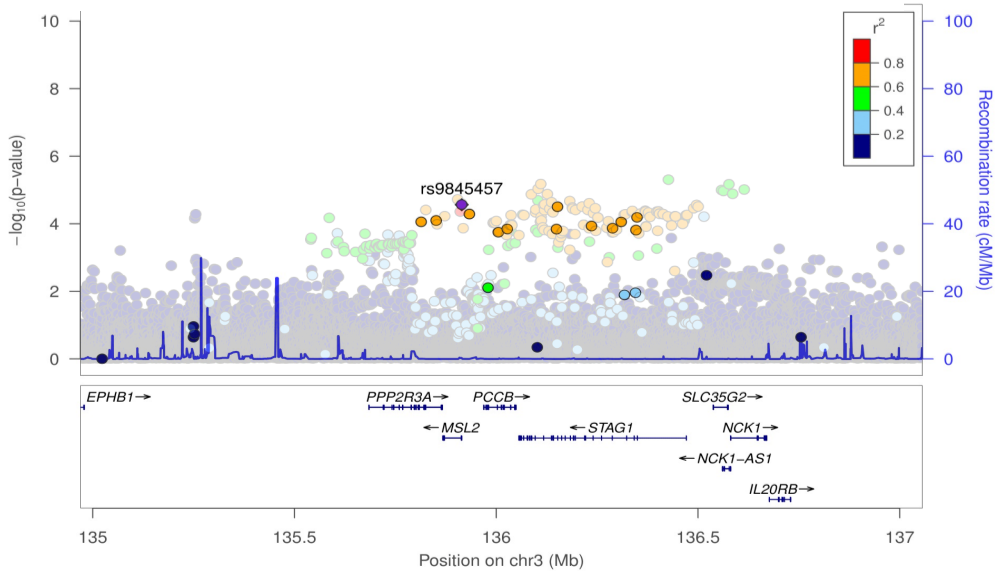
E1) QT interval *PCCB* – Whole blood



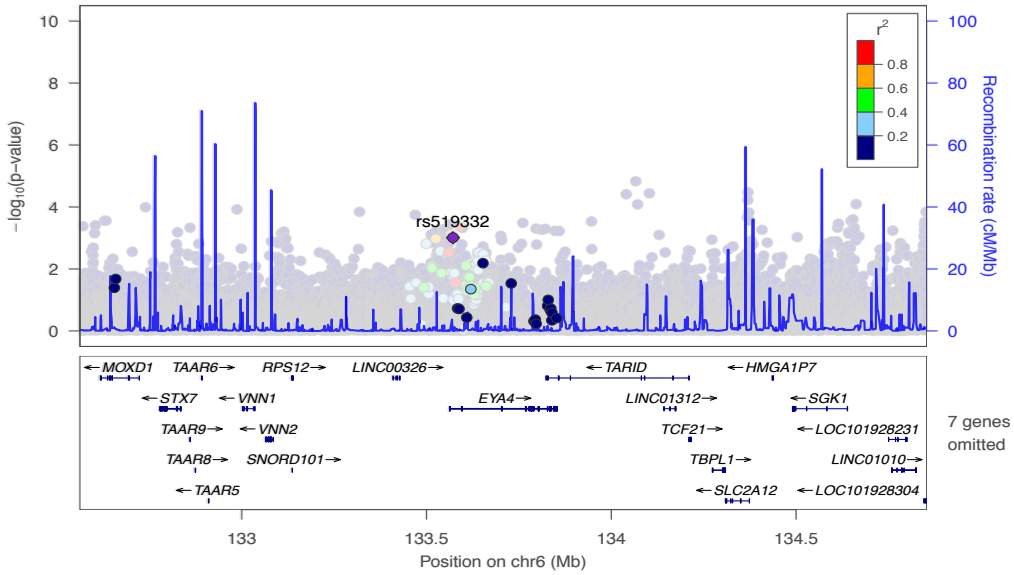
E2) QT interval *PCCB* – Heart atrial appendage



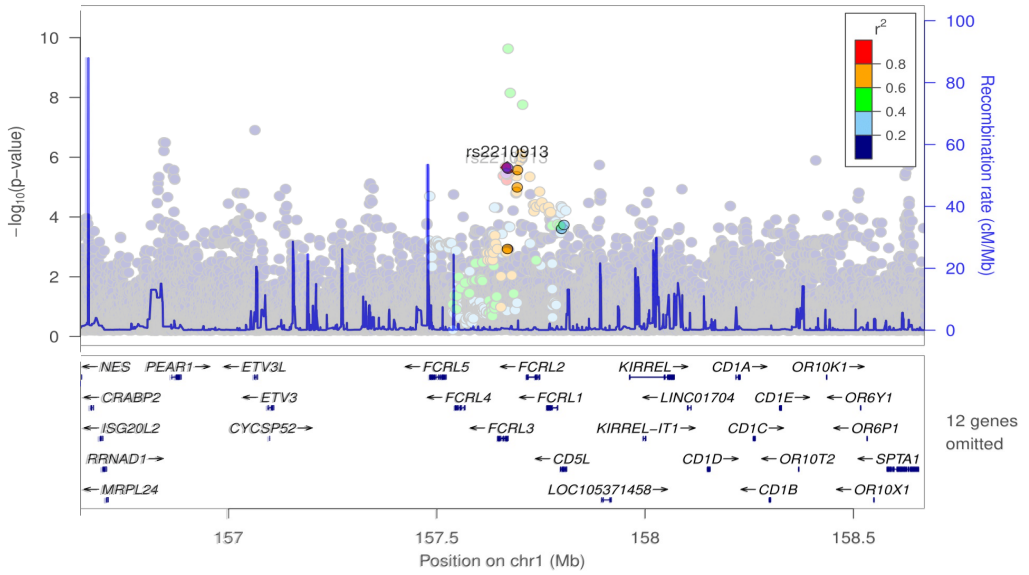
E3) QT interval *PCCB* – Artery tibial



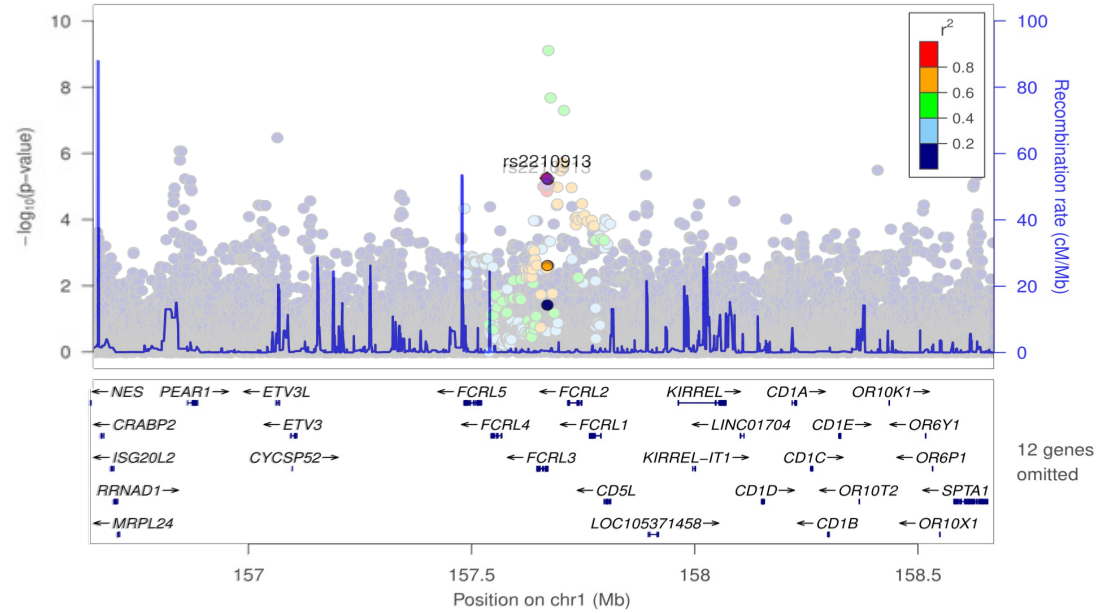
F) QT interval *EYA4* – Artery Aorta



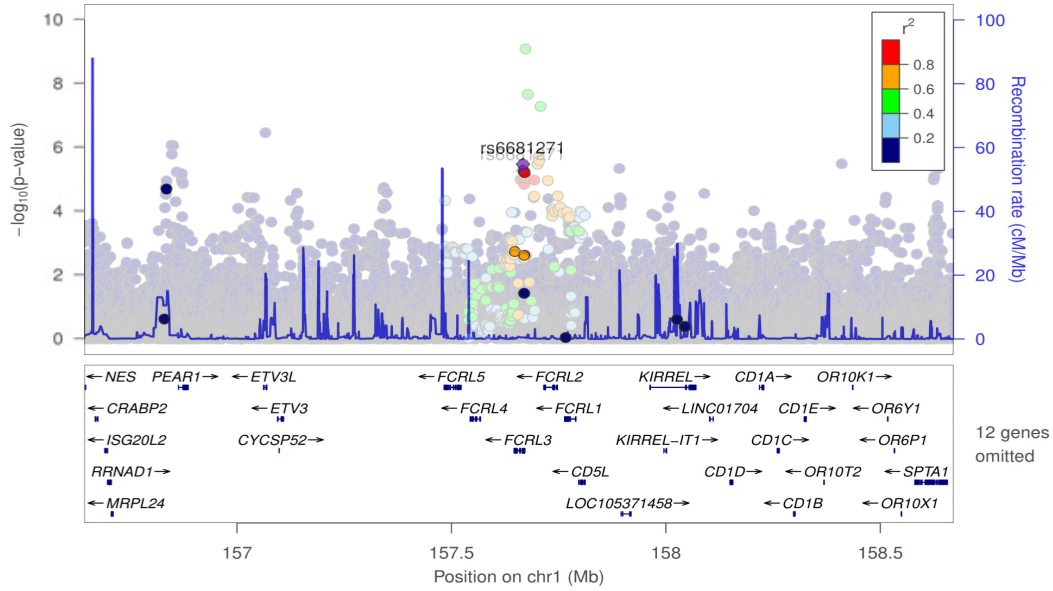
G1) CRP *FCRL3* – Liver



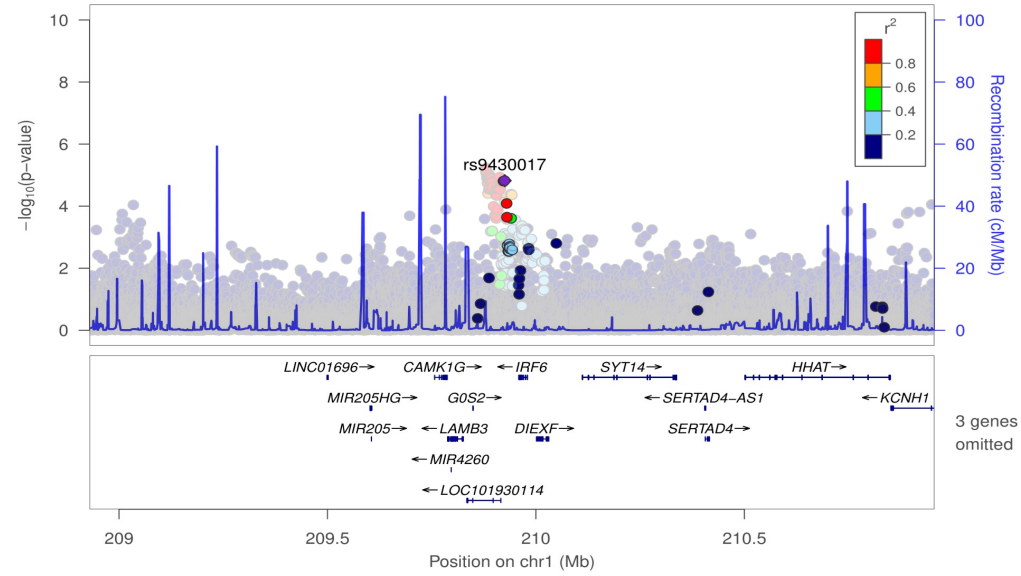
G2) CRP *FCRL3* – Spleen



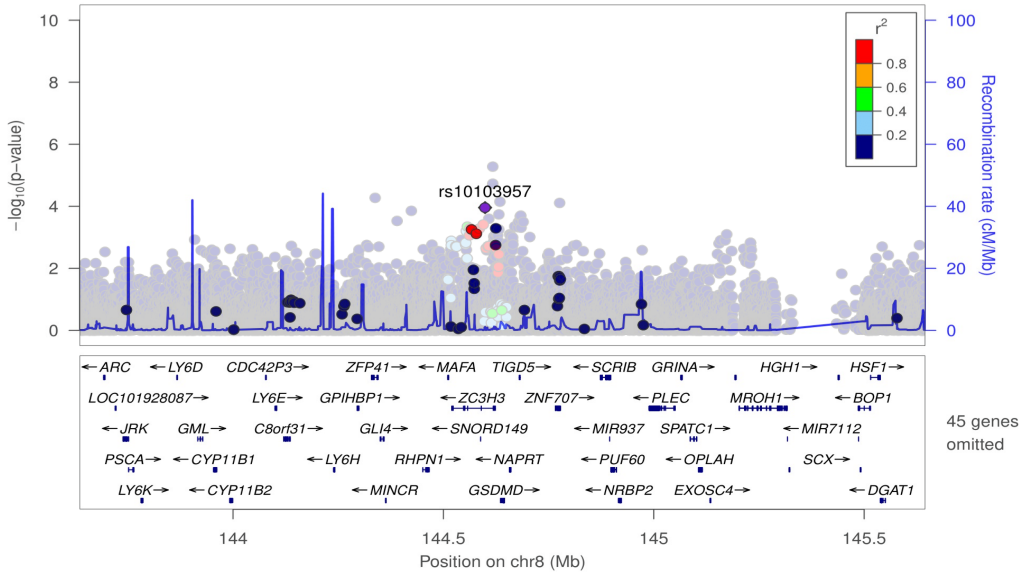
G3) CRP *FCRL3* – Whole blood



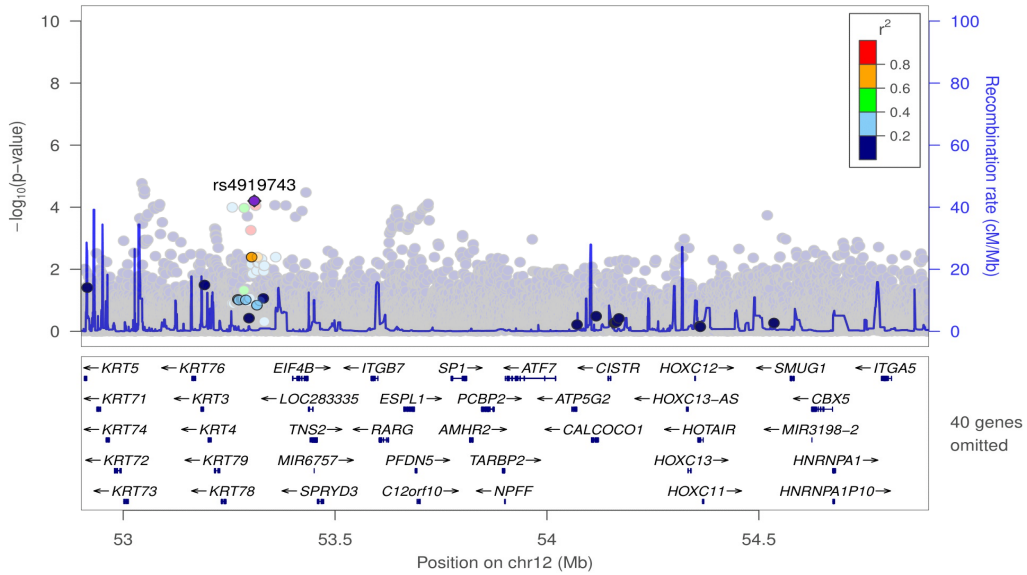
H) CRP *TRAF3IP3* – Whole blood



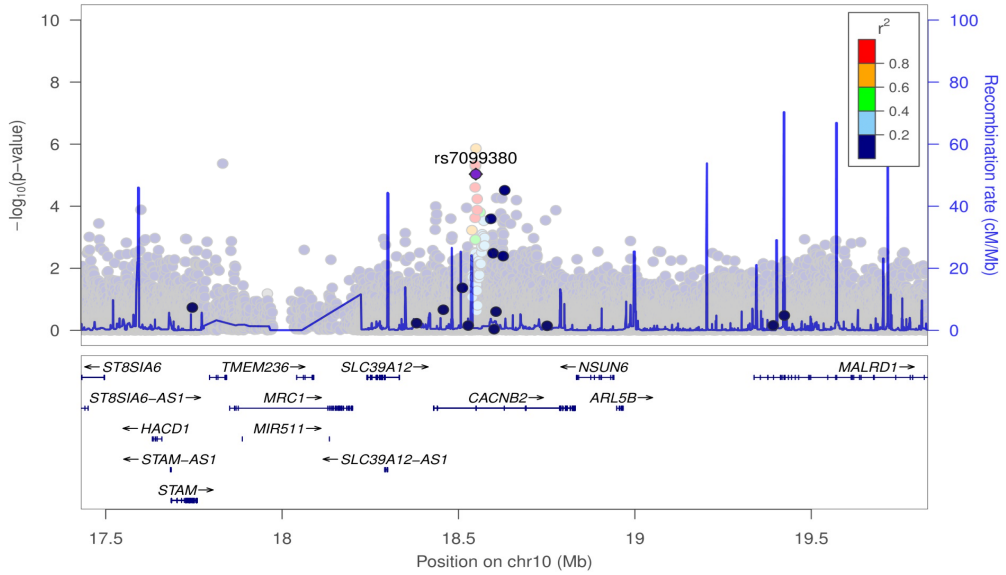
I) CRP *GSDMD* – Whole blood



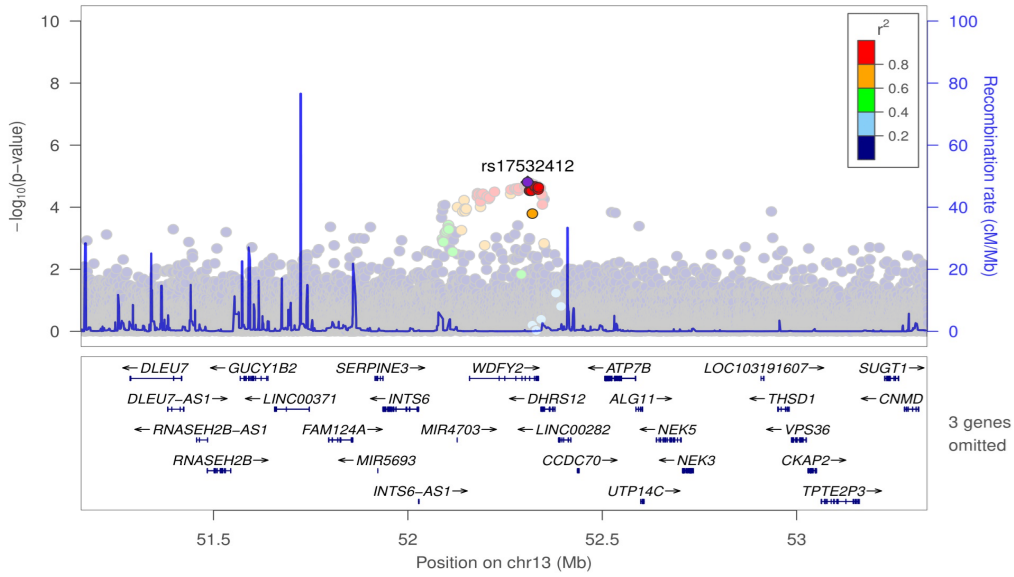
J) eGFR *NPFF* – Artery tibial



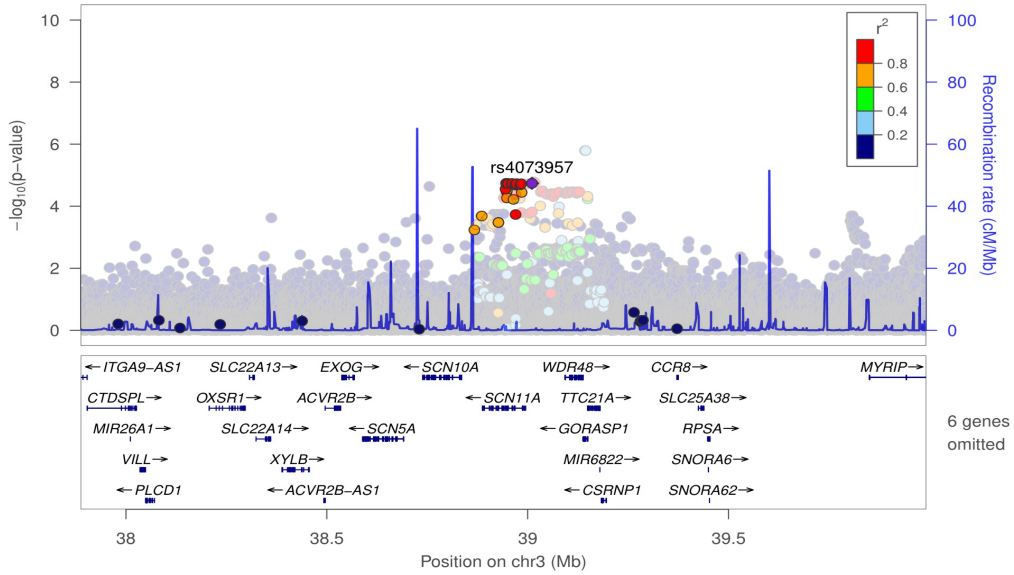
K) BMI *CACNB2* – Muscle Skeletal



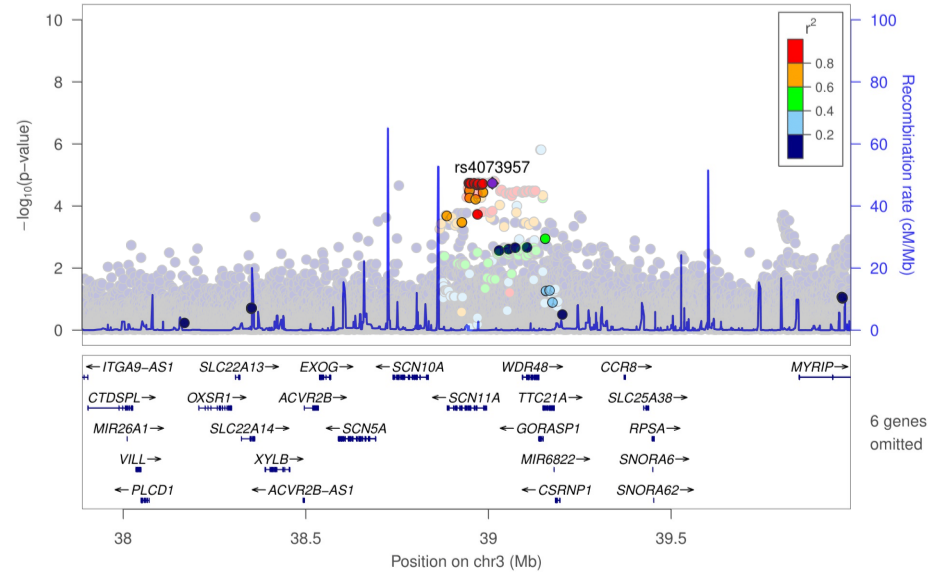
L) BMI *WDFY2* – Adipose Visceral Omentum



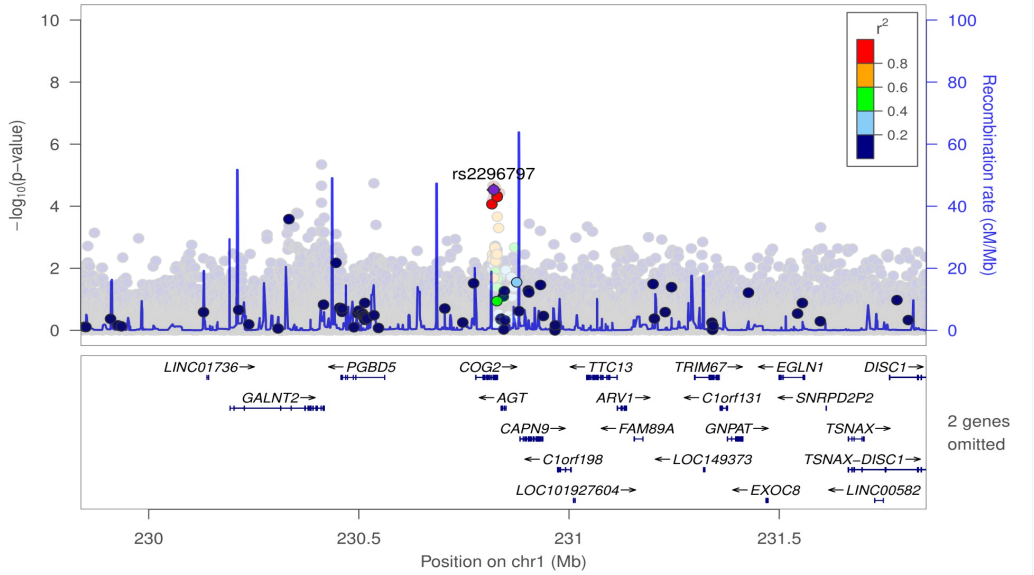
M1) Triglycerides *SCN11A* – Adipose Subcutaneous



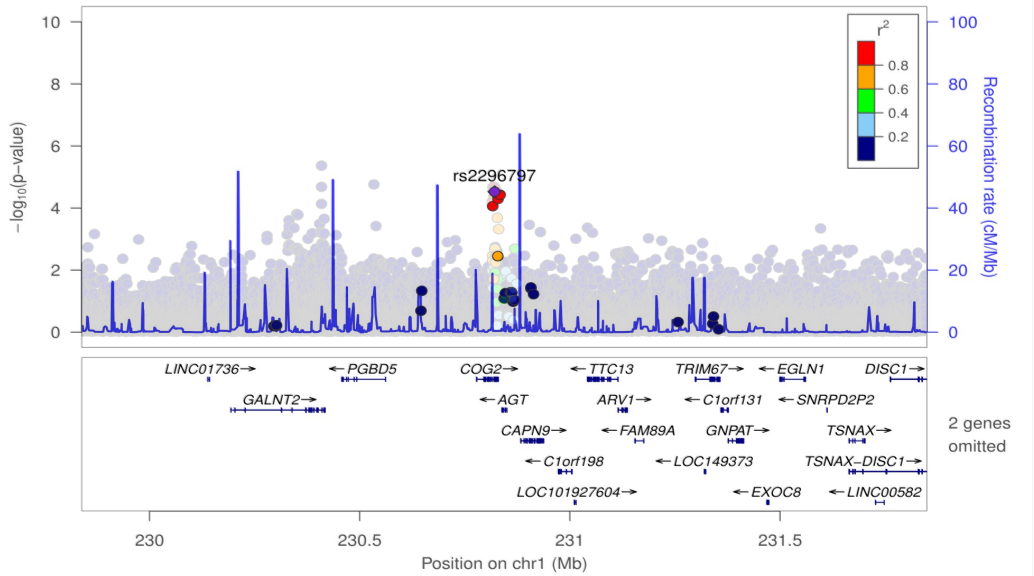
M2) Triglycerides *SCN11A* – Adipose Visceral Omentum



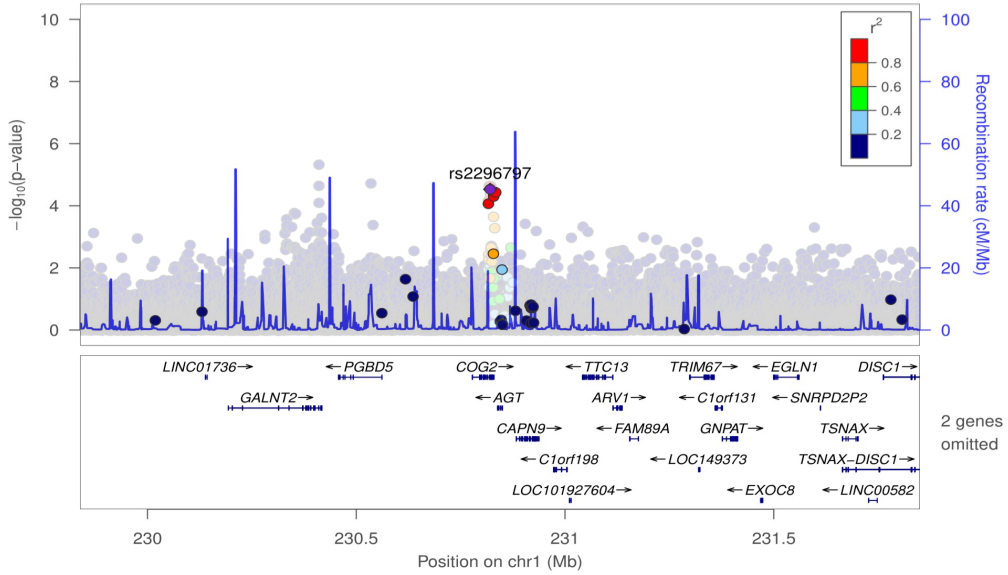
N1) Cups of coffee per day *AGT* – Brain Nucleus accumbens basal ganglia



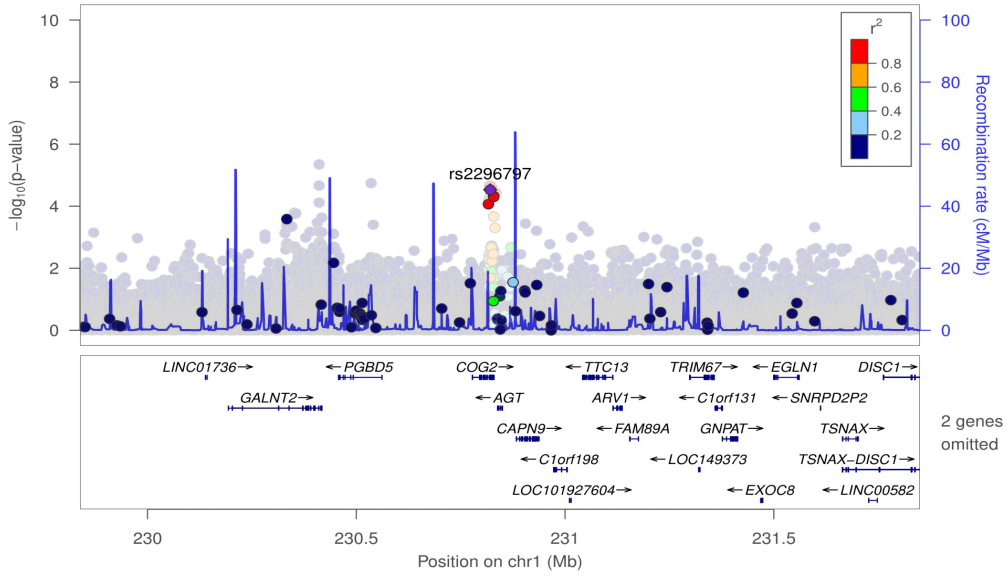
N2) Cups of coffee per day *AGT* – Brain Anterior cingulate cortex BA24



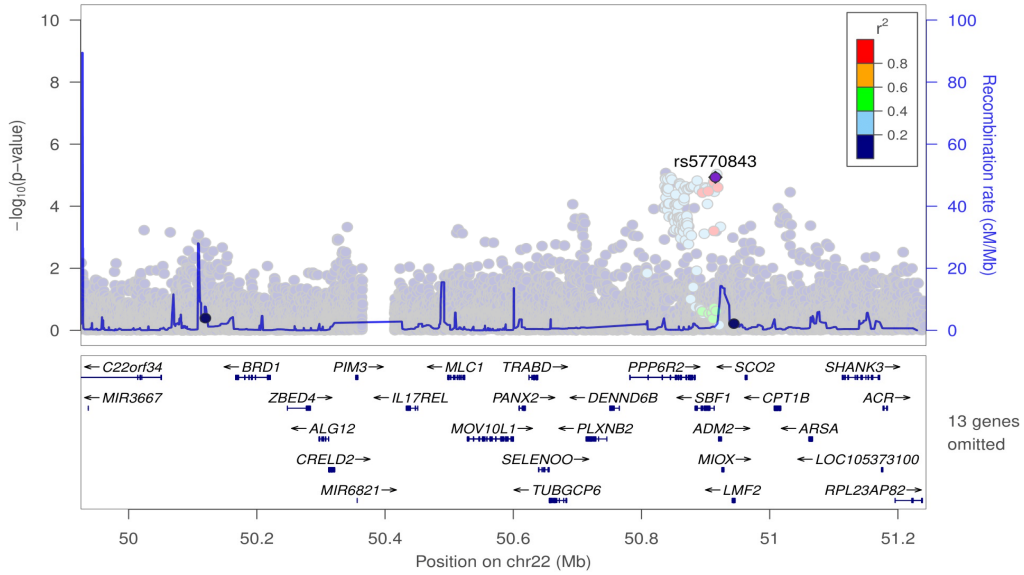
N3) Cups of coffee per day *AGT* – Brain Cortex



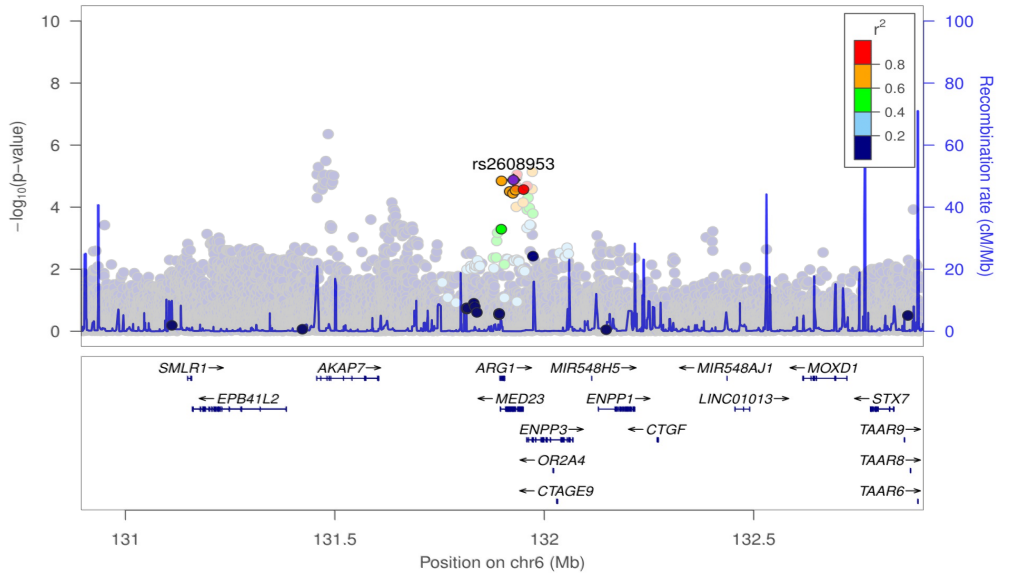
N4) Cups of coffee per day *AGT* – Brain Putamen basal ganglia



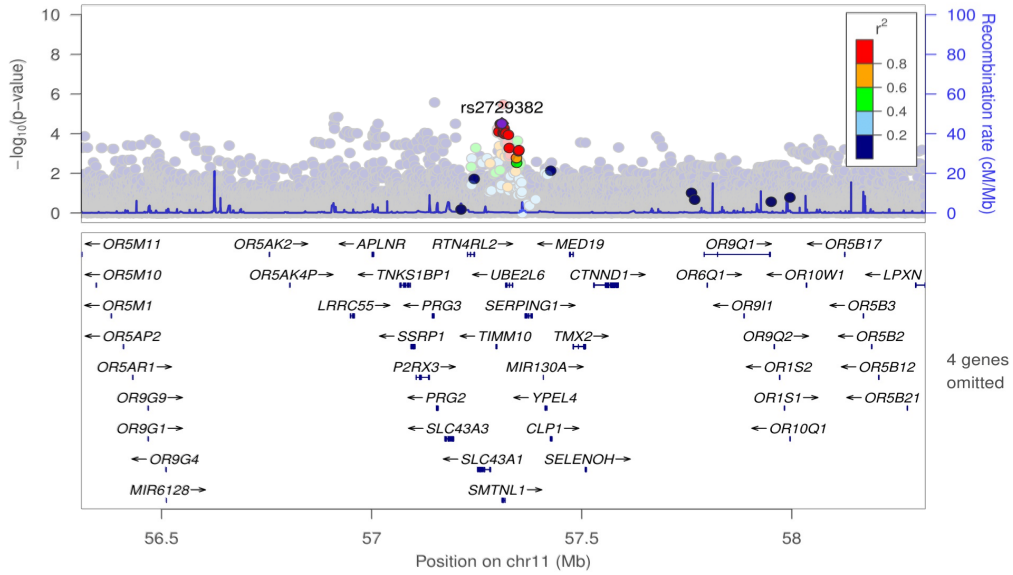
O) Hemoglobin A1c (adjusted for BMI) *MIOX* – Pancreas



P) Type 2 Diabetes (adjusted for BMI) *ARG1* – Whole blood



Q1) Fasting Glucose (adjusted for BMI) *SMTNL1* – Adipose Visceral Omentum



Q2) Fasting Glucose (adjusted for BMI) *SMTNL1* –Whole blood

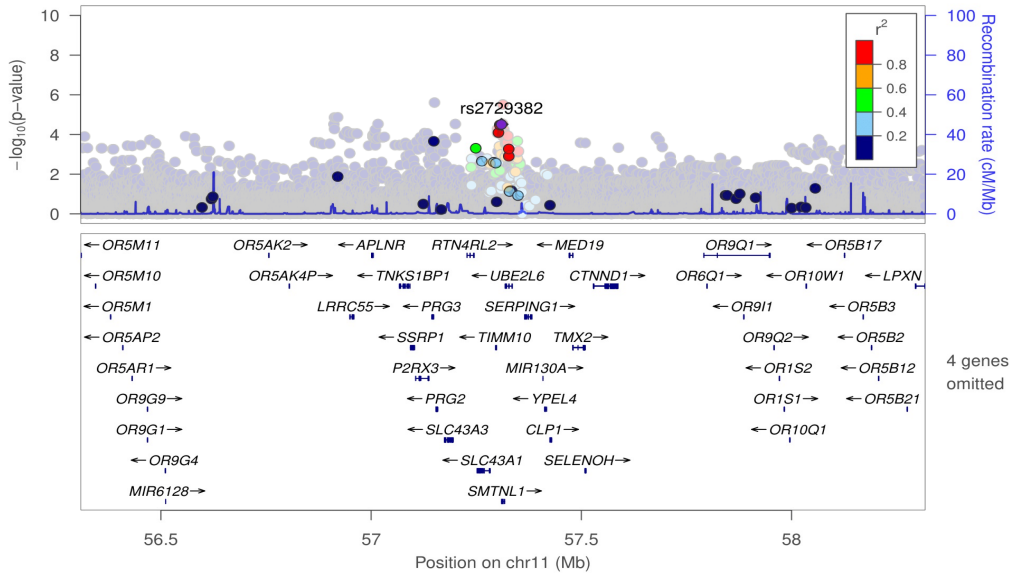


Figure S2. Single variant associations in the combined PAGE samples near novel loci. Variants included in the PrediXcan model are opaque, other variants are translucent. Coloring of points reflects average LD across the PAGE samples. A) *FRMD4B* expression in whole blood associated with PR interval, B) *ITGB3* expression in heart left ventricle associated with PR interval, C) *TMEM87B* expression in adrenal gland associated with PR interval, D) *CISD1* expression in whole blood associated with QRS interval, E1-3) *PCCB* expression in whole blood, heart atrial appendage, and tibial artery, respectively, associated with QT interval, F) *EYA4* expression in aortic artery associated with QT interval, G1-3) *FCRL3* expression in liver, spleen, and whole blood, respectively, associated with C-reactive protein (CRP), H) *TRAF3IP3* expression in whole blood associated with CRP, I) *GSDMD* expression in whole blood associated with CRP, J) *NPFF* expression in tibial artery associated with estimated glomerular filtration rate (eGFR), K) *CACNB2* expression in skeletal muscle associated with BMI, L) *WDFY2* expression in visceral omentum adipose associated with BMI, M1-2) *SCN11A* expression in subcutaneous adipose, and visceral omentum adipose, respectively associated with triglycerides, N1-4) *AGT* expression in brain nucleus accumbens basal ganglia, brain anterior cingulate cortex BA24, brain cortex, and brain putamen basal ganglia, respectively, associated with cups of coffee per day, O) *MIOX* expression in pancreas associated with hemoglobin A1c adjusted for BMI, P) *ARG1* expression in whole blood associated with Type 2 diabetes adjusted for BMI, and Q1-2) *SMTNL1* expression in visceral omentum adipose and whole blood, respectively, associated with fasting glucose adjusted for BMI.

Supplemental Methods

Data sets

BioMe Biobank: The Charles Bronfman Institute for Personalized Medicine at Mount Sinai Medical Center (MSMC), BioMe™ BioBank (BioMe) is an EMR-linked biorepository drawing from more than 70,000 inpatients and 800,000 outpatients annually [1]. The MSMC serves diverse local communities of upper Manhattan, including Central Harlem (86% African American), East Harlem (88% Hispanic/Latino), and the Upper East Side (88% European American) with broad health disparities. BioMe™ enrolled over 26,500 participants between September 2007 and August 2013, with 25% African American, 36% Hispanic/Latino (primarily of Caribbean origin), 30% European American, and 9% of Other ancestry. The BioMe™ population reflects community-level disease burdens and health disparities with broad public health impact. Biobank operations are fully integrated in clinical care processes, with direct recruitment from clinical sites, waiting areas, and phlebotomy stations by dedicated Biobank recruiters independent of clinical care providers, prior to or following a clinician standard of care visit. Recruitment currently occurs at over 30 clinical care sites. BioMe™ participants of self-reported African American and Hispanic/Latino ancestry were included in this analysis. (dbGaP study accession number: phs000925).

HCHS/SOL: The Hispanic Community Health Study / Study of Latinos (HCHS/SOL) is a prospective, community-based cohort study of 16,415 self-identified Hispanic/Latino adults from Miami, FL, San Diego, CA, Bronx, NY, and Chicago, IL. The goals of HCHS/SOL are to identify risk factors for chronic conditions including CVD, diabetes, pulmonary disease, and sleep disorders. The HCHS/SOL survey design consists of a two-stage probability sample of households at each recruitment center [2,3]. Census block groups were selected in defined communities near each center, and then households were sampled within block groups. Households with HL surnames and residents aged 45+ years were oversampled to increase representation of the target population and achieve a more uniform age distribution. Sampling weights were then calculated for each individual. Participants aged 18-74 years underwent extensive psychosocial and clinical assessments between 2008-2011. A re-examination of the HCHS/SOL cohort was conducted between 2015-2017. Annual telephone follow-up interviews are ongoing since study inception to determine health outcomes of interest. (dbGaP study accession number: phs000555).

MEC: The Multiethnic Cohort (MEC) is a population-based prospective cohort study including approximately 215,000 ancestrally diverse adults from Hawaii and California. All participants were 45-75 years of age at baseline, and primarily of 5 ancestries: Japanese Americans, African Americans, European Americans, Hispanic/Latinos, and Native Hawaiians [4,5]. MEC was designed to examine lifestyle risk factors and genetic susceptibility to cancer. All eligible cohort members completed baseline and follow-up questionnaires, and has case control data related to breast, prostate, and colorectal cancer, diabetes, and obesity; common traits that are risk factors for these diseases (e.g., body mass index / weight, waist-to-hip ratio, height), and relevant disease-associated biomarkers (e.g., fasting insulin and lipids, steroid hormones). For this project, MEC contributed African American, Japanese American, and Native Hawaiian samples. (dbGaP study accession number: phs000220).

WHI: The Women's Health Initiative (WHI) is a long-term, prospective, multicenter cohort study investigating post-menopausal women's health in the US [6]. WHI designed to study strategies to prevent heart disease, breast cancer, colon cancer, and osteoporotic fractures in women 50-79 years of age. WHI involves 161,808 women recruited between 1993 and 1998 at 40 centers across the US. The study consists of two parts: the WHI Clinical Trial which was a randomized clinical trial of hormone therapy, dietary modification, and calcium/Vitamin D supplementation; and the WHI Observational Study, which focused on many of the inequities in women's health research and provided practical information about incidence, risk factors, and interventions related to heart disease, cancer, and osteoporotic fractures. For this project, women who self-identified as non-European were included in the analysis. (dbGaP study accession number: phs000227).

Supplemental Acknowledgements

The data and materials included in this report result from collaboration between the following studies and organizations:

CALiCo

Funding support for the Genetic Epidemiology of Causal Variants Across the Life Course (CALiCo) program was provided through the NHGRI PAGE program (U01 HG007416, U01 HG004803). The following studies contributed to this manuscript and are funded by the following agencies: The Atherosclerosis Risk in Communities (ARIC) Study is carried out as a collaborative study supported by National Heart, Lung, and Blood Institute contracts (HHSN268201100005C, HHSN268201100006C, HHSN268201100007C, HHSN268201100008C, HHSN268201100009C, HHSN268201100010C, HHSN268201100011C, and HHSN268201100012C). The Hispanic Community Health Study/Study of Latinos was carried out as a collaborative study supported by contracts from the National Heart, Lung, and Blood Institute (NHLBI) to the University of North Carolina (N01-HC65233), University of Miami (N01-HC65234), Albert Einstein College of Medicine (N01-HC65235), Northwestern University (N01-HC65236), and San Diego State University (N01-HC65237). The following Institutes/Centers/Offices contribute 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.The Coronary Artery Risk Development in Young Adults Study (CARDIA) is conducted and supported by the National Heart, Lung, and Blood Institute (NHLBI) in collaboration with the University of Alabama at Birmingham (HHSN268201800005I & HHSN268201800007I), Northwestern University (HHSN268201800003I), University of Minnesota (HHSN268201800006I), and Kaiser Foundation Research Institute (HHSN268201800004I). Genotyping was funded as part of the NHLBI Candidate-gene Association Resource (N01-HC-65226) and the NHGRI Gene Environment Association Studies (GENEVA) (U01-HG004729, U01-HG04424, and U01-HG004446). The Strong Heart Study

(SHS) is supported by NHLBI grants U01 HL65520, U01 HL41642, U01 HL41652, U01 HL41654, U01 HL65521 and R01 HL109301. The datasets used for the analyses described in this manuscript were obtained from dbGaP under accession phs000223 (ARIC), phs000236, (CARDIA), phs000301 (CHS), phs000555 (HCHS/SOL).

IPM Biobank

Samples and data of The Charles Bronfman Institute for Personalized Medicine (IPM) BioMe Biobank used in this study were provided by The Charles Bronfman Institute for Personalized Medicine at the Icahn School of Medicine at Mount Sinai (New York). Phenotype data collection was supported by The Andrea and Charles Bronfman Philanthropies. Funding support for the Population Architecture Using Genomics and Epidemiology (PAGE) IPM BioMe Biobank study was provided through the National Human Genome Research Institute (U01 HG007417). The datasets used for the analyses described in this manuscript were obtained from dbGaP under accession phs000925.

MEC

The Multiethnic Cohort study (MEC) characterization of epidemiological architecture is funded through the NHGRI Population Architecture Using Genomics and Epidemiology (PAGE) program (U01 HG007397, U01HG004802 and its NHGRI ARRA supplement). The MEC study is funded through the National Cancer Institute (R37CA54281,R01 CA63, P01CA33619, U01CA136792, and U01CA98758). The datasets used for the analyses described in this manuscript were obtained from dbGaP under accession phs000220.

WHI

Funding support for the “Exonic variants and their relation to complex traits in minorities of the WHI ” study is provided through the NHGRI PAGE program (U01HG007376, U01HG004790). The WHI program is funded by the National Heart, Lung, and Blood Institute, National Institutes of Health, U.S. Department of Health and Human Services through contracts HHSN268201100046C, HHSN268201100001C, HHSN268201100002C, HHSN268201100003C, HHSN268201100004C, and HHSN271201100004C. The authors thank the WHI investigators and staff for their dedication, and the study participants for making the program possible. The datasets used for the analyses described in this manuscript were obtained from dbGaP under accession phs000227. A listing of WHI investigators can be found at: https://www.whi.org/researchers/Documents%20%20Write%20a%20Paper/WHI%20Investigator%20Short%20List.pdf

Stanford Global Reference Panel

The Stanford Global Reference Panel comprises multiple datasets from multiple researchers across the world designed to provide a resource for any researchers interested in diverse population data on the Multi-Ethnic Global Array (MEGA), funded by the NHGRI PAGE program (U01HG007419). The authors thank the researchers and research participants who made this dataset available to the community. The specific datasets are:

Mexico: Samples of indigenous origin in Oaxaca were kindly provided by Drs. Karla Sandoval Mendoza, Samuel Canizales Quinteros, and Victor Acuña Alonzo.

Peru: Individuals from a primarily Quechuan and Aymaran-speaking community in Puno, Peru were kindly provided by Drs. Julie Baker and Carlos Bustamante, with funding support from the Burroughs Welcome Fund.

Rapa Nui (Easter Island): Samples were kindly provided by Drs. Karla Sandoval Mendoza and Andres Moreno Estrada with funding from the Charles Rosenkranz Prize for Health Care Research in Developing Countries.

South Africa: Samples of KhoeSan individuals from the ǂKhomani and Nama communities were kindly provided by Drs. Brenna Henn and Christopher Gignoux with funding from the Morrison Institute for Population and Resource Studies.

Honduras and Colombia: Samples from communities in Honduras and Colombia were kindly provided by Dr. Kathleen Barnes (University of Colorado, Denver), Edwin Herraro-Paz (Universidad Católica de Honduras, San Pedro Sula, Honduras), Alvaro Mayorga (Universidad Católica de Honduras, San Pedro Sula, Honduras), Luis Caraballo (University of Cartagena), Javier Marrugo (university of Cartagena)

Additional global samples: The following datasets are open access and available through the lab website of Carlos Bustamante (https://bustamantelab.stanford.edu/). The Human Genome Diversity Panel (HGDP-CEPH) is a group of cell lines maintained by the Centre d'Étude du Polymorphisme Humain, Fondation Jean Dausset (Paris, France) comprising 52 diverse populations across the world (Africa, Near East, Europe, South Asia, Central Asia, East Asia, Oceania and the Americas). Additional information on these datasets can be found on the CEPH website (http://www.cephb.fr/en/hgdp_panel.php), or originally at http://www.ncbi.nlm.nih.gov/pubmed/11954565 and http://www.ncbi.nlm.nih.gov/pubmed/12493913, with numerous subsequent publications. Samples were filtered to include the H952 unrelated individuals as published here: http://www.ncbi.nlm.nih.gov/pubmed/17044859.

Also available on the Bustamante Lab website is genotype data for the Maasai from Kinyawa, Kenya (MKK) samples maintained by the Coriell Institute for Medical Research (https://catalog.coriell.org/1/NHGRI/Collections/HapMap-Collections/Maasai-in-Kinyawa-Kenya-MKK) and genotyped as part of the International HapMap Project Phase 3(http://hapmap.ncbi.nlm.nih.gov/, http://www.sanger.ac.uk/resources/downloads/human/hapmap3.html) . We have genotyped a subset of unrelated individuals using the filters recommended in http://www.ncbi.nlm.nih.gov/pubmed/20869033.

Supplemental References

1. Gottesman, O., Kuivaniemi, H., Tromp, G., Faucett, W.A., Li, R., Manolio, T.A., Sanderson, S.C., Kannry, J., Zinberg, R., Basford, M.A., et al. (2013). The Electronic Medical Records and Genomics (eMERGE) Network: past, present, and future. Genet. Med. *15*, 761–771.

2. Sorlie, P.D., Avilés-Santa, L.M., Wassertheil-Smoller, S., Kaplan, R.C., Daviglus, M.L., Giachello, A.L., Schneiderman, N., Raij, L., Talavera, G., Allison, M., et al. (2010). Design and implementation of the Hispanic Community Health Study/Study of Latinos. Ann. Epidemiol. *20*, 629–641.

3. Lavange, L.M., Kalsbeek, W.D., Sorlie, P.D., Avilés-Santa, L.M., Kaplan, R.C., Barnhart, J., Liu, K., Giachello, A., Lee, D.J., Ryan, J., et al. (2010). Sample design and cohort selection in the Hispanic Community Health Study/Study of Latinos. Ann. Epidemiol. *20*, 642–649.

4. Lim, U., Ernst, T., Buchthal, S.D., Latch, M., Albright, C.L., Wilkens, L.R., Kolonel, L.N., Murphy, S.P., Chang, L., Novotny, R., et al. (2011). Asian women have greater abdominal and visceral adiposity than Caucasian women with similar body mass index. Nutr. Diabetes *1*, e6.

5. Kolonel, L.N., Henderson, B.E., Hankin, J.H., Nomura, A.M., Wilkens, L.R., Pike, M.C., Stram, D.O., Monroe, K.R., Earle, M.E., and Nagamine, F.S. (2000). A multiethnic cohort in Hawaii and Los Angeles: baseline characteristics. Am. J. Epidemiol. *151*, 346–357.

6. (1998). Design of the Women's Health Initiative clinical trial and observational study. The Women's Health Initiative Study Group. Control. Clin. Trials *19*, 61–109.