Rare coding variants in 35 genes associate with circulating lipid levels—A multi-ancestry analysis of 170,000 exomes

George Hindy,[1,2,3] Peter Dornbos,[4,5,6] Mark D. Chaffin,[1,7,8] Dajiang J. Liu,[9] Minxian Wang,[1,7,10] Margaret Sunitha Selvaraj,[1,7] David Zhang,[11,12] Joseph Park,[11,12] Carlos A. Aguilar-Salinas,[13] Lucinda Antonacci-Fulton,[14,15] Diego Ardissino,[16,17,18] Donna K. Arnett,[19] Stella Aslibekyan,[20] Gil Atzmon,[21,22] Christie M. Ballantyne,[23,24] Francisco Barajas-Olmos,[25] Nir Barzilai,[21] Lewis C. Becker,[26] Lawrence F. Bielak,[27] Joshua C. Bis,[28] John Blangero,[29] Eric Boerwinkle,[30,31] Lori L. Bonnycastle,[32] Erwin Bottinger,[33,34] Donald W. Bowden,[35] Matthew J. Bown,[36,37] Jennifer A. Brody,[28] Jai G. Broome,[38] Noël P. Burtt,[4] Brian E. Cade,[39,40] Federico Centeno-Cruz,[25] Edmund Chan,[41] Yi-Cheng Chang,[42] Yii-Der I. Chen,[43] Ching-Yu Cheng,[44,45,46] Won Jung Choi,[47] Rajiv Chowdhury,[48,49] Cecilia Contreras-Cubas,[25] Emilio J. Córdova,[25] Adolfo Correa,[50]

(Author list continued on next page)

Summary

Large-scale gene sequencing studies for complex traits have the potential to identify causal genes with therapeutic implications. We performed gene-based association testing of blood lipid levels with rare (minor allele frequency < 1%) predicted damaging coding variation by using sequence data from >170,000 individuals from multiple ancestries: 97,493 European, 30,025 South Asian, 16,507 African, 16,440 Hispanic/Latino, 10,420 East Asian, and 1,182 Samoan. We identified 35 genes associated with circulating lipid levels; some of these genes have not been previously associated with lipid levels when using rare coding variation from population-based samples. We prioritize 32 genes in array-based genome-wide association study (GWAS) loci based on aggregations of rare coding variants; three (*EVI5*, *SH2B3*, and *PLIN1*) had no prior association of rare coding variants with lipid levels. Most of our associated genes showed evidence of association among multiple ancestries. Finally, we observed an enrichment of gene-based associations for low-density lipoprotein cholesterol drug target genes and for genes closest to GWAS index single-nucleotide polymorphisms (SNPs). Our results demonstrate that gene-based associations can be beneficial for drug target development and provide evidence that the gene closest to the array-based GWAS index SNP is often the functional gene for blood lipid levels.

[1]Program in Medical and Population Genetics, Broad Institute of Harvard and MIT, Cambridge, MA 02142, USA; [2]Department of Clinical Sciences, Lund University, Malmö, Sweden; [3]Department of Population Medicine, Qatar University College of Medicine, QU Health, Doha, Qatar; [4]Programs in Metabolism and Medical & Population Genetics, Broad Institute, Cambridge, MA 02142, USA; [5]Department of Pediatrics, Harvard Medical School, Boston, MA 02115, USA; [6]Division of Genetics and Genomics, Boston Children's Hospital, Boston, MA 02115, USA; [7]Center for Genomic Medicine, Department of Medicine, Massachusetts General Hospital, Boston, MA 02114, USA; [8]Cardiovascular Research Center, Massachusetts General Hospital, Boston, MA 02114, USA; [9]Department of Public Health Sciences, Penn State College of Medicine, Hershey, PA 17033, USA; [10]Cardiovascular Disease Initiative, Broad Institute of Harvard and MIT, Cambridge, MA 02142, USA; [11]Department of Genetics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104, USA; [12]Department of Medicine, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104, USA; [13]Instituto Nacional de Ciencias Medicas y Nutricion, Mexico City, Mexico; [14]Department of Genetics, Washington University in St. Louis, St. Louis, MO 63110, USA; [15]The McDonnell Genome Institute, Washington University School of Medicine, St. Louis, MO 63108, USA; [16]ASTC: Associazione per lo Studio Della Trombosi in Cardiologia, Pavia, Italy; [17]Azienda Ospedaliero-Universitaria di Parma, Parma, Italy; [18]Universit, degli Studi di Parma, Parma, Italy; [19]Dean's Office, College of Public Health, University of Kentucky, Lexington, KY 40536, USA; [20]Department of Epidemiology, University of Alabama at Birmingham, Birmingham, AL 35233, USA; [21]Departments of Medicine and Genetics, Albert Einstein College of Medicine, Bronx, NY 10461, USA; [22]University of Haifa, Faculty of Natural Science, Haifa, Israel; [23]Houston Methodist Debakey Heart and Vascular Center, Houston, TX 77030, USA; [24]Section of Cardiovascular Research, Baylor College of Medicine, Houston, TX 77030, USA; [25]Instituto Nacional de Medicina Genómica, Mexico City, Mexico; [26]Department of Medicine, Johns Hopkins University School of Medicine, Baltimore, MD 21287, USA; [27]Department of Epidemiology, School of Public Health, University of Michigan, Ann Arbor, MI 49109, USA; [28]Cardiovascular Health Research Unit, Department of Medicine, University of Washington, Seattle, WA 98101, USA; [29]Department of Human Genetics and South Texas Diabetes and Obesity Institute, University of Texas Rio Grande Valley School of Medicine, Brownsville, TX 78520, USA; [30]Human Genetics Center, Department of Epidemiology, Human Genetics, and Environmental Sciences, School of Public Health, The University of Texas Health Science Center at Houston, Houston, TX 77030, USA; [31]Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX 77030, USA; [32]Medical Genomics and Metabolic Genetics Branch, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD 20892, USA; [33]Hasso Plattner Institute for Digital Health at Mount Sinai, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA; [34]Digital Health Center, Hasso Plattner Institute, University of Potsdam, Potsdam, Germany; [35]Department of Biochemistry, Wake Forest School of Medicine, Winston-Salem, NC 27157, USA; [36]Department of Cardiovascular Sciences, University of Leicester, Leicester, UK; [37]NIHR Leicester Biomedical Research Centre, Glenfield Hospital, Leicester, UK; [38]Department of Biostatistics, University of Washington, Seattle, WA 98195, USA; [39]Department of Medicine, Brigham and Women's Hospital, Boston, MA 02115, USA; [40]Division of Sleep Medicine, Harvard Medical School, Boston, MA 02115, USA; [41]Department of Medicine, Yong Loo Lin School of Medicine, National University of Singapore, National University Health System, Singapore; [42]Institute of Biomedical Sciences, Academia Sinica, Taiwan; [43]The Institute for Translational Genomics and Population Sciences, Department of Pediatrics, The Lundquist Institute for Biomedical Innovation at Harbor-UCLA Medical Center, Torrance, CA 90502, USA; [44]Department of Ophthalmology, Yong Loo Lin

(Affiliations continued on next page)

L. Adrienne Cupples,[51,52] Joanne E. Curran,[29] John Danesh,[48,53] Paul S. de Vries,[30] Ralph A. DeFronzo,[54] Harsha Doddapaneni,[31] Ravindranath Duggirala,[29] Susan K. Dutcher,[14,15] Patrick T. Ellinor,[8,10] Leslie S. Emery,[38] Jose C. Florez,[1,7,55,56] Myriam Fornage,[57] Barry I. Freedman,[58] Valentin Fuster,[59,60] Ma. Eugenia Garay-Sevilla,[61] Humberto García-Ortiz,[25] Soren Germer,[62] Richard A. Gibbs,[31,63] Christian Gieger,[64,65,66] Benjamin Glaser,[67] Clicerio Gonzalez,[68] Maria Elena Gonzalez-Villalpando,[69] Mariaelisa Graff,[70] Sarah E. Graham,[71] Niels Grarup,[72] Leif C. Groop,[73,74] Xiuqing Guo,[43] Namrata Gupta,[1] Sohee Han,[75] Craig L. Hanis,[76] Torben Hansen,[72,77] Jiang He,[78,79] Nancy L. Heard-Costa,[52,80] Yi-Jen Hung,[81] Mi Yeong Hwang,[75] Marguerite R. Irvin,[82] Sergio Islas-Andrade,[83] Gail P. Jarvik,[84] Hyun Min Kang,[85] Sharon L.R. Kardia,[27] Tanika Kelly,[78] Eimear E. Kenny,[59,86,87] Alyna T. Khan,[38] Bong-Jo Kim,[75] Ryan W. Kim,[47] Young Jin Kim,[75] Heikki A. Koistinen,[88,89,90] Charles Kooperberg,[91] Johanna Kuusisto,[92] Soo Heon Kwak,[93] Markku Laakso,[92] Leslie A. Lange,[94] Jiwon Lee,[39] Juyoung Lee,[75] Seonwook Lee,[47] Donna M. Lehman,[54] Rozenn N. Lemaitre,[28] Allan Linneberg,[95,96] Jianjun Liu,[41,97] Ruth J.F. Loos,[98,99] Steven A. Lubitz,[8,10] Valeriya Lyssenko,[73,100] Ronald C.W. Ma,[101,102,103] Lisa Warsinger Martin,[104] Angélica Martínez-Hernández,[25] Rasika A. Mathias,[26] Stephen T. McGarvey,[105] Ruth McPherson,[106] James B. Meigs,[56,107,108] Thomas Meitinger,[109,110] Olle Melander,[73,111] Elvia Mendoza-Caamal,[25] Ginger A. Metcalf,[31] Xuenan Mi,[78] Karen L. Mohlke,[112] May E. Montasser,[113] Jee-Young Moon,[114]

(Author list continued on next page)

School of Medicine, National University of Singapore, National University Health System, Singapore; [45]Ophthalmology & Visual Sciences Academic Clinical Program (Eye ACP), Duke-NUS Medical School, Singapore; [46]Singapore Eye Research Institute, Singapore National Eye Centre, Singapore; [47]Psomagen, Inc. (formerly Macrogen USA), 1330 Piccard Drive Ste 103, Rockville, MD 20850, USA; [48]MRC/BHF Cardiovascular Epidemiology Unit, Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK; [49]Centre for Non-Communicable Disease Research, Bangladesh; [50]Department of Medicine, University of Mississippi Medical Center, Jackson, MS 39216, USA; [51]Department of Biostatistics, Boston University School of Public Health, Boston, MA 02118, USA; [52]NHLBI Framingham Heart Study, Framingham, MA 01702, USA; [53]The National Institute for Health Research Blood and Transplant Research Unit in Donor Health and Genomics at the University of Cambridge, Cambridge, UK; [54]Department of Medicine, University of Texas Health Science Center, San Antonio, TX 78229, USA; [55]Diabetes Research Center (Diabetes Unit), Department of Medicine, Massachusetts General Hospital, Boston, MA 02114, USA; [56]Department of Medicine, Harvard Medical School, Boston, MA 02115, USA; [57]Institute of Molecular Medicine, University of Texas Health Science Center at Houston, Houston, TX 770030, USA; [58]Department of Internal Medicine, Section on Nephrology, Wake Forest School of Medicine, Winston-Salem, NC 27157, USA; [59]Department of Medicine, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA; [60]Centro Nacional de Investigaciones Cardiovasculares Carlos III, Madrid, Spain; [61]Department of Medical Science, Division of Health Science, University of Guanajuato, Guanajuanto, Mexico; [62]New York Genome Center, New York, NY 10013, USA; [63]Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX 77030, USA; [64]Research Unit of Molecular Epidemiology, Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg, Germany; [65]Institute of Epidemiology, Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg, Germany; [66]German Center for Diabetes Research, Neuherberg, Germany; [67]Endocrinology and Metabolism Service, Hadassah-Hebrew University Medical Center, Jerusalem, Israel; [68]Unidad de Diabetes y Riesgo Cardiovascular, Instituto Nacional de Salud Pœblica, Cuernavaca, Morelos, Mexico; [69]Centro de Estudios en Diabetes, Mexico City, Mexico; [70]Department of Epidemiology, University of North Carolina at Chapel Hill, Chapel Hill, NC 27514, USA; [71]Department of Internal Medicine, University of Michigan, Ann Arbor, MI 48109, USA; [72]Novo Nordisk Foundation Center for Basic Metabolic Research, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark; [73]Department of Clinical Sciences, Diabetes and Endocrinology, Lund University Diabetes Centre, Malmö, Sweden; [74]Finnish Institute for Molecular Genetics, University of Helsinki, Helsinki, Finland; [75]Division of Genome Science, Department of Precision Medicine, Chungcheongbuk-do, Republic of Korea; [76]Human Genetics Center, School of Public Health, The University of Texas Health Science Center at Houston, Houston, TX 77030, USA; [77]Faculty of Health Sciences, University of Southern Denmark, Odense, Denmark; [78]Department of Epidemiology, Tulane University School of Public Health and Tropical Medicine, New Orleans, LA 70112, USA; [79]Tulane University Translational Science Institute, New Orleans, LA 70112, USA; [80]Department of Neurology, Boston University School of Medicine, Boston, MA 02118, USA; [81]Division of Endocrine and Metabolism, Tri-Service General Hospital Songshan Branch, Taipei, Taiwan; [82]Department of Epidemiology, School of Public Health, UAB, Birmingham, AL 35294, USA; [83]Dirección de Investigación, Hospital General de México "Dr. Eduardo Liceaga," Secretaría de Salud, Mexico City, Mexico; [84]Departments of Medicine (Medical Genetics) and Genome Sciences, University of Washington Medical Center, Seattle, WA 98195, USA; [85]Department of Biostatistics and Center for Statistical Genetics, University of Michigan, Ann Arbor, MI 48109, USA; [86]Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA; [87]Institute for Genomic Health, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA; [88]Department of Public Health Solutions, Finnish Institute for Health and Welfare, Helsinki, Finland; [89]Minerva Foundation Institute for Medical Research, Helsinki, Finland; [90]University of Helsinki and Department of Medicine, Helsinki University Central Hospital, Helsinki, Finland; [91]Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, WA 98103, USA; [92]Institute of Clinical Medicine, University of Eastern Finland, and Kuopio University Hospital, Kuopio, Finland; [93]Department of Internal Medicine, Seoul National University Hospital, Seoul, Republic of Korea; [94]Department of Medicine, University of Colorado Denver, Anschutz Medical Campus, Aurora, CO 80045, USA; [95]Center for Clinical Research and Prevention, Bispebjerg and Frederiksberg Hospital, The Capital Region, Copenhagen, Denmark; [96]Department of Clinical Medicine, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark; [97]Genome Institute of Singapore, Agency for Science, Technology and Research, Singapore; [98]Charles R. Bronfman Institute of Personalized Medicine, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA; [99]The Mindich Child Health and Development Institute, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA; [100]University of Bergen, Bergen, Norway; [101]Department of Medicine and Therapeutics, The Chinese University of Hong Kong, Hong Kong, China; [102]Hong Kong Institute of Diabetes and Obesity, The Chinese University of Hong Kong, Hong Kong, China; [103]Li Ka Shing Institute of Health Sciences, The Chinese University of Hong Kong, Hong Kong, China; [104]Division of Cardiology, Department of Medicine, George Washington University, Washington, DC 20037, USA; [105]Department of Epidemiology and International Health Institute, Brown University School of Public Health, Providence, RI 02912, USA; [106]Ruddy Canadian Cardiovascuar Genetics Centre, University of Ottawa Heart Institute, Ottawa, ON, Canada; [107]General Medicine Division, Massachusetts General Hospital, Boston, MA 02114, USA; [108]Broad Institute of MIT and Harvard, Cambridge, MA 02142, USA; [109]Deutsches Forschungszentrum fŸr Herz-Kreislauferkrankungen, Partner Site Munich Heart Alliance, Munich, Germany; [110]Institute of Human Genetics, Technische Universität München, Munich, Germany; [111]Department of Emergency and Internal Medicine, SkŒne University Hospital, Malmö, Sweden; [112]Department of Genetics, University of North Carolina, Chapel Hill, NC 27514, USA; [113]University of Maryland School of Medicine, Division of Endocrinology, Diabetes and Nutrition and Program for

(Affiliations continued on next page)

Hortensia Moreno-Macías,[115] Alanna C. Morrison,[30] Donna M. Muzny,[31] Sarah C. Nelson,[38] Peter M. Nilsson,[2] Jeffrey R. O'Connell,[113] Marju Orho-Melander,[2] Lorena Orozco,[25] Colin N.A. Palmer,[116] Nicholette D. Palmer,[35] Cheol Joo Park,[47] Kyong Soo Park,[93,117,118] Oluf Pedersen,[72] Juan M. Peralta,[29] Patricia A. Peyser,[27] Wendy S. Post,[119] Michael Preuss,[98] Bruce M. Psaty,[28,120,121] Qibin Qi,[114] D.C. Rao,[122] Susan Redline,[39,40] Alexander P. Reiner,[123] Cristina Revilla-Monsalve,[124] Stephen S. Rich,[125] Nilesh Samani,[36,37] Heribert Schunkert,[126] Claudia Schurmann,[33,34,98] Daekwan Seo,[47] Jeong-Sun Seo,[47] Xueling Sim,[127] Rob Sladek,[128,129,130] Kerrin S. Small,[131] Wing Yee So,[101,102,103] Adrienne M. Stilp,[38] E. Shyong Tai,[41,127,132] Claudia H.T. Tam,[101,102,103] Kent D. Taylor,[43] Yik Ying Teo,[127,133,134] Farook Thameem,[135] Brian Tomlinson,[136] Michael Y. Tsai,[137] Tiinamaija Tuomi,[138,139,140] Jaakko Tuomilehto,[141,142,143] Teresa Tusié-Luna,[13,144] Miriam S. Udler,[1,55,56] Rob M. van Dam,[127,145] Ramachandran S. Vasan,[52,146] Karine A. Viaud Martinez,[147] Fei Fei Wang,[38] Xuzhi Wang,[51] Hugh Watkins,[148] Daniel E. Weeks,[149,150] James G. Wilson,[151] Daniel R. Witte,[152,153] Tien-Yin Wong,[44,45,46] Lisa R. Yanek,[26] AMP-T2D-GENES, Myocardial Infarction Genetics Consortium, NHLBI Trans-Omics for Precision Medicine (TOPMed) Consortium, NHLBI TOPMed Lipids Working Group, Sekar Kathiresan,[1,7,56,154] Daniel J. Rader,[11,12,155] Jerome I. Rotter,[43] Michael Boehnke,[85] Mark I. McCarthy,[156,157,160] Cristen J. Willer,[71,158,159] Pradeep Natarajan,[1,8,10,56] Jason A. Flannick,[4,5,6] Amit V. Khera,[1,7,8] and Gina M. Peloso[51,*]

Introduction

Blood lipid levels are heritable complex risk factors for atherosclerotic cardiovascular diseases.[1] Array-based genome-wide association studies (GWASs) have identified >400 loci as associated with blood lipid levels, explaining 9%–12% of the phenotypic variance of lipid traits.[2–8] These studies have identified mostly common (minor allele frequency [MAF] > 1%) noncoding variants with modest effect sizes and have been instrumental in defining the causal roles of lipid fractions on cardiovascular disease.[9–13] Despite these advances, the mechanisms and causal genes for most of the identified variants and loci can be difficult to determine.

Genetic association studies testing rare coding variants have potential to directly implicate causal genes. Advances in next-generation sequencing over the last decade have facilitated increasingly larger studies with improved power to detect associations of rare variants with complex diseases and traits.[14,15] However, most exome sequencing studies to date have been insufficiently powered for rare variant discovery; for example, Flannick et al. estimated

Personalized and Genomic Medicine, Baltimore, MD 21201, USA; [114]Department of Epidemiology and Population Health, Albert Einstein College of Medicine, Bronx, NY 10461, USA; [115]Universidad Autónoma Metropolitana, Mexico City, Mexico; [116]Pat Macpherson Centre for Pharmacogenetics and Pharmacogenomics, Division of Population Health and Genomics, University of Dundee, Ninewells Hospital and Medical School, Dundee, UK; [117]Department of Internal Medicine, Seoul National University College of Medicine, Seoul, Republic of Korea; [118]Department of Molecular Medicine and Biopharmaceutical Sciences, Graduate School of Convergence Science and Technology, Seoul National University, Seoul, Republic of Korea; [119]Division of Cardiology, Department of Medicine, Johns Hopkins University, Baltimore, MD 21287, USA; [120]Department of Epidemiology, University of Washington, Seattle, WA 98101, USA; [121]Department of Health Services, University of Washington, Seattle, WA 98101, USA; [122]Division of Biostatistics, Washington University School of Medicine, St. Louis, MO 63108, USA; [123]University of Washington, Seattle, WA 98101, USA; [124]Instituto Mexicano del Seguro Social SXXI, Mexico City, Mexico; [125]Center for Public Health Genomics, University of Virginia, Charlottesville, VA 22908, USA; [126]Deutsches Herzzentrum München, Technische UniversitŠt München, Deutsches Zentrum fŸr Herz-Kreislauf-Forschung, München, Germany; [127]Saw Swee Hock School of Public Health, National University of Singapore and National University Health System, Singapore; [128]Department of Human Genetics, McGill University, Montreal, QC, Canada; [129]Division of Endocrinology and Metabolism, Department of Medicine, McGill University, Montreal, QC, Canada; [130]McGill University and Génome Québec Innovation Centre, Montreal, QC, Canada; [131]Department of Twin Research and Genetic Epidemiology, King's College London, London, UK; [132]Duke-NUS Medical School Singapore, Singapore; [133]Department of Statistics and Applied Probability, National University of Singapore, Singapore; [134]Life Sciences Institute, National University of Singapore, Singapore; [135]Department of Biochemistry, Faculty of Medicine, Health Science Center, Kuwait University, Safat, Kuwait; [136]Faculty of Medicine, Macau University of Science & Technology, Macau, China; [137]Department of Laboratory Medicine and Pathology, University of Minnesota, Minneapolis, MN 55455, USA; [138]Department of Endocrinology, Abdominal Centre, Helsinki University Hospital, Helsinki, Finland; [139]Folkhälsan Research Centre, Helsinki, Finland; [140]Research Programs Unit, Diabetes and Obesity, University of Helsinki, Helsinki, Finland; [141]Public Health Promotion Unit, Finnish Institute for Health and Welfare, Helsinki, Finland; [142]Department of Public Health, University of Helsinki, Helsinki, Finland; [143]Diabetes Research Group, King Abdulaziz University, Jeddah, Saudi Arabia; [144]Departamento de Medicina Genómica y Toxicología, Instituto de Investigaciones Biomédicas, Universidad Nacional Autónoma de México/ Instituto Nacional de Ciencias Médicas y Nutrición Salvador Zubirán, Mexico City, Mexico; [145]Department of Nutrition, Harvard School of Public Health, Boston, MA 02115, USA; [146]Departments of Medicine & Epidemiology, Boston University Schools of Medicine & Public Health, Boston, MA 02118, USA; [147]Illumina Laboratory Services, Illumina Inc., San Diego, CA 92122, USA; [148]Cardiovascular Medicine, Radcliffe Department of Medicine and the Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK; [149]Department of Biostatistics, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, PA 15261, USA; [150]Department of Human Genetics, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, PA 15261, USA; [151]Division of Cardiovascular Medicine, Beth Israel Deaconess Medical Center, Boston, MA 02215, USA; [152]Department of Public Health, Aarhus University, Aarhus, Denmark; [153]Steno Diabetes Center Aarhus, Aarhus, Denmark; [154]Verve Therapeutics, Cambridge, MA 02139, USA; [155]Institute for Translational Medicine and Therapeutics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104, USA; [156]Oxford Centre for Diabetes, Endocrinology and Metabolism, Radcliffe Department of Medicine, University of Oxford, Oxford, UK; [157]Wellcome Centre for Human Genetics, University of Oxford, Oxford, UK; [158]Department of Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, MI 48109, USA; [159]Department of Human Genetics, University of Michigan, Ann Arbor, MI 48109, USA
[160]Present address: Genentech, 1 DNA Way, South San Francisco, CA 94080, USA
*Correspondence: gpeloso@bu.edu
https://doi.org/10.1016/j.ajhg.2021.11.021.

that it would require 75,000 to 185,000 sequenced cases of type 2 diabetes (T2D) to detect associations at known drug target genes at exome-wide significance.[15]

Identifying rare variants with impact on protein function has helped elucidate biological pathways underlying dyslipidemia and atherosclerotic diseases such as coronary artery disease (CAD).[14,16–25] Successes with this approach have led to the development of novel therapeutic targets to modify blood lipid levels and lower risk of atherosclerotic diseases.[26,27]

The vast majority of participants in previous studies have been of European ancestry, highlighting the need for more diverse study sample. Such diversity can identify associated variants absent or present at very low frequencies in European populations and help implicate new genes with generalizability extending to all populations.

We have assembled exome sequence data from >170,000 individuals across multiple ancestries and systematically tested the association of rare variants in each gene with six circulating lipid phenotypes: low-density lipoprotein cholesterol (LDL-C), high density lipoprotein cholesterol (HDL-C), non-HDL-C, total cholesterol (TC), triglycerides (TG), and the ratio of TG to HDL-C (TG:HDL). We find 35 genes associated with blood lipid levels, show evidence of gene-based signals in array-based GWAS loci, show enrichment of lipid gene-based associations in LDL-C drug targets and genes in close proximity of GWAS index variants, and test lipid genes for association with CAD, T2D, and liver enzymes.

Subjects and methods

Study overview

Our study samples were derived from four major data sources with exome or genome sequence data and blood lipid levels: CAD case-control studies from the Myocardial Infarction Genetics Consortium[28,29] (MIGen, n = 44,208) and a UK Biobank (UKB) nested case-control study of CAD[28] (n = 10,689); T2D cases-control studies from the AMP-T2D-GENES exomes[15] (n = 32,486); population-based studies from the TOPMed project[30,31] freeze 6a data (n = 44,101) restricted to the exome; and the UKB first tranche of exome sequence data[32,33] (n = 40,586) (see supplemental information). Informed consent was obtained from all subjects, and committees approving the studies are available in the supplemental information.

Within each data source, individuals were excluded if they failed study-specific sequencing quality metrics, lacked lipid phenotype data, or were duplicated in other sources. Sequencing and quality control performed in each study is available in the supplemental methods. We additionally removed first- and second-degree relatives across data sources while we kept relatives within each data source because we were able to adjust for relatedness within each data source by using kinship matrices in linear mixed models. If samples from the same study were present in different data sources, we used the samples in the data source that has the largest sample size from the study and removed the overlapping set from the other data source. For instance, samples from the Atherosclerosis Risk in Communities (ARIC) Study were removed from TOPMed and kept in MIGen, which had more sequenced samples from ARIC. Similarly, samples from the Jackson Heart Study were kept in TOPMed and removed from MIGen. To obtain duplicate and kinship information across data sources, we used 14,834 common (MAF > 1%) and no more than weakly dependent ($r^2 < 0.2$) variants by using the make-king flag in PLINK v2.0.

Single-variant association analyses were performed within each data source, case status, and ancestry combination. The data were sequenced and variant calling was performed separately by data source, and this allowed us to look for effects by case status and genetically inferred and/or reported ancestry groups. We performed gene-based meta-analyses by combining single-variant summary statistics and covariance matrices generated from RVTESTS.[34] We performed ancestry-specific gene-based meta-analyses by combining single-variant summary data from five major ancestries with >10,000 individuals across all data sources: European, South Asian, African, Hispanic, and East Asian ancestries.

Phenotypes

We studied six lipid phenotypes; total cholesterol (TC), LDL-C, HDL-C, non-HDL-C, triglycerides (TG), and TG:HDL. TC was adjusted by dividing the value by 0.8 in individuals reporting lipid-lowering medication use after 1994 or statin use at any time point. If LDL-C levels were not directly measured, then they were calculated via the Friedewald equation for individuals with TG levels < 400 mg/dL with adjusted TC levels. If LDL-C levels were directly measured then, their values were divided by 0.7 in individuals reporting lipid-lowering medication use after 1994 or statin use at any time point.[5] TG and TG:HDL levels were natural logarithm transformed. Non-HDL-C was obtained by subtracting HDL-C from adjusted TC levels. Residuals for each trait in each cohort, ancestry, and case status grouping were created after adjustment for age, age^2, sex, principal components, sequencing platform, and fasting status (when available) in a linear regression model. We then inverse-normal transformed the residuals and multiplied them by the standard deviation of the trait to scale the effect sizes to the interpretable units.

Variant annotation

We compiled autosomal variants with call rate > 95% within each case and ancestry-specific analysis dataset with minor allele count (MAC) ≥ 1 (across the combined data). Variants were annotated with the Ensembl Variant Effect Predictor[35] and its associated Loss-of-Function Transcript Effect Estimator (LOFTEE)[36] and the dbNSFP[37] version 3.5a plugins. We limited our annotations to the canonical transcripts. The LOFTEE plugin assesses stop-gained, frameshift, and splice-site-disrupting variants. Loss-of-function variants are classified as either high confidence or low confidence. The dbNSFP is a database that provides functional prediction data and scores for non-synonymous variants by using multiple algorithms.[37] We used this database to classify missense variants as damaging by using two different definitions based on bioinformatic prediction algorithms. The first is based on MetaSVM,[38] which is derived from ten different component scores (SIFT, PolyPhen-2 HDIV, PolyPhen-2 HVAR, GERP++, MutationTaster, Mutation Assessor, FATHMM, LRT, SiPhy, PhyloP). The second is based on five variant prediction algorithms including SIFT, PolyPhen-2 HumVar, PolyPhen-2 HumDiv, MutationTaster, and LRT scores. Additionally, we ran a deep neural

network analysis (Splice AI) to predict splice-site-altering variants.[39] Variant descriptive analysis was performed with a maximal set of variants that were used for analysis of the lipid phenotype with the largest sample size. The counts and proportions of variants—annotated according to the different predicted consequences described above—were obtained out of an overall set of variants.

Single-variant association analysis

Each data source was sub-categorized on the basis of ancestry and CAD or T2D case status in the studies ascertained by disease status. Subgrouping data sources yielded a total of 23 distinct sample subcategories. As relatives were kept within each sub-group, we performed generalized linear mixed models to analyze the association of single autosomal variants with standard-deviation corrected-inverse-normal transformed traits by using RVTESTS.[34] We used RVTESTS to generate summary statistics and covariance matrices using 500 kb sliding windows. To obtain the single-variant associations, we performed a fixed-effects inverse-variance weighted meta-analysis for multi-ancestry and within each of the five major ancestries. We used an exome-wide significance threshold of $p < 7.2 \times 10^{-8}$ (Bonferroni correction for six traits and with previously recommended threshold for coding variants $p < 4.3 \times 10^{-7}$)[40] to determine significant coding variants.

Gene-based association analysis

We used summary level score statistics and covariance matrices from autosomal single-variant association results to perform gene-based meta-analyses among all individuals and within each ancestry by using RAREMETALS version 7.2.[41] Samoan individuals only contributed to the overall analysis. Gene-based association testing aggregates variants within each gene unit by using burden tests and sequence kernel association tests (SKATs), which allows variable variant effect direction and size.[42] The "rareMETALS.range.group" function was used with MAF < 1%, which filters out all variants with combined MAF > 1% in all meta-analytic datasets. All variants with call rates < 95% and not annotated as loss of function (LOF) via LOFTEE, splice-site variants or damaging missense as defined by MetaSVM or by all SIFT, PolyPhen-2 HumVar, PolyPhen-2 HumDiv, MutationTaster, and LRT prediction algorithms (damaging 5 out of 5) were excluded in the gene-based meta-analyses.

We used six different variant groupings to determine the set of damaging variants within each gene, (1) high-confidence LOF via LOFTEE, (2) LOF and predicted splice-site-altering variants, (3) LOF and MetaSVM missense variants, (4) LOF, MetaSVM missense, and predicted splice-site-altering variants, (5) LOF and damaging 5 out of 5 missense variants, and (6) LOF, damaging 5 out of 5 missense, and predicted splice-site-altering variants. We used an exome-wide significance threshold of $p < 4.3 \times 10^{-7}$, Bonferroni corrected for the maximum number of annotated genes (n = 19,540) and six lipid traits, to determine significant coding variants. Two gene transcripts, *DOCK6* and *DOCK7*, that overlap with two well-studied lipid genes, *ANGPTL8* and *ANGPTL3*, respectively, met our exome-wide significance threshold. After excluding variation observed in *ANGPTL8* and *ANGPTL3*, *DOCK6* and *DOCK7*, respectively, were no longer significant and have been excluded as associated genes.

We performed a series of sensitivity analyses for our results. We repeated the multi-ancestry gene-based analyses by using an MAF < 0.1% and compared our exome-wide significant gene-based results by using an MAF < 1% to using an MAF < 0.1%. We compared the single variants in our top gene-based associations with respective traits by using GWAS summary data.[8] Gene-based tests were repeated excluding variants identified in GWASs with $p < 5 \times 10^{-8}$. Furthermore, all single variants included in each of the top gene-based associations were analyzed in relation to the respective trait. For each exome-wide significant gene-based association, we obtained the association of each single variant within the gene-specific variant groups with the respective phenotype. Then we determined—out of each gene's overall set of variants—those that had p values at different significance thresholds to identify the percentages of variants contributing to each gene-based signal. To assess whether the most significant variant within each gene was driving the association, we repeated gene-based analyses after removing the respective top single variant from gene-specific variant groups.

To understand whether variants contributing to top gene-based signals were similar or different across different ancestries, we determined the degree of overlap across ancestries for all variants incorporated and then for those with $p < 0.05$. Finally, we checked for overlap across the most significant (lowest p value) variant from each of the gene-based signals.

Heterogeneity of gene-based estimates in all gene-trait-variant grouping combinations passing exome-wide significant levels was assessed across the five main ancestries (European, South Asian, African, Hispanic, and East Asian) and between T2D and CAD cases and controls via Cochran's Q.

We performed replication of our top gene-based associations with blood lipid levels in the Penn Medicine BioBank (PMBB) and UK Biobank samples that did not contribute to the discovery analysis (see supplemental methods).

Gene-based analysis of GWAS loci and drug targets

We obtained variants associated with LDL-C, HDL-C, and TG from a recent GWAS in the Million Veterans Program.[8] Then we identified genes within ±200 kb of each GWAS index variant and performed gene-based analysis for each of those genes by using the six variant groups. *In-silico* lookup of gene-based associations for respective lipid traits was then performed for all genes within defined GWAS loci. Drug target genes were obtained from the drug bank database[43] with the following search categories: "hypolipidemic agents," "lipid regulating agents," "anticholesteremic agents," "lipid modifying agents," and/or "hypercholesterolemia." A liberal definition for drug targets was used—drugs with any number of targets and targets targeted by any number of drugs—and then *in-silico* lookups were performed for gene-based associations.

Gene-set enrichment analysis

Gene-set enrichment analyses were performed for sets of Mendelian-, protein-altering- and non-protein-altering GWASs, and drug target genes with LDL-C, HDL-C, and TG. 21 genes associated with Mendelian lipid conditions were included on the basis of previous literature:[2] *LDLR*, *APOB*, *PCSK9*, *LDLRAP1*, *ABCG5*, *ABCG8*, *CETP*, *LIPC*, *LIPG*, *APOC3*, *ABCA1*, *APOA1*, *LCAT*, *APOA5*, *APOE*, *LPL*, *APOC2*, *GPIHBP1*, *LMF1*, *ANGPTL3*, and *ANGPTL4*. We analyzed GWAS gene sets on the basis of their coding status and their proximity to the most significant signal in the GWAS. Coding variants were defined as missense, frameshift, or stop-gained variants. Gene sets for coding or non-coding variants were then stratified into three categories on the basis of proximity to the

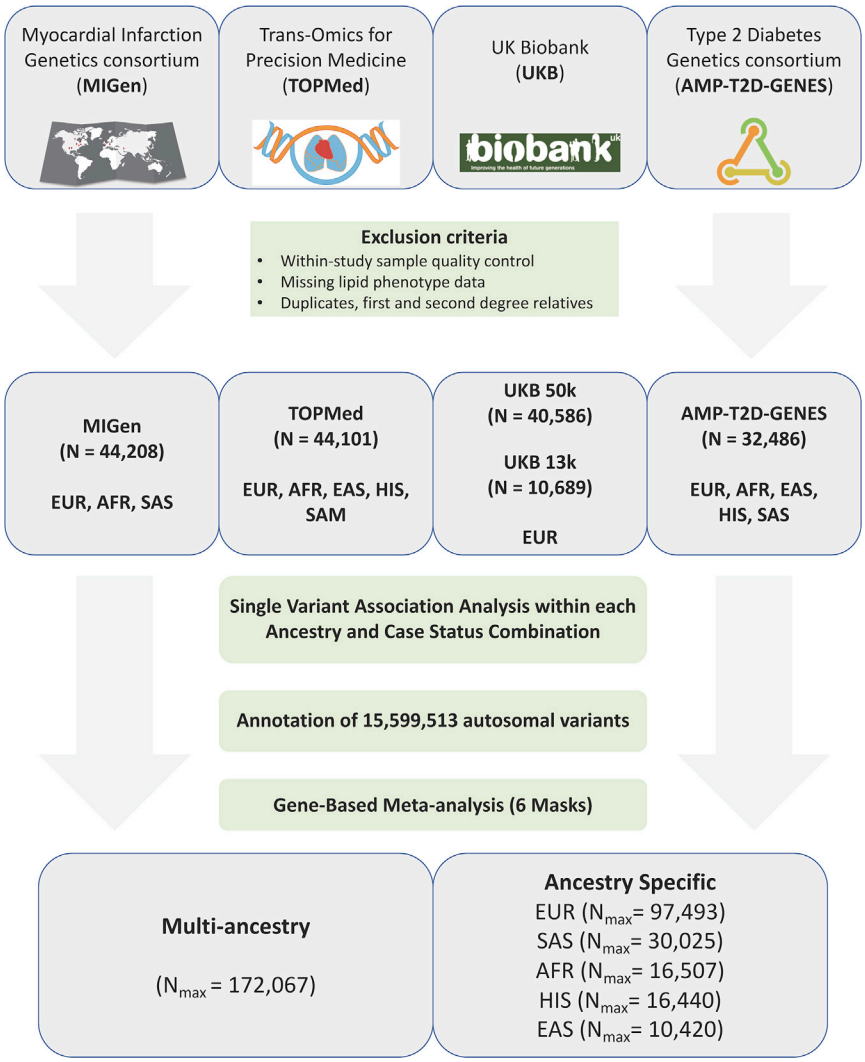

Figure 1. Study samples and design
Flow chart of the different stages of the study. Exome sequence genotypes were derived from four major data sources: the Myocardial Infarction Genetics consortium (MIGen), the Trans-Omics from Precision Medicine (TOPMed), the UK Biobank, and the Type 2 Diabetes Genetics (AMP-T2D-GENES) consortium. Single-variant association analyses were performed by ancestry and case status in case-control studies and meta-analyzed. Single-variant summary estimates and covariance matrices were used in gene-based analyses with six different variant groups and in multi-ancestry and each of the five main ancestries. AFR, African ancestry; EAS, East Asian ancestry; EUR, European ancestry; HIS, Hispanic ancestry; SAM, Samoan ancestry; SAS, South Asian ancestry.

most significant variant within each locus—closest, second closest, and greater than second closest gene. For each gene within each set, we obtained the most significant association in the multi-ancestry or ancestry-specific meta-analysis set by using any of the six different variant groups. Then each gene within each gene set was matched to ten other genes on the basis of sample size, total number of variants, cumulative MAC, and variant grouping nearest neighbors via the matchit R function. Then we compared the proportions by using Fisher's exact test between the main and matched gene sets by applying different p value thresholds.

Association of lipid genes with CAD and T2D data and liver fat/markers

We determined the associations of 40 genes identified in the main and GWAS loci analyses with CAD, T2D, and glycemic and liver enzyme blood measurements. The association with T2D was obtained from the latest gene-based exome association data from the AMP-T2D-GENES consortium.[15] The reported associations were obtained from different variant groups on the basis of their previous analyses. We additionally performed gene-based association analyses with CAD by using the MIGen case-control, UKB case-control, and UKB cohort samples with the variant groups described above. Further, six traits including fasting plasma glucose, HbA1c, alanine aminotransferase, aspartate aminotransferase, gamma glutamyl transferase, and albumin were analyzed in the UKB dataset. Single-variant association analyses were performed with RVTESTS. We used linear mixed models incorporating kinship matrices to adjust for relatedness within each study. Covariance matrices were generated with 500 kb sliding windows. We used RAREMETALS to assess associations between aggregated variants (MAF < 1%) in SKATs and burden tests with CAD and each of the six quantitative traits. We used six different variant groupings to determine the set of damaging variants within each gene, (1) high-confidence LOF with LOFTEE, (2) LOF and predicted splice-site-altering variants, (3) LOF and MetaSVM missense variants, (4) LOF, MetaSVM missense, and predicted splice-site-altering variants, (5) LOF and damaging 5 out of 5 missense variants, and (6) LOF, damaging 5 out of 5 missense, and predicted splice-site-altering variants.

Results

Sample and variant characteristics

Individual-level, quality-controlled data were obtained from four sequenced study sources with circulating lipid data for individuals of multiple ancestries (Figure 1). Characteristics of the study samples are detailed in Table S1. We analyzed data on up to 172,000 individuals with LDL-C, non-HDL-C (a calculated measure of TC minus HDL-C), TC, HDL-C, TG, and TG:HDL ratio (a proxy for insulin resistance).[44,45] 56.7% (n = 97,493) of the sample are of European ancestry, 17.4% (n = 30,025) South Asian, 9.6% (n = 16,507) African American, 9.6% (n = 16,440) Hispanic, 6.1% (n = 10,420) East Asian, and 0.7% (n = 1,182) Samoan, based on genetically estimated and/or self-reported ancestry.

After sequencing, we observed 15.6 million variants across all studies; we classified 5.0 million (32.6%) as transcript-altering coding variants on the basis of an annotation of frameshift, missense, nonsense, or splice-site acceptor/donor by using the Variant Effect Predictor (VEP).[35] A total of 340,214 (6.7%) of the coding variants were annotated as high-confidence LOF via the LOFTEE VEP plugin,[36] 238,646 (4.7%) as splice-site-altering identified by Splice AI,[39] 729,098 (14.3%) as damaging missense as predicted by the MetaSVM algorithm,[38] and 1,106,309 (21.8%) as damaging missense as predicted by consensus in all five prediction algorithms (SIFT, PolyPhen-2 HumVar, PolyPhen-2 HumDiv, MutationTaster, and LRT).[37] As expected, we observed a trend of decreasing proportions of putatively deleterious variants with increasing allele count (Figure S2, Table S3).

Single-variant association

We performed inverse-variance weighted fixed-effects meta-analyses of single-variant association results of LDL-C, non-HDL-C, TC, HDL-C, TG, and TG:HDL ratio from each consortium and ancestry group. Meta-analysis results were well controlled with genomic inflation factors ranging between 1.01 and 1.04 (Table S4). Single-variant results were limited to the 425,912 protein-altering coding variants with a total MAC > 20 across all 172,000 individuals. We defined significant associations by a previously established exome-wide significance threshold for coding variants ($p < 4.3 \times 10^{-7}$)[40] that was additionally corrected for testing six traits ($p = 4.3 \times 10^{-7}$ divided by 6) within all study samples or within each of the five major ancestries (Tables S5–S10); this yielded in each analysis a significance threshold of $p < 7.2 \times 10^{-8}$. A total of 104 rare coding variants in 57 genes were associated with LDL-C, 95 in 54 genes with non-HDL-C, 109 in 65 genes with TC, 92 in 56 genes with HDL-C, 61 in 36 genes with TG, and 68 in 42 genes with TG:HDL. We identified six missense variants in six genes (*TRIM5* p.Val112Phe, *ADH1B* p.His48Arg, *CHUK* p.Val268Ile, *ERLIN1* p.Ile291Val [rs2862954], *TMEM136* p.Gly77Asp, and *PPARA* p.Val227Ala) >1 Mb away from any index variant previously associated with a lipid phenotype (LDL-C, HDL-C, TC, or TG) in previous genetic discovery efforts (Tables S5–S10).[3,7,8] *PPARA* p.Val227Ala has previously been associated with blood lipids at a nominal significance level in East Asians ($p < 0.05$), where it is more common than in other ancestries.[46] Both *TRIM5* and *ADH1B* LDL-C increasing alleles have been associated with higher risk of CAD in a recent GWAS from CARDIOGRAM (odds ration [OR]: 1.08, $p = 2 \times 10^{-9}$; OR = 1.08, $p = 4 \times 10^{-4}$).[47] Single-variant associations were further performed in each of the five main ancestries (Table S11).

Gene-based association

Next, we performed gene-based testing of transcript-altering variants in aggregated SKATs and burden tests[42] in all study participants and within each of the five ancestries for six lipid traits: LDL-C, HDL-C, non-HDL-C, TC, TG, and TG:HDL. We excluded the Samoans from the single-ancestry analysis given the small number of individuals. We limited attention to variants with MAF ≤ 1% for each of six variant groups: (1) LOF, (2) LOF and predicted splice-site-altering variants via Splice AI, (3) LOF and MetaSVM missense variants, (4) LOF, MetaSVM missense, and predicted splice-site-altering variants, (5) LOF and damaging 5 out of 5 missense variants, and (6) LOF, damaging 5 out of 5 missense, and predicted splice-site-altering variants. Meta-analyses results were well controlled (Table S12).

We identified 35 genes reaching exome-wide significance ($p = 4.3 \times 10^{-7}$) for at least one of the six variant groupings (Tables S13–S19). Most of the significant results were from the multi-ancestry analysis where multiple ancestries contributed to the top signals (Figure 2A), and most of the 35 genes were associated with more than one lipid phenotype (Figure 2B). Ten of the 35 genes did not have prior evidence of gene-based links with blood lipid phenotypes (Table 1), and seven genes, including *ALB*, *SRSF2*, *CREB3L3*, *NR1H3*, *PLA2G12A*, *PPARG*, and *STAB1*, have evidence for a biological connection to circulating lipid levels (Box 1).

We performed a series of sensitivity analyses on our results. To determine whether low-frequency variants between 0.1%–1% frequency were driving our gene-based association results, we performed the gene-based multi-ancestry meta-analyses by using a maximum MAF threshold of 0.1% instead of 1%. We observed exome-wide significant associations ($p < 4.3 \times 10^{-7}$) for 29 genes with a 0.1% MAF threshold, all observed in our primary analyses with an MAF threshold of 1% (Table S20). We then intersected our 35 lipid-associated genes from 85 gene-based associations observed in the primary analysis with our results with an MAF threshold of 0.1%. All genes remained at least nominally significant ($p < 0.05$) with an 0.1% MAF threshold, except the *A1CF* and *TMEM136* associations (Table S21). Furthermore, we determined whether those signals were driven by previously reported GWAS hits. We identified a total of seven HDL-C associated variants in six genes, seven LDL-C variants in three genes, three TC variants in one gene, and seven TG variants in six genes that were previously found to be genome-wide significant in the Million Veterans Program (MVP) GWAS results (Table S22).[8] Respective gene-based analyses were repeated without those variants. Gene-based signals at *A1CF* and *BUD13* were lost after removal of one variant in each of those genes (Table S23).

The *JAK2* signal was further investigated after splitting the 136 contributing variants into those annotated as somatic via the Catalogue of Somatic Mutations in Cancer (COSMIC)[64] database and not annotated as a somatic variant. We observed an association only among a set of 26 variants annotated as somatic, while we observed no association when using the remaining 110 variants (Table S24). We also observed that after removal of the most

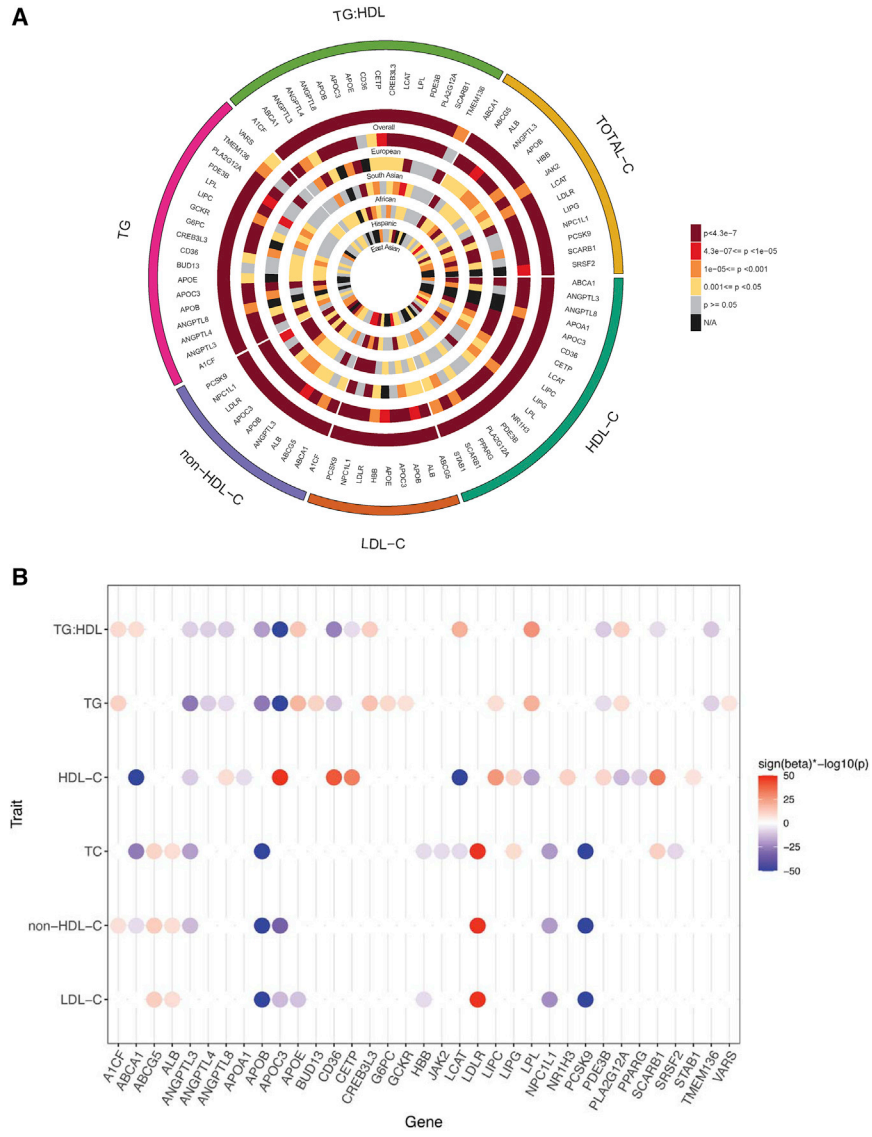

Figure 2. Exome-wide significant associations with blood lipid phenotypes

(A) Circular plot highlighting the evidence of association between the exome-wide significant 35 genes with any of the six different lipid traits ($p < 4.3 \times 10^{-7}$). The most significant associations from any of the six different variant groups are plotted. For almost all of the genes, the most significant associations were obtained from the multi-ancestry meta-analysis.

(B) Strength of association of the 35 exome-wide significant genes based on the most significant variant grouping and ancestry across the six lipid phenotypes studied. Beta (effect size) is obtained from the corresponding burden test for SKAT results. Most of the genes indicated associations with more than one phenotype. Sign(beta)*−log10(p value) displayed for associations that reached a $p < 4.3 \times 10^{-7}$. When the Sign(beta)*−log10(p) > 50, they were trimmed to 50.

significant variant in *JAK2* (p.Val617Phe; rs77375493), a somatic variant, there is no association between *JAK2* and total cholesterol ($p = 0.10$, Table S13).

We also determined which of the 35 genes were outside GWAS regions defined as those within ±200 kb flanking regions of GWAS-indexed single-nucleotide polymorphisms (SNPs) for TC (487 SNPs), LDL-C (531 SNPs), HDL-C, and TG (471 SNPs).[8] We identified 1,295 unique genes included in these lipid GWAS regions. Eight out of the 35 associated genes (23%) were not within a GWAS region (Table S13).

To understand whether the gene-based signals were driven by variants that could be identified through single-variant analyses, we looked at the proportion of the 35 genes that were associated with each trait that have at least one single contributing variant that passed the genome-wide significance threshold of 5×10^{-8}. Seventeen genes were associated with HDL-C at exome-wide significance (Table S13); eight genes had at least one variant with $p < 5 \times 10^{-8}$ (Table S8). Similarly, we observed 4/9 for LDL-C, 4/10 non-HDL-C, 4/14 TC, 7/18 TG, and 6/17 TG:HDL genes with at least one genome-wide significant variant (Tables S5–S10).

For genes with both gene-based and single-variant signals, we determined the variants that were driving these signals and determined the single-variant associations for all variants contributing to the top 35 genes (Table S25). From a total of 85 gene-based associations, 33 had at least one and 19 had only one single variant with $p < 5 \times 10^{-8}$ (Tables S25 and S26). All of the 19 had at least two variants passing nominal significance ($p < 0.05$) and 13 had at least ten variants with $p < 0.05$. Finally, gene-based associations in *A1CF*, *BUD13*, *JAK2*, and *TMEM136* were lost after removal of the respective most-significant single variant from the group of variants aggregated in each gene-based association (Table S13).

Comparison of gene-based associations across ancestries

We determined the overlap between single variants included in gene-based signals, which additionally were nominally significant ($p < 0.05$) in each of the five main ancestries. A large proportion of variants from each ancestry did not overlap with any other ancestry (Figure S3). For example, a total of four genes (*CETP*, *ABCA1*, *CD36*, and *LCAT*) were observed to have significant gene-based associations with HDL-C in multi-ancestry meta-analyses. A total 68% of variants from European ancestry samples that contributed to HDL-C gene-based associations did not

Table 1. Genes associated with blood lipids identified in this study

Gene	Name	Trait	N	cMAC	nVAR	β	SE	p	Mask	Test	Ancestry	UKBB replication	PMBB replication
ALB	albumin	LDL-C	165,003	51	18	29.51	5.11	7.76×10^{-9}	LOF	burden	multi-ancestry	<0.005	N/A
ALB	albumin	non-HDL-C	166,327	50	17	33.91	6.07	2.27×10^{-8}	LOF	burden	multi-ancestry	N/A	N/A
ALB	albumin	TC	172,103	54	18	33.37	5.89	1.48×10^{-8}	LOF	burden	multi-ancestry	N/A	N/A
SRSF2	serine and arginine rich splicing factor 2	TC	172,103	59	14	−30.59	5.49	2.46×10^{-8}	LOF/DAM5of5/SPLICE AI	burden	multi-ancestry	N/A	<0.005
JAK2	janus kinase 2	TC	975,33	441	136	−7.10	1.98	1.71×10^{-7}	LOF/DAM5of5/SPLICE AI	SKAT	EUR	<0.05	<0.05
CREB3L3	camp responsive element binding protein 3 like 3	TG	170,239	874	71	0.12	0.02	2.43×10^{-15}	LOF/DAM5of5/SPLICE AI	burden	multi-ancestry	<0.005	<0.005
CREB3L3	camp responsive element binding protein 3 like 3	TG/HDL-C	165,380	855	69	0.14	0.02	5.76×10^{-13}	LOF/DAM5of5/SPLICE AI	burden	multi-ancestry	N/A	N/A
TMEM136	transmembrane protein 136	TG	29,571	157	24	−0.15	0.04	3.39×10^{-9}	LOF/DAM5of5/SPLICE AI	SKAT	SAS	N/A	N/A
TMEM136	transmembrane protein 136	TG/HDL-C	29,517	157	24	−0.20	0.05	1.76×10^{-11}	LOF/DAM5of5/SPLICE AI	SKAT	SAS	N/A	N/A
VARS	valyl-trna synthetase 1	TG	56,140	67	51	0.32	0.06	4.30×10^{-7}	LOF/MetaSVM	burden	EUR	N/A	N/A
NR1H3	nuclear receptor subfamily 1 group h member 3	HDL-C	93,044	521	111	3.47	0.60	1.45×10^{-11}	LOF/MetaSVM/SPLICE AI	SKAT	EUR	<0.005	<0.05
PLA2G12A	phospholipase a2 group xiia	HDL-C	166,441	1,975	47	−2.28	0.31	8.12×10^{-14}	LOF/DAM5of5	burden	multi-ancestry	<0.005	<0.005
PLA2G12A	phospholipase a2 group xiia	TG	170,239	2,047	47	0.06	0.01	1.17×10^{-8}	LOF/DAM5of5	burden	multi-ancestry	N/A	N/A
PLA2G12A	phospholipase a2 group xiia	TG/HDL-C	165,380	1,969	46	0.11	0.01	7.56×10^{-13}	LOF/DAM5of5	burden	multi-ancestry	N/A	N/A
PPARG	peroxisome proliferator activated receptor gamma	HDL-C	166,441	147	72	−6.24	1.07	4.71×10^{-9}	LOF/DAM5of5/SPLICE AI	burden	multi-ancestry	<0.005	<0.005
STAB1	stabilin 1	HDL-C	166,441	6,550	804	0.83	0.16	2.58×10^{-7}	LOF/MetaSVM/SPLICE AI	burden	multi-ancestry	<0.005	N/A

cMAC, cumulative minor allele count; nVAR, number of variants in test; EUR, European ancestry; SAS, South Asian ancestry; N/A, not applicable.

Box 1. Genes with biological links to lipid metabolism

ALB

The association between mutations in the albumin gene and elevated cholesterol levels has been previously observed in rare cases of congenital analbuminemia.[48] This has been mainly suggested to result from compensatory increases in hepatic production of other non-albumin plasma proteins to maintain colloid osmotic pressure, particularly apolipoprotein B-100, leading to elevations in TC and LDL-C but normal HDL-C levels—which is consistent with our findings—although the exact mechanisms remain uncertain.[49] A lipodystrophy-like phenotype has also been linked to analbuminemia, which is consistent with the suggestive tendency for increased risk of T2D with LOF and predicted damaging variants in albumin in the population (OR = 1.85; p = 0.007) (Table S30).

SRSF2

SRSF2 encodes a highly conserved serine/arginine-rich splicing factor and has previously been linked to acute liver failure in liver-specific knockout in mice with accumulation of TC in the mutant liver.[50] Thus, this gene could be linked to a non-alcoholic fatty liver phenotype with accumulation of lipids in the liver as observed with other genes as *PNPLA3* and *TM6SF2*.[7] Therefore, we looked at association with liver function markers and we found an association between *SRSF2* and higher albumin levels ($p = 1 \times 10^{-4}$) and a suggestive tendency for higher gamma glutamyl transferase (GGT) (p = 0.05), consistent with potential liver involvement (Tables S46–S49).

CREB3L3

The association between *CREB3L3* and higher TG supports previous evidence from a single family and cohorts with severe hypertriglyceridemia but not sufficient evidence to be classified as a Mendelian lipid gene.[51–53] This has been additionally supported by functional studies where *Creb3l3*-knockout mice showed hypertriglyceridemia partly due to deficient expression of lipoprotein lipase coactivators (*Apoc2*, *Apoa4*, and *Apoa5*) and increased expression of activator *Apoc3*.[52]

NR1H3

The observed association of *NR1H3* with higher HDL-C and lower TG is supported by previous evidence of a role in non-alcoholic fatty liver disease in mice.[54] This gene encodes a liver X receptor alpha (LXRα), which is a nuclear receptor that acts as a cholesterol sensor and protects from cholesterol overload.[55,56] It has previously been shown that disrupting the LXRα phosphorylation at Ser196 in mice prevents non-alcoholic fatty liver disease.[54]

PLA2G12A

PLA2G12A is in the secretory phospholipase A2 ($sPLA_2$) family, which liberates fatty acids in the −sn2 position of phospholipids. This pattern suggests a previously unreported possible lipolytic role of this phospholipase in a manner similar to another member of the adipose-specific phospholipases, *PLA2G16*, which has been shown to have a lipolytic role in mice.[57,58] Further studies are needed to confirm whether *PLA2G12A* has a lipolytic role.

PPARG

Rare LOF mutations in *PPARG* have been previously found to be associated with reduced adipocyte differentiation, lipodystrophy, and increased risk of T2D.[59–61]

STAB1

STAB1 is a scavenger receptor that has been shown to mediate uptake of oxidized LDL-C.[62,63] There was a suggestive association between LOF variants and higher LDL-C ($\beta = 4.3$ mg/dL, $p = 2 \times 10^{-3}$), consistent with its role in LDL-C uptake.

overlap with any other ancestry and nor did 62% in South Asian, 44% in African, 41% in Hispanic, and 59% in East Asian ancestry. When restricted to variants with $p < 0.05$ in the multi-ancestry meta-analysis, the overlap among ancestries increased (Figure S4). A total of 61% of variants from European ancestry did not overlap with any other ancestry and nor did 46% in South Asian, 27% in African, 27% in Hispanic, and 32% in East Asian ancestry. Finally, we determined the top single variant contributing to each gene-based association (Figure S5). Out of the four HDL-C or the three LDL-C genes, none of the top variants overlapped among any of the ancestries, and at least one out of three variants from the TG genes was shared between two ancestries.

But, the gene-based associations were mostly consistent across the five ancestry groupings: European, South Asian, African, Hispanic, and East Asian. Three of the 17 HDL-C genes showed association in at least two different ancestries at exome-wide significance level ($p = 4.3 \times 10^{-7}$). Similarly, 3/9 LDL-C, 4/10 non-HDL-C, 5/14 TC, 2/18 TG, and 2/17 TG:HDL genes showed association in at least two different ancestries at an exome-wide significance level. Using a less stringent significance level ($p < 0.01$), across the six lipid traits, 59%–89% of associated genes from the joint analysis were associated in at least two different ancestries.

We tested the top 35 genes for heterogeneity across all 303 gene-trait-variant grouping combinations passing the exome-wide significance threshold ($p < 4.3 \times 10^{-7}$). We

observed heterogeneity in effect estimates ($p_{Het} < 1.7 \times 10^{-4}$, accounting for 303 combinations) in 19 (6%) different gene-trait-variant grouping combinations and in six different genes: *LIPC*, *LPL*, *LCAT*, *ANGPTL3*, *APOB*, and *LDLR* (Table S27). Although the LOF gene-based effect sizes were largely consistent across ancestries, there were differences in the cumulative frequencies of LOF variants for several genes, including *PCSK9*, *NPC1L1*, *HBB*, and *ABCG5* (Figures S6–S8).

We observed LOF and predicted-damaging variants in *TMEM136* associated with TG and TG:HDL only among individuals of South Asian ancestry ($p_{SKAT} = 3 \times 10^{-9}$ and 2×10^{-11}, respectively) (Table 1, Figure 2A). With the same variant grouping and ancestry, we observed associations with reduced TG by burden tests ($\beta = -15\%$, $p = 3 \times 10^{-4}$) and TG:HDL ($\beta = -20\%$, $p = 6 \times 10^{-5}$) (Tables S18 and S19). Additionally, a single missense variant was associated only among South Asians (rs760568794, 11:120327605-G/A, p.Gly77Asp) with TG ($\beta = -36.9\%$, $p = 2 \times 10^{-8}$) (Table S9). This variant was present only among individuals with South Asian (MAC = 24) and Hispanic ancestry (MAC = 8) but showed no association among the Hispanic population ($p = 0.86$). This gene encodes a transmembrane protein of unknown function.

Replication of gene-based associations

We performed replication by using the PMBB and UKB samples that did not contribute to the initial analysis. In PMBB, we observed four out of ten genes without prior evidence of gene-based links with blood lipid phenotypes to have a $p < 0.005$ (Bonferroni correction for testing ten genes) and in the same direction as the discovery (*SRSF2*, *CREB3L3*, *PLA2G12A*, *PPARG*) with their respective blood lipids with an additional two genes that met a nominal significance level ($p < 0.05$; *JAK2* and *NR1H3*). For *TMEM136*, we found an association of nominal significance for TG and TG:HDL as well but with a beta in the opposite and positive direction. For the other three genes, *ALB*, *VARS*, and *STAB1*, we did not find associations at a nominal significance level for their respective blood lipid traits (Table S28). In UKB, we found six of the ten genes were associated at a $p < 0.005$ and in the same direction of effect as the discovery analysis (*ALB*, *CREB3L3*, *NR1H3*, *PLA2G12A*, *PPARG*, *STAB1*) (Table S29) with *JAK2* reaching a nominal significance threshold ($p < 0.05$). The only two genes that did not show any evidence of replication in at least one of the replication studies were *TMEM136* and *VARS*. This may indicate these associations are false positives or that we lack power for replication for these associations. Our replication studies did not include individuals of South Asian ancestry, and we observed that our association of *TMEM136* with TG and TG:HDL is driven by individuals of South Asian ancestry.

Comparison of gene-based associations by case status

We analyzed heterogeneity by CAD or T2D case status for the top 35 genes. The top 85 signals presented in Table S13 determined in case-status-specific meta-analyses for CAD and T2D. Out of the 85 different gene-based associations, we observed minimal heterogeneity in the results by case status. *LDLR*, *LCAT*, and *LPL* showed significant heterogeneity by CAD case status and *LCAT* and *ANGPTL4* by T2D status ($p_{Het} < 6 \times 10^{-4}$) (Tables S30 and S31).

Gene-based associations in GWAS loci

We determined whether genes near lipid array-based GWAS signals[8] were associated with the corresponding lipid measure by using gene-based tests of rare variants with the same traits. We obtained genes from 200 kb flanking regions on both sides of each GWAS signal: 487 annotated to LDL-C GWAS signals, 531 to HDL-C signals, and 471 to TG signals. We analyzed genes within these three sets for gene-based associations with their associated traits. A total of 13, 19, and 13 genes were associated ($p < 3.4 \times 10^{-5}$, corrected for the number of genes tested for the three traits) with LDL-C, HDL-C, or TG, and 32 unique genes were identified in the GWAS loci (Tables S32–S37).

Three of the 32 genes had no prior aggregate rare variant evidence of blood lipid association. Variants annotated as LOF or predicted damaging in *EVI5* were associated with LDL-C ($p_{SKAT} = 2 \times 10^{-5}$). The burden test showed association with higher LDL-C levels ($\beta = 1.9$ mg/dL, $p = 0.008$) (Table S32). Variants annotated as LOF or predicted damaging in *SH2B3* were associated with lower HDL-C ($\beta = -2.5$ mg/dL, $p = 1 \times 10^{-6}$) among Europeans, and variants that were annotated as LOF in *PLIN1* were associated with higher HDL-C ($\beta = 3.9$ mg/dL, $p = 1 \times 10^{-5}$) (Table S33). Other genes in the regions of *EVI5*, *SH2B3*, and *PLIN1* did not show an association with the corresponding lipid traits ($p > 0.05$) in multi-ancestry analyses. A previous report implicated two heterozygous frameshift mutations in *PLIN1* in three families with partial lipodystrophy.[65] The gene encodes perilipin, the most abundant protein that coats adipocyte lipid droplets and is critical for optimal TG storage.[66] We observed a nominal associations of *PLIN1* with TG ($\beta = -7.0\%$, $p = 0.02$). Our finding is contrary to what would be expected with hypertriglyceridemia in a lipodystrophy phenotype given the association with lower TG. This gene has an additional role where silencing in cow adipocytes has been shown to inhibit TG synthesis and promote lipolysis,[67] which may explain those contradictions.

Enrichment of Mendelian, GWAS, and drug targets genes

We next sought to test the utility of genes that showed some evidence for association but did not reach exome-wide significance. Within the genes that reached a sub-threshold level of significant association in this study via SKATs or burden tests ($p < 0.005$), we determined the enrichment of (1) Mendelian dyslipidemia (N = 21 genes);[2] (2) lipid GWAS (N = 487 for LDL-C, N = 531 for HDL-C, and N = 471 for TG);[8] and (3) drug target genes (N = 53).[43] We stratified genes in GWAS loci according to coding status of the index SNP and proximity to the index SNP (nearest gene,

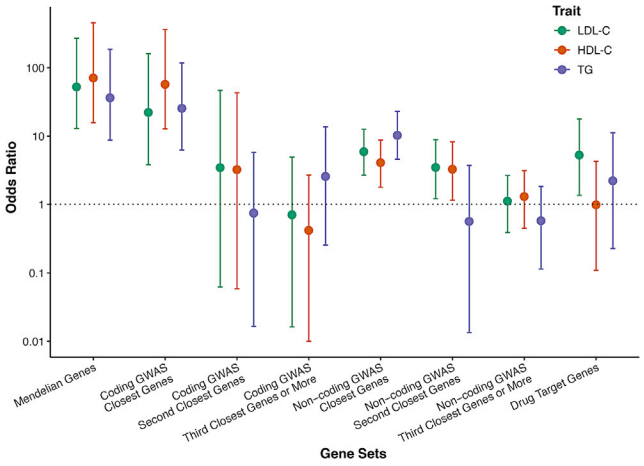

Figure 3. Enrichment of Mendelian, GWAS, and drug target genes in the gene-based lipid associations

Enrichment of gene sets of Mendelian genes (n = 21), GWAS loci for LDL-C (n = 487), HDL-C (n = 531), and triglycerides (TG) (n = 471) genes, and drug target genes (n = 53). Error bars denote 95% confidence intervals.

second nearest gene, and genes further away). We tested for enrichment of gene-based signals ($p < 0.005$) in the gene sets compared to matched genes (Figure 3). For each gene within each gene set, the most significant association in the multi-ancestry or an ancestry-specific analysis was obtained and then matched to ten genes on the basis of sample size, total number of variants, cumulative MAC, and variant grouping. The strongest enrichment was observed for Mendelian dyslipidemia genes within the genes that reached $p < 0.005$ in our study. For example, 52% of the HDL-C Mendelian genes versus 1.4% of the matched set reached $p < 0.005$ (OR: 71, 95% CI: 16–455). We also observed that 45.5% of the set of genes closest to an HDL-C protein-altering GWAS variant reached $p < 0.005$ versus 1.4% in the matched gene set (OR: 57, 95% CI: 13–362). Results were significant but much less striking for genes at non-coding index variants. We observed that 8.9% of the set of genes closest to an HDL-C non-protein-altering GWAS variant reached $p < 0.005$ versus 2.3% in the matched set (OR: 4.1, 95% CI: 1.8–8.7), while 8% of the set of genes in the second closest to an HDL-C non-protein-altering GWAS variant reached $p < 0.005$ versus 2.6% in the matched set (OR: 3, 95% CI: 1.1–8.3). There was no significant enrichment in second closest or ≥ third closest genes to protein-altering GWAS signals and in ≥ third closest genes to non-protein-altering GWAS signals. Drug target genes were significantly enriched in LDL-C gene-based associations (OR: 5.3, 95% CI: 1.4–17.8) but not in TG (OR: 2.2, 95% CI: 0.2–11.2) or HDL-C (OR: 1.0, 95% CI: 0.1–4.3) (Figure 3 and Tables S38–S41).

Association of lipid genes with CAD, T2D, glycemic traits, and liver enzymes

We tested the genes identified through our discovery (35 genes) and GWAS loci genes (32 genes) for associations with CAD or T2D in our gene-based analyses (40 genes across the two sets). The CAD analyses were restricted to a subset of the overall exome sequence data with information on CAD status, which included the MIGen CAD case-control, UKB CAD nested case-control, and the UKB cohort with a total of 32,981 cases and 79,879 controls. We observed four genes significantly associated with CAD ($p_{CAD} < 0.00125$, corrected for 40 genes). The four genes associated with lipids and CAD were all primarily associated with LDL-C: *LDLR* (OR: 2.97, $p = 7 \times 10^{-24}$), *APOB* ($p_{SKAT} = 4 \times 10^{-5}$), *PCSK9* (OR: 0.5, $p = 2 \times 10^{-4}$), and *JAK2* ($p_{SKAT} = 0.001$). Several other known CAD-associated genes (*NPC1L1*, *CETP*, *APOC3*, and *LPL*) showed nominal significance for association with lipids ($p < 0.05$). We observed nominal associations with CAD for two of the newly identified lipid genes: *PLIN1* ($p_{SKAT} = 0.002$) and *EVI5* (OR: 1.29, $p = 0.002$; Table S42). None of the 40 lipid genes reached significance for association with T2D in the latest AMP-T2D exome sequence results. We observed nominal associations of T2D with *STAB1* (OR: 1.05, $p_{T2D} = 0.002$) and *APOB* (OR: 1.08, $p_{T2D} = 0.005$) (Table S43).[15]

We additionally tested the 40 genes for association with six glycemic and liver biomarkers in the UKB: blood glucose, HbA1c, alanine aminotransferase (ALT), aspartate aminotransferase (AST), gamma glutamyl transferase (GGT), and albumin (Tables S44–S49). Using a significance threshold of $p = 0.0012$, we found associations between *PDE3B* and elevated blood glucose, *JAK2* and *SH2B3* and lower HbA1c, and *APOC3* and higher HbA1c. However, JAK2 was no longer associated with Hba1c after removal of the p.Val617Phe missense variant that is known to frequently occur as a somatic mutation ($\beta = 0.22$, SE = 0.40, $p = 0.47$).

We found associations between *CREB3L3* and lower ALT and ALB and higher AST and between *A1CF* and higher GGT. *ALB* and *SRSF2* were associated with lower and higher albumin levels, respectively (Tables S44–S49).

Discussion

We conducted a large multi-ancestry study to identify genes in which protein-altering variants demonstrated association with blood lipid levels. First, we confirm previous associations of genes with blood lipid levels and show that we detect associations across multiple ancestries. Second, we identified gene-based associations that were not observed previously. Third, we show that along with Mendelian lipid genes, the genes closest to both protein-altering and non-protein-altering GWAS signals and LDL-C drug target genes have the highest enrichment of gene-based associations. Fourth, of the new gene-based lipid associations, *PLIN1* and *EVI5* showed suggestive evidence of an association with CAD.

Our study found evidence of gene-based associations for the same gene in multiple ancestries. The heterogeneity in

genetic association with common traits and complex diseases has been discussed extensively. A recent study has shown significant heterogeneity across different ancestries in the effect sizes of multiple GWAS-identified variants.[68] However, our study shows that gene-based signals are detected in multiple ancestries with limited heterogeneity in the effect sizes.

Our study highlights enrichment of gene-based associations for Mendelian dyslipidemia genes, genes with protein-altering variants identified by GWASs, and genes that are closest to non-protein-altering GWAS index variants. A previous transcriptome-wide Mendelian randomization study of eQTL variants indicated that most of the genes closest to top GWAS signals (>71%) do not show significant association with the respective phenotype.[69] In contrast, our study provides evidence from sequence data that the closest gene to each top non-coding GWAS signal is most likely to be the causal one, indicating an allelic series in associated loci. This has implications for GWAS results, suggesting the prioritization of the closest genes for follow-up studies. We also observed enrichment of drug target genes only among LDL-C gene-based associations and not for HDL-C and TG gene-based associations, consistent with the fact that most approved therapeutics for cardiovascular disease target LDL-C

The gene-based analyses of lipid genes with CAD confirmed previously reported and known associations (*LDLR*, *APOB*, and *PCSK9)*. Using a nominal p threshold of 0.05, we also confirmed associations with *NPC1L1*, *CETP*, *APOC3*, and *LPL*. Of the identified lipid-associated genes, we observed borderline significant signals with *EVI5* and higher risk of CAD and between *PLIN1* and lower risk of CAD. The putative cardio-protective role of PLIN1 deficiency is supported by previous evidence in mice, which has indicated reduced atherosclerotic lesions with Plin1 deficiency in bone-marrow-derived cells.[70] This suggests *PLIN1* as a putative target for CAD prevention; however, replication of the CAD association would be needed for confirmation of those signals.

There are limitations to our results. First, we had lower sample sizes for the non-European ancestries, limiting our power to detect ancestry-specific associations and detect replication for *TMEM136* that was driven by a variant in South Asians. However, we find consistency of results across ancestries, and when we relax our significance threshold, the majority of associations (59%–89%) are observed in more than one ancestry. Second, it has been reported that there was an issue with the UKB functionally equivalent WES calling.[71] This mapping issue may have resulted in under-calling alternative alleles and therefore should not increase false positive findings. Third, we relied on a meta-analysis approach by using summary statistics to perform our gene-based testing because of differences in sequencing platforms and genotyping calling within the multiple consortia contributing to the results. This approach has been shown to be equivalent to a pooled approach for continuous outcomes.[41]

In summary, we demonstrated association between rare protein-altering variants with circulating lipid levels in >170,000 individuals of diverse ancestries. We identified 35 genes associated with blood lipids, including ten genes not previously shown to have gene-based signals. Our results support the hypothesis that genes closest to a GWAS index SNP are enriched for evidence of association.

Data and code availability

Controlled access of the individual-level data is available through dbGAP (please refer to the supplemental information), and the individual-level UK Biobank data are available upon application to the UK Biobank. Summary association results are available on the downloads page of the Cardiovascular Disease Knowledge Portal (broadcvdi.org).

Supplemental information

Supplemental information can be found online at https://doi.org/10.1016/j.ajhg.2021.11.021.

Acknowledgements

This work was supported by a grant from the Swedish Research Council (2016-06830) and grants from the National Heart, Lung, and Blood Institute (NHLBI): R01HL142711 and R01HL127564. Please refer to the supplemental information for the full acknowledgements.

Declaration of interests

The authors declare no competing interests for the present work. P.N. reports investigator-initiated grants from Amgen, Apple, and Boston Scientific; is a scientific advisor to Apple, Blackstone Life Sciences, and Novartis; and has spousal employment at Vertex, all unrelated to the present work. A.V.K. has served as a scientific advisor to Sanofi, Medicines Company, Maze Pharmaceuticals, Navitor Pharmaceuticals, Verve Therapeutics, Amgen, and Color; received speaking fees from Illumina, MedGenome, Amgen, and the Novartis Institute for Biomedical Research; received sponsored research agreements from the Novartis Institute for Biomedical Research and IBM Research; and reports a patent related to a genetic risk predictor (20190017119). C.J.W.'s spouse is employed at Regeneron. L.E.S. is currently an employee of Celgene/Bristol Myers Squibb. Celgene/Bristol Myers Squibb had no role in the funding, design, conduct, and interpretation of this study. M.E.M. receives funding from Regeneron unrelated to this work. E.E.K. has received speaker honoraria from Illumina, Inc and Regeneron Pharmaceuticals. B.M.P. serves on the Steering Committee of the Yale Open Data Access Project funded by Johnson & Johnson. L.A.C. has consulted with the Dyslipidemia Foundation on lipid projects in the Framingham Heart Study. P.T.E. is supported by a grant from Bayer AG to the Broad Institute focused on the genetics and therapeutics of cardiovascular disease. P.T.E. has consulted for Bayer AG, Novartis, MyoKardia, and Quest Diagnostics. S.A.L. receives sponsored research support from Bristol Myers Squibb/Pfizer, Bayer AG, Boehringer Ingelheim, Fitbit, and IBM and has consulted for Bristol Myers Squibb/Pfizer, Bayer AG, and Blackstone Life Sciences. The views expressed in this article are those of the author(s) and not necessarily those of the NHS,

the NIHR, or the Department of Health. M.I.M. has served on advisory panels for Pfizer, NovoNordisk, and Zoe Global and has received honoraria from Merck, Pfizer, Novo Nordisk, and Eli Lilly and research funding from Abbvie, Astra Zeneca, Boehringer Ingelheim, Eli Lilly, Janssen, Merck, NovoNordisk, Pfizer, Roche, Sanofi Aventis, Servier, and Takeda. As of June 2019, M.I.M. is an employee of Genentech and a holder of Roche stock. M.E.J. holds shares in Novo Nordisk A/S. H.M.K. is an employee of Regeneron Pharmaceuticals; he owns stock and stock options for Regeneron Pharmaceuticals. M.E.J. has received research grants form Astra Zeneca, Boehringer Ingelheim, Amgen, and Sanofi. S.K. is founder of Verve Therapeutics.

Received: August 18, 2021
Accepted: November 21, 2021
Published: December 20, 2021

References

1. Di Angelantonio, E., Sarwar, N., Perry, P., Kaptoge, S., Ray, K.K., Thompson, A., Wood, A.M., Lewington, S., Sattar, N., Packard, C.J., et al. (2009). Major lipids, apolipoproteins, and risk of vascular disease. JAMA *302*, 1993–2000.
2. Teslovich, T.M., Musunuru, K., Smith, A.V., Edmondson, A.C., Stylianou, I.M., Koseki, M., Pirruccello, J.P., Ripatti, S., Chasman, D.I., Willer, C.J., et al. (2010). Biological, clinical and population relevance of 95 loci for blood lipids. Nature *466*, 707–713.
3. Willer, C.J., Schmidt, E.M., Sengupta, S., Peloso, G.M., Gustafsson, S., Kanoni, S., Ganna, A., Chen, J., Buchkovich, M.L., Mora, S., et al. (2013). Discovery and refinement of loci associated with lipid levels. Nat. Genet. *45*, 1274–1283.
4. Chasman, D.I., Paré, G., Mora, S., Hopewell, J.C., Peloso, G., Clarke, R., Cupples, L.A., Hamsten, A., Kathiresan, S., Mälarstig, A., et al. (2009). Forty-three loci associated with plasma lipoprotein size, concentration, and cholesterol content in genome-wide analysis. PLoS Genet. *5*, e1000730.
5. Peloso, G.M., Auer, P.L., Bis, J.C., Voorman, A., Morrison, A.C., Stitziel, N.O., Brody, J.A., Khetarpal, S.A., Crosby, J.R., Fornage, M., et al. (2014). Association of low-frequency and rare coding-sequence variants with blood lipids and coronary heart disease in 56,000 whites and blacks. Am. J. Hum. Genet. *94*, 223–232.
6. Asselbergs, F.W., Guo, Y., van Iperen, E.P., Sivapalaratnam, S., Tragante, V., Lanktree, M.B., Lange, L.A., Almoguera, B., Appelman, Y.E., Barnard, J., et al. (2012). Large-scale gene-centric meta-analysis across 32 studies identifies multiple lipid loci. Am. J. Hum. Genet. *91*, 823–838.
7. Liu, D.J., Peloso, G.M., Yu, H., Butterworth, A.S., Wang, X., Mahajan, A., Saleheen, D., Emdin, C., Alam, D., Alves, A.C., et al. (2017). Exome-wide association study of plasma lipids in >300,000 individuals. Nat. Genet. *49*, 1758–1766.
8. Klarin, D., Damrauer, S.M., Cho, K., Sun, Y.V., Teslovich, T.M., Honerlaw, J., Gagnon, D.R., DuVall, S.L., Li, J., Peloso, G.M., et al. (2018). Genetics of blood lipids among ~300,000 multi-ethnic participants of the Million Veteran Program. Nat. Genet. *50*, 1514–1523.
9. Voight, B.F., Peloso, G.M., Orho-Melander, M., Frikke-Schmidt, R., Barbalic, M., Jensen, M.K., Hindy, G., Hólm, H., Ding, E.L., Johnson, T., et al. (2012). Plasma HDL cholesterol and risk of myocardial infarction: a mendelian randomisation study. Lancet *380*, 572–580.
10. Do, R., Willer, C.J., Schmidt, E.M., Sengupta, S., Gao, C., Peloso, G.M., Gustafsson, S., Kanoni, S., Ganna, A., Chen, J., et al. (2013). Common variants associated with plasma triglycerides and risk for coronary artery disease. Nat. Genet. *45*, 1345–1352.
11. Hindy, G., Engström, G., Larsson, S.C., Traylor, M., Markus, H.S., Melander, O., Orho-Melander, M.; and Stroke Genetics Network (SiGN) (2018). Role of Blood Lipids in the Development of Ischemic Stroke and its Subtypes: A Mendelian Randomization Study. Stroke *49*, 820–827.
12. Smith, J.G., Luk, K., Schulz, C.A., Engert, J.C., Do, R., Hindy, G., Rukh, G., Dufresne, L., Almgren, P., Owens, D.S., et al. (2014). Association of low-density lipoprotein cholesterol-related genetic variants with aortic valve calcium and incident aortic stenosis. JAMA *312*, 1764–1771.
13. Afshar, M., Luk, K., Do, R., Dufresne, L., Owens, D.S., Harris, T.B., Peloso, G.M., Kerr, K.F., Wong, Q., Smith, A.V., et al. (2017). Association of Triglyceride-Related Genetic Variants With Mitral Annular Calcification. J. Am. Coll. Cardiol. *69*, 2941–2948.
14. Dewey, F.E., Murray, M.F., Overton, J.D., Habegger, L., Leader, J.B., Fetterolf, S.N., O'Dushlaine, C., Van Hout, C.V., Staples, J., Gonzaga-Jauregui, C., et al. (2016). Distribution and clinical impact of functional variants in 50,726 whole-exome sequences from the DiscovEHR study. Science *354*, aaf6814.
15. Flannick, J., Mercader, J.M., Fuchsberger, C., Udler, M.S., Mahajan, A., Wessel, J., Teslovich, T.M., Caulkins, L., Koesterer, R., Barajas-Olmos, F., et al. (2019). Exome sequencing of 20,791 cases of type 2 diabetes and 24,440 controls. Nature *570*, 71–76.
16. Do, R., Stitziel, N.O., Won, H.H., Jørgensen, A.B., Duga, S., Angelica Merlini, P., Kiezun, A., Farrall, M., Goel, A., Zuk, O., et al. (2015). Exome sequencing identifies rare LDLR and APOA5 alleles conferring risk for myocardial infarction. Nature *518*, 102–106.
17. Pollin, T.I., Damcott, C.M., Shen, H., Ott, S.H., Shelton, J., Horenstein, R.B., Post, W., McLenithan, J.C., Bielak, L.F., Peyser, P.A., et al. (2008). A null mutation in human APOC3 confers a favorable plasma lipid profile and apparent cardioprotection. Science *322*, 1702–1705.
18. Crosby, J., Peloso, G.M., Auer, P.L., Crosslin, D.R., Stitziel, N.O., Lange, L.A., Lu, Y., Tang, Z.Z., Zhang, H., Hindy, G., et al. (2014). Loss-of-function mutations in APOC3, triglycerides, and coronary disease. N. Engl. J. Med. *371*, 22–31.
19. Jørgensen, A.B., Frikke-Schmidt, R., Nordestgaard, B.G., and Tybjærg-Hansen, A. (2014). Loss-of-function mutations in APOC3 and risk of ischemic vascular disease. N. Engl. J. Med. *371*, 32–41.
20. Musunuru, K., Pirruccello, J.P., Do, R., Peloso, G.M., Guiducci, C., Sougnez, C., Garimella, K.V., Fisher, S., Abreu, J., Barry, A.J., et al. (2010). Exome sequencing, ANGPTL3 mutations, and familial combined hypolipidemia. N. Engl. J. Med. *363*, 2220–2227.
21. Dewey, F.E., Gusarova, V., Dunbar, R.L., O'Dushlaine, C., Schurmann, C., Gottesman, O., McCarthy, S., Van Hout, C.V., Bruse, S., Dansky, H.M., et al. (2017). Genetic and Pharmacologic Inactivation of ANGPTL3 and Cardiovascular Disease. N. Engl. J. Med. *377*, 211–221.
22. Dewey, F.E., Gusarova, V., O'Dushlaine, C., Gottesman, O., Trejos, J., Hunt, C., Van Hout, C.V., Habegger, L., Buckler, D., Lai, K.M., et al. (2016). Inactivating Variants in ANGPTL4 and Risk of Coronary Artery Disease. N. Engl. J. Med. *374*, 1123–1133.

23. Cohen, J., Pertsemlidis, A., Kotowski, I.K., Graham, R., Garcia, C.K., and Hobbs, H.H. (2005). Low LDL cholesterol in individuals of African descent resulting from frequent nonsense mutations in PCSK9. Nat. Genet. *37*, 161–165.
24. Cohen, J.C., Boerwinkle, E., Mosley, T.H., Jr., and Hobbs, H.H. (2006). Sequence variations in PCSK9, low LDL, and protection against coronary heart disease. N. Engl. J. Med. *354*, 1264–1272.
25. Kathiresan, S.; and Myocardial Infarction Genetics Consortium (2008). A PCSK9 missense variant associated with a reduced risk of early-onset myocardial infarction. N. Engl. J. Med. *358*, 2299–2300.
26. Sabatine, M.S., Giugliano, R.P., Keech, A.C., Honarpour, N., Wiviott, S.D., Murphy, S.A., Kuder, J.F., Wang, H., Liu, T., Wasserman, S.M., et al. (2017). Evolocumab and Clinical Outcomes in Patients with Cardiovascular Disease. N. Engl. J. Med. *376*, 1713–1722.
27. Gaudet, D., Alexander, V.J., Baker, B.F., Brisson, D., Tremblay, K., Singleton, W., Geary, R.S., Hughes, S.G., Viney, N.J., Graham, M.J., et al. (2015). Antisense Inhibition of Apolipoprotein C-III in Patients with Hypertriglyceridemia. N. Engl. J. Med. *373*, 438–447.
28. Nomura, A., Emdin, C.A., Won, H.H., Peloso, G.M., Natarajan, P., Ardissino, D., Danesh, J., Schunkert, H., Correa, A., Bown, M.J., et al. (2020). Heterozygous *ABCG5* Gene Deficiency and Risk of Coronary Artery Disease. Circ. Genom. Precis. Med. *13*, 417–423.
29. Peloso, G.M., Nomura, A., Khera, A.V., Chaffin, M., Won, H.H., Ardissino, D., Danesh, J., Schunkert, H., Wilson, J.G., Samani, N., et al. (2019). Rare Protein-Truncating Variants in APOB, Lower Low-Density Lipoprotein Cholesterol, and Protection Against Coronary Heart Disease. Circ. Genom. Precis. Med. *12*, e002376.
30. Taliun, D., Harris, D.N., Kessler, M.D., Carlson, J., Szpiech, Z.A., Torres, R., Taliun, S.A.G., Corvelo, A., Gogarten, S.M., Kang, H.M., et al. (2019). Sequencing of 53,831 diverse genomes from the NHLBI TOPMed Program. bioRxiv. https://doi.org/10.1101/563866.
31. Natarajan, P., Peloso, G.M., Zekavat, S.M., Montasser, M., Ganna, A., Chaffin, M., Khera, A.V., Zhou, W., Bloom, J.M., Engreitz, J.M., et al. (2018). Deep-coverage whole genome sequences and blood lipids among 16,324 individuals. Nat. Commun. *9*, 3391.
32. Szustakowski, J.D., Balasubramanian, S., Kvikstad, E., Khalid, S., Bronson, P.G., Sasson, A., Wong, E., Liu, D., Wade Davis, J., Haefliger, C., et al. (2021). Advancing human genetics research and drug discovery through exome sequencing of the UK Biobank. Nat. Genet. *53*, 942–948.
33. Van Hout, C.V., Tachmazidou, I., Backman, J.D., Hoffman, J.D., Liu, D., Pandey, A.K., Gonzaga-Jauregui, C., Khalid, S., Ye, B., Banerjee, N., et al. (2020). Exome sequencing and characterization of 49,960 individuals in the UK Biobank. Nature *586*, 749–756.
34. Zhan, X., Hu, Y., Li, B., Abecasis, G.R., and Liu, D.J. (2016). RVTESTS: an efficient and comprehensive tool for rare variant association analysis using sequence data. Bioinformatics *32*, 1423–1426.
35. McLaren, W., Gil, L., Hunt, S.E., Riat, H.S., Ritchie, G.R., Thormann, A., Flicek, P., and Cunningham, F. (2016). The Ensembl Variant Effect Predictor. Genome Biol. *17*, 122.
36. Karczewski, K.J., Francioli, L.C., Tiao, G., Cummings, B.B., Alföldi, J., Wang, Q., Collins, R.L., Laricchia, K.M., Ganna, A., Birnbaum, D.P., et al. (2020). The mutational constraint spectrum quantified from variation in 141,456 humans. Nature *581*, 434–443.
37. Liu, X., Wu, C., Li, C., and Boerwinkle, E. (2016). dbNSFP v3.0: A One-Stop Database of Functional Predictions and Annotations for Human Nonsynonymous and Splice-Site SNVs. Hum. Mutat. *37*, 235–241.
38. Dong, C., Wei, P., Jian, X., Gibbs, R., Boerwinkle, E., Wang, K., and Liu, X. (2015). Comparison and integration of deleteriousness prediction methods for nonsynonymous SNVs in whole exome sequencing studies. Hum. Mol. Genet. *24*, 2125–2137.
39. Jaganathan, K., Kyriazopoulou Panagiotopoulou, S., McRae, J.F., Darbandi, S.F., Knowles, D., Li, Y.I., Kosmicki, J.A., Arbelaez, J., Cui, W., Schwartz, G.B., et al. (2019). Predicting Splicing from Primary Sequence with Deep Learning. Cell *176*, 535–548.e24.
40. Sveinbjornsson, G., Albrechtsen, A., Zink, F., Gudjonsson, S.A., Oddson, A., Másson, G., Holm, H., Kong, A., Thorsteinsdottir, U., Sulem, P., et al. (2016). Weighting sequence variants based on their annotation increases power of whole-genome association studies. Nat. Genet. *48*, 314–317.
41. Liu, D.J., Peloso, G.M., Zhan, X., Holmen, O.L., Zawistowski, M., Feng, S., Nikpay, M., Auer, P.L., Goel, A., Zhang, H., et al. (2014). Meta-analysis of gene-level tests for rare variant association. Nat. Genet. *46*, 200–204.
42. Wu, M.C., Lee, S., Cai, T., Li, Y., Boehnke, M., and Lin, X. (2011). Rare-variant association testing for sequencing data with the sequence kernel association test. Am. J. Hum. Genet. *89*, 82–93.
43. Wishart, D.S., Feunang, Y.D., Guo, A.C., Lo, E.J., Marcu, A., Grant, J.R., Sajed, T., Johnson, D., Li, C., Sayeeda, Z., et al. (2018). DrugBank 5.0: a major update to the DrugBank database for 2018. Nucleic Acids Res. *46* (D1), D1074–D1082.
44. McLaughlin, T., Abbasi, F., Cheal, K., Chu, J., Lamendola, C., and Reaven, G. (2003). Use of metabolic markers to identify overweight individuals who are insulin resistant. Ann. Intern. Med. *139*, 802–809.
45. Li, C., Ford, E.S., Meng, Y.X., Mokdad, A.H., and Reaven, G.M. (2008). Does the association of the triglyceride to high-density lipoprotein cholesterol ratio with fasting serum insulin differ by race/ethnicity? Cardiovasc. Diabetol. *7*, 4.
46. Chan, E., Tan, C.S., Deurenberg-Yap, M., Chia, K.S., Chew, S.K., and Tai, E.S. (2006). The V227A polymorphism at the PPARA locus is associated with serum lipid concentrations and modulates the association between dietary polyunsaturated fatty acid intake and serum high density lipoprotein concentrations in Chinese women. Atherosclerosis *187*, 309–315.
47. van der Harst, P., and Verweij, N. (2018). Identification of 64 Novel Genetic Loci Provides an Expanded View on the Genetic Architecture of Coronary Artery Disease. Circ. Res. *122*, 433–443.
48. Minchiotti, L., Galliano, M., Caridi, G., Kragh-Hansen, U., and Peters, T., Jr. (2013). Congenital analbuminaemia: molecular defects and biochemical and clinical aspects. Biochim. Biophys. Acta *1830*, 5494–5502.
49. Koot, B.G., Houwen, R., Pot, D.J., and Nauta, J. (2004). Congenital analbuminaemia: biochemical and clinical implications. A case report and literature review. Eur. J. Pediatr. *163*, 664–670.

50. Cheng, Y., Luo, C., Wu, W., Xie, Z., Fu, X., and Feng, Y. (2016). Liver-Specific Deletion of SRSF2 Caused Acute Liver Failure and Early Death in Mice. Mol. Cell. Biol. *36*, 1628–1638.
51. Cefalù, A.B., Spina, R., Noto, D., Valenti, V., Ingrassia, V., Giammanco, A., Panno, M.D., Ganci, A., Barbagallo, C.M., and Averna, M.R. (2015). Novel CREB3L3 Nonsense Mutation in a Family With Dominant Hypertriglyceridemia. Arterioscler. Thromb. Vasc. Biol. *35*, 2694–2699.
52. Lee, J.H., Giannikopoulos, P., Duncan, S.A., Wang, J., Johansen, C.T., Brown, J.D., Plutzky, J., Hegele, R.A., Glimcher, L.H., and Lee, A.H. (2011). The transcription factor cyclic AMP-responsive element-binding protein H regulates triglyceride metabolism. Nat. Med. *17*, 812–815.
53. Dron, J.S., Dilliott, A.A., Lawson, A., McIntyre, A.D., Davis, B.D., Wang, J., Cao, H., Movsesyan, I., Malloy, M.J., Pullinger, C.R., et al. (2020). Loss-of-Function *CREB3L3* Variants in Patients With Severe Hypertriglyceridemia. Arterioscler. Thromb. Vasc. Biol. *40*, 1935–1941.
54. Becares, N., Gage, M.C., Voisin, M., Shrestha, E., Martin-Gutierrez, L., Liang, N., Louie, R., Pourcet, B., Pello, O.M., Luong, T.V., et al. (2019). Impaired LXRalpha Phosphorylation Attenuates Progression of Fatty Liver Disease. Cell Rep. *26*, 984–995.e6.
55. Zhao, C., and Dahlman-Wright, K. (2010). Liver X receptor in cholesterol metabolism. J. Endocrinol. *204*, 233–240.
56. Hong, C., and Tontonoz, P. (2014). Liver X receptors in lipid metabolism: opportunities for drug discovery. Nat. Rev. Drug Discov. *13*, 433–444.
57. Jaworski, K., Ahmadian, M., Duncan, R.E., Sarkadi-Nagy, E., Varady, K.A., Hellerstein, M.K., Lee, H.Y., Samuel, V.T., Shulman, G.I., Kim, K.H., et al. (2009). AdPLA ablation increases lipolysis and prevents obesity induced by high-fat feeding or leptin deficiency. Nat. Med. *15*, 159–168.
58. Quach, N.D., Arnold, R.D., and Cummings, B.S. (2014). Secretory phospholipase A2 enzymes as pharmacological targets for treatment of disease. Biochem. Pharmacol. *90*, 338–348.
59. Barroso, I., Gurnell, M., Crowley, V.E., Agostini, M., Schwabe, J.W., Soos, M.A., Maslen, G.L., Williams, T.D., Lewis, H., Schafer, A.J., et al. (1999). Dominant negative mutations in human PPARgamma associated with severe insulin resistance, diabetes mellitus and hypertension. Nature *402*, 880–883.
60. Agostini, M., Schoenmakers, E., Mitchell, C., Szatmari, I., Savage, D., Smith, A., Rajanayagam, O., Semple, R., Luan, J., Bath, L., et al. (2006). Non-DNA binding, dominant-negative, human PPARgamma mutations cause lipodystrophic insulin resistance. Cell Metab. *4*, 303–311.
61. Majithia, A.R., Flannick, J., Shahinian, P., Guo, M., Bray, M.A., Fontanillas, P., Gabriel, S.B., Rosen, E.D., Altshuler, D.; GoT2D Consortium; NHGRI JHS/FHS Allelic Spectrum Project; SIGMA T2D Consortium; and T2D-GENES Consortium (2014). Rare variants in PPARG with decreased activity in adipocyte differentiation are associated with increased risk of type 2 diabetes. Proc. Natl. Acad. Sci. USA *111*, 13127–13132.
62. Goerdt, S., Walsh, L.J., Murphy, G.F., and Pober, J.S. (1991). Identification of a novel high molecular weight protein preferentially expressed by sinusoidal endothelial cells in normal human tissues. J. Cell Biol. *113*, 1425–1437.
63. Li, R., Oteiza, A., Sørensen, K.K., McCourt, P., Olsen, R., Smedsrød, B., and Svistounov, D. (2011). Role of liver sinusoidal endothelial cells and stabilins in elimination of oxidized low-density lipoproteins. Am. J. Physiol. Gastrointest. Liver Physiol. *300*, G71–G81.
64. Tate, J.G., Bamford, S., Jubb, H.C., Sondka, Z., Beare, D.M., Bindal, N., Boutselakis, H., Cole, C.G., Creatore, C., Dawson, E., et al. (2019). COSMIC: the Catalogue Of Somatic Mutations In Cancer. Nucleic Acids Res. *47* (D1), D941–D947.
65. Gandotra, S., Le Dour, C., Bottomley, W., Cervera, P., Giral, P., Reznik, Y., Charpentier, G., Auclair, M., Delépine, M., Barroso, I., et al. (2011). Perilipin deficiency and autosomal dominant partial lipodystrophy. N. Engl. J. Med. *364*, 740–748.
66. Brasaemle, D.L., Subramanian, V., Garcia, A., Marcinkiewicz, A., and Rothenberg, A. (2009). Perilipin A and the control of triacylglycerol metabolism. Mol. Cell. Biochem. *326*, 15–21.
67. Zhang, S., Liu, G., Xu, C., Liu, L., Zhang, Q., Xu, Q., Jia, H., Li, X., and Li, X. (2018). Perilipin 1 Mediates Lipid Metabolism Homeostasis and Inhibits Inflammatory Cytokine Synthesis in Bovine Adipocytes. Front. Immunol. *9*, 467.
68. Wojcik, G.L., Graff, M., Nishimura, K.K., Tao, R., Haessler, J., Gignoux, C.R., Highland, H.M., Patel, Y.M., Sorokin, E.P., Avery, C.L., et al. (2019). Genetic analyses of diverse populations improves discovery for complex traits. Nature *570*, 514–518.
69. Porcu, E., Rüeger, S., Lepik, K., Santoni, F.A., Reymond, A., Kutalik, Z.; eQTLGen Consortium; and BIOS Consortium (2019). Mendelian randomization integrating GWAS and eQTL data reveals genetic determinants of complex and clinical traits. Nat. Commun. *10*, 3300.
70. Zhao, X., Gao, M., He, J., Zou, L., Lyu, Y., Zhang, L., Geng, B., Liu, G., and Xu, G. (2015). Perilipin1 deficiency in whole body or bone marrow-derived cells attenuates lesions in atherosclerosis-prone mice. PLoS ONE *10*, e0123738.
71. Jia, T., Munson, B., Lango Allen, H., Ideker, T., and Majithia, A.R. (2020). Thousands of missing variants in the UK Biobank are recoverable by genome realignment. Ann. Hum. Genet. *84*, 214–220.

The American Journal of Human Genetics, Volume *109*

Supplemental information

Rare coding variants in 35 genes associate with circulating lipid levels—A multi-ancestry analysis of 170,000 exomes

George Hindy, Peter Dornbos, Mark D. Chaffin, Dajiang J. Liu, Minxian Wang, Margaret Sunitha Selvaraj, David Zhang, Joseph Park, Carlos A. Aguilar-Salinas, Lucinda Antonacci-Fulton, Diego Ardissino, Donna K. Arnett, Stella Aslibekyan, Gil Atzmon, Christie M. Ballantyne, Francisco Barajas-Olmos, Nir Barzilai, Lewis C. Becker, Lawrence F. Bielak, Joshua C. Bis, John Blangero, Eric Boerwinkle, Lori L. Bonnycastle, Erwin Bottinger, Donald W. Bowden, Matthew J. Bown, Jennifer A. Brody, Jai G. Broome, Noël P. Burtt, Brian E. Cade, Federico Centeno-Cruz, Edmund Chan, Yi-Cheng Chang, Yii-Der I. Chen, Ching-Yu Cheng, Won Jung Choi, Rajiv Chowdhury, Cecilia Contreras-Cubas, Emilio J. Córdova, Adolfo Correa, L. Adrienne Cupples, Joanne E. Curran, John Danesh, Paul S. de Vries, Ralph A. DeFronzo, Harsha Doddapaneni, Ravindranath Duggirala, Susan K. Dutcher, Patrick T. Ellinor, Leslie S. Emery, Jose C. Florez, Myriam Fornage, Barry I. Freedman, Valentin Fuster, Ma. Eugenia Garay-Sevilla, Humberto García-Ortiz, Soren Germer, Richard A. Gibbs, Christian Gieger, Benjamin Glaser, Clicerio Gonzalez, Maria Elena Gonzalez-Villalpando, Mariaelisa Graff, Sarah E. Graham, Niels Grarup, Leif C. Groop, Xiuqing Guo, Namrata Gupta, Sohee Han, Craig L. Hanis, Torben Hansen, Jiang He, Nancy L. Heard-Costa, Yi-Jen Hung, Mi Yeong Hwang, Marguerite R. Irvin, Sergio Islas-Andrade, Gail P. Jarvik, Hyun Min Kang, Sharon L.R. Kardia, Tanika Kelly, Eimear E. Kenny, Alyna T. Khan, Bong-Jo Kim, Ryan W. Kim, Young Jin Kim, Heikki A. Koistinen, Charles Kooperberg, Johanna Kuusisto, Soo Heon Kwak, Markku Laakso, Leslie A. Lange, Jiwon Lee, Juyoung Lee, Seonwook Lee, Donna M. Lehman, Rozenn N. Lemaitre, Allan Linneberg, Jianjun Liu, Ruth J.F. Loos, Steven A. Lubitz, Valeriya Lyssenko, Ronald C.W. Ma, Lisa Warsinger Martin, Angélica Martínez-Hernández, Rasika A. Mathias, Stephen T. McGarvey, Ruth McPherson, James B. Meigs, Thomas Meitinger, Olle Melander, Elvia Mendoza-Caamal, Ginger A. Metcalf, Xuenan Mi, Karen L. Mohlke, May E. Montasser, Jee-Young Moon, Hortensia Moreno-Macías, Alanna C. Morrison, Donna M. Muzny, Sarah C. Nelson, Peter M. Nilsson, Jeffrey R. O'Connell, Marju Orho-Melander, Lorena Orozco, Colin N.A. Palmer, Nicholette D. Palmer, Cheol Joo Park, Kyong Soo Park, Oluf Pedersen, Juan M. Peralta, Patricia A. Peyser, Wendy S. Post, Michael Preuss, Bruce M. Psaty, Qibin Qi, D.C. Rao, Susan Redline, Alexander P. Reiner, Cristina Revilla-Monsalve, Stephen S. Rich, Nilesh Samani, Heribert Schunkert, Claudia Schurmann, Daekwan Seo, Jeong-Sun Seo, Xueling Sim, Rob Sladek, Kerrin S. Small, Wing Yee So, Adrienne M. Stilp, E. Shyong Tai, Claudia H.T. Tam, Kent D. Taylor, Yik Ying Teo, Farook Thameem, Brian Tomlinson, Michael Y. Tsai, Tiinamaija Tuomi, Jaakko Tuomilehto, Teresa Tusié-Luna, Miriam S. Udler, Rob M. van Dam, Ramachandran S. Vasan, Karine A. Viaud Martinez, Fei Fei Wang, Xuzhi Wang, Hugh Watkins, Daniel E. Weeks, James G.

The American Journal of Human Genetics, Volume *109*

Supplemental information

Rare coding variants in 35 genes associate with circulating lipid levels—A multi-ancestry analysis of 170,000 exomes

Hindy, et al.

Figure S1. Comparison of effect sizes and p-value in UK Biobank including and excluding individuals on statin treatment.

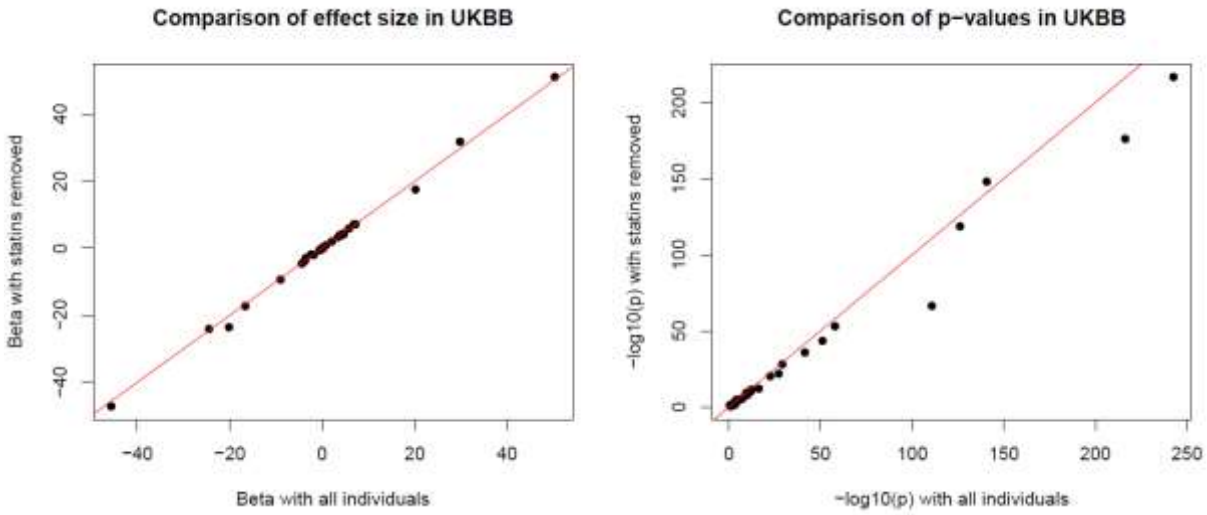

Figure S2. Descriptive variant characteristics by type, ancestry, and minor allele count.

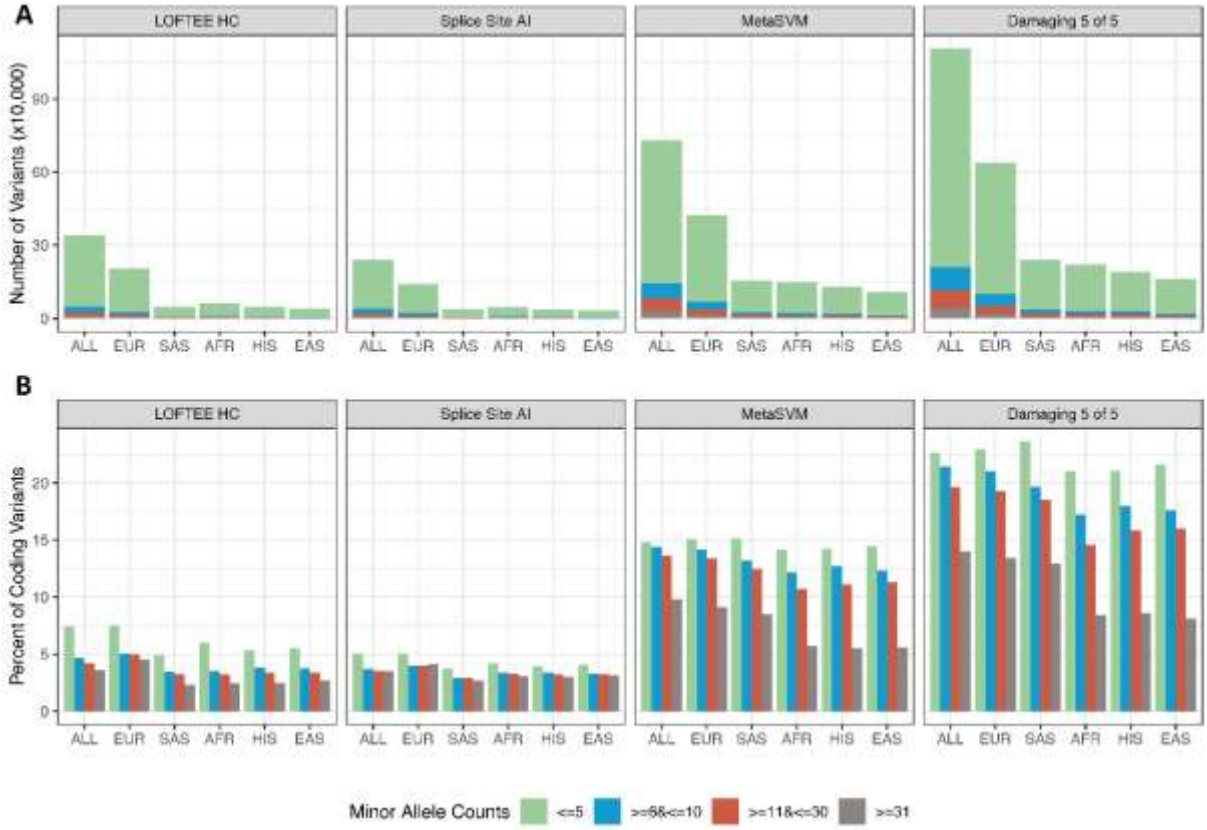

Figure S2: A) Our study included 15,599,513 genetic variants. Variants were annotated as high confidence loss-of-function by LOFTEE (n=340,214), splice site altering variants using a deep neural network prediction (SPLICE AI) (n=238,646), damaging missense variants according to the MetaSVM algorithm (n=729,098) and damaging missing in 5 out of 5 prediction algorithms (n=1,106,309). Most of the variants had a minor allele count of less than 5 in all (n=1,171,5189) and within each of the four different annotations. **B)** The proportion of specific annotations out of the total number of variants that were annotated as coding (n=5,085,712). Each of the four annotations demonstrated the highest enrichment among the variants with the lowest frequency. ALL=multi-ancestry, AFR=African ancestry, EAS=East Asian ancestry, EUR=European ancestry, HIS=Hispanic ancestry, SAS=South Asian ancestry.

Figure S3. Overlap among different ancestries for all variants contributing to significant gene-based associations with HDL-C, TG and LDL-C

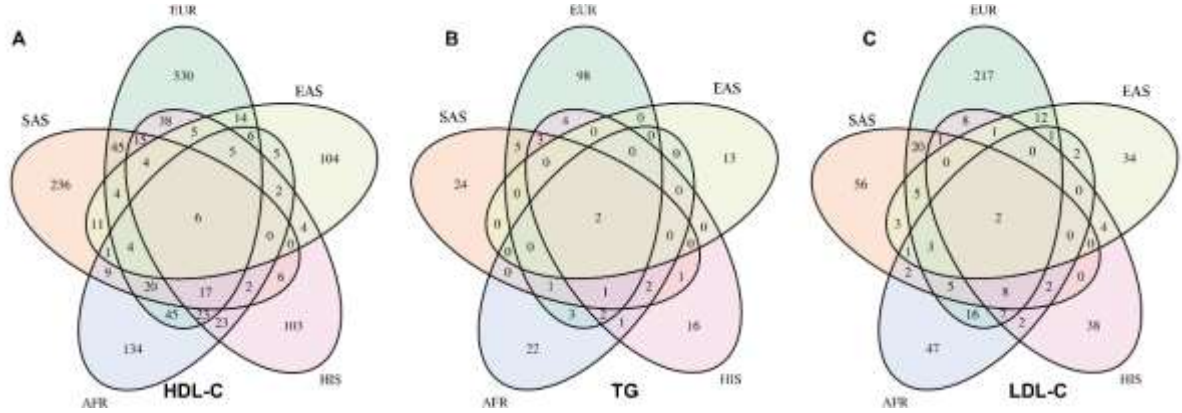

Figure S3. Venn Diagram for the overlap of all variants included in the significant gene-based association analysis among different ancestries. **A)** A total of 4 genes (*CETP*, *ABCA1*, *CD36*, and *LCAT*) showed significant gene-based associations in multi-ancestry and each of the 5 different ancestries analyses with P < 0.05 with HDL cholesterol (HDL-C). Ancestry-specific single-variant contributions included a total of 781 from European-, 380 from South Asian-, 302 African-, 253 Hispanic- and 175 East Asian ancestries. **B)** A total of 3 genes (*APOC3*, *ANGPTL3*, and *APOB*) showed significant gene-based associations in multi-ancestry and each of the 5 different ancestries analyses with P < 0.05 with triglycerides (TG). Ancestry-specific single-variant contributions included a total of 119 from European-, 39 from South Asian-, 34 African-, 32 Hispanic- and 15 East Asian ancestries. **C)** A total of 3 genes (*LDLR*, *PCSK9*, and *APOB*) showed significant gene-based associations in multi-ancestry and each of the 5 different ancestries analyses with P < 0.05 with LDL cholesterol (LDL-C). Ancestry-specific single-variant contributions included a total of 306 from European-, 108 from South Asian-, 98 African-, 73 Hispanic- and 68 East Asian ancestries.

Figure S4. Overlap among different ancestries for variants contributing to significant gene-based associations with HDL-C, TG and LDL-C with P value less than 0.05

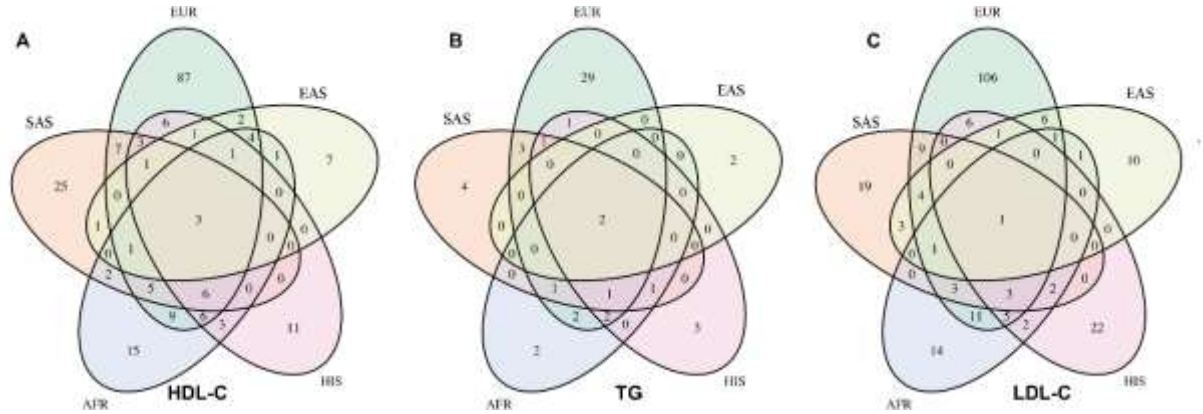

Figure S4. Venn Diagram for the overlap of variants with P<0.05 included in the significant gene-based association analysis among different ancestries. **A)** A total of 4 genes (*CETP*, *ABCA1*, *CD36*, and *LCAT*) showed significant gene-based associations in multi-ancestry and each of the 5 different ancestries analyses with P < 0.05 with HDL cholesterol (HDL-C). Ancestry-specific single-variant contributions included a total of 142 from European-, 54 from South Asian-, 56 African-, 41 Hispanic- and 22 East Asian ancestries. **B)** A total of 3 genes (*APOC3*, *ANGPTL3*, and *APOB*) showed significant gene-based associations in multi-ancestry analyses with triglycerides (TG). Ancestry-specific single-variant contributions included a total of 42 from European-, 13 from South Asian-, 11 African-, 11 Hispanic- and 4 East Asian ancestries. **C)** A total of 3 genes (*LDLR*, *PCSK9*, and *APOB*) showed significant gene-based associations in multi-ancestry and each of the 5 different ancestries analyses with P < 0.05 with LDL cholesterol (LDL-C). Ancestry-specific single-variant contributions included a total of 157 from European-, 45 from South Asian-, 44 African-, 42 Hispanic- and 28 East Asian ancestries.

Figure S5. Overlap among different ancestries for the top variant contributing to significant gene-based associations with HDL-C, TG and LDL-C

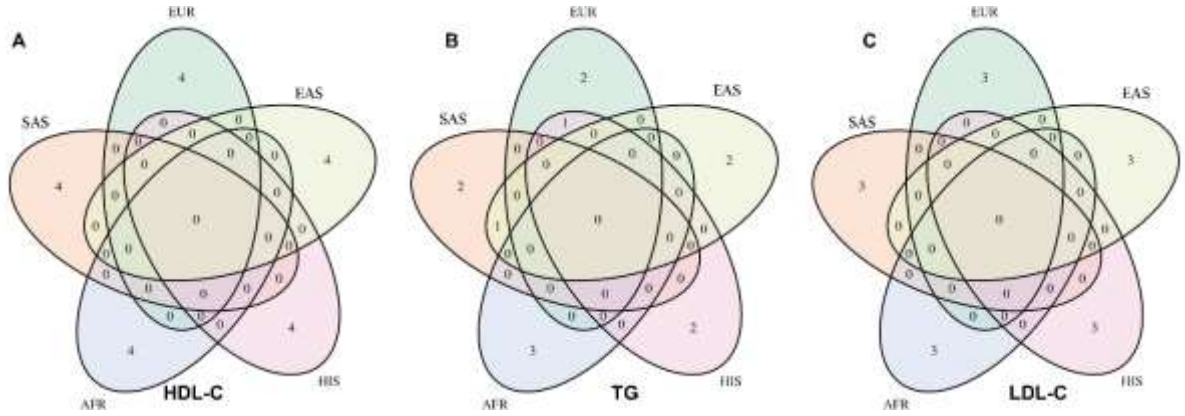

Figure S5. Venn Diagram for the overlap of the top variant included in each of the significant gene-based association analysis among different ancestries. **A)** 4 genes (*CETP*, *ABCA1*, *CD36*, and *LCAT*) showed significant gene-based associations in multi-ancestry and each of the 5 different ancestries analyses with P < 0.05 with HDL cholesterol (HDL-C). **B)** A total of 3 genes (*APOC3*, *ANGPTL3*, and *APOB*) showed significant gene-based associations in multi-ancestry analyses with triglycerides (TG). **C)** A total of 3 genes (*LDLR*, *PCSK9*, and *APOB*) showed significant gene-based associations in multi-ancestry and each of the 5 different ancestries analyses with P < 0.05 with LDL cholesterol (LDL-C).

Figure S6. Cumulative loss-of-function minor allele count and effect size on LDL cholesterol by ancestry

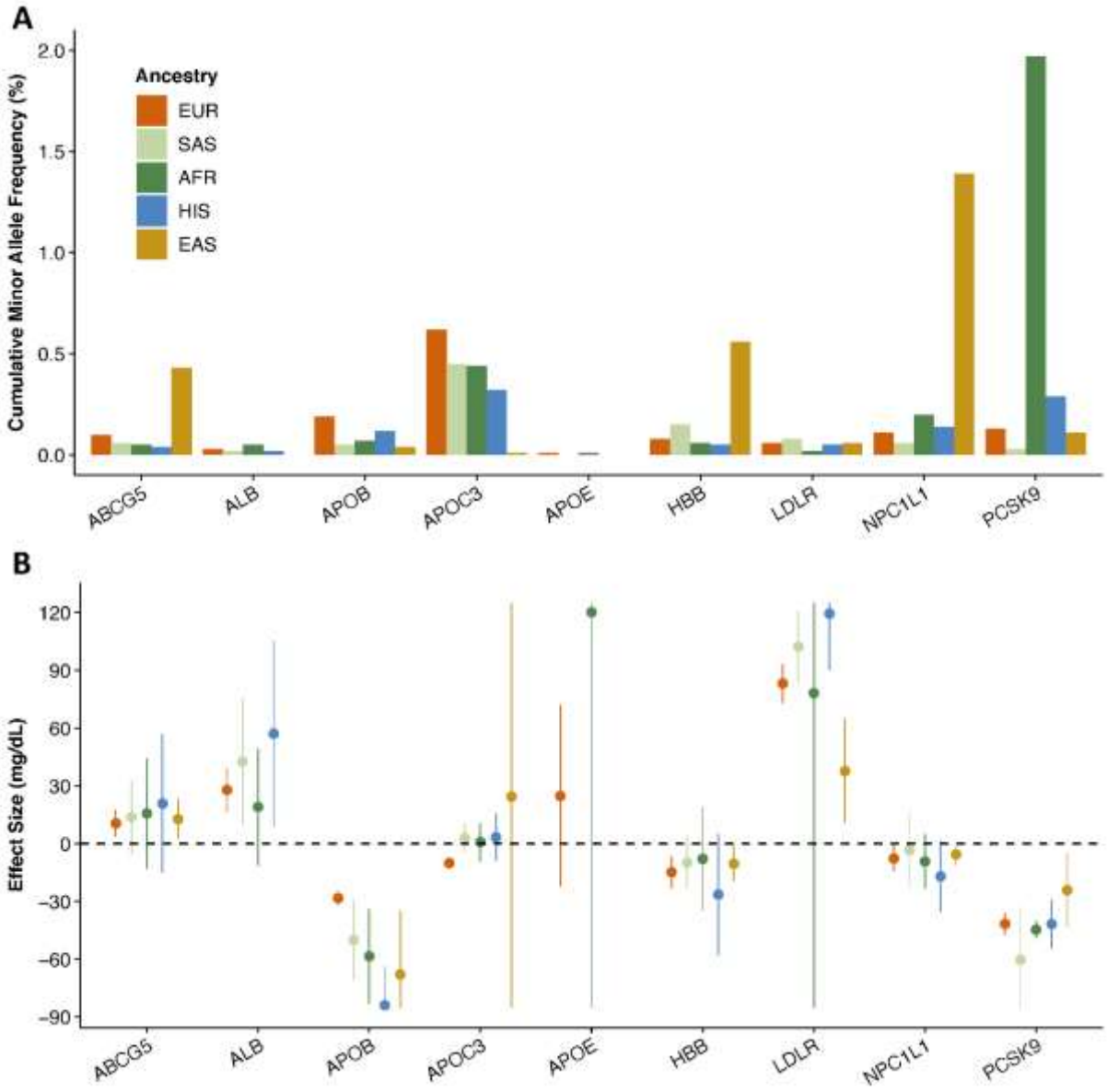

Figure S6: A) Cumulative minor allele frequencies and **B)** burden test effect sizes on LDL cholesterol levels for exome-wide significant genes ($P < 4.3\times10^{-7}$) within each of the five major ancestries using variants from the high confidence loss-of-function grouping (LOFTEE). AFR=African ancestry, EAS=East Asian ancestry, EUR=European ancestry, HIS=Hispanic ancestry, SAS=South Asian ancestry.

Figure S7. Cumulative loss-of-function minor allele count and effect size on triglycerides by ancestry

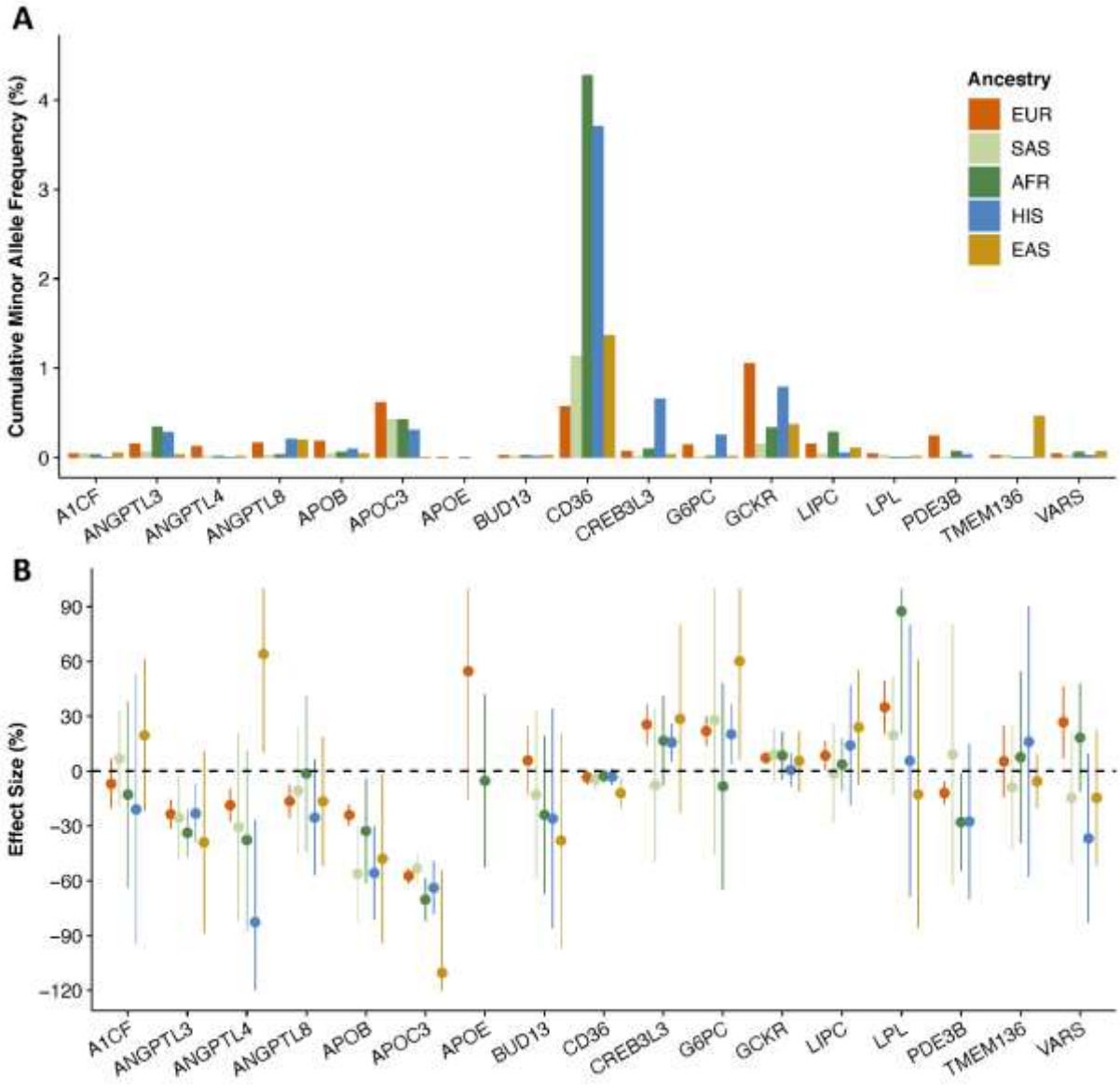

Figure S7: A) Cumulative minor allele frequencies and **B)** burden test effect sizes on triglyceride levels for exome-wide significant genes ($P < 4.3\times10^{-7}$) within each of the five major ancestries using variants from the high confidence loss-of-function grouping (LOFTEE). AFR=African ancestry, EAS=East Asian ancestry, EUR=European ancestry, HIS=Hispanic ancestry, SAS=South Asian ancestry.

Figure S8. Cumulative loss-of-function minor allele count and effect size on HDL cholesterol by ancestry

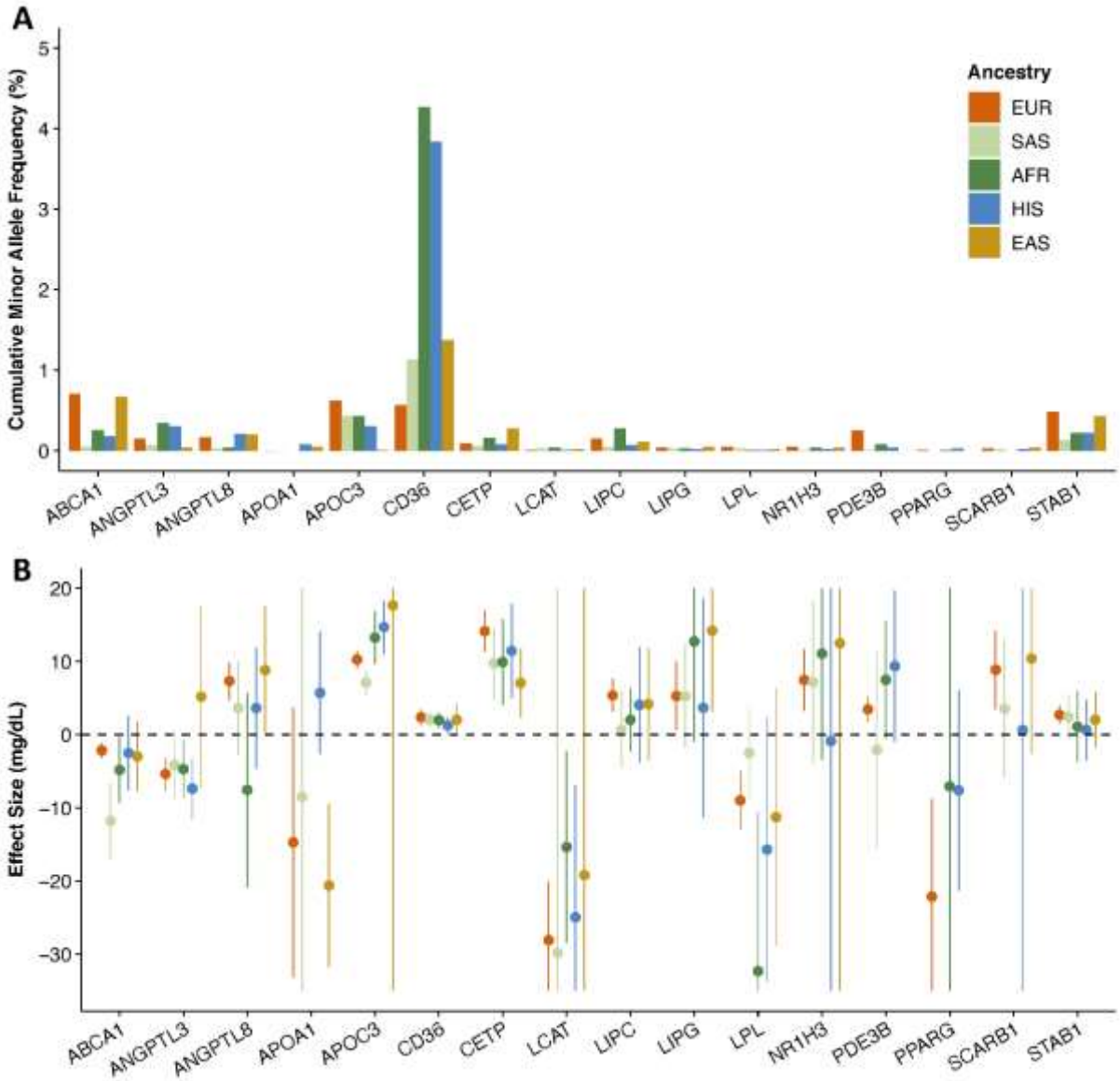

Figure S8: A) Cumulative minor allele frequencies and **B)** burden test effect sizes on HDL cholesterol levels for exome-wide significant genes ($P < 4.3 \times 10^{-7}$) within each of the five major ancestries using variants from the high confidence loss-of-function grouping (LOFTEE). AFR=African ancestry, EAS=East Asian ancestry, EUR=European ancestry, HIS=Hispanic ancestry, SAS=South Asian ancestry.

Supplemental Methods

Sequencing and Quality Control

Myocardial Infarction Genetics Consortium (MIGen)

A set of common variants was extracted for sample quality control including relative inference, principal component analysis, and estimation of heterozygosity. SNPs on autosomes and not in low complexity regions or segmental duplications were extracted. SNPs with quality of depth (QD)> 2, call rate >98%, self-reported-race-specific Hardy-Weinberg equilibrium (HWE) p-value $>1\times10^{-8}$, Variant Quality Score Recalibration (VQSR) of PASS and MAF>1% were retained. Sample relatedness was estimated with KING and duplicate samples removed. Genetically inferred ancestry was assigned to each individual by calculating principal components jointly with 1000 Genomes phase 3 version 5 and building a 5-Nearest Neighbor classifier[1] using the top 6 principal components. Heterozygosity was estimated within each genetic ancestry group and samples with F statistic above 0.3 were removed. Genetic sex was inferred based on high quality X-chromosome variation including variants with call rate >0.95, MAF>2%, a PASS VQSR, QD>3 if the variant is an insertion or deletion and QD>2 if it is SNP. Samples with discordant phenotypic sex and genetic sex were removed. Finally, sample quality control metrics were calculated using Hail and samples with call rate<0.9a mean depth (DP)<30 and mean genotype quality (GQ)<0.8 were excluded. A total of 44,240 samples with lipid data measurements were included after further excluding duplicates and relatives with other data sources (**Table S1**).

Variant quality control was performed amongst remaining samples and a total of 8,716,575 autosomal variants were included after removing those that fail HWE as

calculated by genetic ancestry group (p-value<1×10^{-8}), lie in low complexity regions or segmental duplications, with inbreeding coefficient< -0.3, are insertions or deletions with QD ≤ 3 or SNPs with QD ≤ 2 or variants where VQSR does not PASS with the exception of singletons where variants with VQSRTrancheSNP99.60to99.80 were retained.

Trans-Omics for Precision Medicine (TOPMed)

Whole genome sequencing at 30X mean depth was performed at one of six sequencing centers: Broad Institute of MIT and Harvard, Northwest Genomics Center, New York Genome Center, Illumina Laboratory Services, Psomagen, Inc. (formerly Macrogen USA), Baylor College of Medicine Human Genome Sequencing Center. For most studies, all individuals in the study were sequenced at the same center. Sequence reads were aligned to human genome build GRCh37 or GRCh38 at each center using similar, but not identical, processing pipelines. The resulting sequence data files were transferred from all centers to the TOPMed Informatics Research Center (IRC), where they were re-aligned to build GRCh38, using a common pipeline to produce a set of 'harmonized' .cram files. Processing was coordinated and managed by the 'GotCloud' processing pipeline. The IRC performed joint genotype calling on all samples. Quality control was performed at each stage of the process by the Sequencing Centers, the IRC, and the TOPMed Data Coordinating Center (DCC). Only samples that passed QC were included in the call set.

The two sequence quality criteria that were used to pass sequence data on for joint variant discovery and genotyping are: estimated DNA sample contamination below

3%, and fraction of the genome covered at least 10x 95% or above. DNA sample contamination was estimated from the sequencing center read mapping using software verifyBamId.[2]

The genotype used for analysis are from “freeze 6a” of the variant calling pipeline performed by the TOPMed Informatics Research Center (Center for Statistical Genetics, University of Michigan, Hyun Min Kang, Tom Blackwell and Gonçalo Abecasis). Variant detection (SNPs and indels) from each sequenced (and aligned) genome was performed by the vt discover2 software tool. The variant calling software tools are under active development; updated versions can be accessed at http://github.com/atks/vt, http://github.com/hyunminkang/apigenome, and https://github.com/statgen/topmed_variant_calling.

One individual from duplicate pairs identified by the DCC was removed, retaining the individual with lipid levels available when one did not have lipid levels. If both individuals had lipid levels, one individual was randomly selected. Individuals were excluded when their genotype determined sex did not match phenotype reported sex (n=6) and individuals <18 years old were excluded (n=865). Ancestry was defined as reported ancestry, which showed, generally, good concordance with PCs

AMP-T2D-GENES

Sequencing and quality control of the AMP-T2D-GENES study has been previously described.[3] Sequencing reads were processed and aligned to the human genome (build hg19) using the Picard (broadinstitute.github.io/picard/), BWA, and GATKsoftware packages, following best-practice pipelines. Single nucleotide and

short indel variants were then called using a series of GATK commands (version nightly-2015-07-31-g3c929b0): ApplyRecalibration, CombineGVCFs, CombineVariants, GenotypeGVCFs, HaplotypeCaller, SelectVariants, and VariantFiltration. Variants were called within 50bp of any region targeted for capture in any sequenced cohort. Following variant calling, all sites were then lifted over to build GRCh38 using CrossMap.

To perform data quality control, we first calculated a range of metrics measuring sample sequencing quality. We then stratified samples by ancestry and sequence capture technology and excluded from further analysis samples that were outliers according to any metric, based on visual inspection by comparison to other samples within the same stratum. After exclusion of samples, we calculated an additional set of variant metrics and excluded any variant with overall call rate <0.3, heterozygosity of 1, or heterozygote allele balance of 0 or 1 (i.e. 100% or 0% of reads called non-reference for heterozygous genotypes). After these initial quality control steps, 49,484 samples and 7.02M variants remained in our dataset.

Following initial sample and variant quality control, we performed additional exclusions of samples from association analysis. First, we computed a set of "transethnic" SNPs for use in identity-by-descent (IBD) and principal component (PC) analysis. We began this analysis with variants in the clean dataset (a) with genotype call rate >95%, (b) with minor allele frequency (MAF) >1% in each ancestry, and (c) further than 250Kb from the HLA region or an established T2D association signal. We LD-pruned variants using PLINK based on maximum r2 = 0.2 (parameters –indep-pairwise 50 5 0.2). We used the remaining 171K variants to estimate pairwise

individual IBD using PLINK, and the top 10 PCs of genetic ancestry using EIGENSTRAT. For each pair of individuals with IBD>0.9, we excluded the individual with the lower call rate (337 duplicate exclusions). We then excluded, for each of the five ancestries, any individual who appeared, based on visual inspection of the first two transethnic PCs, to lie outside of the main PC cluster corresponding to that ancestry (133 ethnic outliers). Finally, we used the subset of transethnic ancestry SNPs on the X chromosome to compare genetic sex to reported sex, using PLINK, and excluded all discordant individuals (273 sex discordances). Exclusion of the samples failing quality control, and variants that became monomorphic as a result of these sample exclusions, yielded a dataset of 45,231 individuals and 6.33M variants.

After these three rounds of sample exclusions, we identified five sets of ancestry-specific “ancestry” SNPs. We used the same procedure as for the transethnic SNPs (described above), except that we applied the MAF threshold only within the appropriate ancestry. We used these ancestry SNPs to estimate, for each ancestry, pairwise IBD values, genetic relatedness matrices (GRMs), and PCs for use in downstream association analysis. Additionally, from the IBD values, we generated a list of unrelated individuals within each ancestry by excluding the individual with the lower call rate in any pair of individuals with IBD>0.3 (leading to 2,157 excluded individuals). The final “unrelated analysis” set consisted of 43,090 individuals and yielded 6.29M non-monomorphic variants.

UK Biobank

We used two UKB datasets with exome sequence data. The first is a CAD case control study with 12,938 individuals. 29 samples were removed as they had

discordant genotypes with genotyping array data, 17 showed mismatch between the reported and genetically inferred sex, 4 had excess heterozygosity and 6 had a call rate <95%. To perform the sex-mismatch analyses, variants on the X-chromosome were selected after filtering out low quality genotypes, call rate<95%, MAF<2%, low QD score (3 for INDELs and 2 for SNPs), low confidence regions and segmental duplications and those that do not have PASS VQSR. A set of high quality common autosomal variants were extracted for relative inference, principal component analysis, and estimation of heterozygosity after removing low confidence regions and segmental duplications, low quality genotypes, QD<2, call rate<98%, self-reported ancestry-specific HWE $p>1x10^{-6}$ among controls, MAF<1% and do not have PASS VQSR. Heterozygosity was estimated within each ancestry and samples with F statistic>2 were removed. Genetically inferred ancestry was obtained using the 1000 Genomes as reference. Sample QC metrics were then calculated in HAIL using autosomal variants after filtering out low-quality genotypes, variants with ancestry-specific HWE $p<1\times10^{-6}$, low confidence regions and segmental duplications, low QD score (3 for INDELs and 2 for SNPs) and those that do not have PASS VQSR. Samples with call rate below 95%, mean DP below 30 and mean GQ below 80 were removed. Variant QC was done through filtering out monomorphic variants, call rate below 95%, those with HWE ($p < 1 \times 10^{-6}$), lie in low confidence regions or segmental duplications, are insertions or deletions with QD <= 3 or SNPs with QD <= 2 or variants where VQSR does not PASS unless singleton in which case retain those with VQSRTrancheSNP99.60to99.80. A total of 11,216 PC-identified European ancestry participants were included after additional removal of duplicates and relatives across data sources. A total of 2,734,519 variants were included.

The second UKB data set is a population-based dataset. Samples were filtered out if they showed mismatch between genetically determined and reported sex, high rates of heterozygosity or contamination (D-stat > 0.4), low sequence coverage (<85% of targeted bases achieving >20X coverage), duplicates, and exome sequence variants discordant with genotyping chip. More details are described elsewhere.[4] The "Functionally Equivalent" (FE) call set was used.[5] A total of 43,243 PC-identified European ancestry individuals were included after additional removal of duplicates and relatives across data sources.

Replication of gene-based associations

We performed replication of our top gene-based associations with blood lipid levels in the Penn Medicine BioBank (PMBB) and UK Biobank samples that did not contribute to the discovery analysis.

The PMBB is a repository of genotype and phenotype data for 43,731 patients at the University of Pennsylvania Perelman School of Medicine. All individuals recruited for PMBB are patients of clinical practice sites of the University of Pennsylvania Health System. Appropriate consent was obtained from each participant regarding storage of biological specimens, genetic sequencing, and access to all available EHR data. The study was approved by the Institutional Review Board of the University of Pennsylvania and complied with the principles set out in the Declaration of Helsinki. The six lipid phenotypes studied were HDL-C (n=21,247), LDL-C (n=21,040), non-HDL-C (n=21,087), TC (n=21,153), TG (n=21,418), and TG:HDL (n=21,213). All available lipid trait measurements up to July 2020 were included. HDL-C, LDL-C, TC, and TG levels were measured directly and accessible via PMBB. Non-HDL-C levels

were obtained by subtracting HDL-C from TC levels. TG and TG:HDL levels were logarithmically transformed to normalize their distribution for association testing. Due to the clinical nature of the biobank, samples often had multiple phenotype values corresponding to a patient's various clinical appointments. Gene-based associations were performed on the minimum, median, and maximum phenotype values to account for both potentially protective and pathogenic effects. Conceptually, the idea behind using minimum, median, and maximum phenotype values is to better capture the full range of phenotypes, given that lipid levels can vary over time, including the effects of lipid-lowering medications. For example, in the common situation in which a patient has initiated statin therapy during the course of their EHR record, the maximum LDL-C is more likely to reflect the untreated 'basal' level than the median or the minimum LDL-C. Genetic variants that elevate a specific lipid phenotype are likely to be stronger for maximum values, while genetic variants that reduce a specific lipid phenotype are likely to be stronger for minimum values. For the gene-based association analysis, 10 different variant groupings were used to determine the set of damaging variants within each gene including the six groupings used in the initial study. The additional four groupings used predicted loss-of-function (pLOF) variants that included frameshift, stop gain, and splicing variants as annotated by RefGene. Missense variants were annotated using Rare Exome Variant Ensemble Learner (REVEL) and filtered for those with a pathogenicity score>0.5. The four additional groupings consisted of, 1) pLOF, MAF≤0.1%, 2) pLOF, MAF≤0.1%, REVEL missense, 3) pLOF, MAF≤1%, and 4) pLOF, MAF≤1%, and REVEL missense. Each of the 10 groupings were used in a gene-based association test with the minimum, median, and maximum values of the 6 lipid phenotypes. Furthermore, ancestry-specific associations were also performed to elucidate any potential ancestry-specific

effects. This included associations among African and European ancestries separately, and then the two populations meta-analyzed. All associations were adjusted for sex, age, and principal components. The number of PCs chosen for each ancestry were determined according to ancestry-specific scree plots. The first 5 principal components were used for African ancestry associations, and the first 10 principal components were used for European ancestry associations.

In UK Biobank, we analyzed the association of rare variant aggregates from the 10 genes against four lipid phenotypes in the UK biobank whole exome sequencing (WES) data. Variant aggregates were obtained for the following four categories 1) LOFTEE – HC 2) LOFTEE - HC & predicted splice site altering 3) LOFTEE - HC & deleterious-METAsvm 4) LOFTEE - HC & deleterious-METAsvm & predicted splice site altering. We removed UK Biobank individuals used in the discovery analysis, resulting in 150,694 individuals for replication. The phenotypes were adjusted for lipid lowering medications, where total cholesterol was adjusted by dividing by 0.8 and LDL-C by dividing by 0.7. Triglycerides were natural log transformed for analysis. The phenotypes were inverse rank normalized and scaled by the standard deviation of the trait and adjusted for covariates (sex, age, age2, PC1-PC10, if British ancestry). Rare variant aggregate test was conducted using STAAR[6] with a MAF of 0.01 for the four lipids. Effect estimates were calculated using glmm.wald burden test.

References for Supplemental Methods:

1. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., et al. (2011). Scikit-learn: Machine Learning in Python. J Mach Learn Res 12, 2825–2830.
2. Jun, G., Flickinger, M., Hetrick, K.N., Romm, J.M., Doheny, K.F., Abecasis, G.R., Boehnke, M., and Kang, H.M. (2012). Detecting and estimating contamination

of human DNA samples in sequencing and array-based genotype data. Am J Hum Genet 91, 839-848.
3. Flannick, J., Mercader, J.M., Fuchsberger, C., Udler, M.S., Mahajan, A., Wessel, J., Teslovich, T.M., Caulkins, L., Koesterer, R., Barajas-Olmos, F., et al. (2019). Exome sequencing of 20,791 cases of type 2 diabetes and 24,440 controls. Nature 570, 71-76.
4. Van Hout, C.V., Tachmazidou, I., Backman, J.D., Hoffman, J.X., Ye, B., Pandey, A.K., Gonzaga-Jauregui, C., Khalid, S., Liu, D., Banerjee, N., et al. (2019). Whole exome sequencing and characterization of coding variation in 49,960 individuals in the UK Biobank. bioRxiv, 572347.
5. Regier, A.A., Farjoun, Y., Larson, D.E., Krasheninina, O., Kang, H.M., Howrigan, D.P., Chen, B.J., Kher, M., Banks, E., Ames, D.C., et al. (2018). Functional equivalence of genome sequencing analysis pipelines enables harmonized variant calling across human genetics projects. Nature communications 9, 4038.
6. Li, X., Li, Z., Zhou, H., Gaynor, S.M., Liu, Y., Chen, H., Sun, R., Dey, R., Arnett, D.K., Aslibekyan, S., et al. (2020). Dynamic incorporation of multiple in silico functional annotations empowers rare variant association analysis of large whole-genome sequencing studies at scale. Nat Genet 52, 969-983.

Study Participant Descriptions

Myocardial Infarction Genetics Consortium (MIGen) study participants

MIGen studies included the Atherosclerosis Risk in Communities study (ARIC), Italian Atherosclerosis Thrombosis and Vascular Biology (ATVB) study,[1] Bangladesh Risk of Acute Vascular Events study (BRAVE),[2] the Exome Sequencing Project Early-Onset Myocardial Infarction (ESP-EOMI) study,[3] a nested case-control cohort of the Jackson Heart Study (JHS),[4] the South German Myocardial Infarction study,[5] the Ottawa Heart Study (OHS),[6] the Precocious Coronary Artery Disease Study (PROCARDIS),[7] the Pakistan Risk of Myocardial Infarction Study (PROMIS),[8] the Registre Gironi del COR (Gerona Heart Registry or REGICOR) study,[9] the Leicester Myocardial Infarction study,[10] and the North German Myocardial Infarction study[11] (**Supplemental Table 37**). Clinical data were assessed in each study.

All participants in the study provided written informed consent for genetic studies. The institutional review boards at the Broad Institute and each participating institution approved the study protocol.

In order to minimize the possibility of unintentionally sharing information that can be used to re-identify private information, a subset of the data generated for this study are available at dbGaP and can be accessed at through dbGaP Study Accessions: phs000090.v1.p1 (ARIC), phs000814.v1.p1 (ATVB), phs001398.v1.p1 (BRAVE), phs000279.v2.p1 (EOMI), phs001098.v1.p1 (JHS), phs001000.v1.p1 (Leicester), phs000990.v1.p1 (NorthGermanMI), phs000916.v1.p1 (SouthGermanMI), phs000806.v1.p1 (OHS), phs000883.v1.p1 (PROCARDIS), phs000917.v1.p1 (PROMIS), phs000902.v1.p1(Regicor).

TOPMed program study participants

Atherosclerosis Risk in Communities study (ARIC, 2868)

TOPMed dbGaP accession#: phs001211, Parent dbGaP accession#: phs000280

ARIC is a large population-based prospective longitudinal cohort study (began 1987) from four U.S. communities: Forsyth County, NC; Jackson, MS; the northwest suburbs of Minneapolis, MN; and Washington County, MD. ARIC was designed to investigate the etiology and natural history of atherosclerosis, its consequences, and related medical care by race, gender, location, and time as previously described.[12] A total of 15,792 participants (55% female and 27% African American) aged 45-64 years were recruited between 1987 and 1989 and received extensive examination, including medical, social and demographic data. The baseline visit was conducted between 1987 and1989, the second visit in 1990-1992, the third visit in 1993-1995, the fourth visit in 1996-1998, the fifth visit in 2011-2013, the sixth visit in 2016-2017, and the seventh visit in 2018-2019. Follow-up is also conducted semi-annually since 2012 (annually prior to that) by telephone to maintain contact with participants and to assess the health status of the cohort.

The Atherosclerosis Risk in Communities study has been funded in whole or in part with Federal funds from the National Heart, Lung, and Blood Institute, National Institutes of Health, Department of Health and Human Services (contract numbers HHSN268201700001I, HHSN268201700002I, HHSN268201700003I, HHSN268201700004I and HHSN268201700005I). The authors thank the staff and participants of the ARIC study for their important contributions.

Whole genome sequencing (WGS) for the Trans-Omics in Precision Medicine (TOPMed) program was supported by the National Heart, Lung and Blood Institute (NHLBI). WGS for “NHLBI TOPMed: Atherosclerosis Risk in Communities (ARIC)” (phs001211) was performed at the Baylor College of Medicine Human Genome Sequencing Center (HHSN268201500015C and 3U54HG003273-12S2) and the Broad Institute for MIT and Harvard (3R01HL092577-06S1). Centralized read mapping and genotype calling, along with variant quality metrics and filtering were provided by the TOPMed Informatics Research Center (3R01HL-117626-02S1). Phenotype harmonization, data management, sample-identity QC, and general study coordination, were provided by the TOPMed Data Coordinating Center (3R01HL-120393-02S1). We gratefully acknowledge the studies and participants who provided biological samples and data for TOPMed.

The Genome Sequencing Program (GSP) was funded by the National Human Genome Research Institute (NHGRI), the National Heart, Lung, and Blood Institute (NHLBI), and the National Eye Institute (NEI). The GSP Coordinating Center (U24 HG008956) contributed to cross-program scientific initiatives and provided logistical and general study coordination. The Centers for Common Disease Genomics (CCDG) program was supported by NHGRI and NHLBI, and whole genome sequencing was performed at the Baylor College of Medicine Human Genome Sequencing Center (UM1 HG008898 and R01HL059367).

Old Order Amish (Amish, 1,083)

TOPMed dbGaP accession#: phs000956, Parent dbGaP accession#: phs000391

The Amish Complex Disease Research Program includes a set of large community-based studies focused largely on cardiometabolic health carried out in the Old Order Amish (OOA) community of Lancaster, Pennsylvania.[13] The OOA population of Lancaster County, PA immigrated to the Colonies from Western Europe in the early 1700's. There are now over 30,000 OOA individuals in the Lancaster area, nearly all of whom can trace their ancestry back 12-14 generations to approximately 700 founders. Investigators at the University of Maryland School of Medicine have been studying the genetic determinants of cardiometabolic health in this population since 1993. To date, over 7,000 Amish adults have participated in one or more of our studies.

The Amish studies upon which these data are based were supported by NIH grants R01 AG18728, U01 HL072515, R01 HL088119, R01 HL121007, U01 HL137181, and P30 DK072488, American Heart Association grant AHA 17GRNT33661168 WGS for “NHLBI TOPMed: Genetics of Cardiometabolic Health in the Amish” (phs000956) was performed at the Broad Institute of MIT and Harvard (3R01HL121007-01S1).

Mt Sinai BioMe Biobank (BioMe, 3257)

TOPMed dbGaP accession#: phs001644, Parent dbGaP accession#: phs000925

The Mount Sinai Institute for Personalized Medicine BioMe Biobank is a consented, EMR-linked medical care setting biorepository of the Mount Sinai Medical Center drawing from a population of over 70,000 inpatients and 800,000 outpatient visits annually.[14] The Mount Sinai Medical Center services diverse local communities of upper Manhattan, including Central Harlem (86% African American), East Harlem

(88% Hispanic Latino), and Upper East Side (88% European ancestry/white) with broad health disparities. Biobank operations are fully integrated in clinical care processes, including 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 a broad spectrum of over 30 clinical care sites.

The Mount Sinai BioMe Biobank has been supported by The Andrea and Charles Bronfman Philanthropies and in part by Federal funds from the NHLBI and NHGRI (U01HG00638001; U01HG007417; X01HL134588). WGS for "NHLBI TOPMed: Mount Sinai BioMe Biobank" (phs001644) was performed at the Baylor College of Medicine Human Genome Sequencing Center (HHSN268201600033I). We thank all participants in the Mount Sinai Biobank. We also thank all our recruiters who have assisted and continue to assist in data collection and management and are grateful for the computational resources and staff expertise provided by Scientific Computing at the Icahn School of Medicine at Mount Sinai.

Coronary Artery Risk Development in Young Adults (CARDIA, 2724)

TOPMed dbGaP accession#: phs001612, Parent dbGaP accession#: phs000285

The Coronary Artery Risk Development in Young Adults Study (CARDIA) is a study examining the etiology and natural history of cardiovascular disease beginning in young adulthood.[15] In 1985-1986, a cohort of 5115 healthy black and white men and women aged 18-30 years were selected to have approximately the same number of people in subgroups of age (18-24 and 25-30), sex, race, and education (high school or less and more than high school) within each of four US Field Centers. These same participants were asked to participate in follow-up examinations during 1987-1988 (Year 2), 1990-1991 (Year 5), 1992-1993 (Year 7), 1995-1996 (Year 10), 2000-2001 (Year 15), 2005-2006 (Year 20), 2010-2011 (Year 25); and 2015-2016 (Year 30). A majority of the group has been examined at each of the follow-up examinations (91%, 86%, 81%, 79%, 74%, 72%, 72%, and 71%, respectively). In addition to the follow-up examinations, participants are contacted regularly for the ascertainment of information on out-patient procedures and hospitalizations experienced between contacts.

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). CARDIA was also partially supported by the Intramural Research Program of the National Institute on Aging (NIA) and an intra-agency agreement between NIA and NHLBI (AG0005). WGS for "NHLBI TOPMed: Coronary Artery Risk Development in Young Adults" (phs001612) was performed at the Baylor College of Medicine Human Genome Sequencing Center (HHSN268201600033I).

Cleveland Family Study (CFS, 532)

TOPMed dbGaP accession#: phs000954, Parent dbGaP accession#: phs000284

The Cleveland Family Study (CFS) is a family-based study of sleep apnea, comprising of 2,284 individuals (46% African American) from 361 families studied up to 4 occasions over 16 years, 1990-2006.[16-19] Index probands (n=275) were recruited from 3 area hospital sleep labs if they had a confirmed diagnosis of sleep apnea and at least 2 first-degree relatives available to be studied. In the first 5 years of the study, neighborhood control probands (n=87) with at least 2 living relatives available for study were selected at random from a list provided by the index family and also studied. All available first-degree relatives and spouses of the case and control probands also were recruited. Second-degree relatives, including half-sibs, aunts, uncles and grandparents, were also included if they lived near the first-degree relatives (cases or controls), or if the family had been found to have two or more relatives with sleep apnea. Blood was sampled and DNA isolated for participants seen in the last two exam cycles (n=1,447).

CFS is supported by grants from the NHLBI (HL046389, HL113338, and 1R35HL135818). WGS for “NHLBI TOPMed: Cleveland Family Study - WGS Collaboration” (phs000954) was performed at the University of Washington Northwest Genomics Center (3R01HL098433-05S1 and HHSN268201600032I).

Cardiovascular Health Study (CHS, 2070)

TOPMed dbGaP accession#: phs001368, Parent dbGaP accession#: phs000287

The Cardiovascular Health Study (CHS) originated in 1988 and is a study of risk factors for development and progression of coronary heart disease and stroke in people aged 65 years and older.[20-22] The 5,888 study participants were recruited from four U.S. communities and have undergone extensive clinic examinations for evaluation of markers of subclinical cardiovascular disease. The original cohort totaled 5,201 participants. A new cohort was recruited in 1992. The 687 participants in the new cohort are predominately African-American and were recruited at three of the four field centers. Starting in 1989, and continuing through 1999, participants underwent annual extensive clinical examinations. Measurements included traditional risk factors such as blood pressure and lipids as well as measures of subclinical disease, including echocardiography of the heart, carotid ultrasound, and cranial magnetic-resonance imaging (MRI). At six-month intervals between clinic visits, and once clinic visits ended, participants were contacted by phone to ascertain hospitalizations and health status. The main outcomes are coronary heart disease (CHD), angina, heart failure (HF), stroke, transient ischemic attack (TIA), claudication, and mortality. Participants continue to be contacted by phone every 6 months.

This CHS research was supported by NHLBI contracts HHSN268201200036C, HHSN268200800007C, HHSN268201800001C, N01HC55222, N01HC85079, N01HC85080, N01HC85081, N01HC85082, N01HC85083, N01HC85086; and NHLBI grants U01HL080295, R01HL087652, R01HL105756, R01HL103612, R01HL120393, R01HL130114, and R01 HL059367, with additional contribution from the National Institute of Neurological Disorders and Stroke (NINDS). Additional support was provided through R01AG023629 from the National Institute on Aging (NIA). A full list of principal CHS investigators and institutions can be found at CHS-NHLBI.org. WGS for “NHLBI TOPMed: Cardiovascular Health Study” (phs001368) was performed at the Baylor College of Medicine Human Genome Sequencing Center (3U54HG003273-12S2,

HHSN268201500015C, and HHSN268201600033I). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Diabetes Heart Study (DHS, 345)

TOPMed dbGaP accession#: phs001412, Parent dbGaP accession#: phs001012

The Diabetes Heart Study (DHS) is a family-based study enriched for type 2 diabetes (T2D).[23] The cohort included 1443 European American and African American participants from 564 families with multiple cases of type 2 diabetes. The cohort was recruited between 1998 and 2006. Participants were extensively phenotyped for measures of subclinical CVD and other known CVD risk factors. Primary outcomes were quantified burden of vascular calcified plaque in the coronary artery, carotid artery, and abdominal aorta all determined from non-contrast computed tomography scans.

This work was supported by R01 HL92301, R01 HL67348, R01 NS058700, R01 AR48797, R01 DK071891, R01 AG058921, the General Clinical Research Center of the Wake Forest University School of Medicine (M01 RR07122, F32 HL085989), the American Diabetes Association, and a pilot grant from the Claude Pepper Older Americans Independence Center of Wake Forest University Health Sciences (P60 AG10484). WGS for “NHLBI TOPMed: Diabetes Heart Study” (phs001412) was performed at the Broad Institute of MIT and Harvard (HHSN268201500014C).

Framingham Heart Study (FHS, 3,961)

TOPMed dbGaP accession#: phs000974, Parent dbGaP accession#: phs000007

The Framingham Heart Study (FHS) is a prospective cohort study of 3 generations of subjects who have been followed up to 65 years to evaluate risk factors for cardiovascular disease.[24-27] Its large sample of ~15,000 men and women who have been extensively phenotyped with repeated examinations make it ideal for the study of genetic associations with cardiovascular disease risk factors and outcomes. DNA samples have been collected and immortalized since the mid-1990s and are available on ~8000 study participants in 1037 families. These samples have been used for collection of GWAS array data and exome chip data in nearly all with DNA samples, and for targeted sequencing, deep exome sequencing and light coverage whole genome sequencing in limited numbers. Additionally, mRNA and miRNA expression data, DNA methylation data, metabolomics and other 'omics data are available on a sizable portion of study participants. This project will focus on deep whole genome sequencing (mean 30X coverage) in ~4100 subjects and imputed to all with GWAS array data to more fully understand the genetic contributions to cardiovascular, lung, blood and sleep disorders.

FHS acknowledges the support of contracts NO1-HC-25195 and HHSN268201500001I from the National Heart, Lung and Blood Institute and grant supplement R01 HL092577-06S1 for this research. WGS for “NHLBI TOPMed: Whole Genome Sequencing and Related Phenotypes in the Framingham Heart Study” (phs000974) was performed at the Broad Institute of MIT and Harvard (HHSN268201500014C, 3R01HL092577-06S1, and 3U54HG003067-12S2). We also

acknowledge the dedication of the FHS study participants without whom this research would not be possible.

Genetic Epidemiology Network of Arteriopathy (GENOA, 391)

TOPMed dbGaP accession#: phs001345, Parent dbGaP accession#: phs001238

The Genetic Epidemiology Network of Arteriopathy (GENOA) is one of four networks in the NHLBI Family-Blood Pressure Program (FBPP).28 GENOA's long-term objective is to elucidate the genetics of target organ complications of hypertension, including both atherosclerotic and arteriolosclerotic complications involving the heart, brain, kidneys, and peripheral arteries.29 The longitudinal GENOA Study recruited European-American and African-American sibships with at least 2 individuals with clinically diagnosed essential hypertension before age 60 years. All other members of the sibship were invited to participate regardless of their hypertension status. Participants were diagnosed with hypertension if they had either 1) a previous clinical diagnosis of hypertension by a physician with current anti-hypertensive treatment, or 2) an average systolic blood pressure ≥ 140 mm Hg or diastolic blood pressure ≥ 90 mm Hg based on the second and third readings at the time of their clinic visit. Only participants of the African-American Cohort were sequenced through TOPMed.

Support for GENOA was provided by the National Heart, Lung and Blood Institute (HL054457, HL054464, HL054481, and HL087660) of the National Institutes of Health. WGS for “NHLBI TOPMed: Genetic Epidemiology Network of Arteriopathy” (phs001345) was performed at the Broad Institute of MIT and Harvard (HHSN268201500014C) and the University of Washington Northwest Genomics Center (3R01HL055673-18S1).

Genetics of Lipid-Lowering Drugs and Diet Network (GOLDN, 594)

TOPMed dbGaP accession#: phs001359, Parent dbGaP accession#: phs000741

The Genetics of Lipid-Lowering Drugs and Diet Network (GOLDN) study was initiated to assess how genetic factors interact with environmental (diet and drug) interventions to influence blood levels of triglycerides and other atherogenic lipid species and inflammation markers (registered at clinicaltrials.gov, number NCT00083369).[30] The study recruited participants of European ancestry primarily from three-generational pedigrees from two NHLBI Family Heart Study (FHS) field centers (Minneapolis, MN and Salt Lake City, UT).[31] Only families with at least two siblings were recruited and only participants who did not take lipid-lowering agents (pharmaceuticals or nutraceuticals) for at least 4 weeks prior to the initial visit were included. The diet intervention followed the protocol of Patsch et al.[32] The whipping cream (83% fat) meal had 700 Calories/m2 body surface area (2.93 mJ/m2 body surface area): 3% of calories were derived from protein (instant nonfat dry milk) and 14% from carbohydrate (sugar). The ratio of polyunsaturated to saturated fat was 0.06 and the cholesterol content of the average meal was 240 mg. The mixture was blended with ice and flavorings. Blood samples were drawn immediately before (fasting) and at 3.5 and 6 hours after consuming the high-fat meal. The diet intervention was administered at baseline as well as after a 3-week treatment with 160 mg micronized fenofibrate. Participants were given the option to complete one or

both (diet and drug) interventions. Of all participants, 1079 had phenotypic data and provided appropriate consent, and underwent whole genome sequencing through the TOPMed program.

GOLDN biospecimens, baseline phenotype data, and intervention phenotype data were collected with funding from National Heart, Lung and Blood Institute (NHLBI) grant U01 HL072524. WGS for “NHLBI TOPMed: Genetics of Lipid Lowering Drugs and Diet Network” (phs001359) was performed at the University of Washington Northwest Genomics Center (3R01HL104135-04S1 and R01 HL104135).

Genetic Epidemiology Network of Salt Sensitivity (GenSalt, 1,749)

TOPMed dbGaP accession#: phs001217, Parent dbGaP accession#: phs000784

The Genetic Epidemiology Network of Salt-Sensitivity (GenSalt) study, using a family feeding-study design, aims to identify genes which interact with dietary sodium and potassium intake to influence blood pressure in Han Chinese participants from rural north China.[33] The dietary intervention included a 7-day low-sodium feeding (51.3 mmol/day), a 7-day high-sodium feeding (307.8 mmol/day) and a 7-day high-sodium feeding with an oral potassium supplementation (60 mmol/day). Microsatellite markers for genome-wide linkage scan and single nucleotide polymorphism (SNP) markers in candidate genes will be genotyped. Overall, 3153 participants from 658 families were recruited for GenSalt. Whole genome sequencing has been conducted for 1,860 participants as a part of TOPMed.

GenSalt was supported by research grants (U01HL072507, R01HL087263, and R01HL090682) from the National Heart, Lung and Blood Institute, National Institutes of Health, Bethesda, MD. WGS for “NHLBI TOPMed: Genetic Epidemiology Network of Salt Sensitivity” (phs001217) was performed at the Baylor College of Medicine Human Genome Sequencing Center (HHSN268201500015C).

Genetic Studies of Atherosclerosis Risk (GeneSTAR, 1,749)

TOPMed dbGaP accession#: phs001218, Parent dbGaP accession#: phs000375

GeneSTAR began in 1982 as the Johns Hopkins Sibling and Family Heart Study, a prospective longitudinal family-based study conducted originally in healthy adult siblings of people with documented early onset coronary disease under 60 years of age.[34,35] Commencing in 2003, the siblings, their offspring, and the coparent of the offspring participated in a 2 week trial of aspirin 81 mg/day with pre and post ex vivo platelet function assessed using multiple agonists in whole blood and platelet rich plasma. Extensive additional cardiovascular testing and risk assessment was done at baseline and serially. Follow-up was carried out to determine incident cardiovascular disease, stroke, peripheral arterial disease, diabetes, cancer, and related comorbidities, from 5 to 30 years after study entry. The goal of several additional phenotyping and interventional substudies has been to discover and amplify understanding of the mechanisms of atherogenic vascular diseases and attendant comorbidities.

GeneSTAR was supported by grants from the National Institutes of Health/National Heart, Lung, and Blood Institute (U01 HL72518, HL087698, HL49762, HL58625, HL071025, HL112064), the

National Institutes of Health/National Institute of Nursing Research (NR0224103), and by a grant from the National Institutes of Health/National Center for Research Resources (M01-RR000052) to the Johns Hopkins General Clinical Research Center. WGS for “NHLBI TOPMed: Genetic Studies of Atherosclerosis Risk” (phs001218) was performed at the Broad Institute of MIT and Harvard (HHSN268201500014C), the Macrogen Corp. (3R01HL112064-04S1), and Illumina (R01HL112064).

Hispanic Community Health Study - Study of Latinos (HCHS/SOL, 2540)

TOPMed dbGaP accession#: phs001395, Parent dbGaP accession#: phs000810

The Hispanic Community Health Study / Study of Latinos (HCHS/SOL) is a multi-center epidemiologic study in Hispanic/Latino populations to determine the role of acculturation in the prevalence and development of disease, and to identify risk factors playing a protective or harmful role in Hispanics/Latinos.[36] The goals of the HCHS/SOL include studying the prevalence and development of disease in Hispanics/Latinos, including the role of acculturation, and identifying disease risk factors that play protective or harmful roles in Hispanics/Latinos. A total of 16,415 persons of Cuban, Dominican, Mexican, Puerto Rican, Central American, and South American backgrounds were recruited through four Field Centers affiliated with San Diego State University, Northwestern University in Chicago, Albert Einstein College of Medicine in the Bronx area of New York, and the University of Miami. Seven additional academic centers serve as scientific and logistical support centers. Study participants aged 18-74 years took part in an extensive clinic exam and assessments to ascertain socio-demographic, cultural, environmental and biomedical characteristics. Annual follow-up interviews are conducted to determine a range of health outcomes.

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 Center on Minority Health and Health Disparities, the National Institute of Deafness and Other Communications Disorders, the National Institute of Dental and Craniofacial Research, the National Institute of Diabetes and Digestive and Kidney Diseases, the National Institute of Neurological Disorders and Stroke, and the Office of Dietary Supplements. WGS for “NHLBI TOPMed: Hispanic Community Health Study - Study of Latinos” (phs001395) was performed at the Baylor College of Medicine Human Genome Sequencing Center (HHSN268201600033I).

Hypertension Genetic Epidemiology Network and Genetic Epidemiology Network of Arteriopathy (HyperGEN, 1,797)

TOPMed dbGaP accession#: phs001293, Parent dbGaP accession#: phs001293

The Hypertension Genetic Epidemiology Network Study (HyperGEN) - Genetics of Left Ventricular (LV) Hypertrophy is a familial study aimed to understand genetic risk factors for LV hypertrophy by conducting genetic studies of continuous traits from echocardiography exams.[37] The originating HyperGEN study aimed to understand genetic risk factors for hypertension.[38] HyperGEN recruited 470 multiply-affected population-based hypertensive AA sibships (N=1224 siblings) from 1996-1999. HyperGEN probands were ascertained by early onset hypertension (i.e., before 60 years); to participate, they had to have at least one hypertensive sibling who was also willing to participate. Data from detailed clinical exams as well as genotyping data for linkage studies, candidate gene studies and GWAS have been collected and is shared between HyperGEN and the ancillary HyperGEN - Genetics of LV Hypertrophy study.

The HyperGEN Study is part of the National Heart, Lung, and Blood Institute (NHLBI) Family Blood Pressure Program; collection of the data represented here was supported by grants U01 HL054472 (MN Lab), U01 HL054473 (DCC), U01 HL054495 (AL FC), and U01 HL054509 (NC FC). The HyperGEN: Genetics of Left Ventricular Hypertrophy Study was supported by NHLBI grant R01 HL055673 with whole-genome sequencing made possible by supplement -18S1. WGS for “NHLBI TOPMed: Hypertension Genetic Epidemiology Network” (phs001293) was performed at the University of Washington Northwest Genomics Center (3R01HL055673-18S1).

Jackson Heart Study (JHS, 1722)

TOPMed dbGaP accession#: phs000964, Parent dbGaP accession#: phs000286

The purpose of the Jackson Heart Study (JHS) is to explore the reasons for heightened cardiovascular disease prevalence among African Americans and to uncover new approaches to reduce it. The JHS is a large, community-based, observational study whose 5,306 participants were recruited from among the non-institutionalized African-American adults from urban and rural areas of the three counties (Hinds, Madison, and Rankin) that make up the Jackson, MS, metropolitan statistical area (MSA).[4,39,40] The JHS design included participants from the Jackson ARIC study who had originally been recruited through random selection from a drivers' license registry. New JHS participants were chosen randomly from the Accudata America commercial listing, which provides householder name, address, zip code, phone number (if available), age group in decades, and family components. In addition, a family component was included in the JHS. The sampling frame for the family study was a participant in any one of the ARIC, random, or volunteer samples whose family size met eligibility requirements. Recruitment was limited to persons 35-84 years old except in the family cohort, where those 21 years old and above were eligible.

The Jackson Heart Study (JHS) is supported and conducted in collaboration with Jackson State University (HHSN268201800013I), Tougaloo College (HHSN268201800014I), the Mississippi State Department of Health (HHSN268201800015I) and the University of Mississippi Medical Center (HHSN268201800010I, HHSN268201800011I and HHSN268201800012I) contracts from the National Heart, Lung, and Blood Institute (NHLBI) and the National Institute on Minority Health and Health Disparities (NIMHD). WGS for “NHLBI TOPMed: The Jackson Heart Study” (phs000964) was performed at the University of Washington

Northwest Genomics Center (HHSN268201100037C). The authors also wish to thank the staffs and participants of the JHS.

Multi-Ethnic Study of Atherosclerosis (MESA, 5,185)

TOPMed dbGaP accession#: phs001416, Parent dbGaP accession#: phs000209

The Multi-Ethnic Study of Atherosclerosis (MESA) is a study of the characteristics of subclinical cardiovascular disease (disease detected non-invasively before it has produced clinical signs and symptoms) and the risk factors that predict progression to clinically overt cardiovascular disease or progression of the subclinical disease.[41] MESA researchers study a diverse, population-based sample of 6,814 asymptomatic men and women aged 45-84. Thirty-eight percent of the recruited participants are white, 28 percent African-American, 22 percent Hispanic, and 12 percent Asian, predominantly of Chinese descent. Participants were recruited from six field centers across the United States: Wake Forest University, Columbia University, Johns Hopkins University, University of Minnesota, Northwestern University and University of California - Los Angeles. Each participant received an extensive exam and determination of coronary calcification, ventricular mass and function, flow-mediated endothelial vasodilation, carotid intimal-medial wall thickness and presence of echogenic lucencies in the carotid artery, lower extremity vascular insufficiency, arterial wave forms, electrocardiographic (ECG) measures, standard coronary risk factors, sociodemographic factors, lifestyle factors, and psychosocial factors. Selected repetition of subclinical disease measures and risk factors at follow-up visits allows study of the progression of disease. Blood samples have been assayed for putative biochemical risk factors and stored for case-control studies. DNA has been extracted and lymphocytes cryopreserved (for possible immortalization) for study of candidate genes and possibly, genome-wide scanning, expression, and other genetic techniques. Participants are being followed for identification and characterization of cardiovascular disease events, including acute myocardial infarction and other forms of coronary heart disease (CHD), stroke, and congestive heart failure; for cardiovascular disease interventions; and for mortality.

Whole genome sequencing (WGS) for the Trans-Omics in Precision Medicine (TOPMed) program was supported by the National Heart, Lung and Blood Institute (NHLBI). WGS for “NHLBI TOPMed: Multi-Ethnic Study of Atherosclerosis (MESA)” (phs001416.v1.p1) was performed at the Broad Institute of MIT and Harvard (3U54HG003067-13S1). Centralized read mapping and genotype calling, along with variant quality metrics and filtering were provided by the TOPMed Informatics Research Center (3R01HL-117626-02S1). Phenotype harmonization, data management, sample-identity QC, and general study coordination, were provided by the TOPMed Data Coordinating Center (3R01HL-120393-02S1). MESA and the MESA SHARe project are conducted and supported by the National Heart, Lung, and Blood Institute (NHLBI) in collaboration with MESA investigators. The MESA project is conducted and supported by the National Heart, Lung, and Blood Institute (NHLBI) in collaboration with MESA investigators. Support for MESA is provided by contracts 75N92020D00001, HHSN268201500003I, N01-HC-95159, 75N92020D00005, N01-HC-95160, 75N92020D00002, N01-HC-95161, 75N92020D00003, N01-HC-95162, 75N92020D00006, N01-HC-95163, 75N92020D00004, N01-HC-95164, 75N92020D00007, N01-HC-95165, N01-HC-95166, N01-HC-95167, N01-HC-95168, N01-HC-95169, UL1-TR-000040, UL1-TR-001079, UL1-TR-001420. Support is

provided by grants and contracts R01HL071051, R01HL071205, R01HL071250, R01HL071251, R01HL071258, R01HL071259, by the National Center for Research Resources, Grant UL1RR033176. The provision of genotyping data was supported in part by the National Center for Advancing Translational Sciences, CTSI grant UL1TR001881, and the National Institute of Diabetes and Digestive and Kidney Disease Diabetes Research Center (DRC) grant DK063491 to the Southern California Diabetes Endocrinology Research Center.

Massachusetts General Hospital Atrial Fibrillation Study (MGH AF, 682)

TOPMed dbGaP accession#: phs001062, Parent dbGaP accession#: phs001001

The Massachusetts General Hospital (MGH) Atrial Fibrillation Study was initiated in 2001.[42,43] The study has enrolled serial probands, unaffected and affected family members with atrial fibrillation. At enrollment participants undergo a structured interview to systematically capture their past medical history, AF treatments, and family history. An electrocardiogram is performed; the results of an echocardiogram are obtained; and blood samples are obtained. For the TOPMed whole genome sequencing project only early-onset atrial fibrillation cases were sequenced. Early-onset atrial fibrillation was defined as an age of onset prior to 66 years of age.

The MGH AF Study was supported by grants to Dr. Ellinor from the Fondation Leducq (14CVD01), the National Institutes of Health to Dr. Ellinor (1RO1HL092577, R01HL128914, K24HL105780) and Dr. Lubitz (1R01HL139731) and by grants from the American Heart Association to Dr. Ellinor (18SFRN34110082) and to Dr. Lubitz (18SFRN34250007). WGS for “NHLBI TOPMed: Massachusetts General Hospital Atrial Fibrillation Study” (phs001062) was performed at the Broad Institute of MIT and Harvard (3R01HL092577-06S1, 3U54HG003067-12S2, 3U54HG003067-13S1, and 3UM1HG008895-01S2)

San Antonio Family Study (SAFS, 575)

TOPMed dbGaP accession#: phs001215, Parent dbGaP accession#: phs000462

The San Antonio Family Heart Study is a complex pedigree-based mixed longitudinal study designed to identify low frequency or rare variants influencing susceptibility to cardiovascular disease, using whole genome sequence (WGS) information from 3,000 individuals in large Mexican American pedigrees from San Antonio, Texas.[44] The major objectives of this study are to identify low frequency or rare variants in and around known common variant signals for CVD, as well as to find novel low frequency or rare variants influencing susceptibility to CVD. The study began in 1991, and included 1,431 individuals in 42 extended families at baseline. Probands were 40 to 60 year old low-income Mexican Americans selected at random without regard to presence or absence of disease, almost exclusively from Mexican American census tracts in San Antonio, Texas. All first, second, and third -degree relatives of the proband and of the proband's spouse, aged 16 years or above, were eligible to participate in the study. 1,200 WGS at 30X WGS were obtained through Illumina funded by a supplement as part of the NHLBI's TOPMed program.

Collection of the San Antonio Family Study data was supported in part by National Institutes of Health (NIH) grants R01 HL045522, MH078143, MH078111 and MH083824; and whole genome sequencing of SAFS subjects was supported by

U01 DK085524 and R01 HL113323. We are very grateful to the participants of the San Antonio Family Study for their continued involvement in our research programs. WGS for “NHLBI TOPMed: Whole Genome Sequencing to Identify Causal Genetic Variants Influencing CVD Risk - San Antonio Family Studies” (phs001215) was performed at Illumina (3R01HL113323-03S1 and R01HL113322).

Samoan Adiposity Study (Samoan, 1,182)

TOPMed dbGaP accession#: phs000972, Parent dbGaP accession#: phs000914

The research goal of the Samoan Adiposity Study is to identify genetic variation that increases susceptibility to obesity and cardiometabolic phenotypes among adult Samoans using genome-wide association (GWAS) methods.[45,46] DNA from peripheral blood and phenotypic information were collected from 3,119 adult Samoans, 23 to 70 years of age. The participants reside throughout the independent nation of Samoa, which is experiencing economic development and the nutrition transition. Genotyping was performed with the Affymetrix Genome-Wide Human SNP 6.0 Array using a panel of approximately 900,000 SNPs. Anthropometric, fasting blood biomarkers and detailed dietary, physical activity, health and socio-demographic variables were collected. Whole genome sequencing of a subset was motivated by the opportunity to create a Samoan-specific reference panel for imputation into the larger parent study.

Data collection was funded by NIH grant R01-HL093093 and R01-HL133040. WGS for “NHLBI TOPMed: Samoan Adiposity Study” (phs000972) was performed at the University of Washington Northwest Genomics Center (HHSN268201100037C and HHSN268201500016C). We thank the Samoan participants of the study and local village authorities. We acknowledge the support of the Samoan Ministry of Health and the Samoa Bureau of Statistics for their support of this research.

The Samoan Obesity, Lifestyle and Genetic Adaptations Study (OLaGA) Group:

Ranjan Deka, Dept. of Environmental Health, University of Cincinnati;

Nicola L. Hawley, Dept. of Chronic Disease Epidemiology, Yale University;

Stephen T McGarvey, Dept. of Epidemiology and International Health Institute, and Dept. of Anthropology, Brown University;

Ryan L Minster, Dept. of Human Genetics, University of Pittsburgh;

Take Naseri, Ministry of Health, Government of Samoa;

Muagututi‘a Sefuiva Reupena, Lutia I Puava Ae Mapu I Fagalele;

Daniel E. Weeks, Depts. of Human Genetics and Biostatistics, University of Pittsburgh

Taiwan Study of Hypertension using Rare Variants (THRV, 1,979)

TOPMed dbGaP accession#: phs001387, Parent dbGaP accession#: phs001387

The THRV-TOPMed study consists of three cohorts: The SAPPHIRe Family cohort (N=1,271), TSGH (Tri-Service General Hospital, a hospital-based cohort, N=160), and TCVGH (Taichung Veterans General Hospital, another hospital-based cohort, N=922), all based in Taiwan.[47,48] 1,271 subjects were previously recruited as

part of the NHLBI-sponsored SAPPHIRe Network (which is part of the Family Blood Pressure Program, FBPP). The SAPPHIRe families were recruited to have two or more hypertensive sibs, some families also with one normotensive/hypotensive sib. The two Hospital-based cohorts (TSGH and TCVGH) both recruited unrelated subjects with different recruitment criteria (matched with SAPPHIRe subjects for age, sex, and BMI category).

The Rare Variants for Hypertension in Taiwan Chinese (THRV) is supported by the National Heart, Lung, and Blood Institute (NHLBI) grant (R01HL111249) and its participation in TOPMed is supported by an NHLBI supplement (R01HL111249-04S1). THRV is a collaborative study between Washington University in St. Louis, LA BioMed at Harbor UCLA, University of Texas in Houston, Taichung Veterans General Hospital, Taipei Veterans General Hospital, Tri-Service General Hospital, National Health Research Institutes, National Taiwan University, and Baylor University. THRV is based (substantially) on the parent SAPPHIRe study, along with additional population-based and hospital-based cohorts. SAPPHIRe was supported by NHLBI grants (U01HL54527, U01HL54498) and Taiwan funds, and the other cohorts were supported by Taiwan funds. WGS for “NHLBI TOPMed: Taiwan Study of Hypertension using Rare Variants” (phs001387) was performed at the Baylor College of Medicine Human Genome Sequencing Center (3R01HL111249-04S1, HHSN26820150015C)

Women’s Health Initiative (WHI, 8,188)

TOPMed dbGaP accession#: phs001237, Parent dbGaP accession#: phs000200

The Women's Health Initiative (WHI) is a long-term national health study that has focused on strategies for preventing heart disease, breast and colorectal cancer, and osteoporotic fractures in postmenopausal women (clinicaltrials.gov NCT00000611).[49-51] The original WHI study included 161,808 postmenopausal women enrolled between 1993 and 1998. The Fred Hutchinson Cancer Research Center in Seattle, WA serves as the WHI Clinical Coordinating Center for data collection, management, and analysis of the WHI. The WHI has two major parts: a partial factorial randomized Clinical Trial (CT) and an Observational Study (OS); both were conducted at 40 Clinical Centers nationwide. The CT enrolled 68,132 postmenopausal women between the ages of 50-79 into trials testing three prevention strategies. If eligible, women could choose to enroll in one, two, or all three of the trial components. The components are: hormone therapy trials, dietary modification trial, and calcium / vitamin D trial. The Observational Study (OS) examines the relationship between lifestyle, environmental, medical and molecular risk factors and specific measures of health or disease outcomes. This component involves tracking the medical history and health habits of 93,676 women not participating in the CT. Recruitment for the observational study was completed in 1998 and participants were followed annually for 8 to 12 years.

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 HHSN268201600018C, HHSN268201600001C, HHSN268201600002C, HHSN268201600003C, and HHSN268201600004C. WGS for “NHLBI TOPMed: Women's Health Initiative” (phs001237) was performed at the Broad Institute of MIT and Harvard (HHSN268201500014C)

UK Biobank (external to TOPMed)

The UK Biobank analyses were conducted using the UK Biobank resource under application 7089.

References for Study Participant Descriptions

1 Atherosclerosis, Thrombosis, and Vascular Biology Italian Study Group. No evidence of association between prothrombotic gene polymorphisms and the development of acute myocardial infarction at a young age. *Circulation* **107**, 1117-1122 (2003).

2 Chowdhury, R. *et al.* The Bangladesh Risk of Acute Vascular Events (BRAVE) Study: objectives and design. *European journal of epidemiology* **30**, 577-587, doi:10.1007/s10654-015-0037-2 (2015).

3 Do, R. *et al.* Exome sequencing identifies rare LDLR and APOA5 alleles conferring risk for myocardial infarction. *Nature* **518**, 102-106, doi:10.1038/nature13917 (2015).

4 Taylor, H. A., Jr. *et al.* Toward resolution of cardiovascular health disparities in African Americans: design and methods of the Jackson Heart Study. *Ethnicity & disease* **15**, S6-4-17 (2005).

5 Tg *et al.* Loss-of-function mutations in APOC3, triglycerides, and coronary disease. *N Engl J Med* **371**, 22-31, doi:10.1056/NEJMoa1307095 (2014).

6 McPherson, R. *et al.* A common allele on chromosome 9 associated with coronary heart disease. *Science* **316**, 1488-1491, doi:10.1126/science.1142447 (2007).

7 Clarke, R. *et al.* Genetic variants associated with Lp(a) lipoprotein level and coronary disease. *N Engl J Med* **361**, 2518-2528, doi:10.1056/NEJMoa0902604 (2009).

8 Saleheen, D. *et al.* The Pakistan Risk of Myocardial Infarction Study: a resource for the study of genetic, lifestyle and other determinants of myocardial infarction in South Asia. *Eur J Epidemiol* **24**, 329-338, doi:10.1007/s10654-009-9334-y (2009).

9 Senti, M., Tomas, M., Marrugat, J., Elosua, R. & Investigators, R. Paraoxonase1-192 polymorphism modulates the nonfatal myocardial infarction risk associated with decreased HDLs. *Arterioscler Thromb Vasc Biol* **21**, 415-420 (2001).

10 Samani, N. J. *et al.* Genomewide association analysis of coronary artery disease. *The New England journal of medicine* **357**, 443-453, doi:10.1056/NEJMoa072366 (2007).

11 Myocardial Infarction, G. *et al.* Coding Variation in ANGPTL4, LPL, and SVEP1 and the Risk of Coronary Disease. *The New England journal of medicine* **374**, 1134-1144, doi:10.1056/NEJMoa1507652 (2016).

12 The Atherosclerosis Risk in Communities (ARIC) Study: design and objectives. The ARIC investigators. *American journal of epidemiology* **129**, 687-702 (1989).

13 Mitchell, B. D. *et al.* The genetic response to short-term interventions affecting cardiovascular function: rationale and design of the Heredity and Phenotype Intervention (HAPI) Heart Study. *American heart journal* **155**, 823-828, doi:10.1016/j.ahj.2008.01.019 (2008).

14 Nadkarni, G. N. *et al.* Apolipoprotein L1 Variants and Blood Pressure Traits in African Americans. *Journal of the American College of Cardiology* **69**, 1564-1574, doi:10.1016/j.jacc.2017.01.040 (2017).
15 Hughes, G. H. *et al.* Recruitment in the Coronary Artery Disease Risk Development in Young Adults (Cardia) Study. *Control Clin Trials* **8**, 68S-73S (1987).
16 Redline, S. *et al.* Risk factors for sleep-disordered breathing in children. Associations with obesity, race, and respiratory problems. *Am J Respir Crit Care Med* **159**, 1527-1532, doi:10.1164/ajrccm.159.5.9809079 (1999).
17 Dean, D. A., 2nd *et al.* Scaling Up Scientific Discovery in Sleep Medicine: The National Sleep Research Resource. *Sleep* **39**, 1151-1164, doi:10.5665/sleep.5774 (2016).
18 Redline, S. *et al.* The familial aggregation of obstructive sleep apnea. *Am J Respir Crit Care Med* **151**, 682-687, doi:10.1164/ajrccm.151.3.7881656 (1995).
19 Zhang, G. Q. *et al.* The National Sleep Research Resource: towards a sleep data commons. *J Am Med Inform Assoc* **25**, 1351-1358, doi:10.1093/jamia/ocy064 (2018).
20 Fried, L. P. *et al.* The Cardiovascular Health Study: design and rationale. *Annals of epidemiology* **1**, 263-276 (1991).
21 O'Leary, D. H. *et al.* Carotid-artery intima and media thickness as a risk factor for myocardial infarction and stroke in older adults. Cardiovascular Health Study Collaborative Research Group. *The New England journal of medicine* **340**, 14-22, doi:10.1056/NEJM199901073400103 (1999).
22 O'Leary, D. H. *et al.* Use of sonography to evaluate carotid atherosclerosis in the elderly. The Cardiovascular Health Study. CHS Collaborative Research Group. *Stroke; a journal of cerebral circulation* **22**, 1155-1163 (1991).
23 Bowden, D. W. *et al.* Review of the Diabetes Heart Study (DHS) family of studies: a comprehensively examined sample for genetic and epidemiological studies of type 2 diabetes and its complications. *Rev Diabet Stud* **7**, 188-201, doi:10.1900/RDS.2010.7.188 (2010).
24 Castelli, W. P. Epidemiology of coronary heart disease: the Framingham study. *Am J Med* **76**, 4-12 (1984).
25 Kannel, W. B., Castelli, W. P., Gordon, T. & McNamara, P. M. Serum cholesterol, lipoproteins, and the risk of coronary heart disease. The Framingham study. *Annals of internal medicine* **74**, 1-12 (1971).
26 Kannel, W. B., Dawber, T. R., Kagan, A., Revotskie, N. & Stokes, J., 3rd. Factors of risk in the development of coronary heart disease--six year follow-up experience. The Framingham Study. *Annals of internal medicine* **55**, 33-50 (1961).
27 Kannel, W. B., Feinleib, M., McNamara, P. M., Garrison, R. J. & Castelli, W. P. An investigation of coronary heart disease in families. The Framingham offspring study. *American journal of epidemiology* **110**, 281-290 (1979).
28 FBPP Investigators. Multi-center genetic study of hypertension: The Family Blood Pressure Program (FBPP). *Hypertension* **39**, 3-9 (2002).
29 Daniels, P. R. *et al.* Familial aggregation of hypertension treatment and control in the Genetic Epidemiology Network of Arteriopathy (GENOA) study. *Am J Med* **116**, 676-681, doi:10.1016/j.amjmed.2003.12.032 (2004).
30 Liu, Y. *et al.* Pharmacogenetic association of the APOA1/C3/A4/A5 gene cluster and lipid responses to fenofibrate: the genetics of lipid-lowering drugs

and diet network study. *Pharmacogenet Genomics* **19**, 161-169, doi:10.1097/FPC.0b013e32831e030e (2009).
31 Higgins, M. *et al.* NHLBI Family Heart Study: objectives and design. *American journal of epidemiology* **143**, 1219-1228 (1996).
32 Patsch, J. R. *et al.* Relation of triglyceride metabolism and coronary artery disease. Studies in the postprandial state. *Arteriosclerosis and thrombosis : a journal of vascular biology / American Heart Association* **12**, 1336-1345 (1992).
33 GenSalt Collaborative Research, G. GenSalt: rationale, design, methods and baseline characteristics of study participants. *J Hum Hypertens* **21**, 639-646, doi:10.1038/sj.jhh.1002207 (2007).
34 Kral, B. G. *et al.* A common variant in the CDKN2B gene on chromosome 9p21 protects against coronary artery disease in Americans of African ancestry. *J Hum Genet* **56**, 224-229, doi:10.1038/jhg.2010.171 (2011).
35 Bray, P. F. *et al.* Heritability of platelet function in families with premature coronary artery disease. *J Thromb Haemost* **5**, 1617-1623, doi:10.1111/j.1538-7836.2007.02618.x (2007).
36 Sorlie, P. D. *et al.* Design and implementation of the Hispanic Community Health Study/Study of Latinos. *Annals of epidemiology* **20**, 629-641, doi:10.1016/j.annepidem.2010.03.015 (2010).
37 Arnett, D. K. *et al.* Sibling correlation of left ventricular mass and geometry in hypertensive African Americans and whites: the HyperGEN study. Hypertension Genetic Epidemiology Network. *Am J Hypertens* **14**, 1226-1230, doi:10.1016/s0895-7061(01)02200-2 (2001).
38 Williams, R. R. *et al.* NHLBI family blood pressure program: methodology and recruitment in the HyperGEN network. Hypertension genetic epidemiology network. *Annals of epidemiology* **10**, 389-400 (2000).
39 Fuqua, S. R. *et al.* Recruiting African-American research participation in the Jackson Heart Study: methods, response rates, and sample description. *Ethnicity & disease* **15**, S6-18-29 (2005).
40 Wilson, J. G. *et al.* Study design for genetic analysis in the Jackson Heart Study. *Ethnicity & disease* **15**, S6-30-37 (2005).
41 Bild, D. E. *et al.* Multi-Ethnic Study of Atherosclerosis: objectives and design. *American journal of epidemiology* **156**, 871-881 (2002).
42 Ellinor, P. T., Low, A. F., Patton, K. K., Shea, M. A. & Macrae, C. A. Discordant atrial natriuretic peptide and brain natriuretic peptide levels in lone atrial fibrillation. *Journal of the American College of Cardiology* **45**, 82-86, doi:10.1016/j.jacc.2004.09.045 (2005).
43 Ellinor, P. T., Yoerger, D. M., Ruskin, J. N. & MacRae, C. A. Familial aggregation in lone atrial fibrillation. *Human genetics* **118**, 179-184, doi:10.1007/s00439-005-0034-8 (2005).
44 Mitchell, B. D. *et al.* Genetic and environmental contributions to cardiovascular risk factors in Mexican Americans. The San Antonio Family Heart Study. *Circulation* **94**, 2159-2170 (1996).
45 Hawley, N. L. *et al.* Prevalence of adiposity and associated cardiometabolic risk factors in the Samoan genome-wide association study. *Am J Hum Biol* **26**, 491-501, doi:10.1002/ajhb.22553 (2014).
46 Minster, R. L. *et al.* A thrifty variant in CREBRF strongly influences body mass index in Samoans. *Nat Genet* **48**, 1049-1054, doi:10.1038/ng.3620 (2016).

47 Ranade, K. *et al.* A genome scan for hypertension susceptibility loci in populations of Chinese and Japanese origins. *Am J Hypertens* **16**, 158-162, doi:10.1016/s0895-7061(02)03245-4 (2003).
48 Wu, K. D. *et al.* Clustering and heritability of insulin resistance in Chinese and Japanese hypertensive families: a Stanford-Asian Pacific Program in Hypertension and Insulin Resistance sibling study. *Hypertens Res* **25**, 529-536 (2002).
49 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 (1998).
50 Anderson, G. L. *et al.* Effects of conjugated equine estrogen in postmenopausal women with hysterectomy: the Women's Health Initiative randomized controlled trial. *JAMA : the journal of the American Medical Association* **291**, 1701-1712, doi:10.1001/jama.291.14.1701 (2004).
51 Manson, J. E. *et al.* Menopausal Hormone Therapy and Long-term All-Cause and Cause-Specific Mortality: The Women's Health Initiative Randomized Trials. *JAMA : the journal of the American Medical Association* **318**, 927-938, doi:10.1001/jama.2017.11217 (2017).

Acknowledgements

This work was supported by a grant from the Swedish Research Council (2016-06830) to GH. This work was supported by grants from the National Heart, Lung, and Blood Institute (NHLBI): R01HL142711 and R01HL127564 (P.N. and G.M.P.), R03HL141439 (G.M.P.), R01HL148050 and R01HL148565 (PN). P.N. is also supported by a Hassenfeld Scholar Award from the Massachusetts General Hospital, Foundation Leducq (TNE-18CVD04), and additional grants from the National Heart, Lung, and Blood Institute (R01HL148565 and R01HL148050). PSdV is supported by American Heart Association grant number 18CDA34110116. BEC and JL are supported by R35HL135818, HL113338, and HL46380. BEC is also supported by K01HL135405. S.A.L. is supported by NIH grant 1R01HL139731 and American Heart Association 18SFRN34250007. AVK is supported by K08HG010155. JF is supported by R01DK125490. DJL is supported by R01HG008983 and R01GM126479. CJW is supported by R35HL135824, R01HL142023, and R01HL127564. MJB is supported British Heart Foundation CS/14/2/30841. OM is supported by European Research Council ERC-AdG-2019-885003, NNF17OC0026936, and VR 2018-02760. MOM is supported by the European Research Council ERC-CoG-2014-649021 and NNF18OC0034386. RM is supported by Canadian Institutes Health Research FDN 154308. SR is supported by R35HL135818, HL113338, HL46380, and K01HL135405. NDP is supported by R01 HL92301, R01 HL67348, R01 NS058700, R01 AR48797, R01 DK071891, R01 AG058921, the General Clinical Research Center of the Wake Forest University School of Medicine (M01 RR07122, F32 HL085989), the American Diabetes Association, and a pilot grant from the Claude Pepper Older Americans Independence Center of Wake Forest University Health Sciences (P60 AG10484). BIF is supported by R01 NS058700 and R01 DK071891.

QQ is supported by R01-DK119268, R01-HL060712, and R01-HL140976. PTE is supported by the National Institutes of Health (1RO1HL092577, R01HL128914, K24HL105780), the American Heart Association (18SFRN34110082), and by the Foundation Leducq (14CVD01). NLHC is supported by HHSN268201500001. JCF is support by NIDDK U01 DK105554 and NIDDK K24 DK110550. MIM is supported by NIDDK U01-DK105535 and Wellcome 090532, 098381, 106130, 203141, 212259. NPB is supported by NIDDK U01-DK105554. NG is supported by Novo Nordisk Foundation Center for Basic Metabolic Research is an independent Research Center, based at the University of Copenhagen, Denmark and partially funded by an unconditional donation from the Novo Nordisk Foundation (www.cbmr.ku.dk) (Grant number NNF18CC0034900). TLT and HMM are supported by CONACYT, 312688. KLM is supported by US National Institutes of Health DK072193 and DK093757. CLH is supported by DK085591. EST is supported by the National Medical Research Council of Singapore Clinician Scientist Award. NB is supported by The American Federation for Aging Research, the Einstein Glenn Center and the NIA (PO1AG027734, R01AG 046949, 1R01AG042188 and P30AG038072). RCWM is supported by the Hong Kong Research Grants Council Theme-based Research Scheme (T12-402/13N) and Research Impact Fund (R4012-18).

The KARE cohort was supported by grants from Korea Centers for Disease Control and Prevention (4845–301, 4851–302, 4851–307) and intramural grants from the Korea National Institute of Health (2019-NG-053-01). Collection of the San Antonio Family Study data were supported in part by National Institutes of Health (NIH) grants P01 HL045522, R01 MH078143, R01 MH078111 and R01 MH083824; and whole genome sequencing of SAFS subjects was supported by U01 DK085524 and R01 HL113323. We are very grateful to the participants of the San Antonio Family

Study for their continued involvement in our research programs. DGI (Principal investigators Leif Groop and Tiinamaija Tuomi) and the Botnia Study is supported by the Sigrid Juselius Foundation, The Folkhalsan Research Foundation, Nordic Center of Excellence in Disease Genetics, EU (EXGENESIS), Finnish Diabetes Research Foundation, Foundation for Life and Health in Finland, Finnish Medical Society, Helsinki University Central Hospital Research Foundation, Perklén Foundation, Ollqvist Foundation, Narpes Health Care Foundation as well as the Municipal Heath Care Center and Hospital in Jakobstad and Health Care Centers in Vasa, Narpes and Korsholm. The work in Malmö, Sweden, was also funded by a Linné grant from the Swedish Research Council (349-2006-237). The contribution of the Botnia and Skara research teams is gratefully acknowledged. The TwinsUK study was funded by the Wellcome Trust and European Community's Seventh Framework Programme (FP7/2007-2013). The TwinsUK study also receives support from the National Institute for Health Research (NIHR)- funded BioResource, Clinical Research Facility and Biomedical Research Centre based at Guy's and St Thomas' NHS Foundation Trust in partnership with King's College London. The KORA study was initiated and financed by the Helmholtz Zentrum München – German Research Center for Environmental Health, which is funded by the German Federal Ministry of Education and Research (BMBF) and by the State of Bavaria. Furthermore, KORA research was supported within the Munich Center of Health Sciences (MC-Health), Ludwig-Maximilians-Universität, as part of LMUinnovativ. Funded by the Bavarian State Ministry of Health and Care through the research project DigiMed Bayern (www.digimed-bayern.de).

Molecular data for the Trans-Omics in Precision Medicine (TOPMed) program was supported by the National Heart, Lung and Blood Institute (NHLBI). See the TOPMed Omics Support information in the **Supplementary Text** for study specific omics support information. Core support including centralized genomic read mapping and genotype calling, along with variant quality metrics and filtering were provided by the TOPMed Informatics Research Center (3R01HL-117626-02S1; contract HHSN268201800002I). Core support including phenotype harmonization, data management, sample-identity QC, and general program coordination were provided by the TOPMed Data Coordinating Center (R01HL-120393; U01HL-120393; contract HHSN268201800001I). We gratefully acknowledge the studies and participants who provided biological samples and data for TOPMed. The views expressed in this manuscript are those of the authors and do not necessarily represent the views of the National Heart, Lung, and Blood Institute; the National Institutes of Health; or the U.S. Department of Health and Human Services.

AMP-T2D-GENES Consortium

Carlos A. Aguilar-Salinas[1], Gil Atzmon[2,3], Francisco Barajas-Olmos[4], Nir Barzilai[2], John Blangero[5], Michael Boehnke[6], Eric Boerwinkle[7,8], Lori L. Bonnycastle[9], Erwin Bottinger[10,11], Donald W. Bowden[12], Noël P. Burtt[13], Federico Centeno-Cruz[4], John C. Chambers[14-16], Edmund Chan[17], Juliana Chan[18-20], Ching-Yu Cheng[21-23], Yoon Shin Cho[24], Cecilia Contreras-Cubas[4], Emilio J. Córdova[4], Adolfo Correa[25], Ralph A. DeFronzo[26], Ravindranath Duggirala[5], Josée Dupuis[27], Jason A. Flannick[13,28,29], Jose C. Florez[30-33], Ma. Eugenia Garay-Sevilla[34], Humberto García-Ortiz[4], Christian Gieger[35-37], Benjamin Glaser[38], Clicerio Gonzalez[39], Maria Elena Gonzalez-Villalpando[40], Niels Grarup[41], Leif C. Groop[42,43], Myron Gross[44], Christopher Haiman[45], Sohee Han[46], Craig L. Hanis[47], Torben Hansen[41,48], Nancy L. Heard-Costa[49,50], Brian E. Henderson[45], Mi Yeong Hwang[46], Sergio Islas-Andrade[51], Marit E. Jørgensen[52-54], Hyun Min Kang[6], Bong-Jo Kim[46], Young Jin Kim[46], Heikki A. Koistinen[55-57], Jaspal Singh Kooner[58], Johanna Kuusisto[59], Soo Heon Kwak[60], Markku Laakso[59], Leslie A. Lange[61], Jong-Young Lee[62], Juyoung Lee[46], Donna M. Lehman[26], Allan Linneberg[63,64], Jianjun Liu[17,65], Ruth J.F. Loos[66,67], Valeriya Lyssenko[42,68], Ronald C.W. Ma[18-20], Juan Manuel Malacara Hernandez[69], Angélica Martínez-Hernández[4], Mark I. McCarthy[70,71], James B. Meigs[33,72,73], Thomas Meitinger[74,75], Elvia Mendoza-Caamal[4], Karen L. Mohlke[76], Andrew D. Morris[77], Alanna C. Morrison[7], Maggie C.Y. Ng[12,78,79], Peter M. Nilsson[80], Christopher J. O'Donnell[81-84], Lorena Orozco[4], Colin N.A. Palmer[85], Kyong Soo Park[60,86,87], Oluf Pedersen[41], Wendy S. Post[88], Michael Preuss[66], Bruce M. Psaty[89-91], Alexander P. Reiner[92], Cristina Revilla-Monsalve[93], Stephen S. Rich[94], Jerome I. Rotter[95], Danish Saleheen[96,97], Claudia Schurmann[10,11,66], Xueling Sim[98], Rob Sladek[99-101], Kerrin S. Small[102], Wing Yee So[18-20], Xavier Soberón[4], Timothy D. Spector[102], Konstantin Strauch[36,103], Tim M. Strom[75,104], E Shyong Tai[17,98,105], Claudia H.T. Tam[18-20], Yik Ying Teo[98,106,107], Farook Thameem[108], Brian Tomlinson[109], Russell P. Tracy[110,111], Tiinamaija Tuomi[112-114], Jaakko Tuomilehto[115-117], Teresa Tusié-Luna[1,118], Rob M. van Dam[98,119], Ramachandran S. Vasan[49,120], James G. Wilson[121], Daniel R. Witte[122,123], Tien-Yin Wong[21-23]

1 - Instituto Nacional de Ciencias Medicas y Nutricion , Mexico City, Mexico.; 2 - Departments of Medicine and Genetics, Albert Einstein College of Medicine, New York, USA.; 3 - University of Haifa, Faculty of natural science, Haifa, Isarel.; 4 - Instituto Nacional de Medicina Genómica, Mexico City, Mexico.; 5 - Department of Human Genetics and South Texas Diabetes and Obesity Institute, University of Texas Rio Grande Valley School of Medicine, Brownsville, TX 78520, USA.; 6 - Department of Biostatistics and Center for Statistical Genetics, University of Michigan, Ann Arbor, Michigan, USA.; 7 - Human Genetics Center, Department of Epidemiology, Human Genetics, and Environmental Sciences, School of Public Health, The University of Texas Health Science Center at Houston, Houston, TX, USA, 77030.; 8 - Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX, 77030, USA.; 9 - Medical Genomics and Metabolic Genetics Branch, National Human Genome Research Institute, National Institutes of Health, Bethesda, Maryland, USA.; 10 - Hasso Plattner Institute for Digital Health at Mount Sinai, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA.; 11 - Digital Health Center, Hasso Plattner Institute, University of Potsdam, Potsdam, Germany.; 12 - Department of Biochemistry, Wake Forest School of Medicine, Winston-Salem, NC, USA, 27157.; 13 - Programs in Metabolism and Medical & Population Genetics, Broad Institute, Cambridge, Massachusetts, USA.; 14 - Department of Epidemiology and Biostatistics, Imperial College London, London, UK.; 15 - Imperial College Healthcare NHS Trust, Imperial College London, London, UK.; 16 - Department of Cardiology, Ealing Hospital NHS Trust, Southall, Middlesex, UK.; 17 - Department of Medicine, Yong Loo Lin School of Medicine, National University of Singapore, National University Health System, , Singapore.; 18 - Department of Medicine and Therapeutics, The Chinese University of Hong Kong, Hong Kong, China.; 19 - Hong Kong Institute of Diabetes and Obesity, The Chinese University of Hong Kong, Hong Kong, China.; 20 - Li Ka Shing Institute of Health Sciences, The Chinese University of Hong Kong, Hong Kong, China.; 21 - Department of Ophthalmology, Yong Loo Lin School of Medicine, National University of Singapore, National University Health System, , Singapore.; 22 - Ophthalmology & Visual Sciences Academic Clinical Program (Eye ACP), Duke-NUS Medical School, Singapore.; 23 - Singapore Eye Research Institute, Singapore National Eye Centre, Singapore.; 24 - Department of Biomedical Science, Hallym University, Chuncheon, Republic of Korea.; 25 - Department of Medicine, University of Mississippi Medical Center, Jackson, MS, USA, 39216.; 26 - Department of Medicine, University of Texas Health Science Center, San Antonio, Texas, USA.; 27 - Department of Biostatistics, Boston University School of Public Health, Boston, Massachusetts, USA.; 28 - Department of Pediatrics, Harvard Medical School, Boston, MA, USA, 02115.; 29 - Division of Genetics and Genomics, Boston Children's Hospital, Boston, MA, USA, 02115.; 30 - Center for Genomic Medicine, Department of Medicine, Massachusetts General Hospital, Boston, Massachusetts, USA, 02114.; 31 - Diabetes Research Center (Diabetes Unit), Department of Medicine, Massachusetts General Hospital, Boston, Massachusetts, USA.; 32 - Program in Medical and Population Genetics, Broad Institute of Harvard and MIT, Cambridge, MA, USA, 02142.; 33 - Department of Medicine, Harvard Medical School, Boston, MA, USA, 02115.; 34 - Department of Medical Science, Division of Health Science. University of Guanajuato.; 35 - Research Unit of Molecular Epidemiology, Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg, Germany.; 36 - Institute of Epidemiology, Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg, Germany.; 37 - German Center for Diabetes Research (DZD), Neuherberg, Germany.; 38 - Endocrinology and Metabolism Service, Hadassah-Hebrew University Medical Center, Jerusalem, Israel.; 39 - Unidad de Diabetes y Riesgo Cardiovascular, Instituto Nacional de Salud Pœblica, Cuernavaca, Morelos, Mexico.; 40 - Centro de Estudios en Diabetes, Mexico City, Mexico.; 41 - Novo Nordisk Foundation Center for Basic Metabolic Research, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark.; 42 - Department of Clinical Sciences, Diabetes and Endocrinology, Lund University Diabetes Centre, Malmö, Sweden.; 43 - Finnish Institute for Molecular Genetics, University of Helsinki, Helsinki, Finland.; 44 - Department of Laboratory Medicine and Pathology, University of Minnesota, Minneapolis, Minnesota, USA.; 45 - Department of Preventive Medicine, Keck School of Medicine, University of Southern California, Los Angeles, California, USA.; 46 - Division of Genome Science, Department of Precision Medicine, Chungcheongbuk-do, Republic of Korea.; 47 - Human Genetics Center, School of Public Health, The University of Texas Health Science Center at Houston, Houston, Texas, USA.; 48 - Faculty of Health Sciences, University of Southern Denmark, Odense,

Denmark.; 49 - NHLBI Framingham Heart Study, Framingham, Massachusetts, USA.; 50 - Department of Neurology, Boston University School of Medicine, Boston, Massachusetts, USA.; 51 - Dirección de Investigación, Hospital General de México "Dr. Eduardo Liceaga", Secretaría de Salud, Mexico City, Mexico.; 52 - National Institute of Public Health, University of Southern Denmark, Copenhagen, Denmark.; 53 - Steno Diabetes Center Copenhagen, Gentofte, Denmark.; 54 - Greenland Centre for Health Research, University of Greenland, Nuuk, Greenland.; 55 - Department of Public Health Solutions, Finnish Institute for Health and Welfare, Helsinki, Finland.; 56 - Minerva Foundation Institute for Medical Research, Helsinki, Finland.; 57 - University of Helsinki and Department of Medicine, Helsinki University Central Hospital, Helsinki, Finland.; 58 - National Heart and Lung Institute, Cardiovascular Sciences, Hammersmith Campus, Imperial College London, London, UK.; 59 - Institute of Clinical Medicine, University of Eastern Finland, and Kuopio University Hospital, Kuopio, Finland.; 60 - Department of Internal Medicine, Seoul National University Hospital, Seoul, Republic of Korea.; 61 - Department of Medicine, University of Colorado Denver, Anschutz Medical Campus, Aurora, CO, USA 80045.; 62 - Ministry of Health and Welfare, Seoul, Republic of Korea.; 63 - Center for Clinical Research and Prevention, Bispebjerg and Frederiksberg Hospital, The Capital Region, Copenhagen, Denmark.; 64 - Department of Clinical Medicine, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark.; 65 - Genome Institute of Singapore, Agency for Science, Technology and Research, Singapore.; 66 - Charles R. Bronfman Institute of Personalized Medicine, Icahn School of Medicine at Mount Sinai, New York, New York, USA.; 67 - The Mindich Child Health and Development Institute, Icahn School of Medicine at Mount Sinai, New York, NY 10029.; 68 - University of Bergen, Norway.; 69 - Departments of Medicine and Human Genetics, The University of Chicago, Chicago, Illinois, USA.; 70 - Oxford Centre for Diabetes, Endocrinology and Metabolism, Radcliffe Department of Medicine, University of Oxford, Oxford, UK.; 71 - Wellcome Centre for Human Genetics, University of Oxford, Oxford, UK.; 72 - General Medicine Division, Massachusetts General Hospital, Boston, Massachusetts, USA.; 73 - Broad Institute of MIT and Harvard, Cambridge, Massachusetts, USA.; 74 - Deutsches Forschungszentrum fŸr Herz-Kreislauferkrankungen (DZHK), Partner Site Munich Heart Alliance, Munich, Germany.; 75 - Institute of Human Genetics, Technische Universität München, Munich, Germany.; 76 - Department of Genetics, University of North Carolina, Chapel Hill, North Carolina, USA.; 77 - Clinical Research Centre, Centre for Molecular Medicine, Ninewells Hospital and Medical School, Dundee, UK.; 78 - Center for Diabetes Research, Wake Forest School of Medicine, Winston-Salem, North Carolina, USA.; 79 - Center for Genomics and Personalized Medicine Research, Wake Forest School of Medicine, Winston-Salem, North Carolina, USA.; 80 - Department of Clinical Sciences, Lund University, Malmö, Sweden.; 81 - Section of Cardiology , Department of Medicine, VA Boston Healthcare, Boston, Massachusetts, USA.; 82 - Harvard Medical School, Boston, Massachusetts, USA.; 83 - Department of Medicine, Brigham and Women's Hospital, Boston, MA 02115.; 84 - Intramural Administration Management Branch, National Heart, Lung, and Blood Institute, NIH, Framingham, Massachusetts, USA.; 85 - Pat Macpherson Centre for Pharmacogenetics and Pharmacogenomics,Division of Population Health and Genomics, University of Dundee, Ninewells Hospital and Medical School, Dundee, UK.; 86 - Department of Internal Medicine, Seoul National University College of Medicine, Seoul, Republic of Korea.; 87 - Department of Molecular Medicine and Biopharmaceutical Sciences, Graduate School of Convergence Science and Technology, Seoul National University, Seoul, Republic of Korea.; 88 - Division of Cardiology, Department of Medicine, Johns Hopkins University, Baltimore, Maryland, USA.; 89 - Cardiovascular Health Research Unit, Department of Medicine, University of Washington, Seattle, WA, USA, 98101.; 90 - Department of Epidemiology, University of Washington, Seattle, WA, USA 98101.; 91 - Department of Health Services, University of Washington, Seattle, WA, USA 98101.; 92 - University of Washington, Seattle, Washington, USA.; 93 - Instituto Mexicano del Seguro Social SXXI, Mexico City, Mexico.; 94 - Center for Public Health Genomics, University of Virginia, Charlottesville, VA 22908, USA.; 95 - The Institute for Translational Genomics and Population Sciences, Department of Pediatrics, The Lundquist Institute for Biomedical Innovation at Harbor-UCLA Medical Center, Torrance, CA 90502.; 96 - Department of Biostatistics and Epidemiology, University of Pennsylvania, Philadelphia, Pennsylvania, USA.; 97 - Center for Non-Communicable Diseases, Karachi, Pakistan.; 98 - Saw Swee Hock School of Public Health, National University of Singapore and National University Health System, Singapore.; 99 - Department of Human Genetics, McGill University, Montreal, Quebec, Canada.; 100 - Division of Endocrinology and Metabolism, Department of Medicine, McGill University, Montreal, Quebec, Canada.; 101 - McGill University and Génome Québec Innovation Centre, Montreal, Quebec, Canada.; 102 - Department of Twin Research and Genetic Epidemiology, King's College London, London, UK.; 103 - Institute of Genetic Epidemiology, Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg, Germany.; 104 - Institute of Human Genetics, Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg, Germany.; 105 - Duke-NUS Medical School Singapore, Singapore.; 106 - Department of Statistics and Applied Probability, National University of Singapore, Singapore.; 107 - Life Sciences Institute, National University of Singapore, Singapore.; 108 - Department of Biochemistry, Faculty of Medicine, Health Science Center, Kuwait University, Safat, Kuwait.; 109 - Faculty of Medicine, Macau University of Science & Technology, Macau, China.; 110 - Department of Biochemistry, Robert Larner, M.D. College of Medicine, University of Vermont, Burlington, Vermont, USA.; 111 - Department of Pathology and Laboratory Medicine, Robert Larner, M.D. College of Medicine, University of Vermont, Burlington, Vermont, USA.; 112 - Department of Endocrinology, Abdominal Centre, Helsinki University Hospital, Helsinki, Finland.; 113 - Folkhälsan Research Centre, Helsinki, Finland.; 114 - Research Programs Unit, Diabetes and Obesity, University of Helsinki, Helsinki, Finland.; 115 - Public Health Promition Unit, Finnish Institute for Health and Welfare, Helsinki, Finland.; 116 - Department of Public Health, University of Helsinki, Helsinki, Finland.; 117 - Diabetes Research Group, King Abdulaziz University, Jeddah, Saudi Arabia.; 118 - Departamento de Medicina Genómica y Toxicología, Instituto de Investigaciones Biomédicas, Universidad Nacional Autónoma de México/ Instituto Nacional de Ciencias Médicas y Nutrición Salvador Zubirán, Mexico City, Mexico.; 119 - Department of Nutrition, Harvard School of Public Health, Boston, Massachusetts, USA.; 120 - Departments of Medicine & Epidemiology, Boston University Schools of Medicine & Public Health, Boston, MA.; 121 - Division of Cardiovascular Medicine, Beth Israel Deaconess Medical Center, Boston, MA 02215.; 122 - Department of Public Health, Aarhus University, Aarhus, Denmark.; 123 - Steno Diabetes Center Aarhus, Aarhus, Denmark.;

Myocardial Infarction Genetics Consortium

Diego Ardissino[1-3], Usman Baber[4], Matthew J. Bown[5,6], Mark D. Chaffin[7-9], Rajiv Chowdhury[10,11], John Danesh[10,12], Roberto Elosua Llanos[13-15], Valentin Fuster[4,16], Namrata Gupta[7], Sekar Kathiresan[7,8,17,18], Amit V. Khera[7-9], Ruth McPherson[19], Olle Melander[20,21], Marju Orho-Melander[22], Danish Saleheen[23,24], Nilesh Samani[5,6], Heribert Schunkert[25], Minxian X. Wang[7,8,26], Hugh Watkins[27]

1 - ASTC: Associazione per lo Studio Della Trombosi in Cardiologia, Pavia, Italy.; 2 - Azienda Ospedaliero-Universitaria di Parma,ÊParma,Italy.; 3 - Universitˆ degli Studi di Parma, Parma, Italy.; 4 - Department of Medicine, Icahn School of Medicine at Mount Sinai, New York, NY 10029.; 5 - Department of Cardiovascular Sciences, University of Leicester, Leicester, UK.; 6 - NIHR Leicester Biomedical Research Centre, Glenfield Hospital, Leicester UK.; 7 - Program in Medical and Population Genetics, Broad Institute of Harvard and MIT, Cambridge, MA, USA, 02142.; 8 - Center for Genomic Medicine, Department of Medicine, Massachusetts General Hospital, Boston, Massachusetts, USA, 02114.; 9 - Cardiovascular Research Center, Massachusetts General Hospital, Boston, MA, USA, 02114.; 10 - MRC/BHF Cardiovascular Epidemiology Unit, Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK.; 11 - Centre for Non-Communicable disease Research (CNCR), Bangladesh.; 12 - The National Institute for Health Research Blood and Transplant Research Unit (NIHR BTRU) in Donor Health and Genomics at the University of Cambridge, Cambridge, UK.; 13 - Cardiovascular Epidemiology and Genetics, Hospital del Mar Research Institute, Barcelona, Spain.; 14 - CIBER Enfermedades Cardiovasculares (CIBERCV), Barcelona, Spain.; 15 - Facultat de Medicina, Universitat de Vic-Central de Catalu–a, Vic, Spain.; 16 - Centro Nacional de Investigaciones Cardiovasculares Carlos III (CNIC), Madrid, Spain.; 17 - Department of Medicine, Harvard Medical School, Boston, MA, USA, 02115.; 18 - Verve Therapeutics, Cambridge, MA, USA 02139.; 19 - Ruddy Canadian Cardiovascuar Genetics Centre, University of Ottawa Heart Institute, Ottawa, Canada.; 20 - Department of Clinical Sciences, Diabetes and Endocrinology, Lund University Diabetes Centre, Malmö, Sweden.; 21 - Department of Emergency and Internal Medicine, SkŒne University Hospital, Malmö, Sweden.; 22 - Department of Clinical Sciences, Lund University, Malmö, Sweden.; 23 - Department of Biostatistics and Epidemiology, University of Pennsylvania, Philadelphia, Pennsylvania, USA.; 24 - Center for Non-Communicable Diseases, Karachi, Pakistan.; 25 - Deutsches Herzzentrum MŸnchen, Technische UniversitŠt MŸnchen, Deutsches Zentrum fŸr Herz-Kreislauf-Forschung, MŸnchen, Germany.; 26 - Cardiovascular Disease Initaitive, Broad Institute of Harvard and MIT, Cambridge, MA, USA, 02142.; 27 - Cardiovascular Medicine, Radcliffe Department of Medicine and the Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK.;

NHLBI Trans-Omics for Precision Medicine (TOPMed) Consortium

Namiko Abe[1], Gonalo Abecasis[2], Francois Aguet[3], Christine Albert[4], Laura Almasy[5], Alvaro Alonso[6], Seth Ament[7], Peter Anderson[8], Pramod Anugu[9], Deborah Applebaum-Bowden[10], Kristin Ardlie[3], Dan Arking[11], Donna K Arnett[12], Allison Ashley-Koch[13], Stella Aslibekyan[14], Tim Assimes[15], Paul Auer[16], Dimitrios Avramopoulos[11], John Barnard[17], Kathleen Barnes[18], R. Graham Barr[19], Emily Barron-Casella[11], Lucas Barwick[20], Terri Beaty[11], Gerald Beck[21], Diane Becker[22], Lewis Becker[11], Rebecca Beer[23], Amber Beitelshees[7], Emelia Benjamin[24], Takis Benos[25], Marcos Bezerra[26], Larry Bielak[2], Joshua Bis[27], Thomas Blackwell[2], John Blangero[28], Eric Boerwinkle[29], Donald W. Bowden[30], Russell Bowler[31], Jennifer Brody[8], Ulrich Broeckel[32], Jai Broome[8], Karen Bunting[1], Esteban Burchard[33], Carlos Bustamante[34], Erin Buth[35], Brian Cade[36], Jonathan Cardwell[37], Vincent Carey[38], Cara Carty[39], Richard Casaburi[40], James Casella[11], Peter Castaldi[41], Mark Chaffin[3], Christy Chang[7], Yi-Cheng Chang[42], Daniel Chasman[43], Sameer Chavan[37], Bo-Juen Chen[1], Wei-Min Chen[44], Yii-Der Ida Chen[45], Michael Cho[38], Seung Hoan Choi[3], Lee-Ming Chuang[46], Mina Chung[47], Ren-Hua Chung[48], Clary Clish[49], Suzy Comhair[50], Matthew Conomos[35], Elaine Cornell[51], Adolfo Correa[52], Carolyn Crandall[40], James Crapo[53], L. Adrienne Cupples[54], Joanne Curran[55], Jeffrey Curtis[2], Brian Custer[56], Coleen Damcott[7], Dawood Darbar[57], Sayantan Das[2], Sean David[58], Colleen Davis[8], Michelle Daya[37], Michael DeBaun[59], Dawn DeMeo[38], Ranjan Deka[60], Scott Devine[7], Qing Duan[61], Ravi Duggirala[62], Jon Peter Durda[51], Susan Dutcher[63], Charles Eaton[64], Lynette Ekunwe[9], Adel El Boueiz[65], Patrick Ellinor[66], Leslie Emery[8], Serpil Erzurum[17], Charles Farber[44], Tasha Fingerlin[67], Matthew Flickinger[2], Myriam Fornage[29], Nora Franceschini[68], Chris Frazar[8], Mao Fu[7], Stephanie M. Fullerton[8], Lucinda Fulton[63], Stacey Gabriel[3], Weiniu Gan[23], Shanshan Gao[37], Yan Gao[9], Margery Gass[69], Bruce Gelb[70], Xiaoqi (Priscilla) Geng[2], Mark Geraci[71], Soren Germer[1], Robert Gerszten[72], Auyon Ghosh[38], Richard Gibbs[73], Chris Gignoux[15], Mark Gladwin[25], David Glahn[74], Stephanie Gogarten[8], Da-Wei Gong[7], Harald Goring[75], Sharon Graw[18], Daniel Grine[37], C. Charles Gu[63], Yue Guan[7], Xiuqing Guo[45], Namrata Gupta[3], Jeff Haessler[69], Michael Hall[76], Daniel Harris[7], Nicola L. Hawley[77], Jiang He[78], Ben Heavner[35], Susan Heckbert[8], Ryan Hernandez[33], David Herrington[79], Craig Hersh[80], Bertha Hidalgo[14], James Hixson[29], Brian Hobbs[38], John Hokanson[37], Elliott Hong[7], Karin Hoth[81], Chao (Agnes) Hsiung[82], Yi-Jen Hung[83], Haley Huston[84], Chii Min Hwu[85], Marguerite Ryan Irvin[14], Rebecca Jackson[86], Deepti Jain[8], Cashell Jaquish[23], Min A Jhun[2], Jill Johnsen[84], Andrew Johnson[23], Craig Johnson[8], Rich Johnston[6], Kimberly Jones[11], Hyun Min Kang[87], Robert Kaplan[88], Sharon Kardia[2], Sekar Kathiresan[3], Shannon Kelly[56], Eimear Kenny[70], Michael Kessler[7], Alyna Khan[8], Wonji Kim[89], Greg Kinney[90], Barbara Konkle[84], Charles Kooperberg[69], Holly Kramer[91], Christoph Lange[92], Ethan Lange[37], Leslie Lange[37], Cathy Laurie[8], Cecelia Laurie[8], Meryl LeBoff[38], Jonathon LeFaive[2], Jiwon Lee[38], Seunggeun Shawn Lee[2], Wen-Jane Lee[85], David Levine[8], Dan Levy[23], Joshua Lewis[7], Xiaohui Li[45], Yun Li[61], Henry Lin[45], Honghuang Lin[93], Keng Han Lin[2], Xihong Lin[94], Simin Liu[95], Yongmei Liu[96], Yu Liu[97], Ruth J.F. Loos[98], Steven Lubitz[66], Kathryn Lunetta[93], James Luo[23], Michael Mahaney[55], Barry Make[11], Ani Manichaikul[44], JoAnn Manson[38], Lauren Margolin[3], Lisa Martin[99], Susan Mathai[37], Rasika Mathias[11], Susanne May[35], Patrick McArdle[7], Merry-Lynn McDonald[14], Sean McFarland[89], Stephen McGarvey[64], Daniel McGoldrick[8], Caitlin McHugh[35], Hao Mei[9], Luisa Mestroni[18], Deborah A Meyers[100], Julie Mikulla[23], Nancy Min[9], Mollie Minear[23], Ryan L Minster[25], Braxton D. Mitchell[7], Matt Moll[41], May E. Montasser[7], Courtney Montgomery[101], Arden Moscati[70], Solomon Musani[52], Stanford Mwasongwe[9], Josyf C Mychaleckyj[44], Girish Nadkarni[70], Rakhi Naik[11], Take Naseri[102], Pradeep Natarajan[3], Sergei Nekhai[103], Sarah C. Nelson[35], Bonnie Neltner[37], Deborah Nickerson[8], Kari North[61], Jeff O'Connell[104], Tim O'Connor[7], Heather Ochs-Balcom[105], David T. Paik[106], Nicholette Palmer[107], James Pankow[108], George Papanicolaou[23], Afshin Parsa[7], Juan Manuel Peralta[62], Marco Perez[15], James Perry[7], Ulrike Peters[69], Patricia Peyser[2], Lawrence S Phillips[6], Toni Pollin[7], Wendy Post[109], Julia Powers Becker[110], Meher Preethi Boorgula[37], Michael Preuss[70], Bruce Psaty[8], Pankaj Qasba[23], Dandi Qiao[38], Zhaohui Qin[6], Nicholas Rafaels[111], Laura Raffield[112], Vasan S. Ramachandran [93], D.C. Rao[63], Laura Rasmussen-Torvik[113], Aakrosh Ratan[44], Susan Redline[38], Robert Reed[7], Elizabeth Regan[53], Alex Reiner[114], Muagututia Sefuiva Reupena[115], Ken Rice[8], Stephen Rich[44], Dan Roden[116], Carolina Roselli[3], Jerome Rotter[45], Ingo Ruczinski[11], Pamela Russell[37], Sarah Ruuska[84], Kathleen Ryan[7], Ester Cerdeira Sabino[117], Danish Saleheen[19], Shabnam Salimi[7], Steven Salzberg[11], Kevin Sandow[118], Vijay G. Sankaran[119], Christopher Scheller[2], Ellen Schmidt[2], Karen Schwander[63], David Schwartz[37], Frank Sciurba[25], Christine Seidman[120], Jonathan Seidman[121], Vivien Sheehan[122], Stephanie L. Sherman[123], Amol Shetty[7], Aniket Shetty[37], Wayne Hui-Heng Sheu[85], M. Benjamin Shoemaker[124], Brian Silver[125], Edwin Silverman[38], Jennifer Smith[2], Josh Smith[8], Nicholas Smith[126], Tanja Smith[1], Sylvia Smoller[88], Beverly Snively[127], Michael Snyder[15], Tamar Sofer[38], Nona Sotoodehnia[8], Adrienne M. Stilp[8], Garrett Storm[37], Elizabeth Streeten[7], Jessica Lasky Su[38], Yun Ju Sung[63], Jody Sylvia[38], Adam Szpiro[8], Carole Sztalryd[7], Daniel Taliun[2], Hua Tang[128], Margaret Taub[11], Kent D. Taylor[129], Matthew Taylor[18], Simeon Taylor[7], Marilyn Telen[13], Timothy A. Thornton[8], Machiko Threlkeld[130], Lesley Tinker[69], David

Tirschwell[8], Sarah Tishkoff[131], Hemant Tiwari[132], Catherine Tong[133], Russell Tracy[134], Michael Tsai[108], Dhananjay Vaidya[11], David Van Den Berg[135], Peter VandeHaar[2], Scott Vrieze[108], Tarik Walker[37], Robert Wallace[81], Avram Walts[37], Fei Fei Wang[8], Heming Wang[136], Karol Watson[40], Daniel E. Weeks[25], Bruce Weir[8], Scott Weiss[38], Lu-Chen Weng[66], Jennifer Wessel[137], Cristen Willer[138], Kayleen Williams[35], L. Keoki Williams[139], Carla Wilson[38], James Wilson[140], Quenna Wong[8], Joseph Wu[106], Huichun Xu[7], Lisa Yanek[11], Ivana Yang[37], Rongze Yang[7], Norann Zaghloul[7], Maryam Zekavat[3], Yingze Zhang[141], Snow Xueyan Zhao[53], Wei Zhao[142], Degui Zhi[29], Xiang Zhou[2], Xiaofeng Zhu[143], Michael Zody[1], Sebastian Zoellner[2], Mariza de Andrade[144], Lisa de las Fuentes[145]

1 - New York Genome Center, New York, New York, 10013; 2 - University of Michigan, Ann Arbor, Michigan, 48109; 3 - Broad Institute, Cambridge, Massachusetts, 02142; 4 - Cedars Sinai, Boston, Massachusetts, 02114; 5 - Children's Hospital of Philadelphia, University of Pennsylvania, Philadelphia, Pennsylvania, 19104; 6 - Emory University, Atlanta, Georgia, 30322; 7 - University of Maryland, Baltimore, Maryland, 21201; 8 - University of Washington, Seattle, Washington, 98195; 9 - University of Mississippi, Jackson, Mississippi, 38677; 10 - National Institutes of Health, Bethesda, Maryland, 20892; 11 - Johns Hopkins University, Baltimore, Maryland, 21218; 12 - University of Kentucky, Lexington, Kentucky, 40506; 13 - Duke University, Durham, North Carolina, 27708; 14 - University of Alabama, Birmingham, Alabama, 35487; 15 - Stanford University, Stanford, California, 94305; 16 - University of Wisconsin Milwaukee, Milwaukee, Wisconsin, 53211; 17 - Cleveland Clinic, Cleveland, Ohio, 44195; 18 - University of Colorado Anschutz Medical Campus, Aurora, Colorado, 80045; 19 - Columbia University, New York, New York, 10027; 20 - The Emmes Corporation, LTRC, Rockville, Maryland, 20850; 21 - Cleveland Clinic, Quantitative Health Sciences, Cleveland, Ohio, 44195; 22 - Johns Hopkins University, Medicine, Baltimore, Maryland, 21218; 23 - National Heart, Lung, and Blood Institute, National Institutes of Health, Bethesda, Maryland, 20892; 24 - Boston University, Massachusetts General Hospital, Boston University School of Medicine, Boston, Massachusetts, 02114; 25 - University of Pittsburgh, Pittsburgh, Pennsylvania, 15260; 26 - Fundao de Hematologia e Hemoterapia de Pernambuco - Hemope, Recife, 52011-000; 27 - University of Washington, Cardiovascular Health Research Unit, Department of Medicine, Seattle, Washington, 98195; 28 - University of Texas Rio Grande Valley School of Medicine, Human Genetics, Brownsville, Texas, 78520; 29 - University of Texas Health at Houston, Houston, Texas, 77225; 30 - Wake Forest Baptist Health, Department of Biochemistry, Winston-Salem, North Carolina, 27157; 31 - National Jewish Health, National Jewish Health, Denver, Colorado, 80206; 32 - Medical College of Wisconsin, Milwaukee, Wisconsin, 53226; 33 - University of California, San Francisco, San Francisco, California, 94143; 34 - Stanford University, Biomedical Data Science, Stanford, California, 94305; 35 - University of Washington, Biostatistics, Seattle, Washington, 98195; 36 - Brigham & Women's Hospital, Brigham and Women's Hospital, Boston, Massachusetts, 02115; 37 - University of Colorado at Denver, Denver, Colorado, 80204; 38 - Brigham & Women's Hospital, Boston, Massachusetts, 02115; 39 - Washington State University, Pullman, Washington, 99164; 40 - University of California, Los Angeles, Los Angeles, California, 90095; 41 - Brigham & Women's Hospital, Medicine, Boston, Massachusetts, 02115; 42 - National Taiwan University, Taipei, 10617; 43 - Brigham & Women's Hospital, Division of Preventive Medicine, Boston, Massachusetts, 02215; 44 - University of Virginia, Charlottesville, Virginia, 22903; 45 - Lundquist Institute, Torrance, California, 90502; 46 - National Taiwan University, National Taiwan University Hospital, Taipei, 10617; 47 - Cleveland Clinic, Cleveland Clinic, Cleveland, Ohio, 44195; 48 - National Health Research Institute Taiwan, Miaoli County, 350; 49 - Broad Institute, Metabolomics Platform, Cambridge, Massachusetts, 02142; 50 - Cleveland Clinic, Immunity and Immunology, Cleveland, Ohio, 44195; 51 - University of Vermont, Burlington, Vermont, 05405; 52 - University of Mississippi, Medicine, Jackson, Mississippi, 38677; 53 - National Jewish Health, Denver, Colorado, 80206; 54 - Boston University, Biostatistics, Boston, Massachusetts, 02115; 55 - University of Texas Rio Grande Valley School of Medicine, Brownsville, Texas, 78520; 56 - Vitalant Research Institute, San Francisco, California, 94118; 57 - University of Illinois at Chicago, Chicago, Illinois, 60607; 58 - University of Chicago, Chicago, Illinois, 60637; 59 - Vanderbilt University, Nashville, Tennessee, 37235; 60 - University of Cincinnati, Cincinnati, Ohio, 45220; 61 - University of North Carolina, Chapel Hill, North Carolina, 27599; 62 - University of Texas Rio Grande Valley School of Medicine, Edinburg, Texas, 78539; 63 - Washington University in St Louis, St Louis, Missouri, 63130; 64 - Brown University, Providence, Rhode Island, 02912; 65 - Harvard University, Channing Division of Network Medicine, Cambridge, Massachusetts, 02138; 66 - Massachusetts General Hospital, Boston, Massachusetts, 02114; 67 - National Jewish Health, Center for Genes, Environment and Health, Denver, Colorado, 80206; 68 - University of North Carolina, Epidemiology, Chapel Hill, North Carolina, 27599; 69 - Fred Hutchinson Cancer Research Center, Seattle, Washington, 98109; 70 - Icahn School of Medicine at Mount Sinai, New York, New York, 10029; 71 - Indiana University, Medicine, Indianapolis, Indiana, 46202; 72 - Beth Israel Deaconess Medical Center, Boston, Massachusetts, 02215; 73 - Baylor College of Medicine Human Genome Sequencing Center, Houston, Texas, 77030; 74 - Boston Children's Hospital, Harvard Medical School, Department of Psychiatry, Boston, Massachusetts, 02115; 75 - University of Texas Rio Grande Valley School of Medicine, San Antonio, Texas, 78229; 76 - University of Mississippi, Cardiology, Jackson, Mississippi, 39216; 77 - Yale University, Department of Chronic Disease Epidemiology, New Haven, Connecticut, 06520; 78 - Tulane University, New Orleans, Louisiana, 70118; 79 - Wake Forest Baptist Health, Winston-Salem, North Carolina, 27157; 80 - Brigham & Women's Hospital, Channing Division of Network Medicine, Boston, Massachusetts, 02115; 81 - University of Iowa, Iowa City, Iowa, 52242; 82 - National Health Research Institute Taiwan, Institute of Population Health Sciences, NHRI, Miaoli County, 350; 83 - Tri-Service General Hospital National Defense Medical Center; 84 - Blood Works Northwest, Seattle, Washington, 98104; 85 - Taichung Veterans General Hospital Taiwan, Taichung City, 407; 86 - Ohio State University Wexner Medical Center, Internal Medicine, DIvision of Endocrinology, Diabetes and Metabolism, Columbus, Ohio, 43210; 87 - University of Michigan, Biostatistics, Ann Arbor, Michigan, 48109; 88 - Albert Einstein College of Medicine, New York, New York, 10461; 89 - Harvard University, Cambridge, Massachusetts, 02138; 90 - University of Colorado at Denver, Epidemiology, Aurora, Colorado, 80045; 91 - Loyola University, Public Health Sciences, Maywood, Illinois, 60153; 92 - Harvard School of Public Health, Biostats, Boston, Massachusetts, 02115; 93 - Boston University, Boston, Massachusetts, 02215; 94 - Harvard School of Public Health, Boston, Massachusetts, 02115; 95 - Brown University, Epidemiology and Medicine, Providence, Rhode Island, 02912; 96 - Duke University, Cardiology, Durham, North Carolina, 27708; 97 - Stanford University, Cardiovascular Institute, Stanford, California, 94305; 98 - Icahn School of Medicine at Mount Sinai, The Charles Bronfman Institute for Personalized Medicine, New York, New York, 10029; 99 - George Washington University, Washington, District of Columbia, 20052; 100 - University of Arizona, Tucson, Arizona, 85721; 101 - Oklahoma Medical Research Foundation, Genes and Human Disease, Oklahoma City, Oklahoma, 73104; 102 - Ministry of Health, Government of Samoa, Apia; 103 - Howard University, Washington, District of Columbia, 20059; 104 - University of Maryland, Balitmore, Maryland, 21201; 105 - University at Buffalo, Buffalo, New York, 14260; 106 - Stanford University, Stanford Cardiovascular Institute, Stanford, California, 94305; 107 - Wake Forest Baptist Health, Biochemistry, Winston-Salem, North

Carolina, 27157; 108 - University of Minnesota, Minneapolis, Minnesota, 55455; 109 - Johns Hopkins University, Cardiology/Medicine, Baltimore, Maryland, 21218; 110 - University of Colorado at Denver, Medicine, Denver, Colorado, 80204; 111 - University of Colorado at Denver, Denver, Colorado, 80045; 112 - University of North Carolina, Genetics, Chapel Hill, North Carolina, 27599; 113 - Northwestern University, Chicago, Illinois, 60208; 114 - Fred Hutchinson Cancer Research Center, University of Washington, Seattle, Washington, 98109; 115 - Lutia I Puava Ae Mapu I Fagalele, Apia; 116 - Vanderbilt University, Medicine, Pharmacology, Biomedicla Informatics, Nashville, Tennessee, 37235; 117 - Universidade de Sao Paulo, Faculdade de Medicina, Sao Paulo, 01310000; 118 - Lundquist Institute, TGPS, Torrance, California, 90502; 119 - Harvard University, Division of Hematology/Oncology, Cambridge, Massachusetts, 02138; 120 - Harvard Medical School, Genetics, Boston, Massachusetts, 02115; 121 - Harvard Medical School, Boston, Massachusetts, 02115; 122 - Baylor College of Medicine, Pediatrics, Atlanta, Georgia, 30307; 123 - Emory University, Human Genetics, Atlanta, Georgia, 30322; 124 - Vanderbilt University, Medicine/Cardiology, Nashville, Tennessee, 37235; 125 - UMass Memorial Medical Center, Worcester, Massachusetts, 01655; 126 - University of Washington, Epidemiology, Seattle, Washington, 98195; 127 - Wake Forest Baptist Health, Biostatistical Sciences, Winston-Salem, North Carolina, 27157; 128 - Stanford University, Genetics, Stanford, California, 94305; 129 - Lundquist Institute, Institute for Translational Genomics and Populations Sciences, Torrance, California, 90502; 130 - University of Washington, University of Washington, Department of Genome Sciences, Seattle, Washington, 98195; 131 - University of Pennsylvania, Genetics, Philadelphia, Pennsylvania, 19104; 132 - University of Alabama, Biostatistics, Birmingham, Alabama, 35487; 133 - University of Washington, Department of Biostatistics, Seattle, Washington, 98195; 134 - University of Vermont, Pathology & Laboratory Medicine, Burlington, Vermont, 05405; 135 - University of Southern California, USC Methylation Characterization Center, University of Southern California, California, 90033; 136 - Brigham & Women's Hospital, Mass General Brigham, Boston, Massachusetts, 02115; 137 - Indiana University, Epidemiology, Indianapolis, Indiana, 46202; 138 - University of Michigan, Internal Medicine, Ann Arbor, Michigan, 48109; 139 - Henry Ford Health System, Detroit, Michigan, 48202; 140 - Beth Israel Deaconess Medical Center, Cardiology, Cambridge, Massachusetts, 02139; 141 - University of Pittsburgh, Medicine, Pittsburgh, Pennsylvania, 15260; 142 - University of Michigan, Department of Epidemiology, Ann Arbor, Michigan, 48109; 143 - Case Western Reserve University, Department of Population and Quantitative Health Sciences, Cleveland, Ohio, 44106; 144 - Mayo Clinic, Health Sciences Research, Rochester, Minnesota, 55905; 145 - Washington University in St Louis, Department of Medicine, Cardiovascular Division, St. Louis, Missouri, 63110