

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. Burt,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.

¹Program in Medical and Population Genetics, Broad Institute of Harvard and MIT, Cambridge, MA 02142, USA; ²Department of Clinical Sciences, Lund University, Malmö, Sweden; ³Department of Population Medicine, Qatar University College of Medicine, QU Health, Doha, Qatar; ⁴Programs in Metabolism and Medical & Population Genetics, Broad Institute, Cambridge, MA 02142, USA; ⁵Department of Pediatrics, Harvard Medical School, Boston, MA 02115, USA; ⁶Division of Genetics and Genomics, Boston Children's Hospital, Boston, MA 02115, USA; ⁷Center for Genomic Medicine, Department of Medicine, Massachusetts General Hospital, Boston, MA 02114, USA; ⁸Cardiovascular Research Center, Massachusetts General Hospital, Boston, MA 02114, USA; ⁹Department of Public Health Sciences, Penn State College of Medicine, Hershey, PA 17033, USA; ¹⁰Cardiovascular Disease Initiative, Broad Institute of Harvard and MIT, Cambridge, MA 02142, USA; ¹¹Department of Genetics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104, USA; ¹²Department of Medicine, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104, USA; ¹³Instituto Nacional de Ciencias Medicas y Nutricion, Mexico City, Mexico; ¹⁴Department of Genetics, Washington University in St. Louis, St. Louis, MO 63110, USA; ¹⁵The McDonnell Genome Institute, Washington University School of Medicine, St. Louis, MO 63108, USA; ¹⁶ASTC: Associazione per lo Studio Della Trombosi in Cardiologia, Pavia, Italy; ¹⁷Azienda Ospedaliero-Universitaria di Parma, Parma, Italy; ¹⁸Universit, degli Studi di Parma, Parma, Italy; ¹⁹Dean's Office, College of Public Health, University of Kentucky, Lexington, KY 40536, USA; ²⁰Department of Epidemiology, University of Alabama at Birmingham, Birmingham, AL 35233, USA; ²¹Departments of Medicine and Genetics, Albert Einstein College of Medicine, Bronx, NY 10461, USA; ²²University of Haifa, Faculty of Natural Science, Haifa, Israel; ²³Houston Methodist DeBakey Heart and Vascular Center, Houston, TX 77030, USA; ²⁴Section of Cardiovascular Research, Baylor College of Medicine, Houston, TX 77030, USA; ²⁵Instituto Nacional de Medicina Genómica, Mexico City, Mexico; ²⁶Department of Medicine, Johns Hopkins University School of Medicine, Baltimore, MD 21287, USA; ²⁷Department of Epidemiology, School of Public Health, University of Michigan, Ann Arbor, MI 49109, USA; ²⁸Cardiovascular Health Research Unit, Department of Medicine, University of Washington, Seattle, WA 98101, USA; ²⁹Department of Human Genetics and South Texas Diabetes and Obesity Institute, University of Texas Rio Grande Valley School of Medicine, Brownsville, TX 78520, USA; ³⁰Human 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; ³¹Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX 77030, USA; ³²Medical Genomics and Metabolic Genetics Branch, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD 20892, USA; ³³Hasso Plattner Institute for Digital Health at Mount Sinai, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA; ³⁴Digital Health Center, Hasso Plattner Institute, University of Potsdam, Potsdam, Germany; ³⁵Department of Biochemistry, Wake Forest School of Medicine, Winston-Salem, NC 27157, USA; ³⁶Department of Cardiovascular Sciences, University of Leicester, Leicester, UK; ³⁷NIHR Leicester Biomedical Research Centre, Glenfield Hospital, Leicester, UK; ³⁸Department of Biostatistics, University of Washington, Seattle, WA 98195, USA; ³⁹Department of Medicine, Brigham and Women's Hospital, Boston, MA 02115, USA; ⁴⁰Division of Sleep Medicine, Harvard Medical School, Boston, MA 02115, USA; ⁴¹Department of Medicine, Yong Loo Lin School of Medicine, National University of Singapore, National University Health System, Singapore; ⁴²Institute of Biomedical Sciences, Academia Sinica, Taiwan; ⁴³The 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; ⁴⁴Department 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. Kardina,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; ⁴⁵Ophthalmology & Visual Sciences Academic Clinical Program (Eye ACP), Duke-NUS Medical School, Singapore; ⁴⁶Singapore Eye Research Institute, Singapore National Eye Centre, Singapore; ⁴⁷Psonagen, Inc. (formerly MacroGen USA), 1330 Piccard Drive Ste 103, Rockville, MD 20850, USA; ⁴⁸MRC/BHF Cardiovascular Epidemiology Unit, Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK; ⁴⁹Centre for Non-Communicable Disease Research, Bangladesh; ⁵⁰Department of Medicine, University of Mississippi Medical Center, Jackson, MS 39216, USA; ⁵¹Department of Biostatistics, Boston University School of Public Health, Boston, MA 02118, USA; ⁵²NHLBI Framingham Heart Study, Framingham, MA 01702, USA; ⁵³The National Institute for Health Research Blood and Transplant Research Unit in Donor Health and Genomics at the University of Cambridge, Cambridge, UK; ⁵⁴Department of Medicine, University of Texas Health Science Center, San Antonio, TX 78229, USA; ⁵⁵Diabetes Research Center (Diabetes Unit), Department of Medicine, Massachusetts General Hospital, Boston, MA 02114, USA; ⁵⁶Department of Medicine, Harvard Medical School, Boston, MA 02115, USA; ⁵⁷Institute of Molecular Medicine, University of Texas Health Science Center at Houston, Houston, TX 77030, USA; ⁵⁸Department of Internal Medicine, Section on Nephrology, Wake Forest School of Medicine, Winston-Salem, NC 27157, USA; ⁵⁹Department of Medicine, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA; ⁶⁰Centro Nacional de Investigaciones Cardiovasculares Carlos III, Madrid, Spain; ⁶¹Department of Medical Science, Division of Health Science, University of Guanajuato, Guanajuato, Mexico; ⁶²New York Genome Center, New York, NY 10013, USA; ⁶³Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX 77030, USA; ⁶⁴Research Unit of Molecular Epidemiology, Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg, Germany; ⁶⁵Institute of Epidemiology, Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg, Germany; ⁶⁶German Center for Diabetes Research, Neuherberg, Germany; ⁶⁷Endocrinology and Metabolism Service, Hadassah-Hebrew University Medical Center, Jerusalem, Israel; ⁶⁸Unidad de Diabetes y Riesgo Cardiovascular, Instituto Nacional de Salud Pública, Cuernavaca, Morelos, Mexico; ⁶⁹Centro de Estudios en Diabetes, Mexico City, Mexico; ⁷⁰Department of Epidemiology, University of North Carolina at Chapel Hill, Chapel Hill, NC 27514, USA; ⁷¹Department of Internal Medicine, University of Michigan, Ann Arbor, MI 48109, USA; ⁷²Novo Nordisk Foundation Center for Basic Metabolic Research, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark; ⁷³Department of Clinical Sciences, Diabetes and Endocrinology, Lund University Diabetes Centre, Malmö, Sweden; ⁷⁴Finnish Institute for Molecular Genetics, University of Helsinki, Helsinki, Finland; ⁷⁵Division of Genome Science, Department of Precision Medicine, Chungcheongbuk-do, Republic of Korea; ⁷⁶Human Genetics Center, School of Public Health, The University of Texas Health Science Center at Houston, Houston, TX 77030, USA; ⁷⁷Faculty of Health Sciences, University of Southern Denmark, Odense, Denmark; ⁷⁸Department of Epidemiology, Tulane University School of Public Health and Tropical Medicine, New Orleans, LA 70112, USA; ⁷⁹Tulane University Translational Science Institute, New Orleans, LA 70112, USA; ⁸⁰Department of Neurology, Boston University School of Medicine, Boston, MA 02118, USA; ⁸¹Division of Endocrine and Metabolism, Tri-Service General Hospital Songshan Branch, Taipei, Taiwan; ⁸²Department of Epidemiology, School of Public Health, UAB, Birmingham, AL 35294, USA; ⁸³Dirección de Investigación, Hospital General de México "Dr. Eduardo Liceaga," Secretaría de Salud, Mexico City, Mexico; ⁸⁴Departments of Medicine (Medical Genetics) and Genome Sciences, University of Washington Medical Center, Seattle, WA 98195, USA; ⁸⁵Department of Biostatistics and Center for Statistical Genetics, University of Michigan, Ann Arbor, MI 48109, USA; ⁸⁶Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA; ⁸⁷Institute for Genomic Health, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA; ⁸⁸Department of Public Health Solutions, Finnish Institute for Health and Welfare, Helsinki, Finland; ⁸⁹Minerva Foundation Institute for Medical Research, Helsinki, Finland; ⁹⁰University of Helsinki and Department of Medicine, Helsinki University Central Hospital, Helsinki, Finland; ⁹¹Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, WA 98103, USA; ⁹²Institute of Clinical Medicine, University of Eastern Finland, and Kuopio University Hospital, Kuopio, Finland; ⁹³Department of Internal Medicine, Seoul National University Hospital, Seoul, Republic of Korea; ⁹⁴Department of Medicine, University of Colorado Denver, Anschutz Medical Campus, Aurora, CO 80045, USA; ⁹⁵Center for Clinical Research and Prevention, Bispebjerg and Frederiksberg Hospital, The Capital Region, Copenhagen, Denmark; ⁹⁶Department of Clinical Medicine, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark; ⁹⁷Genome Institute of Singapore, Agency for Science, Technology and Research, Singapore; ⁹⁸Charles R. Bronfman Institute of Personalized Medicine, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA; ⁹⁹The Mindich Child Health and Development Institute, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA; ¹⁰⁰University of Bergen, Bergen, Norway; ¹⁰¹Department of Medicine and Therapeutics, The Chinese University of Hong Kong, Hong Kong, China; ¹⁰²Hong Kong Institute of Diabetes and Obesity, The Chinese University of Hong Kong, Hong Kong, China; ¹⁰³Li Ka Shing Institute of Health Sciences, The Chinese University of Hong Kong, Hong Kong, China; ¹⁰⁴Division of Cardiology, Department of Medicine, George Washington University, Washington, DC 20037, USA; ¹⁰⁵Department of Epidemiology and International Health Institute, Brown University School of Public Health, Providence, RI 02912, USA; ¹⁰⁶Ruddy Canadian Cardiovascular Genetics Centre, University of Ottawa Heart Institute, Ottawa, ON, Canada; ¹⁰⁷General Medicine Division, Massachusetts General Hospital, Boston, MA 02114, USA; ¹⁰⁸Broad Institute of MIT and Harvard, Cambridge, MA 02142, USA; ¹⁰⁹Deutsches Forschungszentrum für Herz-Kreislaufkrankungen, Partner Site Munich Heart Alliance, Munich, Germany; ¹¹⁰Institute of Human Genetics, Technische Universität München, Munich, Germany; ¹¹¹Department of Emergency and Internal Medicine, SkEne University Hospital, Malmö, Sweden; ¹¹²Department of Genetics, University of North Carolina, Chapel Hill, NC 27514, USA; ¹¹³University 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 dis-

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

¹⁶⁰Present 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 CAD²⁸ ($n = 10,689$); T2D cases-control studies from the AMP-T2D-GENES exomes¹⁵ ($n = 32,486$); 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 Athero-

sclerosis 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, age², 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 Predictor³⁵ and its associated Loss-of-Function Transcript Effect Estimator (LOFTEE)³⁶ and the dbNSFP³⁷ 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 sub-categories. 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 re-

sults 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 database⁴³ 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

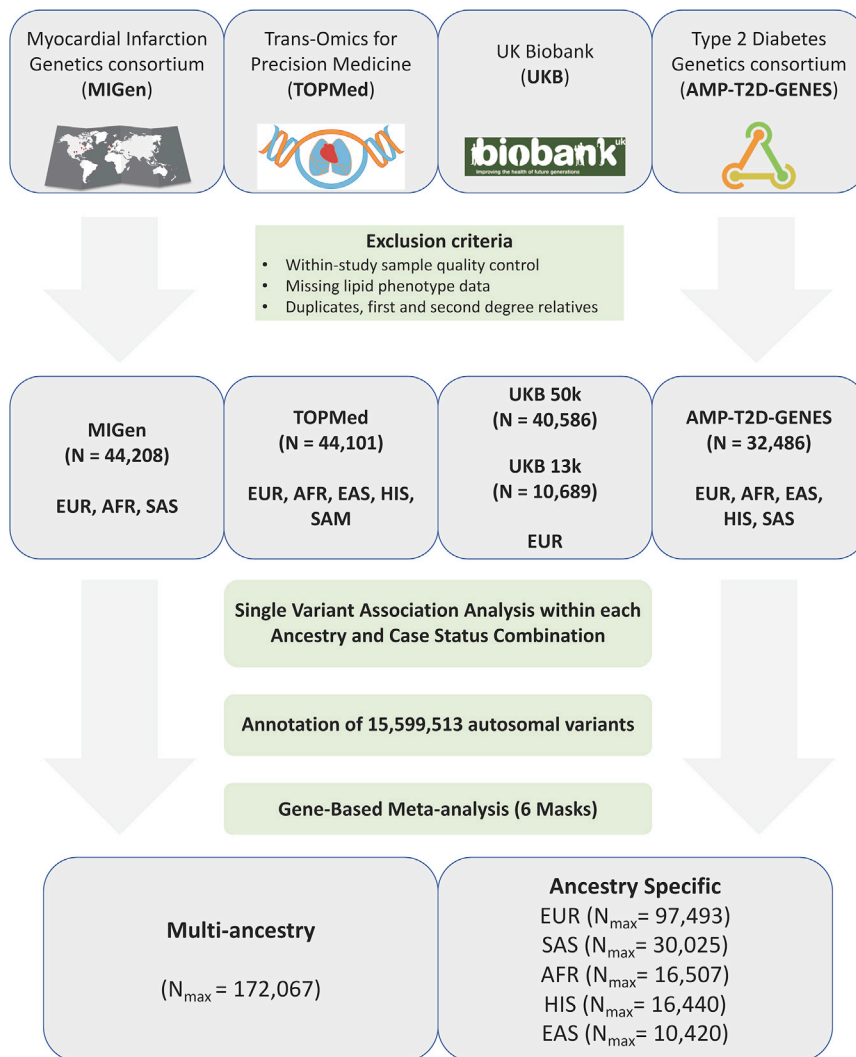


Figure 1. Study samples and design
Flow chart of the different stages of the study. Exome sequence genotypes were derived from four major data sources: the Myocardial Infarction Genetics consortium (MIGen), the Trans-Omics from Precision Medicine (TOPMed), the UK Biobank, and the Type 2 Diabetes Genetics (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.

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)

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

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 ratio [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 tests⁴² in all study participants and within each of the five ances-

tries 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 $\leq 1\%$ 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 *AICF* 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 *AICF* 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

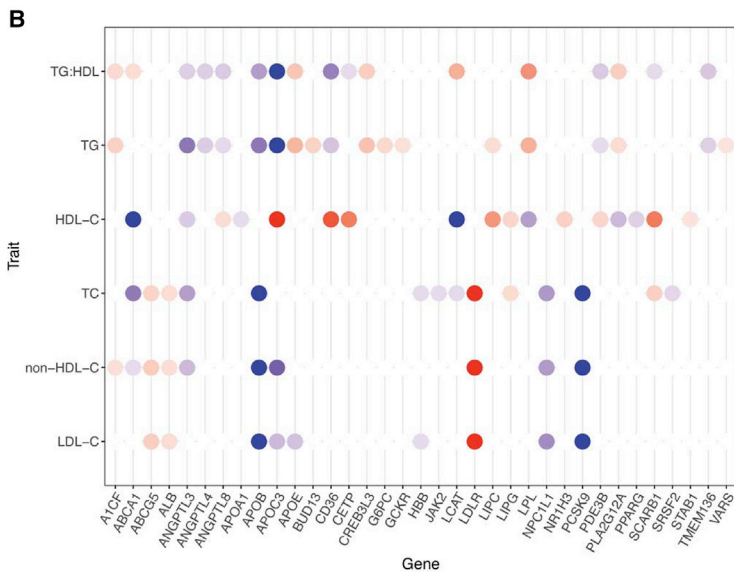
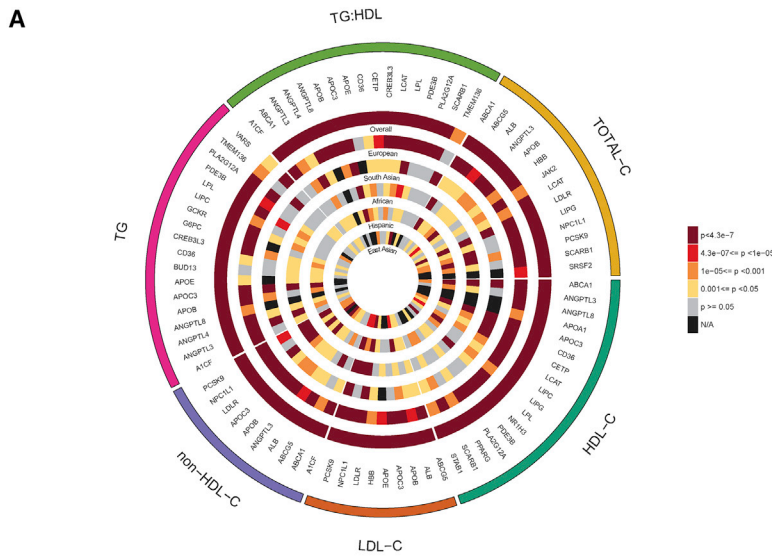


Figure 2. Exome-wide significant associations with blood lipid phenotypes

(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. $\text{Sign}(\beta) \times -\log_{10}(p)$ displayed for associations that reached a $p < 4.3 \times 10^{-7}$. When the $\text{Sign}(\beta) \times -\log_{10}(p) > 50$, they were trimmed to 50.

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

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

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 xiiia	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 xiiia	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 xiiia	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 (LXR α), 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 LXR α phosphorylation at Ser196 in mice prevents non-alcoholic fatty liver disease.⁵⁴

PLA2G1A

PLA2G1A is in the secretory phospholipase A2 (sPLA₂) 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 *PLA2G1A* 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 ($p_{\text{Het}} < 1.7 \times 10^{-4}$, 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 ($p_{\text{SKAT}} = 3 \times 10^{-9}$ 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 ($p_{\text{Het}} < 6 \times 10^{-4}$) (Tables S30 and S31).

Gene-based associations in GWAS loci

We determined whether genes near lipid array-based GWAS signals⁸ 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 *EVIS* were associated with LDL-C ($p_{\text{SKAT}} = 2 \times 10^{-5}$). 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 *EVIS*, *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 target 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,

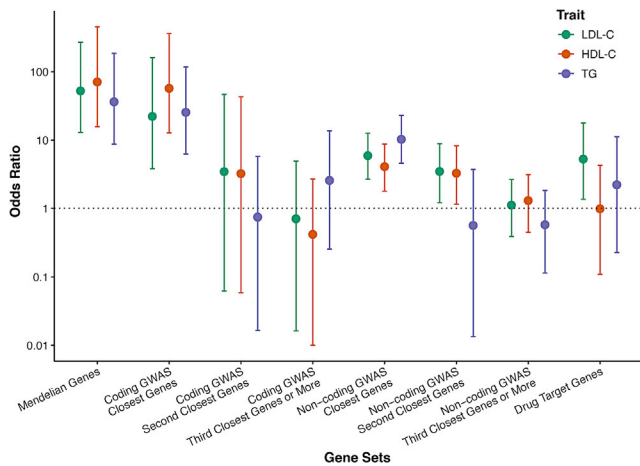


Figure 3. Enrichment of Mendelian, GWAS, and drug target genes in the gene-based lipid associations

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 \geq third closest genes to protein-altering GWAS signals and in \geq 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 ($p_{\text{CAD}} < 0.00125$, 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* ($p_{\text{SKAT}} = 4 \times 10^{-5}$), *PCSK9* (OR: 0.5, $p = 2 \times 10^{-4}$), and *JAK2* ($p_{\text{SKAT}} = 0.001$). 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* ($p_{\text{SKAT}} = 0.002$) and *EVIS* (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, $p_{\text{T2D}} = 0.002$) and *APOB* (OR: 1.08, $p_{\text{T2D}} = 0.005$) (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 *EVIS* 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 *EVIS* 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 from 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

- 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.
- 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.
- 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.
- 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.
- Peloso, G.M., Auer, P.L., Bis, J.C., Voorman, A., Morrison, A.C., Stitzel, 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.
- Asselbergs, F.W., Guo, Y., van Iperen, E.P., Sivapalaratnam, S., Tragante, V., Lanktree, M.B., Lange, L.A., Almqvister, B., Appelmann, 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.
- 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.
- 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.
- 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.
- 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.
- 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.
- 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.
- 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.
- 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.
- 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.
- Do, R., Stitzel, 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.
- 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.
- Crosby, J., Peloso, G.M., Auer, P.L., Crosslin, D.R., Stitzel, 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.
- Jørgensen, A.B., Frikke-Schmidt, R., Nordestgaard, B.G., and Tybjaerg-Hansen, A. (2014). Loss-of-function mutations in APOC3 and risk of ischemic vascular disease. *N. Engl. J. Med.* 371, 32–41.
- 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.
- 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.
- 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., Alfoldi, 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.L., Kosmicki, J.A., Arbelez, 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., Måsson, 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., Dilliot, 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 LX α 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., Schaffer, A.J., et al. (1999). Dominant negative mutations in human *PPARG* 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 *PPARG* 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., Rieger, 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.