frontiers | Frontiers in Genetics

TYPE Original Research
PUBLISHED 02 November 2023
DOI 10.3389/fgene.2023.1235337

Check for updates

OPEN ACCESS

EDITED BY
Lindsay Fernandez-Rhodes,
The Pennsylvania State University (PSU),
United States

REVIEWED BY
Mahdi Vajdi,
Isfahan University of Medical
Sciences, Iran
Deng-Feng Zhang,
Chinese Academy of Sciences (CAS),
China

*CORRESPONDENCE
Lisa de las Fuentes,
lfuentes@wustl.edu

†These authors share first authorship

‡These authors share senior authorship

RECEIVED 06 June 2023
ACCEPTED 27 September 2023
PUBLISHED 02 November 2023

Gene-educational attainment interactions in a multi-population genome-wide meta-analysis identify novel lipid loci

Lisa de las Fuentes[1,2*†], Karen L. Schwander[2,3†], Michael R. Brown[4†], Amy R. Bentley[5†], Thomas W. Winkler[6†], Yun Ju Sung[2,7†], Patricia B. Munroe[8,9†], Clint L. Miller[10,11,12†], Hugo Aschard[13,14], Stella Aslibekyan[15], Traci M. Bartz[16,17], Lawrence F. Bielak[18], Jin Fang Chai[19], Ching-Yu Cheng[20,21], Rajkumar Dorajoo[22], Mary F. Feitosa[3], Xiuqing Guo[23], Fernando P. Hartwig[24,25], Andrea Horimoto[26], Ivana Kolčić[27,28], Elise Lim[29], Yongmei Liu[30], Alisa K. Manning[31,32], Jonathan Marten[33], Solomon K. Musani[34], Raymond Noordam[35], Sandosh Padmanabhan[36], Tuomo Rankinen[37], Melissa A. Richard[38], Paul M. Ridker[39,40], Albert V. Smith[41,42], Dina Vojinovic[43,44], Alan B. Zonderman[45,46], Maris Alver[47], Mathilde Boissel[48,49], Kaare Christensen[50], Barry I. Freedman[51], Chuan Gao[52], Franco Giulianini[39], Sarah E. Harris[53,54], Meian He[55], Fang-Chi Hsu[56], Brigitte Kühnel[57,58], Federica Laguzzi[59], Xiaoyin Li[60,61], Leo-Pekka Lyytikäinen[62,63], Ilja M. Nolte[64], Alaitz Poveda[65], Rainer Rauramaa[66], Muhammad Riaz[67,68], Antonietta Robino[69], Tamar Sofer[70], Fumihiko Takeuchi[71], Bamidele O. Tayo[72], Peter J. van der Most[73], Niek Verweij[74], Erin B. Ware[75], Stefan Weiss[76,77], Wanqing Wen[78], Lisa R. Yanek[79], Yiqiang Zhan[80], Najaf Amin[43], Dan E. Arking[81], Christie Ballantyne[82,83], Eric Boerwinkle[4,84], Jennifer A. Brody[16], Ulrich Broeckel[85], Archie Campbell[86,87], Mickaël Canouil[48,49], Xiaoran Chai[88], Yii-Der Ida Chen[23], Xu Chen[80], Kumaraswamy Naidu Chitrala[45], Maria Pina Concas[69], Ulf de Faire[59], Renée de Mutsert[89], H. Janaka de Silva[90], Paul S. de Vries[4], Ahn Do[2,3], Jessica D. Faul[75], Virginia Fisher[29], James S. Floyd[16], Terrence Forrester[91], Yechiel Friedlander[92], Giorgia Girotto[69], C. Charles Gu[2], Göran Hallmans[93], Sami Heikkinen[94], Chew-Kiat Heng[95,96], Georg Homuth[76], Steven Hunt[97,98], M. Arfan Ikram[43], David R. Jacobs Jr[99], Maryam Kavousi[43], Chiea Chuen Khor[22], Tuomas O. Kilpeläinen[100,101], Woon-Puay Koh[102,103], Pirjo Komulainen[66], Carl D. Langefeld[56], Jingjing Liang[60], Kiang Liu[104], Jianjun Liu[22], Kurt Lohman[30], Reedik Mägi[47],

Ani W. Manichaikul[10], Colin A. McKenzie[91], Thomas Meitinger[105,106], Yuri Milaneschi[107], Matthias Nauck[77,108], Christopher P. Nelson[67,68], Jeffrey R. O'Connell[109,110], Nicholette D. Palmer[111], Alexandre C. Pereira[26], Thomas Perls[112], Annette Peters[58,113], Ozren Polašek[27,28], Olli T. Raitakari[114,115,116], Kenneth Rice[17], Treva K. Rice[2], Stephen S. Rich[10], Charumathi Sabanayagam[20,21], Pamela J. Schreiner[99], Xiao-Ou Shu[78], Stephen Sidney[117], Mario Sims[118], Jennifer A. Smith[18,75], John M. Starr[54,119], Konstantin Strauch[120,121], E. Shyong Tai[19,122,123], Kent D. Taylor[23], Michael Y. Tsai[124], André G. Uitterlinden[43,125], Diana van Heemst[35], Melanie Waldenberger[57,58,126], Ya-Xing Wang[127], Wen-Bin Wei[127], Gregory Wilson[128], Deng Xuan[29], Jie Yao[23], Caizheng Yu[55], Jian-Min Yuan[129,130], Wei Zhao[18], Diane M. Becker[79], Amélie Bonnefond[48,49,131], Donald W. Bowden[111], Richard S. Cooper[72], Ian J. Deary[53,54], Jasmin Divers[56], Tõnu Esko[47,132], Paul W. Franks[65,133,134], Philippe Froguel[48,49,131], Christian Gieger[57,135], Jost B. Jonas[127,136,137], Norihiro Kato[71], Timo A. Lakka[66,94,138], Karin Leander[59], Terho Lehtimäki[62,63], Patrik K. E. Magnusson[80], Kari E. North[139], Ioanna Ntalla[8,140], Brenda Penninx[107], Nilesh J. Samani[67,68], Harold Snieder[64], Beatrice Spedicati[141], Pim van der Harst[142], Henry Völzke[77,143], Lynne E. Wagenknecht[56], David R. Weir[75], Mary K. Wojczynski[3], Tangchun Wu[55], Wei Zheng[78], Xiaofeng Zhu[60], Claude Bouchard[37], Daniel I. Chasman[39,40], Michele K. Evans[45,144], Ervin R. Fox[145], Vilmundur Gudnason[42,146], Caroline Hayward[33], Bernardo L. Horta[24], Sharon L. R. Kardia[18], Jose Eduardo Krieger[26], Dennis O. Mook-Kanamori[89,147], Patricia A. Peyser[18], Michael M. Province[3], Bruce M. Psaty[16,148,149], Igor Rudan[150], Xueling Sim[19], Blair H. Smith[151], Rob M. van Dam[19,152], Cornelia M. van Duijn[43,153], Tien Yin Wong[20,21], Donna K. Arnett[154‡], Dabeeru C. Rao[2‡], James Gauderman[155‡], Ching-Ti Liu[29‡], Alanna C. Morrison[4‡], Jerome I. Rotter[23‡] and Myriam Fornage[4,38‡]

[1]Cardiovascular Division, Department of Medicine, Washington University School of Medicine, St. Louis, MO, United States, [2]Division of Biostatistics, Washington University School of Medicine, St. Louis, MO, United States, [3]Division of Statistical Genomics, Department of Genetics, Washington University School of Medicine, St. Louis, MO, United States, [4]Human Genetics Center, Department of Epidemiology, Human Genetics and Environmental Sciences, School of Public Health, The University of Texas Health Science Center at Houston, Houston, TX, United States, [5]Center for Research on Genomics and Global Health, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD, United States, [6]Department of Genetic Epidemiology, University of Regensburg, Regensburg, Germany, [7]Department of Psychiatry, Washington University School of Medicine, St. Louis, MO, United States, [8]Clinical Pharmacology, Queen Mary University of London, London, United Kingdom, [9]National Institute for Health Research Barts Cardiovascular Biomedical Research Unit, Queen Mary University of London, London, United Kingdom, [10]Center for Public Health Genomics, Department of Public Health Sciences, University of Virginia, Charlottesville, VA, United States, [11]Biochemistry and Molecular Genetics, Department of Public Health Sciences, University of Virginia, Charlottesville, VA, United States, [12]Department of Biochemistry and Molecular Genetics, University of Virginia, Charlottesville, VA, United States, [13]Department of Epidemiology, Harvard School of Public Health, Boston, MA, United States, [14]Département de Génomes et Génétique, Institut Pasteur de Lille, Université de Lille, Lille, France, [15]School of Public Health, Epidemiology, University of Alabama at Birmingham, Birmingham, AL,

United States, [16]Cardiovascular Health Research Unit, University of Washington, Seattle, WA, United States, [17]Department of Biostatistics, University of Washington, Seattle, WA, United States, [18]Department of Epidemiology, School of Public Health, University of Michigan, Ann Arbor, MI, United States, [19]Saw Swee Hock School of Public Health, National University of Singapore and National University Health System, Singapore, Singapore, [20]Ocular Epidemiology, Singapore Eye Research Institute, Singapore National Eye Centre, Singapore, Singapore, [21]Ophthalmology and Visual Sciences Academic Clinical Program, Medical School, Duke-National University of Singapore, Singapore, Singapore, [22]Genome Institute of Singapore, Agency for Science, Technology and Research, Singapore, Singapore, [23]Department of Pediatrics, The Institute for Translational Genomics and Population Sciences, The Lundquist Institute for Biomedical Innovation at Harbor-UCLA Medical Center, Los Angeles, CA, United States, [24]Postgraduate Programme in Epidemiology, Faculty of Medicine, Federal University of Pelotas, Pelotas, RS, Brazil, [25]Medical Research Council Integrative Epidemiology Unit, University of Bristol, Bristol, United Kingdom, [26]Laboratory of Genetics and Molecular Cardiology, Heart Institute, University of Sao Paulo Medical School, Sao Paulo, SP, Brazil, [27]University of Split School of Medicine, Split, Croatia, [28]Algebra University College, Zagreb, Croatia, [29]Department of Biostatistics, Boston University School of Public Health, Boston, MA, United States, [30]Division of Cardiology, Department of Medicine, Duke Molecular Physiology Institute, Duke University School of Medicine, Durham, NC, United States, [31]Clinical and Translational Epidemiology Unit, Massachusetts General Hospital, Boston, MA, United States, [32]Department of Medicine, Harvard Medical School, Boston, MA, United States, [33]Medical Research Council Human Genetics Unit, Institute of Genetics and Cancer, University of Edinburgh, Edinburgh, United Kingdom, [34]Jackson Heart Study, Department of Medicine, University of Mississippi Medical Center, Jackson, MS, United States, [35]Section of Gerontology and Geriatrics, Department of Internal Medicine, Leiden University Medical Center, Leiden, Netherlands, [36]Institute of Cardiovascular and Medical Sciences, University of Glasgow, Glasgow, United Kingdom, [37]Human Genomics Laboratory, Pennington Biomedical Research Center, Baton Rouge, LA, United States, [38]Brown Foundation Institute of Molecular Medicine, The University of Texas Health Science Center at Houston, Houston, TX, United States, [39]Division of Preventive Medicine, Brigham and Women's Hospital, Boston, MA, United States, [40]Harvard Medical School, Boston, MA, United States, [41]Department of Biostatistics, School of Public Health, University of Michigan, Ann Arbor, MI, United States, [42]Icelandic Heart Association, Kopavogur, Iceland, [43]Department of Epidemiology, Erasmus MC, University Medical Center, Rotterdam, Netherlands, [44]Molecular Epidemiology, Department of Biomedical Data Sciences, Leiden University Medical Center, Leiden, Netherlands, [45]Laboratory of Epidemiology and Population Sciences, National Institute on Aging, National Institutes of Health, Baltimore, MD, United States, [46]National Institutes of Health, Baltimore, MD, United States, [47]Estonian Genome Center, Insititute of Genomics, University of Tartu, Tartu, Estonia, [48]European Genomic Institute for Diabetes, Institut Pasteur de Lille, Lille, France, [49]University of Lille, Lille University Hospital, Lille, France, [50]Unit of Epidemiology, Biostatistics and Biodemography, Department of Public Health, University of Southern Denmark, Odense, Denmark, [51]Nephrology Division, Department of Internal Medicine, Wake Forest School of Medicine, Winston-Salem, NC, United States, [52]Molecular Genetics and Genomics Program, Wake Forest School of Medicine, Winston-Salem, NC, United States, [53]Department of Psychology, The University of Edinburgh, Edinburgh, United Kingdom, [54]Centre for Cognitive Ageing and Cognitive Epidemiology, The University of Edinburgh, Edinburgh, United Kingdom, [55]Department of Occupational and Environmental Health and State Key Laboratory of Environmental Health for Incubating, School of Public Health, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China, [56]Department of Biostatistics and Data Science, Division of Public Health Sciences, Wake Forest University School of Medicine, Winston-Salem, NC, United States, [57]Research Unit of Molecular Epidemiology, Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg, Germany, [58]Institute of Epidemiology, Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg, Germany, [59]Cardiovascular and Nutritional Epidemiology, Institute of Environmental Medicine, Karolinska Institutet, Stockholm, Sweden, [60]Department of Population and Quantitative Health Sciences, Cleveland, OH, United States, [61]Department of Mathematics and Statistics, St. Cloud State University, St. Cloud, MN, United States, [62]Department of Clinical Chemistry, University of Tampere, Tampere, Finland, [63]Finnish Cardiovascular Research Center, University of Tampere, Tampere, Finland, [64]Unit of Genetic Epidemiology and Bioinformatics, Department of Epidemiology, University Medical Center Groningen, University of Groningen, Groningen, Netherlands, [65]Genetic and Molecular Epidemiology Unit, Department of Clinical Sciences, Skåne University Hospital, Lund University, Malmö, Sweden, [66]Kuopio Research Institute of Exercise Medicine, Kuopio, Finland, [67]Department of Cardiovascular Sciences, University of Leicester, Leicester, United Kingdom, [68]NIHR Leicester Biomedical Research Centre, Glenfield Hospital, Leicester, United Kingdom, [69]Institute for Maternal and Child Health-IRCCS Burlo Garofolo, Trieste, Italy, [70]Biostatistics, Department of Medicine, Brigham and Women's Hospital, Harvard University, Boston, MA, United States, [71]Department of Gene Diagnostics and Therapeutics, Research Institute, National Center for Global Health and Medicine, Tokyo, Japan, [72]Department of Public Health Sciences, Loyola University Chicago, Maywood, IL, United States, [73]Department of Epidemiology, University Medical Center Groningen, University of Groningen, Groningen, Netherlands, [74]Department of Cardiology, University Medical Center Groningen, University of Groningen, Groningen, Netherlands, [75]Survey Research Center, Institute for Social Research, University of Michigan, Ann Arbor, MI, United States, [76]Interfaculty Institute for Genetics and Functional Genomics, University Medicine Greifswald and University of Greifswald, Greifswald, Germany, [77]German Center for Cardiovascular Research, Greifswald, Germany, [78]Division of Epidemiology, Department of Medicine, Vanderbilt University Medical Center, Nashville, TN, United States, [79]Division of General Internal Medicine, Department of Medicine, Johns Hopkins University School of Medicine, Baltimore, MD, United States, [80]Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden, [81]Department of Genetic Medicine, McKusick-Nathans Institute, Johns Hopkins University School of Medicine, Baltimore, MD, United States, [82]Section of Cardiovascular Research, Baylor College of Medicine, Houston, TX, United States, [83]Houston Methodist Debakey Heart and Vascular Center, Houston, TX, United States, [84]Human Genome Sequencing Center, Baylor College of

Medicine, Houston, TX, United States, [85]Section on Genomic Pediatrics, Department of Pediatrics, Medical College of Wisconsin, Milwaukee, WI, United States, [86]Centre for Genomic and Experimental Medicine, Institute of Genetics & Cancer, University of Edinburgh, Edinburgh, United Kingdom, [87]Usher Institute for Population Health Sciences and Informatics, University of Edinburgh, Edinburgh, United Kingdom, [88]Data Science Unit, Singapore Eye Research Institute, Singapore National Eye Centre, Singapore, Singapore, [89]Department of Clinical Epidemiology, Leiden University Medical Center, Leiden, Netherlands, [90]Department of Medicine, Faculty of Medicine, University of Kelaniya, Ragama, Sri Lanka, [91]Tropical Medicine Research Institute, University of the West Indies, Mona, Jamaica, [92]Braun School of Public Health, Hadassah Medical Center, Hebrew University, Jerusalem, Israel, [93]Section for Nutritional Research, Department of Public Health and Clinical Medicine, Umeå University, Umeå, Sweden, [94]Institute of Biomedicine, School of Medicine, University of Eastern Finland, Kuopio, Finland, [95]Department of Paediatrics, Yong Loo Lin School of Medicine, National University of Singapore, Singapore, Singapore, [96]Khoo Teck Puat National University Children's Medical Institute, National University Health System, Singapore, Singapore, [97]Department of Internal Medicine, University of Utah, Salt Lake City, UT, United States, [98]Department of Genetic Medicine, Weill Cornell Medicine in Qatar, Doha, Qatar, [99]Division of Epidemiology and Community Health, School of Public Health, University of Minnesota, Minneapolis, MN, United States, [100]Novo Nordisk Foundation Center for Basic Metabolic Research, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark, [101]Department of Environmental Medicine and Public Health, The Icahn School of Medicine at Mount Sinai, New York, NY, United States, [102]Healthy Longevity Translational Research Programme, Yong Loo Lin School of Medicine, National University of Singapore, Singapore, Singapore, [103]Singapore Institute for Clinical Sciences, Agency for Science Technology and Research (A*STAR), Singapore, Singapore, [104]Epidemiology, Department of Preventive Medicine, Northwestern University Feinberg School of Medicine, Chicago, IL, United States, [105]Institute of Human Genetics, Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg, Germany, [106]Institute of Human Genetics, Technische Universität München, Munich, Germany, [107]Psychiatry, VU Medisch Centrum, Amsterdam, Netherlands, [108]Institute of Clinical Chemistry and Laboratory Medicine, University Medicine Greifswald, Greifswald, Germany, [109]Division of Endocrinology, Diabetes, and Nutrition, University of Maryland School of Medicine, Baltimore, MD, United States, [110]Program for Personalized and Genomic Medicine, University of Maryland School of Medicine, Baltimore, MD, United States, [111]Department of Biochemistry, Wake Forest School of Medicine, Winston-Salem, NC, United States, [112]Geriatrics Section, Department of Medicine, Boston University School of Medicine, Boston, MA, United States, [113]German Center for Cardiovascular Research, Neuherberg, Germany, [114]Centre for Population Health Research, University of Turku and Turku University Hospital, Turku, Finland, [115]Research Centre of Applied and Preventive Cardiovascular Medicine, University of Turku, Turku, Finland, [116]Department of Clinical Physiology and Nuclear Medicine, Turku University Hospital, Turku, Finland, [117]Division of Research, Kaiser Permanente of Northern California, Oakland, CA, United States, [118]Jackson Heart Study, Department of Medicine, University of Mississippi Medical Center, Jackson, MS, United States, [119]Alzheimer Scotland Dementia Research Centre, The University of Edinburgh, Edinburgh, United Kingdom, [120]German Research Center for Environmental Health, Helmholtz Zentrum München, Institute of Genetic Epidemiology, Neuherberg, Germany, [121]Institute of Medical Informatics Biometry and Epidemiology, Ludwig-Maximilians-Universität München, Munich, Germany, [122]Yong Loo Lin School of Medicine, National University of Singapore and National University Health System, Singapore, Singapore, [123]Duke-National University of Singapore Medical School, Singapore, Singapore, [124]Department of Laboratory Medicine and Pathology, Minneapolis, MN, United States, [125]Department of Internal Medicine, Erasmus MC, University Medical Center, Rotterdam, Netherlands, [126]German Center for Cardiovascular Research (DZHK), Partner Site Munich Heart Alliance, Munich, Germany, [127]Beijing Ophthalmology and Visual Science Key Lab, Beijing Tongren Eye Center, Beijing Tongren Hospital, Beijing Institute of Ophthalmology, Capital Medical University, Beijing, China, [128]Jackson Heart Study Graduate Training Center, School of Public, Jackson State University, Jackson, MS, United States, [129]Department of Epidemiology, School of Public Health, University of Pittsburgh, Pittsburgh, PA, United States, [130]Division of Cancer Control and Population Sciences, University of Pittsburgh Medical Center (UPMC) Hillman Cancer Center, Pittsburgh, PA, United States, [131]Department of Metabolism, Digestion and Reproduction, Imperial College London, London, United Kingdom, [132]Broad Institute, Massachusetts Institute of Technology and Harvard University, Boston, MA, United States, [133]Department of Public Health and Clinical Medicine, Umeå University, Umeå, Sweden, [134]Department of Nutrition, Harvard Chan School of Public Health, Boston, MA, United States, [135]German Center for Diabetes Research, Neuherberg, Germany, [136]Department of Ophthalmology, Medical Faculty Mannheim, University Heidelberg, Mannheim, Germany, [137]Institute of Molecular and Clinical Ophthalmology, Basel, Switzerland, [138]Department of Clinical Physiology and Nuclear Medicine, Kuopio University Hospital, Kuopio, Finland, [139]Department of Epidemiology, University of North Carolina at Chapel Hill, Chapel Hill, NC, United States, [140]Celgene, Bristol Myers Squibb, Mississauga, ON, Canada, [141]Department of Medicine, Surgery and Health Sciences, University of Trieste, Trieste, Italy, [142]Division Heart and Lungs, Department of Cardiology, University Medical Center Utrecht, University of Utrecht, Utrecht, Netherlands, [143]Institute for Community Medicine, University Medicine Greifswald, Greifswald, Germany, [144]National Institute on Aging, National Institutes of Health, Bethesda, MD, United States, [145]Division of Cardiology, Department of Medicine, University of Mississippi Medical Center, Jackson, MS, United States, [146]Faculty of Medicine, University of Iceland, Reykjavik, Iceland, [147]Department of Public Health and Primary Care, Leiden University Medical Center, Leiden, Netherlands, [148]Department of Epidemiology, University of Washington, Seattle, WA, United States, [149]Department of Health Systems and Population Health, University of Washington, Seattle, WA, United States, [150]Centre for Global Health, The Usher Institute, The University of Edinburgh, Edinburgh, United Kingdom, [151]Division of Population Health and Genomics, Ninewells Hospital and Medical School, University of Dundee, Dundee, United Kingdom, [152]Department of Exercise and Nutrition Sciences, Milken Institute School of Public Health, The George Washington University, Washington, DC, United States, [153]Nuffield Department of Population Health, University of Oxford, Oxford, United Kingdom, [154]College of Public Health, Dean's Office, University of Kentucky, Lexington, KY, United States, [155]Division of Biostatistics, Population and Public Health Sciences, University of Southern California, Los Angeles, CA, United States

Introduction: Educational attainment, widely used in epidemiologic studies as a surrogate for socioeconomic status, is a predictor of cardiovascular health outcomes.

Methods: A two-stage genome-wide meta-analysis of low-density lipoprotein cholesterol (LDL), high-density lipoprotein cholesterol (HDL), and triglyceride (TG) levels was performed while accounting for gene-educational attainment interactions in up to 226,315 individuals from five population groups. We considered two educational attainment variables: "Some College" (yes/no, for any education beyond high school) and "Graduated College" (yes/no, for completing a 4-year college degree). Genome-wide significant ($p < 5 \times 10^{-8}$) and suggestive ($p < 1 \times 10^{-6}$) variants were identified in Stage 1 (in up to 108,784 individuals) through genome-wide analysis, and those variants were followed up in Stage 2 studies (in up to 117,531 individuals).

Results: In combined analysis of Stages 1 and 2, we identified 18 novel lipid loci (nine for LDL, seven for HDL, and two for TG) by two degree-of-freedom (2 DF) joint tests of main and interaction effects. Four loci showed significant interaction with educational attainment. Two loci were significant only in cross-population analyses. Several loci include genes with known or suggested roles in adipose (*FOXP1, MBOAT4, SKP2, STIM1, STX4*), brain (*BRI3, FILIP1, FOXP1, LINC00290, LMTK2, MBOAT4, MYO6, SENP6, SRGAP3, STIM1, TMEM167A, TMEM30A*), and liver (*BRI3, FOXP1*) biology, highlighting the potential importance of brain-adipose-liver communication in the regulation of lipid metabolism. An investigation of the potential druggability of genes in identified loci resulted in five gene targets shown to interact with drugs approved by the Food and Drug Administration, including genes with roles in adipose and brain tissue.

Discussion: Genome-wide interaction analysis of educational attainment identified novel lipid loci not previously detected by analyses limited to main genetic effects.

KEYWORDS

educational attainment, lipids, cholesterol, triglycerides, genome-wide association study, meta-analysis

1 Introduction

Educational attainment is widely used in epidemiologic studies as an index of socioeconomic status (SES) (Kaplan and Keil, 1993). Many studies have identified educational level and other indices of SES as predictors of health outcomes (Hamad et al., 2019), coronary heart disease (CHD) risk factors (Hamad et al., 2019), and lifestyle choices such as consumption of an atherogenic diet (Shea et al., 1993). Although educational level may not capture a holistic representation of SES (Braveman et al., 2005), higher educational attainment has been shown to have a positive impact on all-cause mortality (Kaplan and Keil, 1993) and cardiovascular risk traits (Leino et al., 1999) such as blood pressure and hypertension (Leng et al., 2015), coronary artery disease (Matthews et al., 1989), coronary calcification (Gallo et al., 2001), metabolic syndrome (Matthews et al., 1989), and lipid levels (Matthews et al., 1989; Metcalf et al., 1998). However, the mitigating effects of higher education on health outcomes are often attenuated in minoritized groups (Braveman et al., 2005; Assari and Bazargan, 2019), even after controlling for other indices of SES (Metcalf et al., 1998). This differential effect raises the possibility that interactions between educational attainment and genetics contribute to the association with health outcomes.

There has been relatively little focus on genetic interactions with educational attainment as determinants of health outcomes, particularly cardiovascular health, although genetic influences on education level itself (Okbay et al., 2016) have been explored. We have previously reported novel blood pressure loci by genome-wide association studies (GWAS) that explicitly modeled genetic interactions with educational attainment (Basson et al., 2014; de las Fuentes et al., 2020). Other studies have identified evidence of gene-environment interactions for a variety of disease traits including neuropsychiatric disorders (Assary et al., 2018; Werme et al., 2021), systemic lupus erythematosus, (Woo et al., 2022), and lung function (Melbourne et al., 2022).

There has been no comprehensive assessment of interactions between genetic variation and educational attainment on lipid levels. Dyslipidemia, a leading contributor to cardiovascular morbidity and mortality, exhibits significant disparity among population groups. Consideration of educational attainment as a genetic modifier may allow identification of novel lipid loci and offer insights into the biological mechanisms that may serve to identify new therapeutic targets. Here, by combining cohorts available in the Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) Gene-Lifestyle Interactions Working Group (Rao et al., 2017), we performed genome-wide meta-analysis of low-density lipoprotein cholesterol (LDL), high-density lipoprotein cholesterol (HDL), and triglyceride (TG) levels while accounting for gene-educational attainment interactions, used as a surrogate for socioeconomic status.

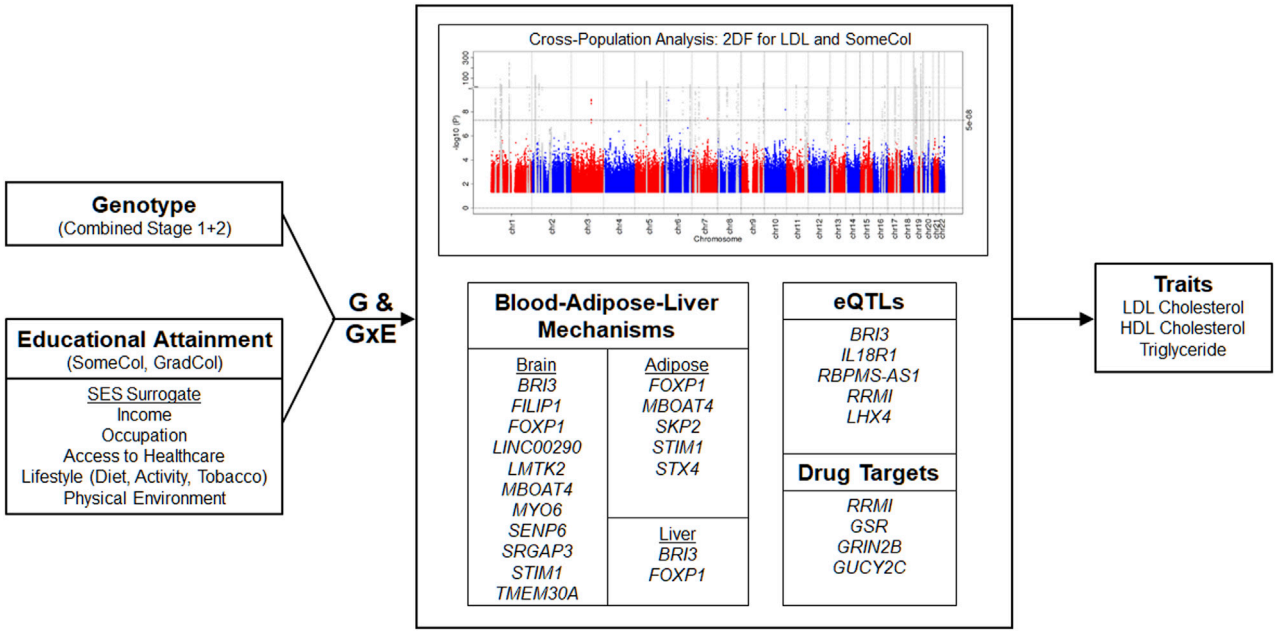


FIGURE 1
Overview. A two-stage meta-analysis of gene-educational attainment interactions on lipid traits considering two educational attainment (a surrogate for socioeconomic status) was performed. Subsequently, a meta-analysis combining results of Stages 1 and 2 was performed to identify known and novel loci for lipid traits. Identified loci include genes with known or suggested roles in brain, adipose, and liver biology. Functional annotation, expression quantitative trait loci (eQTLs), and potential druggability of targets was explored. In the Manhattan plot, known and novel loci are depicted in gray and red/blue, respectively. GxE, gene-environment interaction; GradCol, graduated college; SES, socioeconomic status; SomeCol, some college.

2 Materials and methods

2.1 Participating studies

We performed analyses in two stages (Supplementary Figure S1). A total of 41 cohorts including 108,784 men and women (aged 17–80 years) from European (EUR), African (AFR), East Asian (EAS), Hispanic admixed (HIS), and Brazilian admixed (BRZ) populations contributed to Stage 1 genome-wide interaction analyses (Supplementary Table S1); populations were defined by individual cohorts. An additional 42 cohorts (Supplementary Table S2) including 117,531 individuals contributed to Stage 2 analyses of most promising genetic variants [mostly single nucleotide variants (SNVs), also including a small number of insertions and deletions (indels)] selected from Stage 1. Participating studies are described in the Supplementary Material. Each study obtained informed consent from participants and approval from the appropriate institutional and/or ethical review boards.

2.2 Lipid and educational attainment variables

Both longitudinal and cross-sectional studies were included. In longitudinal cohorts that had multiple clinic visits for each subject, a single visit was chosen that maximized the sample size. Three lipid traits were considered for analyses: low-density lipoprotein cholesterol (LDL), high-density lipoprotein cholesterol (HDL), and triglyceride (TG) (all mg/dL). LDL was directly assayed or calculated via the Friedewald equation (LDL = TC−HDL−[TG/5]) for those with fasting TG ≤ 400 mg/dL (Friedewald et al., 1972). If fasting TG > 400 mg/dL or if TG is non-fasting, LDL was set to missing unless directly assayed. LDL concentrations were adjusted for statin use as described elsewhere (Peloso et al., 2014). Either fasting or non-fasting HDL was acceptable for analysis. Non-fasting TG levels were set to missing. HDL and TG concentrations were natural log-transformed for analysis. Descriptive statistics for these lipid traits are presented in Supplementary Tables S3, S4. For educational attainment, two dichotomous variables were defined in a way that made it possible to harmonize the variable in most cohorts, thereby maximizing the sample size. The first variable, "Some College" (SomeCol), was coded as 1 if the subject received any education beyond high school (i.e., 12 years of combined primary and secondary education), including vocational school, and as 0 if no education beyond high school. The second variable, "Graduated College" (GradCol), was coded as 1 if the subject completed at least a 4-year college degree (i.e., post-secondary or tertiary education, at least 16 years of formal education), and as 0 for any education less than a 4-year degree. Subjects with missing data for lipid levels, educational attainment, or any covariates were excluded from analysis.

2.3 Genotype data

Genotyping was performed by each participating study using Illumina (San Diego, CA, United States) or Affymetrix (Santa Clara, CA, United States) genotyping arrays. Imputation was performed using the 1000 Genomes Project (1000 Genomes Project

TABLE 1 Summary of novel loci.

Locus	rsID[a]	Chr	Pos (hg19)	Coded allele	Coded allele Freq[b]	$Beta_0$[c]	$Beta_1$[d]	2 DF P[e]	1 DF G × E P[f]	N	Genes[g]	RefPop-trait-Exp[h]
1	rs139845754	1	180,258 167	A	0.052	−0.101	−0.06	3.40E-08	2.00E-01	10,913	*LHX4; RP5-11 80C10.2;* ***ACBD6****; XPR1; KIAA1614; STX6*	AFR-TG-SomeCol
2[i]	rs7567133	2	103,239 356	C	0.031	0.102	0.021	1.72E-08	1.16E-03	6,361	*IL1RL1; IL18R1; IL18RAP; SLC9A4;* ***SLC9A2;*** *MFSD9; TMEM182*	AFR-HDL-GradCol
3	rs79367750	3	9,096 107	C	0.096	1.251	−10.279	3.99E-08	1.94E-04	8,855	***SRGAP3***	EAS-LDL-SomeCol
4[i]	rs147731578	3	70,605 665	ATTATT	0.018	−0.145	0.034	1.39E-09	3.51E-09	15,584	*MITF; FOXP1*	EUR-HDL-SomeCol
5[i]	rs74620279	4	179,620 483	C	0.025	−12.24	11.59	4.67E-08	1.39E-06	5,669	*SNORD65*	AFR-LDL-SomeCol
6	rs77249395	4	181,181 245	G	0.035	0.006	−0.026	3.17E-06	2.41E-08	134,413	*LINC00290*	TA-HDL-GradCol EUR-HDL-GradCol
7[i]	rs11132093	4	182,664 815	A	0.061	−1.65	15.84	5.04E-09	4.10E-07	2,809	*RP11-540E16.2; TENM3*	EAS-LDL-GradCol
8	rs190502162	5	37,160 808	C	0.013	9.266	−0.715	3.35E-08	4.04E-07	48,467	*SKP2; NADK2; RANBP3L; SLC1A3; NIPBL;* ***CPLANE1****; NUP155; WDR70; GNDF*	EUR-LDL-SomeCol
9	rs192718305	5	82,514 005	G	0.025	3.481	19.901	2.05E-08	1.60E-03	5,664	*RP11-343L5.2; TMEM167A;* ***XRCC4***	AFR-LDL-SomeCol
10[i]	rs147892694	6	76,693 587	G	0.034	−0.037	−0.087	8.61E-09	1.20E-01	8,599	*TMEM30A; FILIP1; SENP6; MYO6;* ***IMPG1***	AFR-HDL-SomeCol
11	rs7015	7	97,920 623	A	0.203	−1.142	−0.526	4.56E-08	6.18E-02	157,929	*LMTK2; TECPR1;* ***BRI3****; BAIAP2L1; NPTX2*	CPA-LDL-SomeCol
12[j]	rs77655002	8	30,437 023	C	0.042	−9.97	10.73	5.52E-09	7.65E-08	7,166	*MBOAT4; RBPMS-AS1; RBPMS;* ***GTF2E2****; GSR; TEX15*	AFR-LDL-GradCol CPA-LDL-GradCol
13	rs144190766	10	108,927 076	C	0.03	−0.015	−0.103	2.52E-08	5.64E-03	9,677	*SORCS1; RNA5SP326*	AFR-HDL-SomeCol
14	rs116562538	10	129,845 329	A	0.028	−4.573	−9.049	6.79E-09	4.82E-02	20,939	***PTPRE***	CPA-LDL-SomeCol
15	rs35287906	11	4,041 010	C	0.014	3.405	−5.247	1.53E-09	7.71E-11	81,020	*PGAP2; RHOG;* ***STIM1****; RRM1; OR55B1P; RP11-23F23.3*	EUR-LDL-GradCol
16	rs11230661	11	55,451 313	A	0.185	0.01	0.007	3.15E-10	3.90E-01	151,320	*TRIM48; TRIM51; 87* Olfactory Receptor Genes; *APLNR*	EUR-HDL-GradCol EUR-HDL-SomeCol

(Continued on following page)

TABLE 1 (*Continued*) **Summary of novel loci.**

Locus	rsID[a]	Chr	Pos (hg19)	Coded allele	Coded allele Freq[b]	$Beta_0$[c]	$Beta_1$[d]	2 DF P[e]	1 DF G × E P[f]	N	Genes[g]	RefPop-trait-Exp[h]
17[i]	rs148063115	12	14,221 135	T	0.026	0.022	0.222	4.86E-08	4.35E-03	6,168	*GRIN2B; RP11-72J9.1; RP11-298E10.1; RN7SL676P; GUCY2C*	AFR-TG-GradCol
18[i]	rs190746034	13	101,418 063	G	0.022	−0.002	−0.158	1.26E-08	1.06E-05	6,361	*RP11-151A6.4; TMTC4;* ***NALCN-AS1****; ARF4P3*	AFR-HDL-GradCol

Notes:
[a]rsID, based on dbSNP, build 146.
[b]Coded allele frequency.
[c]Effect size (beta) of the unexposed group.
[d]Effect size (beta) of the exposed group.
[e]GWAS, 2 DF *p*-value of the significant lead SNP, for this locus.
[f]GWAS, 1 DF, genetic-educational attainment interaction *p*-value of the significant lead SNP, for this locus.
[g]Nearest gene of all SNVs, in LD (r2>0.2) with lead variant; if SNV, not in FUMA, gene or nearest flanking coding genes noted. Bolded genes reflect intragenic lead SNV.
[h]The reference panel used in FUMA, to obtain functional annotations-trait-exposure; if more than one RefPop-Trait-Exp listed, the data provided is for the more significant association which is listed first.
[i]Only significant in Stage I analyses.
[j]2DF p = 1.83E-08, 1 DF G×E p = 4.14E-08 in CPA-LDL-GradCol analyses.
AFR, african population; EAS, east asian population; EUR, european population; GradCol, graduated college; HDL, high-density lipoprotein cholesterol; LDL, low-density lipoprotein cholesterol; CPA, cross-population analyses; TG, triglyceride.

Consortium et al., 2012) Phase I Integrated Release Version 3 Haplotypes (2010-11 data freeze, 2012-03-14 haplotypes) as a reference panel by most cohorts. Information on genotype platform and imputation for each study is presented in Supplementary Tables S5, S6 and as described in the Supplementary Material.

2.4 Analysis methods

Each study performed population-specific association analyses using the following Model 1 (joint model) that includes the effects of G, educational attainment, and their interaction (see below):

$$E(Y) = \beta_0 + \beta_G\, G + \beta_E\, \text{Education} + \beta_{GE}\, G \times \text{Education} + \boldsymbol{\beta}_C\, \mathbf{C} \quad (1)$$

where Y is the lipid variable (LDL, HDL, or TG), "Education" is the educational variable (SomeCol or GradCol), and G is the dosage of the imputed variant coded additively from 0 to 2. The vector of adjustment covariates (**C**) includes age, sex, indicators of field center (for multi-center studies), and principal components (as many as deemed necessary by study personnel to adjust for population stratification). In addition, studies in Stage 1 performed association analysis using the following Model 2 that includes the effects of G and educational attainment (but not their interaction):

$$E(Y) = \beta_0 + \beta_G\, G + \beta_E\, \text{Education} + \boldsymbol{\beta}_C\, \mathbf{C} \quad (2)$$

For model 1, each study provided the estimated variant effect (β_G), estimated variant-educational attainment interaction effect (β_{GE}), their robust standard errors, and a robust estimate of the covariance between β_G and β_{GE}. We considered the 1 degree of freedom (DF) test of the interaction effect (β_{GE}) and 2 DF joint test of both variant (β_G) and interaction effects (β_{GE}) (Kraft et al., 2007). Population-specific and cross-population inverse-variance weighted meta-analysis was performed for the 1 DF test and joint 2 DF test (Manning et al., 2011), both using METAL (Willer et al., 2010). In Stage 1, EUR, AFR, EAS meta-analyses, variants were included if they were available in more than 5,000 samples or at least 3 cohorts (these filters were not applied to BRZ or HIS because of the limited number/size of the available cohorts included in these meta-analyses). We applied genomic control correction (Devlin and Roeder, 1999) twice in Stage 1, first for study-specific GWAS results and again for meta-analysis results. Genome-wide significant ($p < 5 \times 10^{-8}$) and suggestive ($p < 1 \times 10^{-6}$) variants in Stage 1 were taken forward into Stage 2 analysis. Genomic control correction was not applied to the Stage 2 results as association testing was performed for only selected variants. Results presented reflect meta-analyses combining Stages 1 and 2. Loci were defined by physical distance (±1 Mb around the lead variant of the respective locus).

Extensive quality control was performed, as described in the Supplementary Material. For Stages 1 and 2, to remove unstable study-specific results that reflected small sample size, low minor allele count (MAC), or low imputation quality, we excluded variants for which the "approximate DF" (defined as the minimum of [MAC0, MAC1] × imputation quality) < 20, where MAC0 and MAC1 are the MAC in the two educational attainment strata per exposure variable.

2.5 Characterization of functional roles

Loci were characterized as known (previously reported, as defined in the Supplementary Methods) or novel. A suite of tools implemented in Functional Mapping and Annotation (FUMA) of Genome-Wide Association Studies (Watanabe et al., 2017) (version 1.3.5; described in detail in the Supplementary Material) were used to identify functional roles for the lead variants and nearby variants in linkage disequilibrium (LD; $r^2 \geq 0.2$) in each of the novel lipid loci. LD information was obtained from the 1000 Genomes Project Phase 3 reference genome for the population with the most significant

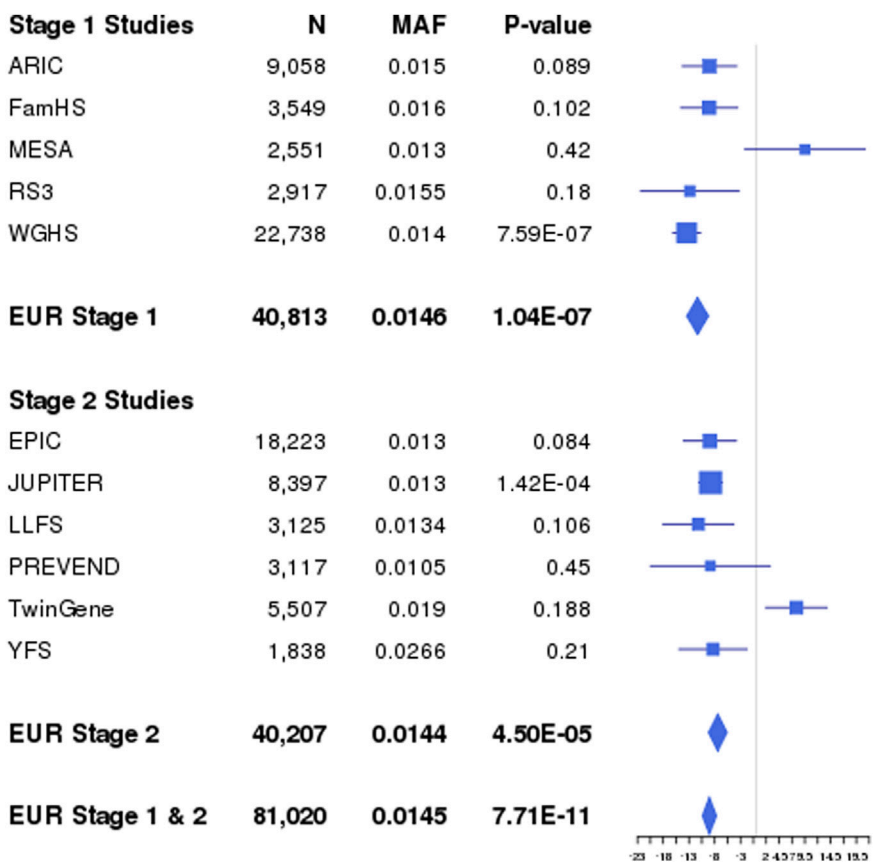


FIGURE 2
Interaction effects of Locus 15 (rs35287906; *STIM1*) identified through combined Stage 1 and Stage 2 interaction effects with GradCol for LDL in EUR and CPA. Forest plots show β values (95% confidence intervals) and *p*-values (1 DF) for the rs35287906 × GradCol interaction term in linear regression models of LDL adjusted for age, sex, field center (for multi-center studies), and principal components. Results shown are for each EUR study, as well as the population-specific combined Stage 1 and 2 meta-analysis results. The interaction effect β_G Educ corresponds to the difference in genetic effects between higher (β_{G1} = −5.25 mg/dL per minor allele) and lower education (β_{G0} = 3.41 mg/dL per minor allele), for a combined interaction effect of −8.66 mg/dL. AF, coded allele frequency; N, sample size.

population-specific association. If the most significant association was in cross-population analyses (CPA), the reference genome for "1000G Phase 3 ALL" was used (Ward and Kellis, 2012). Two lead insertion/deletion loci were not identified in the reference genomes by FUMA and therefore not detailed. Nearest gene annotations were limited to protein coding, long non-coding RNAs (lncRNAs), and non-coding RNAs (ncRNAs) within 10 kb of lead variants and variants in LD ($r^2 \geq 0.2$) with the lead variant (Wang et al., 2010).

For the lead and LD variants, we used FUMA to report the RegulomeDB score, Combined Annotation Dependent Depletion (CADD) scores, the 15-core chromatin state (ChromHMM), and expression quantitative trait loci (eQTLs). Using nearest-gene annotations, FUMA was used to generate tissue-specific gene expression data (GTEx V8 dataset, 53 tissue types).

3 Results

3.1 Overview

We performed a two-stage meta-analysis of gene-educational attainment interactions on lipid traits considering two educational attainment variables, as previously described (Supplementary Figure S1) (de las Fuentes et al., 2020) Herein, we report our findings based on up to 227,850 individuals from five populations. In Stage 1, we pursued genome-wide interrogation in 108,784 individuals of European (EUR; n = 80,379), African (AFR; n = 12,295), East Asian (EAS; n = 11,002), Hispanic/Latino admixed (HIS; n = 1,455), and Brazilian admixed (BRZ; n = 3,653) populations (Supplementary Table S1). We performed genome-wide meta-analyses of approximately 18.8 million SNVs and indels variants imputed using the 1000 Genomes Project reference panel (QQ plots, Supplementary Figures S2A–E). Through the 1 DF test of the interaction effect and the 2 DF joint test of the SNV and interaction effects, we identified 13,851 genome-wide significant ($p < 5 \times 10^{-8}$) and 6,835 suggestive ($p < 1 \times 10^{-6}$ and $\geq 5 \times 10^{-8}$) variants in known or novel loci that were associated with any lipid trait in any population or educational attainment analysis. These were followed-up in 117,531 additional individuals of EUR (n = 92,690), AFR (n = 6,630), EAS (n = 6,589), and HIS (n = 11,622) populations in Stage 2 (Supplementary Table S2).

We then performed meta-analyses combining Stages 1 and 2 (Figure 1; Manhattan Plots; Supplementary Figures S3A–F) and identified 128 significant loci ($p < 5 \times 10^{-8}$): 18 were novel loci and

110 loci were known (previously reported) loci, although the specific index variant often varied based on population (Supplementary Table S7). The majority of the associations with known loci were detected for EUR (99 loci) and cross-population analyses (CPA; 107 loci), reflecting the European-centric composition of prior studies, although known loci were also detected for the other populations: AFR (20 loci), EAS (15 loci), BRZ (3 loci), and HIS (22 loci). Four LDL and five HDL known loci were only identified by CPA. Eight of the 110 known loci were significantly associated with two lipid traits.

Across the lipid fractions, educational exposures, and populations, we identified 18 novel ($p < 5 \times 10^{-8}$) loci located at least 1 Mb away from any known lipid loci (Table 1) using the 1 DF and/or the 2 DF tests. Of these 18 loci, seven were identified by Stage 1 analyses only as Stage 2 analyses did not meet the filter threshold of an "approximate DF" ≥ 20. Among the 18 loci, one was identified only through 1 DF interaction-effect analyses (locus 6), 14 through 2 DF analyses, and three through both 2 DF and 1 DF interaction-effect only analyses (loci 4, 12, and 15). The four loci with significant 1 DF interaction effects include lead variants in or adjacent to forkhead box P1 (*FOXP1*), long intergenic non-protein coding RNA 290 (*LINC00290*), general transcription factor IIE subunit 2 (*GTF2E2*)-membrane bound O-acyltransferase domain containing 4 (*MBOAT4*), and stromal interaction molecule 1 (*STIM1*). For example, at *STIM1* (locus 15), we observed an opposite genetic effect between higher and lower education: the minor allele C was associated with a 0.14 mmol/L lower LDL in higher education (GradCol = 1), whereas it was associated with a 0.09 mmol/L higher LDL in lower education (GradCol = 0), for a combined interaction effect of −0.22 mmol/L (Figure 2).

Among the 18 novel loci, nine were found in LDL analyses, seven in HDL analyses, and two in TG analyses; none of the novel loci reached genome-wide significance for more than one lipid trait. Considering the 17 novel loci that were significant in 2 DF tests, six loci were identified considering "Graduated College" (GradCol), 10 were identified considering "Some College" (SomeCol), and one locus was significant for both "Graduated College" and "Some College" (locus 16). Examining the 17 loci for evidence of support for the non-significant educational attainment exposure (Supplementary Table S8), nine loci had at least nominal significance ($p < 0.05$; loci 1, 2, 3, 4, 7, 8, 11, 14, and 17), four did not have at least nominal significance ($p \geq 0.05$; loci 10, 12, 15, and 18), and three did not have data available for the other educational attainment exposure trait (loci 5, 9, and 13) because of failure to meet filter thresholds (i.e., ≥20 copies of the minor allele in the exposed group).

The LocusZoom plots of these novel loci are presented in Supplementary Figure S4.

3.2 Ancestry-specific and cross-population analyses

Novel loci were identified through separate analyses of AFR (nine loci), EUR (three loci), EAS (two loci), CPA (two loci), and in both EUR and CPA (two loci). Among the 18 novel loci, two loci were identified only through CPA, as none of the population-specific analyses reached genome-wide significance. For example, the SNV (rs7015, locus 11) was only nominally associated with LDL in EUR ($p = 3.04 \times 10^{-4}$), HIS ($p = 9.96 \times 10^{-3}$), EAS ($p = 1.99 \times 10^{-2}$), and AFR ($p = 2.81 \times 10^{-2}$). However, in cross-population analysis combining these four populations, the association reached genome-wide significance ($p = 4.56 \times 10^{-8}$). This SNV resides in the 3' untranslated region (UTR) of an alternatively expressed transcript of brain protein I3 (*BRI3*).

3.3 Functional annotation and eQTL evidence

To obtain functional annotations for the lead variants and nearby variants in linkage disequilibrium (LD; $r^2 \geq 0.2$), we used FUMA (Watanabe et al., 2017). Among the 18 lead variants representing our novel loci, eight variants were intronic to coding genes, one variant was exonic to a non-coding RNA (ncRNA), one variant was intronic-ncRNA, one variant was in a 3′UTR, and five variants were intergenic; two additional variants were indels without available annotation in FUMA. Of the 1,733 annotated variants that include both the lead variants and variants in LD ($r^2 > 0.2$) with available FUMA functional annotation, the majority were intergenic (64%). Among those variants annotated to gene regions (n = 619), 72% were intronic. 8.2% were exonic, and the remaining variants were in UTR and flanking regions (Supplementary Table S9).

Of the 1,769 LD variants, 25 had RegulomeDB scores better than or equal to 3a (17 in AFR, three in EUR, and five in CPA loci), suggesting at least moderate evidence for involvement in transcription regulation (Supplementary Table S9). Sixty-five variants have CADD scores ≥10, representing the top 10% of predicted deleteriousness for SNVs genome-wide (35 in AFR, 22 in EUR, three in EAS, and five in CPA loci). Eight variants in the Chromosome 11 locus including an olfactory receptor cluster have CADD scores ranging 21.2–37.0, placing them in the top 1% of predicted deleteriousness. Two additional variants are notable for high CADD scores, an exonic variant in glutathione-disulfide reductase (*GSR*; CADD score 25.5) and a variant in the *BRI3* 3′UTR region (CADD score 21.4).

The 15-core chromatin state (ChromHMM) was assessed for 127 epigenomes in the 16 lead variants available in FUMA (Supplementary Table S9). Of the lead variants, two had histone chromatin markers consistent with active or flanking active transcription start sites, and three were in regions associated with strong transcription in relevant tissues including brain, adipose tissue, and liver. Among all 1,769 LD variants, 91 had histone chromatin markers characteristic of active or flanking active transcription start sites, 218 had markers consistent with strong transcription, and 57 were in enhancer regions. Among the LD variants, those in five loci were identified as being highly significant *cis*-acting expression trait loci (eQTLs) in the GTEx V8 database: 65 variants in *BRI3* (including index rs7015 SNV in locus 11) expressed in liver [false discovery rate (FDR) *p*-values 2.27×10^{-24}], subcutaneous adipose tissue (FDR *p*-values 9.46×10^{-50}), and brain (FDR *p*-values range 3.75×10^{-18} to 5.82×10^{-19}); eight SNVs in interleukin 18 receptor 1 (*IL18R1;* including index SNV rs7567133 in locus 2) expressed in brain (FDR *p*-values range 8.30×10^{-20} to 5.89×10^{-36}) and subcutaneous adipose tissue (FDR *p*-values 4.48×10^{-7}); one SNV in RBPMS antisense RNA 1 (*RBPMS-AS1;*

locus 12) expressed in subcutaneous adipose tissue (FDR *p*-value 5.71×10^{-28}); seven SNVs in ribonucleotide reductase catalytic subunit M1 (*RRM1;* locus 15) expressed in subcutaneous adipose tissue (FDR *p*-values 1.36×10^{-44}), and three SNVs in LIM homeobox 4 (*LHX4;* locus 1) expressed in brain (FDR *p*-values 8.73×10^{-10}) and subcutaneous adipose tissue (FDR *p*-values range 1.36×10^{-9} to 4.50×10^{-13}).

3.4 Druggability targets

The potential druggability of the identified gene targets was investigated using an integrative approach as previously described (Kavousi et al., 2022). We queried high- and medium-priority candidate gene targets using the Drug-Gene Interaction database (DGIdb), which identified 17 genes annotated as clinically actionable or members of the druggable genome (Supplementary Table S10). We identified eight genes with reported drug interactions and an additional four genes with active ligand interactions in the ChEMBL database. Among these, five gene targets were shown to interact with drugs approved by the Food and Drug Administration (FDA) that have been evaluated in late-stage clinical trials using DrugBank and ClinicalTrials.gov databases (Supplementary Table S11). Among drug targets identified, *RRM1* and *GSR* are both involved in glutathione metabolism and are targets of drugs used to treat various neoplasms; glutamate ionotropic receptor NMDA type subunit 2B (*GRIN2B*) is involved in long-term neuronal potentiation; and guanylate cyclase 2C (*GUCY2C*) modulates gut cyclic GMP signaling. *GRIN2B*, which encodes a *N*-methyl *D*-aspartate (NMDA) receptor GluN2B subunit, is a target of memantine, used to treat moderate to severe dementia in patients with Alzheimer's disease. These results suggest that there are potential drug repurposing opportunities as novel therapies for lipid management.

4 Discussion

4.1 Overview

This study reports a genome-wide meta-analysis of data from up to 226,315 individuals from five population groups. In this study, educational attainment was used as a multidimensional surrogate of SES reflective of a variety of environmental factors such as occupation, wealth, access to quality healthcare, diet, lifestyle, and physical activity. We identified 18 novel loci for LDL, HDL, and TG at genome-wide significance when accounting for gene-educational attainment interactions. The majority of novel loci (nine of 18 loci) were identified in AFR, likely reflecting a lack of population diversity in prior large-scale genome-wide studies. Many of these novel loci include genes with biologic roles in adipose, brain, and hepatic tissue.

Adipose tissue serves a critical role in sequestering circulating free fatty acids as inert triglycerides lipid droplets. Processes that limit the differentiation or subsequent function of adipocytes may contribute to abnormal lipid metabolism. Adipose tissue is also an active endocrine organ that elaborates a variety of adipokines (Ahima, 2006), such as tissue necrosis factor-alpha (TNFα), interleukins (IL)-6, IL-1, leptin, adiponectin, and others. Dysfunctional adipose tissue and pro-inflammatory adipokines can trigger ectopic deposition of fatty acids in other tissues, such as skeletal muscle and liver (Jung and Choi, 2014; Shulman, 2014), which can lead to a variety of metabolic disorders such as insulin resistance, type 2 diabetes, non-alcohol fatty liver disease, and dyslipidemia (Franssen et al., 2011; Jung and Choi, 2014). Increased hepatic fatty acid uptake stimulates synthesis of TG-rich very low-density lipoprotein (VLDL) cholesterol particles that are converted in the bloodstream to small-dense LDL particles through a process that also lowers circulating HDL (Bays et al., 2013). Whereas most of the LDL cholesterol is taken up again by the liver, a small fraction is removed from circulation by endocytosis via LDL receptors located in extrahepatic tissues, including the brain. Brain-adipose-liver communication pathways help maintain homeostasis by integrating peripheral metabolic signals; (Franssen et al., 2011; Gliozzi et al., 2021); miscommunication leads to central dysregulation and metabolic disorders (Yi and Tschop, 2012). For example, in rodents, insulin acts in the brain to enhance hepatic TG secretion via VLDL synthesis (Scherer et al., 2016). There is additional evidence to suggest that circulating plasma cholesterol concentration may play a role in neurodegeneration in susceptible individuals (Dietschy and Turley, 2001). A high-fat, high-cholesterol diet has also been associated with impaired cognition and memory (Ledreux et al., 2016) through mechanisms that may involve brain inflammation (Pistell et al., 2010).

4.2 Novel lipid loci include genes expressed in adipose tissue

Given the important role played by adipose tissue in regulating lipid metabolism, it is notable that two novel loci were identified that include genes with known roles in adipocyte differentiation and/or function. For example, syntaxin 6 (*STX6*; Table 1, locus 1, TG locus in AFR) has been shown to play a role in mediating insulin-stimulated translocation of the glucose transporter-4 (Glut4) in adipose tissue (Perera et al., 2003). After feeding, transgenic mice that overexpress *Glut4* in adipose tissue show reduced activity of lipoprotein lipase (Gnudi et al., 1996), the rate-limiting step for clearing plasma TG (Wang and Eckel, 2009). S-phase kinase associated protein 2 (*SKP2*; locus 8, LDL locus in EUR) plays a role in adipocyte differentiation (Okada et al., 2009); transgenic *Skp2* knock-out mice have a 50% reduction in both subcutaneous and visceral adipocyte numbers (Cooke et al., 2007).

4.3 Novel lipid loci include genes expressed in the brain

Nine novel lipid loci have been identified that include genes responsible for vital functions in the central nervous system. For example, the gene products of lemur tyrosine kinase 2 (*LMTK2*; locus 11, LDL locus in CPA) and myosin VI (*MYO6*; locus 10, HDL locus in AFR) both bind with kinesin-1 light chain in neurons to mediate axonal transportation of a wide variety of cargo including mitochondria and neurotransmitter-containing vesicles, and participate in glutamate receptor endocytosis on the pre-synaptic membrane (Li et al., 2020). Genes with similar functions are often clustered along chromosomes where shared

regulatory domains mediate coexpression (Semon and Duret, 2006). Such may be the case for the locus on chromosome 6 (locus 10, HDL locus in AFR) where a series of genes are expressed in the hippocampus of the brain [transmembrane protein 30A (*TMEM30A*) (Xu et al., 2012), filamin A interacting protein 1 (*FILIP1*), (LoTurco and Bai, 2006), SUMO specific peptidase 6 (*SENP6*), (Loriol et al., 2013), and *MYO6* (Tamaki et al., 2008)]. The hippocampus is responsible for consolidating short-term into long-term memory; modeling of interactions with educational attainment may have facilitated the detection of this novel locus. In murine models, knock-down of *TMEM30A* by small interfering RNAs (siRNAs) reduced neurite outgrowth in the hippocampus (Xu et al., 2012). The expression of *FILIP1*, which produces a negative regulator of filamin A, is required for appropriate neocortical cell migration (LoTurco and Bai, 2006). An LDL locus in EAS (locus 3) includes SLIT-ROBO Rho GTPase activating protein 3 (*SRGAP3*), a SLIT-ROBO activating protein that guides the growth of dendritic spines on cortical neurons (Blockus and Chedotal, 2014). Limited data from a candidate-gene association study also suggest that this locus is associated with total cholesterol, HDL, and apolipoprotein A1 in Maonan Chinese (Miao et al., 2017).

GRIN2B (locus 17, TG locus in AFR), which encodes a *N*-methyl *D*-aspartate (NMDA) receptor subunit, is highly expressed in the hippocampus where it plays critical roles in memory consolidation. Murine models of aging show that transgenic overexpression of *Grin2b* improves learning and memory function (Cao et al., 2007). Notably, individuals with missense mutations in *GRIN2B* develop rare autosomal dominant forms of encephalopathy characterize by intellectual disability, impaired learning, and behavior phenotypes (Swanger et al., 2016; Fedele et al., 2018). Further preclinical and translational studies are warranted to determine the mechanisms by which interactions with educational attainment may modulate lipid levels in humans.

Some novel loci include several genes that are differentially expressed under a variety of conditions that may relate to altered environmental exposures in humans. For example, *MYO6* (Tamaki et al., 2008) (locus 8, LDL locus in EUR) expression is upregulated in animal models of stress; and transmembrane protein 167A (*TMEM167A*; locus 9, LDL locus in AFR) is differentially expressed in the hippocampus of depressed murine models (Zhang et al., 2018). Of additional interest is the long non-coding RNA (lncRNA) *LINC00290* (locus 6, HDL in CPA) which has been proposed as a "human-accelerated element" contributing to primate evolutionary shifts that lead to higher-order human capabilities such as complex language, advanced learning, and long-term planning (Kamm et al., 2013). Given evidence for expression in brain tissues, it is notable that the *LINC00290* locus was only identified through interaction analyses with educational attainment.

4.4 Novel lipid loci include genes with roles in both adipose and brain tissues

There are four additional loci containing genes that have plausible biologic roles in both adipocyte function and in the brain. For example, in locus 12 (LDL locus in AFR), *MBOAT4* has been called a "master switch" for the ghrelin system (Romero et al., 2010). Ghrelin, which is secreted by gastric endocrine cells, is made biologically active when acylated by *MBOAT4*. In addition to playing critical roles in adipogenesis, lipogenesis, and glucose homeostasis, ghrelin also stimulates food intake through actions in the brain (Pradhan et al., 2013). In locus 15 (LDL locus in EUR), the expression of stromal interaction molecule 1 (*STIM1*) negatively regulates adipocyte differentiation, impairing their ability to accumulate lipids (Graham et al., 2009). *Stim1* is also a calcium sensor that plays a critical role in the formation of dendritic spines in developing murine hippocampal cells (Kushnireva et al., 2020). In transgenic mice, overexpression of *Stim1* leads to improved contextual learning and decreased depression- and anxiety-like behaviors (Majewski et al., 2017). solute carrier family 1 member 3 (*SLC1A3*; locus 8, LDL locus in EUR) encodes a high-affinity glutamate reuptake channel in brain astrocytes that terminates excitatory neurotransmission; *Slc1a3* knock-out mice have abnormal sociability (Zhou and Danbolt, 2014). *SLC1A3* is also expressed in adipocytes, although its role in this tissue is not well understood (Krycer et al., 2017).

4.5 Gene with roles in adipose, brain, and liver tissues

Two loci contain genes with plausible biologic roles in adipose, brain, and hepatic tissues. In the brain, *BRI3* (locus 11, LDL locus in CPA) has been implicated in neuronal survival following ischemia/reperfusion injury (Yang et al., 2015) and may be a protective regulator against Alzheimer disease (Matsuda et al., 2009). Several SNVs in this locus are significant eQTLs for *BRI3* expression in both the liver and subcutaneous fat. A variant in *BRI3* is notable for a CADD score predictive of being deleterious. *FOXP1* (locus 4, HDL locus in EUR) is involved in adipocyte differentiation (Liu et al., 2019). In the brain, the transcription factor, FOXP1, heterodimerizes with its paralog, forkhead box P2 (FOXP2), to form a transcription factor; rare mutations in *FOXP2* have been reported in multiple cases of intellectual disability and language impairment (Bacon and Rappold, 2012). In murine models of diabetes, hepatic *FOXP1* expression, a regulator of gluconeogenic gene expression, is downregulated (Zou et al., 2015).

4.6 Limitations

Several limitations are inherent in the design of large-scale multi-population genome-wide association studies such as this one. First, we were unable to validate seven of the 18 novel loci (one EUR, one EAS, and five AFR), largely due to the limited number of non-EUR cohorts available in Stage 2 and variants/cohorts failing to meet quality control thresholds; these loci need further validation. Second, 16 of the 18 novel loci were notable for having minor allele frequencies <0.10 which increases the possibility for type 1 and type 2 errors. Third, the interpretation of educational attainment as a proxy for SES may vary according to gender, population, region, country, and/or birth cohort (Tyroler, 1989; Sorel et al., 1992; Kaplan and Keil, 1993; Metcalf et al., 1998) and dichotomization may fail to capture more nuanced population

differences, in particular for non-minoritized populations. For example, in developing countries where diets are becoming progressively westernized, men and women from higher SES strata are at higher risk for dyslipidemia (Espirito Santo et al., 2019). Fourth, the practice of adjusting LDL concentrations for statin use is based on a method derived from a meta-analysis of largely European-population cohorts (Baigent et al., 2005) which may not be appropriate for other populations. Finally, genome-wide association studies are largely hypothesis generating in scope; findings of association warrant validation in biological systems. While we attempted to enhance potential relevance by reporting of functional annotation and druggability of candidate gene targets, biologic plausibility was extrapolated primarily from animal and *in vitro* data that may not be relevant in human lipid metabolism.

Despite these limitations, this study has multiple strengths such as a sufficiently large sample size of cohorts inclusive across the lifespan and sex and representing diverse populations, the majority which were not selected for lipid abnormalities. Furthermore, consideration of educational attainment is a novel strategy designed to enhance discovery of novel lipid loci. Whereas GWAS studies traditionally identify variants that explain only a fraction of trait variability, even loci associated with modest changes in gene expression or protein function may lay the groundwork for identifying novel drug targets and/or repurposing of existing drugs for lipid management.

4.7 Conclusions

In conclusion, this multi-population meta-analysis of LDL, HDL, and TG identified 18 novel loci by consideration of gene-educational attainment interactions; one locus was identified only through evidence for interaction with educational attainment. Several of the loci include genes with known or suggested roles in adipocyte, brain, and/or liver biology. While findings of gene-environment interactions have generally not yet been translated to clinical practice, the results of this study may identify novel potential therapeutic targets for lipid management, especially those involving central control of lipid metabolism.

Data availability statement

The data analyzed in this study is subject to the following licenses/restrictions: 83 US and international cohorts participated in this study, each subject to different regulations regarding sharing of data. Requests to access these datasets should be directed to LF, lfuentes@wustl.edu.

Ethics statement

The studies involving humans were approved by the Washington University School of Medicine, Human Research Protections Office and by the appropriate institutional and/or ethical review board for each participating cohort (see Supplementary Material). The studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study.

Author contributions

LF wrote the first draft of the manuscript. ARB, CLM, KLS, MRB, PBM, TWW, YJS contributed substantially to the revision of the manuscript and share first authorship. DKA, DCR, WJG, C-TL, ACM, JIR, MF share senior authorship of the manuscript. ACM, CAM, CB, CMB, CMD, CS, C-YC, DH, DIC, DMB, DWB, EB, EST, GG, GHa, GW, HJS, HS, IJD, JBJ, JIR, JMS, J-MY, KC, KEN, LEW, MAI, MF, MH, MKW, MMP, MS, NJS, NK, OTR, PKM, RM, RMD, RR, RSC, SCH, SEH, SFW, SKM, SLK, SSR, TAL, TF, TKR, TL, TR, TTP, TW, TYW, UF, VG, W-BW, W-PK, XS, XZ, and YXW participated in the study concept and design. ARB, AC, ACM, ACP, APe, ARH, AWM, BHS, BIF, BMP, BOT, BP, BS, CAM, CB, CGi, CHC, CMB, C-TL, CY, DH, DKA, DMB, DOM, DRW, DWB, DX, EBW, EL, ERF, EST, GHa, GW, HJS, HV, IJD, IR, JAS, JBJ, JDF, JEK, JIR, JM, DJR, JMS, J-MY, KC, KEN, KLe, KLi, LRY, LCS, MA, MRB, MC, MFF, MH, MK, MKW, MMP, MN, MRB, MS, MW, NJS, NV, OTR, PAP, PF, PH, PJS, PK, PKM, PMR, PWF, RM, RM, RMD, RR, RSC, SA, SCH, SEH, SFW, SKM, SLK, SP, SS, SSR, TE, TF, TL, TM, TR, TTP, UF, VF, VG, W-BW, W-PK, WZhe, XCha, X-OS, XZ, KL, YL, YM, and YXW participated in the phenotype data acquisition and/or quality control. ARB, ABZ, AC, ACM, ACP, AGU, ARH, BHS, BIF, BMP, BOT, CB, CCK, CDL, CGa, CKH, CPN, C-TL, CY, DEA, DIC, DKA, DMB, DOM, DRW, DX, EB, EBW, EL, F-CH, FG, FPH, GHo, GW, IJD, IK, IMN, IN, JAB, JAS, JD, JDF, JEK, JFC, JIR, JLiu, JMS, KDT, KEN, KLe, KLS, KNC, KLS, LFB, L-PL, LRY, LCS, MA, MRB, MC, MF, MFF, MH, MMP, MP, MPC, MRB, MYT, NA, NDP, NJS, NK, NV, OP, PBM, PF, PH, PKM, RDo, RM, RN, SA, SEH, SH, SKM, SLK, SP, SSR, TAL, TE, TKR, TL, TR, TS, UF, VF, WZha, WZhe, X-OS, XS, XZ, Y-DIC, YF, and YJS participated in the genotype data acquisition and/or quality control. ACM, ACP, AD, AKM, APo, AR, ARB, ARH, AVS, BIF, BK, BMP, BOT, CB, CCG, CDL, CGa, CGi, CLM, CPN, C-TL, CY, C-YC, DIC, DKA, DMB, DV, DX, EBW, EL, ERF, F-CH, FG, FL, FPH, FT, GW, HA, IMN, IN, JAB, JAS, JBJ, JD, JDF, JEK, JFC, JLia, JRO, JSF, J-MY, KEN, KLe, KLS, KR, LF, LFB, L-PL, LRY, MA, MAR, MRB, MF, MFF, MH, MKE, MP, MR, MRB, MS, NDP, NV, PAP, PH, PJM, PK, PKM, PSV, RDo, RM, RN, RSC, SA, SEH, SFW, SH, SKM, SLK, TMB, TOK, TR, TW, TWW, TYW, UB, UF, VF, W-BW, WJG, WW, WZha, XChe, XG, XL, XS, YJS, YXW, and YZ participated in data analysis and interpretation. All authors have provided approval for publication of the contents and agree to be accountable for the content of the work.

Funding

This project was largely supported by two grants from the U.S. National Heart, Lung, and Blood Institute (NHLBI), the National Institutes of Health, R01HL118305 and R01HL156991. This research was supported in part by the Intramural Research Program of the National Human Genome Research Institute, National Institutes of Health (AB). An NHLBI career development award (K25HL121091) enabled YS to play an important role on this project. The full set of study-specific funding sources and acknowledgments appear in the Supplementary Material.

Acknowledgments

We gratefully acknowledge the contribution of CAM, JMS, and DMB posthumously.

Conflict of interest

IN is now employed by Celgene. BMP serves on the Steering Committee of the Yale Open Data Access Project funded by Johnson & Johnson. SA is employed by and holds equity in 23andMe, Inc. Authors BK, MW, and CGa were employed by Helmholtz Zentrum München.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2023.1235337/full#supplementary-material

References

Ahima, R. S. (2006). Adipose tissue as an endocrine organ. *Obes. (Silver Spring)* 14, 242S–249S. Suppl 5. doi:10.1038/oby.2006.317

Assari, S., and Bazargan, M. (2019). Minorities' diminished returns of educational attainment on hospitalization risk: national health interview survey (nhis). *Hosp. Pract. Res.* 4 (3), 86–91. doi:10.15171/HPR.2019.17

Assary, E., Vincent, J. P., Keers, R., and Pluess, M. (2018). Gene-environment interaction and psychiatric disorders: review and future directions. *Semin. Cell Dev. Biol.* 77, 133–143. doi:10.1016/j.semcdb.2017.10.016

Bacon, C., and Rappold, G. A. (2012). The distinct and overlapping phenotypic spectra of FOXP1 and FOXP2 in cognitive disorders. *Hum. Genet.* 131 (11), 1687–1698. doi:10.1007/s00439-012-1193-z

Baigent, C., Keech, A., Kearney, P. M., Blackwell, L., Buck, G., Pollicino, C., et al. (2005). Efficacy and safety of cholesterol-lowering treatment: prospective meta-analysis of data from 90,056 participants in 14 randomised trials of statins. *Lancet* 366 (9493), 1267–1278. doi:10.1016/S0140-6736(05)67394-1

Basson, J., Sung, Y. J., Schwander, K., Kume, R., Simino, J., de las Fuentes, L., et al. (2014). Gene-education interactions identify novel blood pressure loci in the Framingham Heart Study. *Am. J. Hypertens.* 27 (3), 431–444. doi:10.1093/ajh/hpt283

Bays, H. E., Toth, P. P., Kris-Etherton, P. M., Abate, N., Aronne, L. J., Brown, W. V., et al. (2013). Obesity, adiposity, and dyslipidemia: a consensus statement from the National Lipid Association. *J. Clin. Lipidol.* 7 (4), 304–383. doi:10.1016/j.jacl.2013.04.001

Blockus, H., and Chedotal, A. (2014). The multifaceted roles of Slits and Robos in cortical circuits: from proliferation to axon guidance and neurological diseases. *Curr. Opin. Neurobiol.* 27, 82–88. doi:10.1016/j.conb.2014.03.003

Braveman, P. A., Cubbin, C., Egerter, S., Chideya, S., Marchi, K. S., Metzler, M., et al. (2005). Socioeconomic status in health research: one size does not fit all. *JAMA* 294 (22), 2879–2888. doi:10.1001/jama.294.22.2879

Cao, X., Cui, Z., Feng, R., Tang, Y. P., Qin, Z., Mei, B., et al. (2007). Maintenance of superior learning and memory function in NR2B transgenic mice during ageing. *Eur. J. Neurosci.* 25 (6), 1815–1822. doi:10.1111/j.1460-9568.2007.05431.x

Cooke, P. S., Holsberger, D. R., Cimafranca, M. A., Meling, D. D., Beals, C. M., Nakayama, K., et al. (2007). The F box protein S phase kinase-associated protein 2 regulates adipose mass and adipocyte number *in vivo*. *Obes. (Silver Spring)* 15 (6), 1400–1408. doi:10.1038/oby.2007.168

de las Fuentes, L., Sung, Y. J., Noordam, R., Winkler, T., Feitosa, M. F., Schwander, K., et al. (2020). Gene-educational attainment interactions in a multi-ancestry genome-wide meta-analysis identify novel blood pressure loci. *Mol. Psychiatry* 26, 2111–2125. doi:10.1038/s41380-020-0719-3

Devlin, B., and Roeder, K. (1999). Genomic control for association studies. *Biometrics* 55 (4), 997–1004. doi:10.1111/j.0006-341x.1999.00997.x

Dietschy, J. M., and Turley, S. D. (2001). Cholesterol metabolism in the brain. *Curr. Opin. Lipidol.* 12 (2), 105–112. doi:10.1097/00041433-200104000-00003

Espirito Santo, L. R., Faria, T. O., Silva, C. S. O., Xavier, L. A., Reis, V. C., Mota, G. A., et al. (2019). Socioeconomic status and education level are associated with dyslipidemia in adults not taking lipid-lowering medication: a population-based study. *Int. Health* 14, 346–353. doi:10.1093/inthealth/ihz089

Fedele, L., Newcombe, J., Topf, M., Gibb, A., Harvey, R. J., and Smart, T. G. (2018). Disease-associated missense mutations in GluN2B subunit alter NMDA receptor ligand binding and ion channel properties. *Nat. Commun.* 9 (1), 957. doi:10.1038/s41467-018-02927-4

Franssen, R., Monajemi, H., Stroes, E. S., and Kastelein, J. J. (2011). Obesity and dyslipidemia. *Med. Clin. North Am.* 95 (5), 893–902. doi:10.1016/j.mcna.2011.06.003

Friedewald, W. T., Levy, R. I., and Fredrickson, D. S. (1972). Estimation of the concentration of low-density lipoprotein cholesterol in plasma, without use of the preparative ultracentrifuge. *Clin. Chem.* 18 (6), 499–502. doi:10.1093/clinchem/18.6.499

Gallo, L. C., Matthews, K. A., Kuller, L. H., Sutton-Tyrrell, K., and Edmundowicz, D. (2001). Educational attainment and coronary and aortic calcification in postmenopaUNITED STATESl women. *Psychosom. Med.* 63 (6), 925–935. doi:10.1097/00006842-200111000-00011

Genomes Project Consortium, Abecasis, G. R., Auton, A., Brooks, L. D., DePristo, M. A., Durbin, R. M., et al. (2012). An integrated map of genetic variation from 1,092 human genomes. *Nature* 491 (7422), 56–65. doi:10.1038/nature11632

Gliozzi, M., Musolino, V., Bosco, F., Scicchitano, M., Scarano, F., Nucera, S., et al. (2021). Cholesterol homeostasis: researching a dialogue between the brain and peripheral tissues. *Pharmacol. Res.* 163, 105215. doi:10.1016/j.phrs.2020.105215

Gnudi, L., Jensen, D. R., Tozzo, E., Eckel, R. H., and Kahn, B. B. (1996). Adipose-specific overexpression of GLUT-4 in transgenic mice alters lipoprotein lipase activity. *Am. J. Physiol.* 270, R785–R792. 4 Pt 2. doi:10.1152/ajpregu.1996.270.4.R785

Graham, S. J., Black, M. J., Soboloff, J., Gill, D. L., Dziadek, M. A., and Johnstone, L. S. (2009). Stim1, an endoplasmic reticulum Ca2+ sensor, negatively regulates 3T3-L1 pre-adipocyte differentiation. *Differentiation* 77 (3), 239–247. doi:10.1016/j.diff.2008.10.013

Hamad, R., Nguyen, T. T., Bhattacharya, J., Glymour, M. M., and Rehkopf, D. H. (2019). Educational attainment and cardiovascular disease in the UNITED STATES: a quasi-experimental instrumental variables analysis. *PLoS Med.* 16 (6), e1002834. doi:10.1371/journal.pmed.1002834

Jung, U. J., and Choi, M. S. (2014). Obesity and its metabolic complications: the role of adipokines and the relationship between obesity, inflammation, insulin resistance, dyslipidemia and nonalcoholic fatty liver disease. *Int. J. Mol. Sci.* 15 (4), 6184–6223. doi:10.3390/ijms15046184

Kamm, G. B., Pisciottano, F., Kliger, R., and Franchini, L. F. (2013). The developmental brain gene NPAS3 contains the largest number of accelerated regulatory sequences in the human genome. *Mol. Biol. Evol.* 30 (5), 1088–1102. doi:10.1093/molbev/mst023

Kaplan, G. A., and Keil, J. E. (1993). Socioeconomic factors and cardiovascular disease: a review of the literature. *Circulation* 88, 1973–1998. 4 Pt 1. doi:10.1161/01.cir.88.4.1973

Kavousi, M., Bos, M. M., Barnes, H. J., Lino Cardenas, C. L., Wong, D., O'Donnell, C. J., et al. (2022). Multi-ancestry genome-wide analysis identifies effector genes and druggable pathways for coronary artery calcification. Available from: https://www.medrxiv.org/content/10.1101/2022.05.02.22273844v1.

Kraft, P., Yen, Y. C., Stram, D. O., Morrison, J., and Gauderman, W. J. (2007). Exploiting gene-environment interaction to detect genetic associations. *Hum. Hered.* 63 (2), 111–119. doi:10.1159/000099183

Krycer, J. R., Fazakerley, D. J., Cater, R. J., Naghiloo, S., and Burchfield, J. G., (2017). The amino acid transporter, SLC1A3, is plasma membrane-localised in adipocytes and its activity is insensitive to insulin. *FEBS Lett.* 591 (2), 322–330. doi:10.1002/1873-3468.12549

Kushnireva, L., Korkotian, E., and Segal, M. (2020). Calcium sensors STIM1 and STIM2 regulate different calcium functions in cultured hippocampal neurons. *Front. Synaptic Neurosci.* 12, 573714. doi:10.3389/fnsyn.2020.573714

Ledreux, A., Wang, X., Schultzberg, M., Granholm, A. C., and Freeman, L. R. (2016). Detrimental effects of a high fat/high cholesterol diet on memory and hippocampal markers in aged rats. *Behav. Brain Res.* 312, 294–304. doi:10.1016/j.bbr.2016.06.012

Leino, M., Raitakari, O. T., Porkka, K. V., Taimela, S., and Viikari, J. S. (1999). Associations of education with cardiovascular risk factors in young adults: the Cardiovascular Risk in Young Finns Study. *Int. J. Epidemiol.* 28 (4), 667–675. doi:10.1093/ije/28.4.667

Leng, B., Jin, Y., Li, G., Chen, L., and Jin, N. (2015). Socioeconomic status and hypertension: a meta-analysis. *J. Hypertens.* 33 (2), 221–229. doi:10.1097/HJH.0000000000000428

Li, S., Xiong, G. J., Huang, N., and Sheng, Z. H. (2020). The cross-talk of energy sensing and mitochondrial anchoring sustains synaptic efficacy by maintaining presynaptic metabolism. *Nat. Metab.* 2 (10), 1077–1095. doi:10.1038/s42255-020-00289-0

Liu, P., Huang, S., Ling, S., Xu, S., Wang, F., Zhang, W., et al. (2019). Foxp1 controls brown/beige adipocyte differentiation and thermogenesis through regulating β3-AR desensitization. *Nat. Commun.* 10 (1), 5070. doi:10.1038/s41467-019-12988-8

Loriol, C., Khayachi, A., Poupon, G., Gwizdek, C., and Martin, S. (2013). Activity-dependent regulation of the sumoylation machinery in rat hippocampal neurons. *Biol. Cell* 105 (1), 30–45. doi:10.1111/boc.201200016

LoTurco, J. J., and Bai, J. (2006). The multipolar stage and disruptions in neuronal migration. *Trends Neurosci.* 29 (7), 407–413. doi:10.1016/j.tins.2006.05.006

Majewski, L., Maciag, F., Boguszewski, P. M., Wasilewska, I., Wiera, G., Wojtowicz, T., et al. (2017). Overexpression of STIM1 in neurons in mouse brain improves contextual learning and impairs long-term depression. *Biochim. Biophys. Acta Mol. Cell Res.* 1864 (6), 1071–1087. doi:10.1016/j.bbamcr.2016.11.025

Manning, A. K., LaValley, M., Liu, C. T., Rice, K., An, P., Liu, Y., et al. (2011). Meta-analysis of gene-environment interaction: joint estimation of SNP and SNP × environment regression coefficients. *Genet. Epidemiol.* 35 (1), 11–18. doi:10.1002/gepi.20546

Matsuda, S., Matsuda, Y., and D'Adamio, L. (2009). BRI3 inhibits amyloid precursor protein processing in a mechanistically distinct manner from its homologue dementia gene BRI2. *J. Biol. Chem.* 284 (23), 15815–15825. doi:10.1074/jbc.M109.006403

Matthews, K. A., Kelsey, S. F., Meilahn, E. N., Kuller, L. H., and Wing, R. R. (1989). Educational attainment and behavioral and biologic risk factors for coronary heart disease in middle-aged women. *Am. J. Epidemiol.* 129 (6), 1132–1144. doi:10.1093/oxfordjournals.aje.a115235

Melbourne, C. A., Mesut Erzurumluoglu, A., Shrine, N., Chen, J., Tobin, M. D., Hansell, A. L., et al. (2022). Genome-wide gene-air pollution interaction analysis of lung function in 300,000 individuals. *Environ. Int.* 159, 107041. doi:10.1016/j.envint.2021.107041

Metcalf, P. A., Sharrett, A. R., Folsom, A. R., Duncan, B. B., Patsch, W., Hutchinson, R. G., et al. (1998). African American-white differences in lipids, lipoproteins, and apolipoproteins, by educational attainment, among middle-aged adults: the Atherosclerosis Risk in Communities Study. *Am. J. Epidemiol.* 148 (8), 750–760. doi:10.1093/oxfordjournals.aje.a009696

Miao, L., Yin, R. X., Wu, J. Z., Yang, S., Lin, W. X., and Pan, S. L. (2017). The SRGAP2 SNPs, their haplotypes and G × E interactions on serum lipid traits. *Sci. Rep.* 7 (1), 11626. doi:10.1038/s41598-017-10950-6

Okada, M., Sakai, T., Nakamura, T., Tamamori-Adachi, M., Kitajima, S., Matsuki, Y., et al. (2009). Skp2 promotes adipocyte differentiation via a p27Kip1-independent mechanism in primary mouse embryonic fibroblasts. *Biochem. Biophys. Res. Commun.* 379 (2), 249–254. doi:10.1016/j.bbrc.2008.12.069

Okbay, A., Beauchamp, J. P., Fontana, M. A., Lee, J. J., Pers, T. H., Rietveld, C. A., et al. (2016). Genome-wide association study identifies 74 loci associated with educational attainment. *Nature* 533 (7604), 539–542. doi:10.1038/nature17671

Peloso, G. M., Auer, P. L., Bis, J. C., Voorman, A., Morrison, A. C., Stitziel, N. O., 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 (2), 223–232. doi:10.1016/j.ajhg.2014.01.009

Perera, H. K., Clarke, M., Morris, N. J., Hong, W., Chamberlain, L. H., and Gould, G. W. (2003). Syntaxin 6 regulates Glut4 trafficking in 3T3-L1 adipocytes. *Mol. Biol. Cell* 14 (7), 2946–2958. doi:10.1091/mbc.e02-11-0722

Pistell, P. J., Morrison, C. D., Gupta, S., Knight, A. G., Keller, J. N., Ingram, D. K., et al. (2010). Cognitive impairment following high fat diet consumption is associated with brain inflammation. *J. Neuroimmunol.* 219 (1-2), 25–32. doi:10.1016/j.jneuroim.2009.11.010

Pradhan, G., Samson, S. L., and Sun, Y. (2013). Ghrelin: much more than a hunger hormone. *Curr. Opin. Clin. Nutr. Metab. Care* 16 (6), 619–624. doi:10.1097/MCO.0b013e328365b9be

Rao, D. C., Sung, Y. J., Winkler, T. W., Schwander, K., Borecki, I., Cupples, L. A., et al. (2017). Multiancestry study of gene-lifestyle interactions for cardiovascular traits in 610 475 individuals from 124 cohorts: design and rationale. *Circ. Cardiovasc Genet.* 10 (3), e001649. doi:10.1161/CIRCGENETICS.116.001649

Romero, A., Kirchner, H., Heppner, K., Pfluger, P. T., Tschop, M. H., and Nogueiras, R. (2010). GOAT: the master switch for the ghrelin system? *Eur. J. Endocrinol.* 163 (1), 1–8. doi:10.1530/EJE-10-0099

Scherer, T., Lindtner, C., O'Hare, J., Hackl, M., Zielinski, E., Freudenthaler, A., et al. (2016). Insulin regulates hepatic triglyceride secretion and lipid content via signaling in the brain. *Diabetes* 65 (6), 1511–1520. doi:10.2337/db15-1552

Semon, M., and Duret, L. (2006). Evolutionary origin and maintenance of coexpressed gene clusters in mammals. *Mol. Biol. Evol.* 23 (9), 1715–1723. doi:10.1093/molbev/msl034

Shea, S., Melnik, T. A., Stein, A. D., Zansky, S. M., Maylahn, C., and Basch, C. E. (1993). Age, sex, educational attainment, and race/ethnicity in relation to consumption of specific foods contributing to the atherogenic potential of diet. *Prev. Med.* 22 (2), 203–218. doi:10.1006/pmed.1993.1017

Shulman, G. I. (2014). Ectopic fat in insulin resistance, dyslipidemia, and cardiometabolic disease. *N. Engl. J. Med.* 371 (12), 1131–1141. doi:10.1056/NEJMra1011035

Sorel, J. E., Ragland, D. R., Syme, S. L., and Davis, W. B. (1992). Educational status and blood pressure: the second national health and nutrition examination survey, 1976-1980, and the hispanic health and nutrition examination survey, 1982-1984. *Am. J. Epidemiol.* 135 (12), 1339–1348. doi:10.1093/oxfordjournals.aje.a116245

Swanger, S. A., Chen, W., Wells, G., Burger, P. B., Tankovic, A., Bhattacharya, S., et al. (2016). Mechanistic insight into NMDA receptor dysregulation by rare variants in the GluN2A and GluN2B agonist binding domains. *Am. J. Hum. Genet.* 99 (6), 1261–1280. doi:10.1016/j.ajhg.2016.10.002

Tamaki, K., Kamakura, M., Nakamichi, N., Taniura, H., and Yoneda, Y. (2008). Upregulation of Myo6 expression after traumatic stress in mouse hippocampus. *Neurosci. Lett.* 433 (3), 183–187. doi:10.1016/j.neulet.2007.12.062

Tyroler, H. A. (1989). Socioeconomic status in the epidemiology and treatment of hypertension. *Hypertension* 13, I94–I97. 5 Suppl. doi:10.1161/01.hyp.13.5_suppl.i94

Wang, H., and Eckel, R. H. (2009). Lipoprotein lipase: from gene to obesity. *Am. J. Physiol. Endocrinol. Metab.* 297 (2), E271–E288. doi:10.1152/ajpendo.90920.2008

Wang, K., Li, M., and Hakonarson, H. (2010). ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* 38 (16), e164. doi:10.1093/nar/gkq603

Ward, L. D., and Kellis, M. (2012). HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. *Nucleic Acids Res.* 40, D930–D934. Database issue. doi:10.1093/nar/gkr917

Watanabe, K., Taskesen, E., van Bochoven, A., and Posthuma, D. (2017). Functional mapping and annotation of genetic associations with FUMA. *Nat. Commun.* 8 (1), 1826. doi:10.1038/s41467-017-01261-5

Werme, J., van der Sluis, S., Posthuma, D., and de Leeuw, C. A. (2021). Genome-wide gene-environment interactions in neuroticism: an exploratory study across 25 environments. *Transl. Psychiatry* 11 (1), 180. doi:10.1038/s41398-021-01288-9

Willer, C. J., Li, Y., and Abecasis, G. R. (2010). METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* 26 (17), 2190–2191. doi:10.1093/bioinformatics/btq340

Woo, J. M. P., Parks, C. G., Jacobsen, S., Costenbader, K. H., and Bernatsky, S. (2022). The role of environmental exposures and gene-environment interactions in the etiology of systemic lupus erythematous. *J. Intern Med.* 291 (6), 755–778. doi:10.1111/joim.13448

Xu, Q., Yang, G. Y., Liu, N., Xu, P., Chen, Y. L., Zhou, Z., et al. (2012). P4-ATPase ATP8A2 acts in synergy with CDC50A to enhance neurite outgrowth. *FEBS Lett.* 586 (13), 1803–1812. doi:10.1016/j.febslet.2012.05.018

Yang, L., Xiong, Y., Hu, X. F., and Du, Y. H. (2015). MicroRNA-323 regulates ischemia/reperfusion injury-induced neuronal cell death by targeting BRI3. *Int. J. Clin. Exp. Pathol.* 8 (9), 10725–10733.

Yi, C. X., and Tschop, M. H. (2012). Brain-gut-adipose-tissue communication pathways at a glance. *Dis. Model Mech.* 5 (5), 583–587. doi:10.1242/dmm.009902

Zhang, Y., Yuan, S., Pu, J., Yang, L., Zhou, X., Liu, L., et al. (2018). Integrated metabolomics and proteomics analysis of Hippocampus in a rat model of depression. *Neuroscience* 371, 207–220. doi:10.1016/j.neuroscience.2017.12.001

Zhou, Y., and Danbolt, N. C. (2014). Glutamate as a neurotransmitter in the healthy brain. *J. Neural Transm. (Vienna)* 121 (8), 799–817. doi:10.1007/s00702-014-1180-8

Zou, Y., Gong, N., Cui, Y., Wang, X., Cui, A., Chen, Q., et al. (2015). Forkhead box P1 (FOXP1) transcription factor regulates hepatic glucose homeostasis. *J. Biol. Chem.* 290 (51), 30607–30615. doi:10.1074/jbc.M115.681627

CITATION
de las Fuentes L, Schwander KL, Brown MR, Bentley AR, Winkler TW, Sung YJ, Munroe PB, Miller CL, Aschard H, Aslibekyan S, Bartz TM, Bielak LF, Chai JF, Cheng C-Y, Dorajoo R, Feitosa MF, Guo X, Hartwig FP, Horimoto A, Kolčić I, Lim E, Liu Y, Manning AK, Marten J, Musani SK, Noordam R, Padmanabhan S, Rankinen T, Richard MA, Ridker PM, Smith AV, Vojinovic D, Zonderman AB, Alver M, Boissel M, Christensen K, Freedman BI, Gao C, Giulianini F, Harris SE, He M, Hsu F-C, Kühnel B, Laguzzi F, Li X, Lyytikäinen L-P, Nolte IM, Poveda A, Rauramaa R, Riaz M, Robino A, Sofer T, Takeuchi F, Tayo BO, van der Most PJ, Verweij N, Ware EB, Weiss S, Wen W, Yanek LR, Zhan Y, Amin N, Arking DE, Ballantyne C, Boerwinkle E, Brody JA, Broeckel U, Campbell A, Canouil M, Chai X, Chen Y-DI, Chen X, Chitrala KN, Concas MP, de Faire U, de Mutsert R, de Silva HJ, de Vries PS, Do A, Faul JD, Fisher V, Floyd JS, Forrester T, Friedlander Y, Girotto G, Gu CC, Hallmans G, Heikkinen S, Heng C-K, Homuth G, Hunt S, Ikram MA, Jacobs DR Jr, Kavousi M, Khor CC, Kilpeläinen TO, Koh W-P, Komulainen P, Langefeld CD, Liang J, Liu K, Liu J, Lohman K, Mägi R, Manichaikul AW, McKenzie CA, Meitinger T, Milaneschi Y, Nauck M, Nelson CP, O'Connell JR, Palmer ND, Pereira AC, Perls T, Peters A, Polašek O, Raitakari OT, Rice K, Rice TK, Rich SS, Sabanayagam C, Schreiner PJ, Shu X-O, Sidney S, Sims M, Smith JA, Starr JM, Strauch K, Tai ES, Taylor KD, Tsai MY, Uitterlinden AG, van Heemst D, Waldenberger M, Wang Y-X, Wei W-B, Wilson G, Xuan D, Yao J, Yu C, Yuan J-M, Zhao W, Becker DM, Bonnefond A, Bowden DW, Cooper RS, Deary IJ, Divers J, Esko T, Franks PW, Froguel P, Gieger C, Jonas JB, Kato N, Lakka TA, Leander K, Lehtimäki T, Magnusson PKE, North KE, Ntalla I, Penninx B, Samani NJ, Snieder H, Spedicati B, van der Harst P, Völzke H, Wagenknecht LE, Weir DR, Wojczynski MK, Wu T, Zheng W, Zhu X, Bouchard C, Chasman DI, Evans MK, Fox ER, Gudnason V, Hayward C, Horta BL, Kardia SLR, Krieger JE, Mook-Kanamori DO, Peyser PA, Province MM, Psaty BM, Rudan I, Sim X, Smith BH, van Dam RM, van Duijn CM, Wong TY, Arnett DK, Rao DC, Gauderman J, Liu C-T, Morrison AC, Rotter JI and Fornage M (2023), Gene-educational attainment interactions in a multi-population genome-wide meta-analysis identify novel lipid loci.
Front. Genet. 14:1235337.
doi: 10.3389/fgene.2023.1235337

Copyright © 2023 de las Fuentes, Schwander, Brown, Bentley, Winkler, Sung, Munroe, Miller, Aschard, Aslibekyan, Bartz, Bielak, Chai, Cheng, Dorajoo, Feitosa, Guo, Hartwig, Horimoto, Kolčić, Lim, Liu, Manning, Marten, Musani, Noordam, Padmanabhan, Rankinen, Richard, Ridker, Smith, Vojinovic, Zonderman, Alver, Boissel, Christensen, Freedman, Gao, Giulianini, Harris, He, Hsu, Kühnel, Laguzzi, Li, Lyytikäinen, Nolte, Poveda, Rauramaa, Riaz, Robino, Sofer, Takeuchi, Tayo, van der Most, Verweij, Ware, Weiss, Wen, Yanek, Zhan, Amin, Arking, Ballantyne, Boerwinkle, Brody, Broeckel, Campbell, Canouil, Chai, Ida Chen, Chen, Chitrala, Concas, de Faire, de Mutsert, de Silva, de Vries, Do, Faul, Fisher, Floyd, Forrester, Friedlander, Girotto, Gu, Hallmans, Heikkinen, Heng, Homuth, Hunt, Ikram, Jacobs, Kavousi, Khor, Kilpeläinen, Koh, Komulainen, Langefeld, Liang, Liu, Liu, Lohman, Mägi, Manichaikul, McKenzie, Meitinger, Milaneschi, Nauck, Nelson, O'Connell, Palmer, Pereira, Perls, Peters, Polašek, Raitakari, Rice, Rice, Rich, Sabanayagam, Schreiner, Shu, Sidney, Sims, Smith, Starr, Strauch, Tai, Taylor, Tsai, Uitterlinden, van Heemst, Waldenberger, Wang, Wei, Wilson, Xuan, Yao, Yu, Yuan, Zhao, Becker, Bonnefond, Bowden, Cooper, Deary, Divers, Esko, Franks, Froguel, Gieger, Jonas, Kato, Lakka, Leander, Lehtimäki, Magnusson, North, Ntalla, Penninx, Samani, Snieder, Spedicati, van der Harst, Völzke, Wagenknecht, Weir, Wojczynski, Wu, Zheng, Zhu, Bouchard, Chasman, Evans, Fox, Gudnason, Hayward, Horta, Kardia, Krieger, Mook-Kanamori, Peyser, Province, Psaty, Rudan, Sim, Smith, van Dam, van Duijn, Wong, Arnett, Rao, Gauderman, Liu, Morrison, Rotter and Fornage. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Supplementary Material

Table of Contents

Supplementary Notes

Supplementary Methods

Stage 1 Study Descriptions

Stage 2 Study Descriptions

Stage 1 Study Acknowledgments

Stage 2 Study Acknowledgments

Supplementary Figures

Figure S1: Flow Chart of Study Design

Figure S2A: QQ Plots of African-Ancestry Cohorts.

Figure S2B: QQ Plots of Asian-Ancestry Cohorts.

Figure S2C: QQ Plots of Brazilian-Ancestry Cohorts.

Figure S2D: QQ Plots of European-Ancestry Cohorts.

Figure S2E: QQ Plots of Hispanic-Ancestry Cohorts.

Figure S3A: Manhattan plots of the combined analyses of Stages 1 and 2 using the 2DF for LDL and GradCol by ancestry.

Figure S3B: Manhattan plots of the combined analyses of Stages 1 and 2 using the 2DF for LDL and SomeCol by ancestry.

Figure S3C: Manhattan plots of the combined analyses of Stages 1 and 2 using the 2DF for HDL and GradCol by ancestry.

Figure S3D: Manhattan plots of the combined analyses of Stages 1 and 2 using the 2DF for HDL and SomeCol by ancestry.

Figure S3E: Manhattan plots of the combined analyses of Stages 1 and 2 using the 2DF for TG and GradCol by ancestry.

Figure S3F: Manhattan plots of the combined analyses of Stages 1 and 2 using the 2DF for TG and SomeCol by ancestry.

Figure S4: LocusZoom plots for 18 novel loci (Table 1).

Supplementary Methods

Human Subjects Research Statement

Informed consent was obtained from all subjects involved in the study by their individual cohorts. The study was conducted according to the guidelines of the Declaration of Helsinki, and approved by the Institutional Review Board (or Ethics Committee) of the institutions named in the Supplementary Information.

Genotype data

Imputation software packages used included MACH,[1] Minimac,[2] IMPUTE2,[3] or BEAGLE[4].

Quality control (QC)

Each participating cohort in Stage 1 excluded variants with minor allele frequency (MAF) < 1% or mapping to sex chromosomes or mitochondria. We performed extensive QC using the R package EasyQC[5] for all cohort-specific and meta-analysis results. For Stages 1 and 2, we excluded all variants with imputation quality measure < 0.5. In addition, to remove unstable study-specific results that reflected small sample size, low minor allele count (MAC), or low imputation quality, we excluded variants for which the "approximate DF" (defined as the minimum of [MAC0, MAC1] x imputation quality) < 20, where MAC0 and MAC1 are the MAC in the two educational attainment strata per exposure variable (either GradCol = 0 and GradCol = 1 or SomeCol = 0 and SomeCol = 1). The allele frequencies provided by each cohort were compared against those from the relevant ancestry-specific 1000 Genomes reference panel. Marker names were harmonized to ensure consistency across cohorts. In addition, we visually compared summary statistics (e.g., mean, median, standard deviation, interquartile range) of all effect estimates, standard errors (SEs), and p-values. We examined SE-N plots[5] and quantile-quantile (QQ) plots to reveal issues with trait transformation or other analytical problems. Any problems encountered during QC steps were resolved through communication with the analysts from the participating studies. More detailed information about the QC steps, including major QC problems encountered and how they were resolved, are described elsewhere.[6]

Characterization of functional roles

Loci were characterized as known (previously reported)[7-32] or novel. A suite of tools implemented in FUMA GWAS[33] (version 1.3.5) were used to identify functional roles for SNVs: RegulomeDB score,[34] which reflects a summary of annotations with known and predicted regulatory elements such as DNase hypersensitivity, binding sites of transcription factors, and promoter regions; Combined Annotation Dependent Depletion (CADD)[35] scores, which predict deleteriousness of variants; the 15-core chromatin state (ChromHMM)[36] for 129 epigenomes (labeled E001-E129) to identify histone modifications consistent with epigenetic regulation of gene expression; expression quantitative trait loci (eQTLs) using the GTEx V8 database;[37] and tissue-specific gene expression data (GTEx V8 dataset, 53 tissue types) with significance determined as a Benjamini-Hochberg false discovery rate (FDR) < 0.05.

Druggability analysis

First, the Drug-Gene Interaction database (DGIdb; v4.2.0) was used to query high or medium priority educational attainment-lipid interacting genes to determine the potentially druggability of the candidate gene targets. Genes were annotated for implicated pathways and functions using the Kyoto Encyclopedia of Genes and Genomes (KEGG) database. We annotated the druggability target categories and queried all interacting drugs reported in 43 databases (BaderLabGenes, CarisMolecularIntelligence, dGene, FoundationOneGenes, GO, HingoraniCasas, HopkinsGroom,

HumanProteinAtlas, IDG, MskImpact, Oncomine, Pharos, RussLampel, Tempus, CGI, CIViC, COSMIC, CancerCommons, ChemblDrugs, ChemblInteractions, ClearityFoundationBiomarkers, ClearityFoundationClinicalTrial, DTC, DoCM, DrugBank, Ensembl, Entrez, FDA, GuideToPharmacology, JACX-CKB, MyCancerGenome, MyCancerGenomeClinicalTrial, NCI, OncoKB, PharmGKB, TALC, TEND, TTD, TdgClinicalTrial, Wikidata). Protein targets were queried for available active ligands in ChEMBL. Gene targets were queried in the druggable genome using the most recent druggable genome list established from the NIH Illuminating the Druggable Genome Project (https://github.com/druggablegenome/IDGTargets) available through the Pharos web platform. Targets were also queried for FDA-approved drugs, late-stage clinical trials, and disease indications in the DrugBank, ChEMBL, ClinicalTrials.gov databases and results provided for the top MESH and DrugBank indications and clinical trials.

References for Supplementary Methods

1. Li Y, Willer CJ, Ding J, Scheet P, Abecasis GR. MaCH: using sequence and genotype data to estimate haplotypes and unobserved genotypes. *Genet Epidemiol* 2010; **34**(8)**:** 816-834.

2. Howie B, Fuchsberger C, Stephens M, Marchini J, Abecasis GR. Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. *Nat Genet* 2012; **44**(8)**:** 955-959.

3. Howie BN, Donnelly P, Marchini J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet* 2009; **5**(6)**:** e1000529.

4. Browning BL, Browning SR. A unified approach to genotype imputation and haplotype-phase inference for large data sets of trios and unrelated individuals. *Am J Hum Genet* 2009; **84**(2)**:** 210-223.

5. Winkler TW, Day FR, Croteau-Chonka DC, Wood AR, Locke AE, Magi R *et al.* Quality control and conduct of genome-wide association meta-analyses. *Nat Protoc* 2014; **9**(5)**:** 1192-1212.

6. de las Fuentes L, Sung YJ, Noordam R, Winkler T, Feitosa MF, Schwander K *et al.* Gene-educational attainment interactions in a multi-ancestry genome-wide meta-analysis identify novel blood pressure loci. *Mol Psychiatry* 2020.

7. Aulchenko YS, Ripatti S, Lindqvist I, Boomsma D, Heid IM, Pramstaller PP *et al.* Loci influencing lipid levels and coronary heart disease risk in 16 European population cohorts. *Nat Genet* 2009; **41**(1)**:** 47-55.

8. Coram MA, Duan Q, Hoffmann TJ, Thornton T, Knowles JW, Johnson NA *et al.* Genome-wide characterization of shared and distinct genetic components that influence blood lipid levels in ethnically diverse human populations. *Am J Hum Genet* 2013; **92**(6)**:** 904-916.

9. Do R, Willer CJ, Schmidt EM, Sengupta S, Gao C, Peloso GM *et al.* Common variants associated with plasma triglycerides and risk for coronary artery disease. *Nat Genet* 2013; **45**(11)**:** 1345-1352.

10. Heid IM, Boes E, Muller M, Kollerits B, Lamina C, Coassin S *et al.* Genome-wide association analysis of high-density lipoprotein cholesterol in the population-based KORA study sheds new light on intergenic regions. *Circ Cardiovasc Genet* 2008; **1**(1)**:** 10-20.

11. Holmen OL, Zhang H, Fan Y, Hovelson DH, Schmidt EM, Zhou W *et al.* Systematic evaluation of coding variation identifies a candidate causal variant in TM6SF2 influencing total cholesterol and myocardial infarction risk. *Nat Genet* 2014; **46**(4)**:** 345-351.

12. Kathiresan S, Melander O, Guiducci C, Surti A, Burtt NP, Rieder MJ *et al.* Six new loci associated with blood low-density lipoprotein cholesterol, high-density lipoprotein cholesterol or triglycerides in humans. *Nat Genet* 2008; **40**(2)**:** 189-197.

13. Kathiresan S, Willer CJ, Peloso GM, Demissie S, Musunuru K, Schadt EE *et al.* Common variants at 30 loci contribute to polygenic dyslipidemia. *Nat Genet* 2009; **41**(1)**:** 56-65.

14. Kim YJ, Go MJ, Hu C, Hong CB, Kim YK, Lee JY *et al.* Large-scale genome-wide association studies in East Asians identify new genetic loci influencing metabolic traits. *Nat Genet* 2011; **43**(10)**:** 990-995.

15. Lanktree MB, Elbers CC, Li Y, Zhang G, Duan Q, Karczewski KJ *et al.* Genetic meta-analysis of 15,901 African Americans identifies variation in EXOC3L1 is associated with HDL concentration. *J Lipid Res* 2015; **56**(9)**:** 1781-1786.

16. Peloso GM, Auer PL, Bis JC, Voorman A, Morrison AC, Stitziel NO *et al.* 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* 2014; **94**(2)**:** 223-232.

17. Ridker PM, Pare G, Parker AN, Zee RY, Miletich JP, Chasman DI. Polymorphism in the CETP gene region, HDL cholesterol, and risk of future myocardial infarction: Genomewide analysis among 18 245 initially healthy women from the Women's Genome Health Study. *Circ Cardiovasc Genet* 2009; **2**(1)**:** 26-33.

18. Sabatti C, Service SK, Hartikainen AL, Pouta A, Ripatti S, Brodsky J *et al.* Genome-wide association analysis of metabolic traits in a birth cohort from a founder population. *Nat Genet* 2009; **41**(1)**:** 35-46.

19. Sandhu MS, Waterworth DM, Debenham SL, Wheeler E, Papadakis K, Zhao JH *et al.* LDL-cholesterol concentrations: a genome-wide association study. *Lancet* 2008; **371**(9611)**:** 483-491.

20. Saxena R, Voight BF, Lyssenko V, Burtt NP, de Bakker PI, Chen H *et al.* Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels. *Science* 2007; **316**(5829)**:** 1331-1336.

21. Spracklen CN, Chen P, Kim YJ, Wang X, Cai H, Li S *et al.* Association analyses of East Asian individuals and trans-ancestry analyses with European individuals reveal new loci associated with cholesterol and triglyceride levels. *Hum Mol Genet* 2017; **26**(9)**:** 1770-1784.

22. Surakka I, Horikoshi M, Magi R, Sarin AP, Mahajan A, Lagou V *et al.* The impact of low-frequency and rare variants on lipid levels. *Nat Genet* 2015; **47**(6)**:** 589-597.

23. Tan A, Sun J, Xia N, Qin X, Hu Y, Zhang S *et al.* A genome-wide association and gene-environment interaction study for serum triglycerides levels in a healthy Chinese male population. *Hum Mol Genet* 2012; **21**(7)**:** 1658-1664.

24. Teslovich TM, Musunuru K, Smith AV, Edmondson AC, Stylianou IM, Koseki M *et al.* Biological, clinical and population relevance of 95 loci for blood lipids. *Nature* 2010; **466**(7307)**:** 707-713.

25. van Leeuwen EM, Karssen LC, Deelen J, Isaacs A, Medina-Gomez C, Mbarek H *et al.* Genome of The Netherlands population-specific imputations identify an ABCA6 variant associated with cholesterol levels. *Nat Commun* 2015; **6:** 6065.

26. Waterworth DM, Ricketts SL, Song K, Chen L, Zhao JH, Ripatti S *et al.* Genetic variants influencing circulating lipid levels and risk of coronary artery disease. *Arterioscler Thromb Vasc Biol* 2010; **30**(11)**:** 2264-2276.

27. Weissglas-Volkov D, Aguilar-Salinas CA, Nikkola E, Deere KA, Cruz-Bautista I, Arellano-Campos O *et al.* Genomic study in Mexicans identifies a new locus for triglycerides and refines European lipid loci. *J Med Genet* 2013; **50**(5)**:** 298-308.

28. Willer CJ, Sanna S, Jackson AU, Scuteri A, Bonnycastle LL, Clarke R *et al.* Newly identified loci that influence lipid concentrations and risk of coronary artery disease. *Nat Genet* 2008; **40**(2)**:** 161-169.

29. Willer CJ, Schmidt EM, Sengupta S, Peloso GM, Gustafsson S, Kanoni S *et al.* Discovery and refinement of loci associated with lipid levels. *Nat Genet* 2013; **45**(11)**:** 1274-1283.

30. Zubair N, Graff M, Luis Ambite J, Bush WS, Kichaev G, Lu Y *et al.* Fine-mapping of lipid regions in global populations discovers ethnic-specific signals and refines previously identified lipid loci. *Hum Mol Genet* 2016; **25**(24)**:** 5500-5512.

31. Klarin D, Damrauer SM, Cho K, Sun YV, Teslovich TM, Honerlaw J *et al.* Genetics of blood lipids among ~300,000 multi-ethnic participants of the Million Veteran Program. *Nat Genet* 2018; **50**(11)**:** 1514-1523.

32. Choudhury A, Brandenburg JT, Chikowore T, Sengupta D, Boua PR, Crowther NJ *et al.* Meta-analysis of sub-Saharan African studies provides insights into genetic architecture of lipid traits. *Nat Commun* 2022; **13**(1)**:** 2578.

33. Watanabe K, Taskesen E, van Bochoven A, Posthuma D. Functional mapping and annotation of genetic associations with FUMA. *Nat Commun* 2017; **8**(1)**:** 1826.

34. Boyle AP, Hong EL, Hariharan M, Cheng Y, Schaub MA, Kasowski M *et al.* Annotation of functional variation in personal genomes using RegulomeDB. *Genome Res* 2012; **22**(9)**:** 1790-1797.

35. Kircher M, Witten DM, Jain P, O'Roak BJ, Cooper GM, Shendure J. A general framework for estimating the relative pathogenicity of human genetic variants. *Nat Genet* 2014; **46**(3)**:** 310-315.

36. Roadmap Epigenomics Consortium, Kundaje A, Meuleman W, Ernst J, Bilenky M, Yen A *et al.* Integrative analysis of 111 reference human epigenomes. *Nature* 2015; **518**(7539)**:** 317-330.

37. GTEx Consortium. Genetic effects on gene expression across human tissues. *Nature* 2017; **550**(7675)**:** 204-213.

STAGE 1 STUDY DESCRIPTIONS

Brief descriptions are provided below for each of stage 1 studies, some of which are based outside the United States:

AGES (Age Gene/Environment Susceptibility Reykjavik Study): The AGES Reykjavik study originally comprised a random sample of 30,795 men and women born in 1907-1935 and living in Reykjavik in 1967. A total of 19,381 people attended, resulting in a 71% recruitment rate. The study sample was divided into six groups by birth year and birth date within month. One group was designated for longitudinal follow up and was examined in all stages; another was designated as a control group and was not included in examinations until 1991. Other groups were invited to participate in specific stages of the study. Between 2002 and 2006, the AGESReykjavik study re-examined 5,764 survivors of the original cohort who had participated before in the Reykjavik Study.

ARIC (Atherosclerosis Risk in Communities): The ARIC study is a population-based prospective cohort study of cardiovascular disease sponsored by the National Heart, Lung, and Blood Institute (NHLBI). ARIC included 15,792 individuals, predominantly European American and African American, aged 45-64 years at baseline (1987-89), chosen by probability sampling from four US communities. Cohort members completed three additional triennial follow-up examinations, a fifth exam in 2011-2013, a sixth exam in 2016-2017, and a seventh exam in 2018-2019. The ARIC study has been described in detail previously (PMC8667593; Wright JD, Folsom AR, Coresh J, et al. The ARIC (Atherosclerosis Risk In Communities) Study: JACC Focus Seminar 3/8. J Am Coll Cardiol. 2021 Jun 15;77(23):2939-2959).

Baependi Heart Study (Brazil): The Baependi Heart Study, is an ongoing family-based cohort conducted in a rural town of the state of Minas Gerais. The study has enrolled approximate 2,200 individuals (over 10% of the town's adult population) and 10-year follow up period of longitudinal data. Briefly, probands were selected at random across 11 out of the 12 census districts in Baependi. After enrolment, the proband's first-degree (parents, siblings, and offspring), second-degree (half-siblings, grandparents/grandchildren, uncles/aunts, nephews/nieces, and double cousins), and third-degree (first cousins, great uncles/aunts, and great nephews/nieces) relatives, and his/her respective spouse's relatives resident both within Baependi (municipal and rural area) and surrounding towns were invited to participate. Only individuals age 18 and older were eligible to participate in the study. The study is conducted from a clinic/office in an easily accessible sector of the town, where the questionnaires were completed. A broad range of phenotypes ranging from cardiovascular, neurocognitive, psychiatric, imaging, physiologic and several layers of endophenotypes like metabolomics and lipidomics have been collected throughout the years Details about follow-up visits and available data can be found in the cohort profile paper (PMID: 18430212). DNA samples were genotyped using the Affymetrix 6.0 genechip. After quality control, the data were prephased using SHAPEIT and imputed using IMPUTE2 based on 1000 Genomes haplotypes.

CARDIA (Coronary Artery Risk Development in Young Adults): CARDIA is a prospective multicenter study with 5,115 Caucasian and African American participants 18-30 years of age at baseline, recruited from four centers in 1985-1986. The recruitment was done from the total community in Birmingham, AL; from selected census tracts in Chicago, IL and Minneapolis, MN; and from the Kaiser Permanente health plan membership in Oakland, CA. The details of the study design for the CARDIA study have been previously published.[1] Nine examinations have been completed since initiation of the study, respectively in the years 0, 2, 5, 7, 10, 15, 20, 25, and 30. Written informed

consent was obtained from participants at each examination and all study protocols were approved by the institutional review boards of the participating institutions.

Age and race were self-reported using standardized questionnaires, as were use of cholesterol-lowering medication, and smoking status (current, former, or never). All participants were asked to fast for 12 hours before each clinic visit. Lipid measures were performed on plasma blood samples drawn from the antecubital vein and stored at −70°C until analyzed. Plasma total cholesterol, HDL-cholesterol, and triglyceride concentrations were measured using enzymatic methods;[2] HDL-cholesterol concentrations were measured after dextran-sulfate-magnesium precipitation of other lipoproteins.[3] LDL-cholesterol concentrations were estimated with the Friedewald equation for individuals with fasting triglyceride values less than 400 mg/dL.[4] The test-retest correlations for total cholesterol, HDL-cholesterol, LDL-cholesterol, and triglycerides were 0.98 to 0.99.[5]

1. Friedman GD, Cutter GR, Donahue RP, Hughes GH, Hully SB, Jacobs DR Jr., Liu K, Savage PJ. CARDIA: study design, recruitment, and some characteristics of the examined subjects. J Clin Epidemiol. 1988;41:1105–1116.

2. Warnick GR. Enzymatic methods for quantification of lipoprotein lipids. Methods Enzymol. 1986;129:101–123.

3. Warnick GR, Benderson J, Albers JJ. Dextran sulfate-Mg+2 precipitation procedure for quantitation of high-density lipoprotein cholesterol. Clin Chem. 1982;28:1379–1388.

4. Friedewald WT, Levy RI, Fredrickson DS. Estimation of the concentration of low-density lipoprotein cholesterol in plasma, without use of the preparative ultracentrifuge. Clin Chem. 1972;18:499–502.

5. Gross M, Steffes M, Jacobs DR Jr., Yu X, Lewis L, Lewis CE, Loria CM. Plasma F2-isoprostanes and coronary artery calcification: the CARDIA Study. Clin Chem. 2005;51:125–131.

CHS (Cardiovascular Health Study): CHS is a population-based cohort study of risk factors for cardiovascular disease in adults 65 years of age or older conducted across four field centers (1). The original predominantly European ancestry cohort of 5,201 persons was recruited in 1989-1990 from random samples of the Medicare eligibility lists and an additional predominately African-American cohort of 687 persons was enrolled in 1992-93 for a total sample of 5,888. Blood samples were drawn from all participants at their baseline examination and DNA was subsequently extracted from available samples. European ancestry participants were excluded from the GWAS study sample due to prevalent coronary heart disease, congestive heart failure, peripheral vascular disease, valvular heart disease, stroke, or transient ischemic attack at baseline. After QC, genotyping was successful for 3271 European ancestry and 823 African-American participants. CHS was approved by institutional review committees at each site and individuals in the present analysis gave informed consent including consent to use of genetic information for the study of cardiovascular disease.

1. Fried LP, Borhani NO, Enright P, Furberg CD, Gardin JM, Kronmal RA, et al. The Cardiovascular Health Study: design and rationale. Ann Epidemiol 1991; 1:263-76.

CROATIA-Korcula: The CROATIA-Korcula study is a family-based, cross-sectional study in the isolated island of Korcula that included 965 examinees aged 18-95. Blood samples were collected in 2007 along with many clinical and biochemical measures and lifestyle and health questionnaires.

CROATIA-Vis: The CROATIA-Vis study is a family-based, cross-sectional study in the isolated island of Vis that included 1,056 examinees aged 8-93. Blood samples were collected in 2003 and 2004 along with many clinical and biochemical measures and lifestyle and health questionnaires.

ERF (Erasmus Rucphen Family study): Erasmus Rucphen Family is a family based study that includes inhabitants of a genetically isolated community in the South-West of the Netherlands, studied as part of the Genetic Research in Isolated Population (GRIP) program.[1,2] The goal of the study is to identify the risk factors in the development of complex disorders. Study population includes approximately 3,000 individuals who are living descendants of 22 couples who lived in the isolate between 1850 and 1900 and had at least six children baptized in the community church. All data were collected between 2002 and 2005. All participants gave informed consent, and the Medical Ethics Committee of the Erasmus University Medical Centre approved the study.

1. Pardo, Luba M., et al. "The effect of genetic drift in a young genetically isolated population." Annals of human genetics 69.3 (2005): 288-295.

2. Aulchenko, Yurii S., et al. "Linkage disequilibrium in young genetically isolated Dutch population." European Journal of Human Genetics 12.7 (2004): 527.

FamHS (Family Heart Study): The NHLBI FamHS study design, collection of phenotypes and covariates as well as clinical examination have been previously described (https://dsgweb.wustl.edu/fhscc/; PMID: 8651220). In brief, the FamHS recruited 1,200 families (approximately 6,000 individuals), half randomly sampled, and half selected because of an excess of coronary heart disease (CHD) or risk factor abnormalities as compared with age- and sex-specific population rates. The participants were sampled from four population-based parent studies: the Framingham Heart Study, the Utah Family Tree Study, and two centers for the Atherosclerosis Risk in Communities study (ARIC: Minneapolis, and Forsyth County, NC). These individuals attended a clinic exam (1994-1996) and a broad range of phenotypes were assessed in the general domains of CHD, atherosclerosis, cardiac and vascular function, inflammation and hemostasis, lipids and lipoproteins, blood pressure, diabetes and insulin resistance, pulmonary function, diet, education, socioeconomic status, habitual behavior, physical activity, anthropometry, medical history and medication use. Approximately 8 years later, study participants belonging to the largest pedigrees were invited for a second clinical exam (2002-04). The most important CHD risk factors were measured again, including lipids, parameters of glucose metabolism, blood pressure, anthropometry, and several biochemical and hematologic markers. In addition, a computed tomography examination provided measures of coronary and aortic calcification, and abdominal and liver fat burden. Medical history and medication use was updated. A total of 2,756 European ancestry subjects in 510 extended random and high CHD risk families were studied. Also, 633 African ancestry subjects were recruited at ARIC field center at the University of Alabama in Birmingham. Informed consent was obtained from all participants.

FHS (Framingham Heart Study): FHS began in 1948 with the recruitment of an original cohort of 5,209 men and women (mean age 44 years; 55 percent women). In 1971 a second generation of study participants was enrolled; this cohort (mean age 37 years; 52% women) consisted of 5,124 children and spouses of children of the original cohort. A third generation cohort of 4,095 children of offspring cohort participants (mean age 40 years; 53 percent women) was enrolled in 2002-2005 and are seen every 4 to 8 years. Details of study designs for the three cohorts are summarized elsewhere. At each clinic visit, a medical history was obtained with a focus on cardiovascular content, and participants underwent a physical examination including measurement of height and weight from which BMI was

calculated. For this study, lipid measurements were used from the first exam of the 2nd generation (1971-1975) and the 3rd generation (2002-2005) cohorts. Fasting levels of total cholesterol, high-density lipoprotein cholesterol, and triglycerides were measured using standard enzymatic methods in accordance with LRC protocols. LDL cholesterol was calculated using the Friedewald formula.

GENOA (Genetic Epidemiology Network of Arteriopathy): GENOA is one of four networks in the NHLBI Family-Blood Pressure Program (FBPP). [The FBPP Investigators, 2002; Daniels et al., 2004] GENOA's long-term objective is to elucidate the genetics of target organ complications of hypertension, including both atherosclerotic and arteriosclerotic complications involving the heart, brain, kidneys, and peripheral arteries. The longitudinal GENOA Study recruited European-American and African-American sibships with at least 2 individuals with clinically diagnosed essential hypertension before age 60 years. All other members of the sibship were invited to participate regardless of their hypertension status. Participants were diagnosed with hypertension if they had either 1) a previous clinical diagnosis of hypertension by a physician with current anti-hypertensive treatment, or 2) an average systolic blood pressure ≥ 140 mm Hg or diastolic blood pressure ≥ 90 mm Hg based on the second and third readings at the time of their clinic visit. Exclusion criteria were secondary hypertension, alcoholism or drug abuse, pregnancy, insulin-dependent diabetes mellitus, or active malignancy. During the first exam (1995-2000), 1,583 European Americans from Rochester, MN and 1,854 African Americans from Jackson, MS were examined. Between 2000 and 2005, 1,241 of the European Americans and 1,482 of the African Americans returned for a second examination. Because African-American probands for GENOA were recruited through the Atherosclerosis Risk in Communities (ARIC) Jackson field center participants, we excluded ARIC participants from analyses.

1. The FBPP Investigators. Multi-center genetic study of hypertension: The Family Blood Pressure Program (FBPP). Hypertension 2002;39:3-9. PubMed PMID: 11799070.

2. Daniels PR, Kardia SL, Hanis CL, Brown CA, Hutchinson R, Boerwinkle E, Turner ST; Genetic Epidemiology Network of Arteriopathy study. Familial aggregation of hypertension treatment and control in the Genetic Epidemiology Network of Arteriopathy (GENOA) study. Am J Med. 2004;116:676-81. PubMed PMID: 15121494.

GS:SFHS (Generation Scotland: Scottish Family Health Study): The GS:SFHS (www.generationscotland.org) is a family-based genetic epidemiology cohort with DNA, other biological samples (serum, urine and cryopreserved whole blood) and socio-demographic and clinical data from approximately 24,000 volunteers, aged 18-98 years, in ~7,000 family groups. An important feature of GS:SFHS is the breadth of phenotype information, including detailed data on cognitive function, personality traits and mental health. Although data collection was cross-sectional, GS:SFHS becomes a longitudinal cohort as a result of the ability to link to routine NHS data, using the community health index (CHI) number.

GOLDN (Genetics of Diet and Lipid Lowering Network): GOLDN is a multi-center family pharmacogenetic study that is investigating gene- environment interactions on lipid profiles. 1,200 subjects in extended pedigrees were measured before and after two environmental exposures: 1) a dietary fat challenge to assess genetic regulators of fat uptake and clearance and 2) a 3-week clinical trial of fenofibrate to assess pharmacogenetic influences on response to treatment. The goals of the study are to identify and characterize genetic loci that predict the lipid profile treatment responses. https://dsgweb.wustl.edu/PROJECTS/MP5.html

HANDLS (Healthy Aging in Neighborhoods of Diversity across the Life Span): HANDLS is a community-based, longitudinal epidemiologic study examining the influences of race and socioeconomic status (SES) on the development of age-related health disparities among a sample of socioeconomically diverse African Americans and whites. This unique study will assess over a 20-year period physical parameters and also evaluate genetic, biologic, demographic, and psychosocial, parameters of African American and white participants in higher and lower SES to understand the driving factors behind persistent black-white health disparities in overall longevity, cardiovascular disease, and cognitive decline. The study recruited 3,722 participants from Baltimore, MD with a mean age of 47.7 years, 2,200 African Americans and 1,522 whites, with 41% reporting household incomes below the 125% poverty delimiter.

Genotyping was done on a subset of self-reporting African American participants by the Laboratory of Neurogenetics, National Institute on Aging, National Institutes of Health (NIH). A larger genotyping effort included a small subset of self-reporting European ancestry samples. This research was supported by the Intramural Research Program of the NIH, NIA and the National Center on Minority Health and Health Disparities.

Health ABC (Health, Aging, and Body Composition): Cohort description: The Health ABC study is a prospective cohort study investigating the associations between body composition, weight-related health conditions, and incident functional limitation in older adults. Health ABC enrolled well-functioning, community-dwelling black (n=1281) and white (n=1794) men and women aged 70-79 years between April 1997 and June 1998. Participants were recruited from a random sample of white and all black Medicare eligible residents in the Pittsburgh, PA, and Memphis, TN, metropolitan areas. Participants have undergone annual exams and semi-annual phone interviews. The current study sample consists of 1559 white participants who attended the second exam in 1998-1999 with available genotyping data.

Genotyping: Genotyping was performed by the Center for Inherited Disease Research (CIDR) using the Illumina Human1M-Duo BeadChip system. Samples were excluded from the dataset for the reasons of sample failure, genotypic sex mismatch, and first-degree relative of an included individual based on genotype data. Genotyping was successful in 1663 Caucasians. Analysis was restricted to SNPs with minor allele frequency ≥ 1%, call rate ≥97% and HWE p≥10-6. Genotypes were available on 914,263 high quality SNPs for imputation based on the HapMap CEU (release 22, build 36) using the MACH software (version 1.0.16). A total of 2,543,888 imputed SNPs were analyzed for association with vitamin D levels.

Association analysis: Linear regression models were used to generate cohort-specific residuals of naturally log transformed vitamin D levels adjusted for age, sex, BMI and season defined as summer (June-August), fall (September-November), winter (December to February) and spring (March to May) standardized to have mean 0 and variance of 1. Association between the additively coded SNP genotypes and the vitamin D residuals standardized was assessed using linear regression models. For imputed SNPs, expected number of minor alleles (i.e. dosage) was used in assessing association with the vitamin D residuals.

HERITAGE (Health, Risk Factors, Exercise Training and Genetics): The HERITAGE is the only known family-based study of exercise intervention to evaluate the role of genes and sequence variants involved in the response to a physically active lifestyle. The current study is based on the data collected at baseline of the study from 99 White families (244 males, 255 females). All subjects were required

to be sedentary and free of chronic diseases at baseline. There are over 18 trait domains (e.g. dietary, lipids and lipoproteins, glucose and insulin metabolism [fasting and IVGTT], steroids, body composition and body fat distribution, cardiorespiratory fitness), for a grand total of over one thousand variables. Moreover, most of the outcome traits were measured twice on two separate days both at baseline and after exercise training was completed. Marker data include a genome-wide linkage scan and GWAS, in addition to a large number of candidate genes.

HyperGEN (Hypertension Genetic Epidemiology Network): HyperGEN is a family-based study that looks at the genetic causes of hypertension and related conditions in EA and AA subjects. HyperGEN recruited hypertensive sibships, along with their normotensive adult offspring, and an age-matched random sample. HyperGEN has collected data on 2,471 Caucasian-American subjects and 2,300 African-American subjects, from five field centers in Alabama, Massachusetts, Minnesota, North Carolina, and Utah.

JHS (Jackson Heart Study): The Jackson Heart Study is a longitudinal, community-based observational cohort study investigating the role of environmental and genetic factors in the development of cardiovascular disease in African Americans. Between 2000 and 2004, a total of 5,306 participants were recruited from a tri-county area (Hinds, Madison, and Rankin Counties) that encompasses Jackson, MS. Details of the design and recruitment for the Jackson Heart Study cohort has been previously published.1-3 Briefly, approximately 30% of participants were former members of the Atherosclerosis Risk in Communities (ARIC) study. The remainder were recruited by either 1) random selection from the Accudata list, 2) commercial listing, 3) a constrained volunteer sample, in which recruitment was distributed among defined demographic cells in proportions designed to mirror those in the overall population, or through the Jackson Heart Study Family Study.

1. Wyatt SB, Diekelmann N, Henderson F, Andrew ME, Billingsley G, Felder SH et al. A community-driven model of research participation: the Jackson Heart Study Participant Recruitment and Retention Study. Ethn Dis 2003; 13(4):438-455.

2. Taylor HA, Jr., Wilson JG, Jones DW, et al. Toward resolution of cardiovascular health disparities in African Americans: design and methods of the Jackson Heart Study. Ethn Dis 2005; 15:S6-17.

3. Fuqua SR, Wyatt SB, Andrew ME, et al. Recruiting African-American research participation in the Jackson Heart Study: methods, response rates, and sample description. Ethn Dis 2005; 15:S6-29.

MESA (Multi-Ethnic Study of Atherosclerosis): The Multi-Ethnic Study of Atherosclerosis (MESA) is a study of the characteristics of subclinical cardiovascular disease and the risk factors that predict progression to clinically overt cardiovascular disease or progression of the subclinical disease. MESA consisted of a diverse, population-based sample of an initial 6,814 asymptomatic men and women aged 45-84. 38 percent of the recruited participants were white, 28 percent African American, 22 percent Hispanic, and 12 percent Asian, predominantly of Chinese descent. Participants were recruited from six field centers across the United States: Wake Forest University, Columbia University, Johns Hopkins University, University of Minnesota, Northwestern University and University of California - Los Angeles. Participants are being followed for identification and characterization of cardiovascular disease events, including acute myocardial infarction and other forms of coronary heart disease (CHD), stroke, and congestive heart failure; for cardiovascular disease interventions; and for mortality. The first examination took place over two years, from July 2000 - July 2002. It was followed

by four examination periods that were 17-20 months in length. Participants have been contacted every 9 to 12 months throughout the study to assess clinical morbidity and mortality.

1. Bild DE, Bluemke DA, Burke GL, Detrano R, Diez Roux AV, Folsom AR, Greenland P, Jacob DR Jr, Kronmal R, Liu K, Nelson JC, O'Leary D, Saad MF, Shea S, Szklo M, Tracy RP. Multi-ethnic study of atherosclerosis: objectives and design. Am J Epidemiol. 2002 Nov 1;156(9):871-81. PubMed PMID: 12397006.

NEO (The Netherlands Epidemiology of Obesity study): The NEO was designed for extensive phenotyping to investigate pathways that lead to obesity-related diseases. The NEO study is a population-based, prospective cohort study that includes 6,671 individuals aged 45–65 years, with an oversampling of individuals with overweight or obesity. At baseline, information on demography, lifestyle, and medical history have been collected by questionnaires. In addition, samples of 24-h urine, fasting and postprandial blood plasma and serum, and DNA were collected. Genotyping was performed using the Illumina HumanCoreExome chip, which was subsequently imputed to the 1000 genome reference panel. Participants underwent an extensive physical examination, including anthropometry, electrocardiography, spirometry, and measurement of the carotid artery intima-media thickness by ultrasonography. In random subsamples of participants, magnetic resonance imaging of abdominal fat, pulse wave velocity of the aorta, heart, and brain, magnetic resonance spectroscopy of the liver, indirect calorimetry, dual energy X-ray absorptiometry, or accelerometry measurements were performed. The collection of data started in September 2008 and completed at the end of September 2012. Participants are currently being followed for the incidence of obesity-related diseases and mortality.

Pelotas Birth Cohort Study (The 1982 Pelotas Birth Cohort Study, Brazil): The maternity hospitals in Pelotas, a southern Brazilian city (current population ~330,000), were visited daily in the year of 1982. The 5,914 liveborns whose families lived in the urban area were examined and their mothers interviewed. Information was obtained for more than 99% of the livebirths. These subjects have been followed-up at the following mean ages: 11.3 months (all children born from January to Abril 1982; n=1457), 19.4 months (entire cohort; n=4934), 43.1 months (entire cohort; n=4742), 13.1 years (random subsample; n=715), 14.7 years (systematic subsample; n=1076); 18.2 (male cohorts attending to compulsory Army recruitment examination; n=2250), 18.9 (systematic subsample; n=1031), 22.8 years (entire cohort; n=4297) and 30.2 years (entire cohort; n=3701). Details about follow-up visits and available data can be found in the two Cohort Profile papers (PMID: 16373375 and 25733577). DNA samples (collected at the mean age of 22.8 years) were genotyped for ~2.5 million of SNPs using the Illumina HumanOmni2.5-8v1 array (which includes autosomal, X and Y chromosomes, and mitochondrial variants). After quality control, the data were prephased using SHAPEIT and imputed using IMPUTE2 based on 1000 Genomes haplotypes.

RS (Rotterdam Study): The Rotterdam Study is a prospective, population-based cohort study among individuals living in the well-defined Ommoord district in the city of Rotterdam in The Netherlands.[1] The aim of the study is to determine the occurrence of cardiovascular, neurological, ophthalmic, endocrine, hepatic, respiratory, and psychiatric diseases in elderly people. The cohort was initially defined in 1990 among 7,983 persons, aged 55 years and older, who underwent a home interview and extensive physical examination at the baseline and during follow-up rounds every 3-4 years (RS-I). Cohort was extended in 2000/2001 (RS-II, 3,011 individuals aged 55 years and older) and 2006/2008 (RS-III, 3,932 subjects, aged 45 and older; since 2016, the cohort was expanded to include persons aged 40 years and older).[1] The Rotterdam Study has been approved by the Medical Ethics Committee of the Erasmus MC (registration number MEC 02.1015) and by the Dutch Ministry of Health, Welfare

and Sport (Population Screening Act WBO, license number 1071272-159521-PG). The Rotterdam Study Personal Registration Data collection is filed with the Erasmus MC Data Protection Officer under registration number EMC1712001. The Rotterdam Study has been entered into the Netherlands National Trial Register (NTR; www.trialregister.nl) and into the WHO International Clinical Trials Registry Platform (ICTRP; www.who.int/ictrp/network/primary/en/) under shared catalogue number NTR6831. All participants provided written informed consent to participate in the study and to have their information obtained from treating physicians.

1. Ikram MA, Brusselle G, Ghanbari M, *et al*. Objectives, design and main findings until 2020 from the Rotterdam Study. Eur J Epidemiol 2020; 35: 483–517. https://doi.org/10.1007/s10654-020-00640-5

SCES (The Singapore Chinese Eye Study) is a population-based, cross-sectional study of Chinese adults residing in the southwestern part of Singapore, with a focus on studying major eye diseases, including diabetic retinopathy, age-related macular degeneration, glaucoma, refractive errors and cataract. The Ministry of Home Affairs provided an initial computer generated list ethnic Chinese names of adults aged 40–80+ years of age. A final sampling frame of 6350 ethnic Chinese residents was derived from this list using an age-stratified random sampling strategy similar to SiMES and SINDI. The ongoing SCES began in February 2009 with an aim to recruit 3300 (75% response rate) participants. PMID: 19995197

SCHS-CHD (Singapore Chinese Health Study - Coronary Heart Disease): SCHS-CHD is a case-control study of coronary heart disease that was nested within the Singapore Chinese Health Study (SCHS), a prospective cohort study of 63,257 Singaporean Chinese men and women aged 45-74 years living in Singapore. We selected cases and controls from participants that provided blood samples and were free of coronary heart disease and stroke at the time of blood collection (N=24,454). Cases (N=760) had acute myocardial infarction (AMI) or died of coronary heart disease. AMI was identified through the Singapore Myocardial Infarction Registry or through the nationwide hospital discharge database followed by confirmation of AMI by cardiologists' review of medical records using the Multi-Ethnic Study of Atherosclerosis criteria (available at: http://www.mesa-nhlbi.org/manuals.aspx). Coronary heart disease deaths were identified through the Singapore Registry of Births and Deaths (ICD9 410-414 as first stated cause of death). Matched controls (N=1,491) were selected using a risk-set sampling strategy. Controls were participants who were alive and free of coronary heart disease at the time of the diagnosis or death of the index cases and were matched for age, sex, dialect group, year of recruitment and date of blood collection. In-person interviews and phlebotomy were conducted before the onset of disease and non-fasting venous blood was stored at -80^{0}C for extraction of DNA and blood biochemistry.

SiMES (Singapore Malay Eye Study): SiMES is a population-based, cross-sectional study of 3280 Malay adults aged 40–79 years, with a focus on studying major eye diseases, including diabetic retinopathy, age-related macular degeneration, glaucoma, refractive errors and cataract. An age-stratified random sampling of all Malay adults, aged 40–80 years, residing in 15 residential districts in the southwestern part of Singapore was drawn from the computer-generated random list of 16069 Malay names provided by the Ministry of Home Affairs. A total of 1400 names from each decade of age (40–49, 50–59, 60–69 and 70–79 years), or 5600 names, were selected. Of these, 4168 individuals (74.4%) were determined to be eligible to participate. A person was considered ineligible if he or she had moved from the residential address, had not lived there in the past 6 months, was deceased, or was terminally ill. Of the 4168 eligible individuals, 3280 participants (78.7%) took part in the study. The study was conducted from August 2004 to June 2006. PMID: 17365815.

SINDI (Singapore Indian Eye Study): SINDI is a population-based, cross-sectional epidemiological study of ethnic Indian adults aged between 40 and 80+ years residing in Singapore, with a focus on studying major eye diseases, including diabetic retinopathy, age-related macular degeneration, glaucoma, refractive errors and cataract. The Ministry of Home Affairs provided an initial computer-generated list of Indian names derived from a simple random sampling of all ethnic Indian adults aged 40–80+ years of age residing in 15 residential districts in southwestern Singapore. From this list, a final sampling frame of 6350 ethnic Indian residents was derived using an age-stratified random sampling strategy similar to SiMES. SINDI was conducted from March 2007 to December 2009 and recruited 3400 (75% response rate) participants. PMID: 19995197

SP2-1M / SP2-610 (Singapore Prospective Study Program): The SP2 (Singapore Prospective Study Program) is a population-based study of diabetes and cardiovascular disease in Singapore. It first surveyed subjects (Chinese, Malay and Indian) from four cross-sectional studies that were conducted in Singapore between 1982 and 1998. Subjects were between the ages of 24-95 years and represented a random sample of the Singapore population. Subjects were re-visited between 2003 and 2007. Among the 10,747 individuals who were eligible, 5,157 subjects completed a questionnaire and the subsequent clinical examinations. Of the 5,517 subjects, 2,434 Chinese were genotyped on a combination of Illumina 610, 1M and 550 arrays. Fasting HDL-C, TC and TG were measured by an automated analyzer autoanalyzer (ADVIA 2400, Bayer Diagnostics). LDL-C was calculated from Friedewald formula. Data from this revisit were utilized for this study.

1. Nang EE, Khoo CM, Tai ES, Lim SC, Tavintharan S, Wong TY, Heng D, Lee J. Is there a clear threshold for fasting plasma glucose that differentiates between those with and without neuropathy and chronic kidney disease?: the Singapore Prospective Study Program. Am J Epidemiol. 2009 Jun 15;169(12):1454-62. doi: 10.1093/aje/kwp076. Epub 2009 Apr 30. PMID: 19406920.

2. Tan KHX, Tan LWL, Sim X, Tai ES, Lee JJ, Chia KS, van Dam RM. Cohort Profile: The Singapore Multi-Ethnic Cohort (MEC) study. Int J Epidemiol. 2018 Jun 1;47(3):699-699j. doi: 10.1093/ije/dyy014. PMID: 29452397.

WGHS (Women's Genome Health Study): WGHS is a prospective cohort of female North American health care professionals representing participants in the Women's Health Study (WHS) trial who provided a blood sample at baseline and consent for blood-based analyses [PMID: 18070814]. Participants in the WHS were 45 years or older at enrollment and free of cardiovascular disease, cancer or other major chronic illness. The current data are derived from 23,294 WGHS participants for whom whole genome genotype information was available at the time of analysis and for whom self-reported European ancestry could be confirmed by multidimensional scaling analysis of 1,443 ancestry informative markers in PLINK v. 1.06. At baseline, BP and lifestyle habits related to smoking, consumption of alcohol, and physical activity as well as other general clinical information were ascertained by a self-reported questionnaire, an approach that has been validated in the WGHS demographic, namely female health care professionals.

STAGE 2 STUDY DESCRIPTIONS

Brief descriptions are provided below for each of stage 2 studies/cohorts:

AA-DHS (African American Diabetes Heart Study): AA-DHS objectives are to improve understanding of ethnic differences in coronary artery calcification and calcified plaque in populations of African and European ancestry. The AA-DHS consists of self-reported African Americans with T2D recruited from two Wake Forest School of Medicine (WFSM) studies: the family-based Diabetes Heart Study (DHS) and unrelated individuals in the AA-DHS. DHS is a cross-sectional study of European American and African American families with siblings concordant for T2D. AA-DHS started after DHS and enrolled unrelated African Americans. The AA-DHS GWAS utilized the Illumina 5M chip with imputation to 1,000 Genomes.

BES (Beijing Eye Study): The Beijing Eye Study is a population-based study that assesses the associated and risk factors of ocular and general diseases in a Chinese population. The study was initialized in 2001 and collected data from 4439 subjects aged ≥ 40 years and living in seven communities in the Beijing area. Three of these communities were located in a rural district and four were located in an urban district. The BES was followed-up in 2006, with 3251 of the original subjects participating, and in 2011, with 2695 subjects returning for the follow-up examination. At the examinations in 2006 and 2011, trained research staffs asked the subjects questions from a standard questionnaire providing information on the family status, level of education, income, quality of life, psychic depression, physical activity, and known major systemic diseases. Fasting blood samples were taken for measurement of concentrations of substances such as blood lipids, glucose, and glycosylated hemoglobin. Individuals were classified as self-reported non-smokers or self-reported current smokers. Alcohol consumption habits based on number of drinks per day were collected. Physical activity was assessed in questions on the number of hours per day and number of days per week spent on intensively or moderately performed sport activities, spent on walking, on riding a bicycle, and spent on sitting. All variables used in analyses were taken from examinations in 2006 or in 2011. The BES subjects were genotyped on two arrays, Illumina Human610-Quad (N = 832) and Illumina OmniExpress (N = 814).

BRIGHT (British Genetics of Hypertension): Participants of the BRIGHT Study are recruited from the Medical Research Council General Practice Framework and other primary care practices in the UK. Each case had a history of hypertension diagnosed prior to 60 years of age with confirmed blood pressure recordings corresponding to seated levels >150/100mmHg (1 reading) or mean of 3 readings >145/95 mmHg. BRIGHT is focused on recruitment of hypertensive individuals with BMI<30. Sample selection for GWAS was based on DNA availability and quantity.

1. Caulfield M, Munroe P, Pembroke J, Samani N, Dominiczak A, Brown M, Benjamin N, Webster J, Ratcliffe P, O'Shea S, Papp J, Taylor E, Dobson R, Knight J, Newhouse S, Hooper J, Lee W, Brain N, Clayton D, Lathrop GM, Farrall M, Connell J, MRC British Genetics of Hypertension Study. Genome-wide mapping of human loci for essential hypertension. Lancet. 2003;361:2118-23.

DESIR (Data from an Epidemiological Study on the Insulin Resistance): The DESIR cohort study aims to: describe and understand the relations between the abnormalities of the syndrome, their evolution, according to age and sex; search for risk factors of insulin resistance, in particular factors associated with the environment, lifestyle and genetic markers; quantify the links between the

syndrome and both cardiovascular disease and diabetes; evaluate the frequency of the syndrome in terms of its consequences on public health.

This work was supported by grants from the French National Research Agency (ANR-10-LABX-46 (European Genomics Institute for Diabetes) and ANR-10-EQPX-07-01 (LIGAN-PM)), and from the National Center for Precision Diabetic Medicine - PreciDIAB, which is jointly supported by the French National Agency for Research (ANR-18-IBHU-0001), by the European Union (FEDER), by the Hauts-de-France Regional Council and by the European Metropolis of Lille (MEL).

DFTJ (Dongfeng-Tongji Cohort Study): The DFTJ-cohort study includes 27,009 retired employees from a state-owned automobile enterprise in China. This study was launched in 2008 and will be followed up every 5 years. In 2013 we conducted the first follow-up. By using semi-structural questionnaire and health examination, those having cancer or severe diseases were excluded. Fasting blood samples and detailed epidemiology data were collected. The main goal of the cohort was to identify the environmental and genetic risk factors and the gene-environment interactions on chronic diseases, and to find novel biomarkers for chronic disease and mortality prediction. Finally, 1,461 included in the present study with GWAS data. All of the participants wrote informed consent and the ethical committees in the Tongji Medical College approved this research project. Detailed information has been described in elsewhere(1).

QC criteria and imputation methods:

We did the GWAS scan on the DFTJ-cohort with Affymetrix Genome-Wide Human SNP Array 6.0 chips. In total, we genotyped 906,703 SNPs among 1,461 subjects. After stringent QC filtering, SNPs with MAF < 0.01, Hardy-Weinberg Equilibrium (HWE) < 0.0001, and SNP call rate < 95% were excluded. Individuals with call rates < 95% were also not included for further analysis. In total, we retained 1,452 subjects with 658,288 autosomal SNPs for statistical analyses, with an overall call rate of 99.68%. We used MACH 1.0 software to impute untyped SNPs using the LD information from the HapMap phase II database (CHB+JPT as a reference set (2007-08_rel22, released 2007-03-02). Imputed SNPs with high genotype information content (Rsq > 0.3 for MACH) were kept for the further association analysis.

1. Wang, F., Zhu, J., Yao, P., Li, X., He, M., Liu, Y., Yuan, J., Chen, W., Zhou, L., Min, X. et al. (2012) Cohort profile: The Dongfeng-Tongji cohort study of retired workers. International journal of epidemiology.

DHS (Diabetes Heart Study): The Diabetes Heart Study (DHS) is an ongoing family-based cohort study investigating the epidemiology and genetics of cardiovascular disease (CVD) in a population-based sample. The DHS recruited T2D-affected siblings without advanced renal insufficiency from 1998 through 2005 in western North Carolina. DHS has collected genetic data on 1,220 self-described European American (EA) individuals from 475 families. Genotyping was completed using an Affymetrix Genome-Wide Human SNP Array 5.0 with imputation of 1,000 Genomes project SNPs from this array using IMPUTE2 and the Phase I v2, cosmopolitan (integrated) reference panel, build 37.

DR's EXTRA (Dose-Responses to Exercise Training): The Dose-Responses to Exercise Training (DR's EXTRA) Study is a 4-year RCT on the effects of regular physical exercise and healthy diet on endothelial function, atherosclerosis and cognition in a randomly selected population sample (n=3000) of Eastern Finnish men and women, identified from the national population register, aged

55-74 years. Of the eligible sample, 1410 individuals were randomized into one of the 6 groups: aerobic exercise, resistance exercise, diet, combined aerobic exercise and diet, combined resistance exercise and diet, or reference group following baseline assessments. During the four year intervention the drop-out rate was 15%.

EGCUT (Estonian Genome Center - University of Tartu (Estonian Biobank)): The Estonian Biobank is the population-based biobank of the Estonian Genome Center at the University of Tartu (www.biobank.ee; EGCUT). The entire project is conducted according to the Estonian Gene Research Act and all of the participants have signed the broad informed consent. The cohort size is up to 51535 individuals from 18 years of age and up, which closely reflects the age, sex and geographical distribution of the Estonian population. All of the subjects are recruited randomly by general practitioners and physicians in hospitals. A Computer Assisted Personal interview is filled within 1-2 hours at a doctor's office, which includes personal, genealogical, educational, occupational history and lifestyle data. Anthropometric measurements, blood pressure and resting heart rate are measured and venous blood taken during the visit. Medical history and current health status is recorded according to ICD-10 codes.

EPIC (European Prospective Investigation into Cancer and Nutrition)-Norfolk: The European Prospective Investigation of Cancer (EPIC)-Norfolk study is a population-based cohort study established to study the links diet, lifestyle factors and cancer and other health outcomes. Participants are men and women who were aged between 40 and 79 when they joined the study and who lived in Norwich, UK and the surrounding towns and rural areas. They have been contributing information about their diet, lifestyle and health through questionnaires and health checks over two decades. The Norwich Local Research Ethics Committee granted ethical approval for the study. All participants gave written informed consent.

The EPIC-InterAct Case-Cohort Study: EPIC- InterAct is a type 2 diabetes case-cohort study nested within the European Prospective Investigation into Cancer and Nutrition (EPIC) study. EPIC was initiated in the late 1980s and involves collaboration between 23 research institutions across Europe in 10 countries. The majority of EPIC cohorts were recruited from the general population, with some exceptions. French cohorts included women who were members of a health insurance scheme for school and university employees; Turin and Ragusa (Italy) and the Spanish centres included some blood donors. Participants from Utrecht (Netherlands) and Florence (Italy) were recruited via a breast cancer screening program. The majority of participants recruited by the EPIC Oxford (UK) centre consisted of vegetarian and "health conscious" volunteers from England, Wales, Scotland, and Northern Ireland. EPIC-InterAct sampled a random sub-cohort and all individuals who subsequently developed incident T2DM over follow up from the full cohort of participants in EPIC who provided blood samples at baseline in Denmark, France, Germany, Italy, the Netherlands, Spain, Sweden and the UK. Smoking status was determined at baseline and lipid measurements at baseline were undertaken in a centralized laboratory on all participants in the case-cohort study.

GeneSTAR (Genetic Studies of Atherosclerosis Risk): GeneSTAR is a family-based prospective study of more than 4000 participants begun in 1983 to determine phenotypic and genetic causes of premature cardiovascular disease. Families were identified from 1983-2006 from probands with a premature coronary disease event prior to 60 years of age who were identified at the time of hospitalization in any of 10 hospitals in the Baltimore, Maryland area. Their apparently healthy 30-59 year old siblings without known coronary disease were recruited and screened between 1983 and 2006. From 2003-2006, adult offspring over 21 years of age of all participating siblings and probands, as well as the coparents of the offspring were recruited and screened. Genotyping was

performed in 3,232 participants on the Illumina 1Mv1_c platform. Demographic information including education level was self-reported by participants. Medical history including medication use was assessed by the study nurse or physician.

GLACIER (Gene x Lifestyle Interactions and Complex Traits Involved in Elevated Disease Risk): The Gene-Lifestyle interactions And Complex traits Involved in Elevated disease Risk (GLACIER) Study is nested within the Västerbotten Intervention Programme, which is part of the Northern Sweden Health and Disease Study, a population-based prospective cohort study from northern Sweden. Participants were genotyped with Illumina CardioMetaboChip array. This array contains ~200,000 variants, the majority being common variants. Systolic and diastolic blood pressures were measured once following a period of five minutes rest with the participant in the supine position using a mercury-gauge sphygmomanometer. Analysis of serum lipids (HDL-C, triglycerides and total cholesterol) were undertaken at the Department of Clinical Chemistry at Umeå University Hospital using routine methods. LDL-C was determined using the Friedewald formula. All participants completed a detailed, optically readable, health and lifestyle questionnaire including questions about smoking status and alcohol intake (FFQ). Cohort description - PMID: 25396097

GRAPHIC (Genetic Regulation of Arterial Pressure of Humans in the Community): The GRAPHIC Study comprises 2024 individuals from 520 nuclear families recruited from the general population in Leicestershire, UK between 2003-2005 for the purpose of investigating the genetic determinants of blood pressure and related cardiovascular traits. A detailed medical history was obtained from study subjects by standardized questionnaires and clinical examination was performed by research nurses following standard procedures. Measurements obtained included height, weight, waist-hip ratio, clinic and ambulatory blood pressure and a 12-lead ECG.

HCHS/SOL (Hispanic Community Health Study/ Study of Latinos): The HCHS/SOL is a community-based cohort study of 16,415 self-identified Hispanic/Latino persons aged 18–74 years and selected from households in predefined census-block groups across four US field centers (in Chicago, Miami, the Bronx, and San Diego). The census-block groups were chosen to provide diversity among cohort participants with regard to socioeconomic status and national origin or background. The HCHS/SOL cohort includes participants who self-identified as having a Hispanic/Latino background; the largest groups are Central American (n = 1,730), Cuban (n = 2,348), Dominican (n = 1,460), Mexican (n = 6,471), Puerto Rican (n = 2,728), and South American (n = 1,068). The HCHS/SOL baseline clinical examination occurred between 2008 and 2011 and included comprehensive biological, behavioral, and sociodemographic assessments. Consenting HCHS/SOL subjects were genotyped at Illumina on the HCHS/SOL custom 15041502 B3 array. The custom array comprised the Illumina Omni 2.5M array (HumanOmni2.5-8v.1-1) ancestry-informative markers, known GWAS hits and drug absorption, distribution, metabolism, and excretion (ADME) markers, and additional custom content including ~150,000 SNPs selected from the CLM (Colombian in Medellin, Colombia), MXL (Mexican Ancestry in Los Angeles, California), and PUR (Puerto Rican in Puerto Rico) samples in the 1000Genomes phase 1 data to capture a greater amount of Amerindian genetic variation. QA/QC procedures yielded a total of 12,803 unique study participants for imputation and downstream association analyses.

HRS (Health & Retirement Study): The Health and Retirement Study (HRS) is a longitudinal survey of a representative sample of Americans over the age of 50. The current sample is over 26,000 persons in 17,000 households. Respondents are interviewed every two years about income and wealth, health and use of health services, work and retirement, and family connections. DNA

was extracted from saliva collected during a face-to-face interview in the respondents' homes. These data represent respondents who provided DNA samples and signed consent forms in 2006, 2008, and 2010. Respondents were removed if they had missing genotype or phenotype data.

1. Juster, F. T., Suzman, R. (1995). An Overview of the Health and Retirement Study, Journal of Human Resources 30:Suppl: S7-S56.

2. Sonnega A, Faul JD, Ofstedal MB, Langa KM, Phillips JWR, Weir DR. Cohort Profile: the Health and Retirement Study (HRS). Int. J. Epidemiol. 2014; 43 (2): 576-585. PMID: 24671021.

3. Crimmins, E.M., Guyer H., Langa K.M., Ofstedal M.B., Wallace R.B., and Weir D.R. (2008).

    - Documentation of Physical Measures, Anthropometrics and Blood Pressure in the Health and Retirement Study. HRS Documentation Report DR-011. http://hrsonline.isr.umich.edu/sitedocs/userg/dr-011.pdf

HyperGEN-AXIOM (Hypertension Genetic Epidemiology Network): HyperGEN is a family-based study that investigates the genetic causes of hypertension and related conditions in EA and AA subjects. HyperGEN recruited hypertensive sibships, along with their normotensive adult offspring, and an age-matched random sample. HyperGEN has collected data on 2,471 Caucasian-American subjects and 2,300 African-American subjects, from five field centers in Alabama, Massachusetts, Minnesota, North Carolina, and Utah. HyperGEN participates as a discovery study using GWAS available in a large subset of the samples. The remaining AA subjects without GWAS data were genotyped on the Affymetrix Axiom chip as part of a HyperGEN admixture mapping ancillary study. After excluding subjects already included in the original HyperGEN (or with family members included), this subset of approximately 450 AA subjects are included in the HyperGEN-AXIOM study which participates in replications.

1. Shetty PB, Tang H, Tayo BO, Morrison AC, Hanis CL, Rao DC, Young JH, Fox ER, Boerwinkle E, Cooper RS, Risch NJ, Zhu X; Candidate Gene Association Resource (CARe) Consortium. Variants in CXADR and F2RL1 are associated with blood pressure and obesity in African-Americans in regions identified through admixture mapping. J Hypertens. 2012 Oct;30(10):1970-6. PMID:22914544

INGI-CARL (INGI-Carlantino) and INGI-FVG (INGI-Friuli Venezia Giulia): INGI-FVG and INGI-CARL studies include samples coming from isolated populations and belong to the ITALIAN NETWORK OF GENETIC ISOLATES (INGI). Carlantino is a small village in the Province of Foggia in southern Italy. Genetic analyses of chromosome Y haplotypes as well as mitochondrial DNA show that Carlantino is a genetically homogeneous population and not only a geographically isolated village (DOI: 10.1038/ejhg.2012.229; DOI: 10.1038/s41431-019-0551-x). Participants were randomly selected in the range of 15 – 90 years of age. Genotyping and phenotypic data are available for 630 individuals. People with age < 18 were excluded from analyses. Subjects gave their written informed consent for participating in these studies. The project was approved by the local administration of Carlantino, the Health Service of Foggia Province, Italy, and the ethical committee of the IRCCS Burlo-Garofolo of Trieste.

The Friuli Venezia Giulia population represents a collection of six villages covering a total area of 7858 km2 in a hilly part of Friuli-Venezia Giulia (FVG) county located in north-eastern Italy. Recent studies characterized this population as a genetic isolate with high level of genomic homozygosity

and elevated linkage disequilibrium (DOI: 10.1038/ejhg.2012.229; DOI: 10.1038/s41431-019-0551-x). The cohort accounts for 1590 genotyped samples. Participants were randomly selected people 3-92 years of age. Genotyping and phenotypic data for 1590 samples are available. People with age < 18 were excluded from analyses. A written informed consent for participation was obtained from all subjects. The project was approved by the Ethical committee of the IRCCS Burlo-Garofolo."

IRAS (Insulin Resistance Atherosclerosis Study): The Insulin Resistance Atherosclerosis Study (IRAS) was an epidemiologic cohort study designed to examine the relationship between insulin resistance and carotid atherosclerosis across a range of glucose tolerance. Individuals of self-reported Mexican-American ethnicity were recruited in San Antonio, TX and San Luis Valley, CO. Recruitment was balanced across age and glucose tolerance status. Inclusion of IRAS data is limited to 194 normoglycemic individuals with genotype data from the Illumina OmniExpress and Omni 1S arrays and imputation to the 1000 Genomes Integrated Reference Panel (phase I).

IRASFS (Insulin Resistance Atherosclerosis Family Study): The IRASFS was a family study designed to examine the genetic and epidemiologic basis of glucose homeostasis traits and abdominal adiposity. Briefly, self-reported Mexican American pedigrees were recruited in San Antonio, TX and San Luis Valley, CO. Probands with large families were recruited from the initial non-family-based IRAS, which was modestly enriched for impaired glucose tolerance and T2D. Inclusion of IRASFS data is limited to 1040 normoglycemic individuals in 88 pedigrees with genotype data from the Illumina OmniExpress and Omni 1S arrays and imputation to the 1000 Genomes Integrated Reference Panel (phase I).

JUPITER (Justification for the Use of Statins in Primary Prevention: An Intervention Trial Evaluating Rosuvastatin): Genetic analysis was performed in a sub-population from JUPITER (Justification for the Use of statins in Prevention: an Intervention Trial Evaluating Rosuvastatin), an international, randomized, placebo-controlled trial of rosuvastatin (20mg/day) in the primary prevention of cardiovascular disease conducted among apparently healthy men and women with LDL-C < 130 mg/dL and hsCRP ≥ 2 mg/L (1, 2). Individuals with diabetes or triglyceride concentration >500mg/dL were excluded. The present analysis includes only individuals who provided consent for genetic analysis, had successfully collected genotype information, and who had either verified European or verified South African black ancestry.

1. Ridker PM, Danielson E, Fonseca FA, Genest J, Gotto AM Jr, Kastelein JJ, Koenig W, Libby P, Lorenzatti AJ, MacFadyen JG, Nordestgaard BG, Shepherd J, Willerson JT, Glynn RJ; JUPITER Study Group. Rosuvastatin to prevent vascular events in men and women with elevated C-reactive protein. N Engl J Med 2008 Nov 20; 359(21):2195-207

2. Chasman DI, Giulianini F, MacFadyen J, Barratt BJ, Nyberg F, Ridker PM. Genetic determinants of statin-induced low-density lipoprotein cholesterol reduction: the Justification for the Use of Statins in Prevention: an Intervention Trial Evaluating Rosuvastatin (JUPITER) trial. Circ Cardiovasc Genet. 2012 Apr 1;5(2):257-64. doi: 10.1161/CIRCGENETICS.111.961144. Epub 2012 Feb 13. Erratum in: Circ Cardiovasc Genet. 2012 Jun;5(3):e27. PubMed PMID: 22331829.

KORA (Cooperative Health Research in the Augsburg Region): The KORA study is a series of independent population-based epidemiological surveys of participants living in the region of Augsburg, Southern Germany. All survey participants are residents of German nationality identified

through the registration office and were examined in 1994/95 (KORA S3) and 1999/2001 (KORA S4). In the KORA S3 and S4 studies 4,856 and 4,261 subjects have been examined implying response rates of 75% and 67%, respectively. 3,006 subjects participated in a 10-year follow-up examination of S3 in 2004/05 (KORA F3), and 3080 of S4 in 2006/2008 (KORA F4). The age range of the participants was 25 to 74 years at recruitment. Informed consent has been given by all participants. The study has been approved by the local ethics committee. Individuals for genotyping in KORA F3 and KORA F4 were randomly selected and these genotypes are taken for the analysis of the phenotypes in KORA S3 and KORA S4.

LBC1936 (Lothian Birth Cohort 1936): LBC1936 consists of 1091 (548 male) relatively healthy individuals who underwent cognitive and medical testing at a mean age of 69.6 years (SD = 0.8). They were born in 1936, most took part in the Scottish Mental Survey of 1947, and almost all lived independently in the Lothian region of Scotland.[1,2]

1. Deary IJ, Gow AJ, Pattie A, Starr JM. Cohort profile: the Lothian Birth Cohorts of 1921 and 1936. Int J Epidemiol 2012;41:1576-1584.

2. Taylor AM, Pattie A, Deary IJ. Cohort Profile Update: The Lothian Birth Cohorts of 1921 and 1936. Int J Epidemiol 2018;47:1042-1042r

Lifelines (Netherlands Biobank): Lifelines (https://lifelines.nl/) is a multi-disciplinary prospective population-based cohort study using a unique three-generation design to examine the health and health-related behaviors of 165,000 persons living in the North East region of The Netherlands. It employs a broad range of investigative procedures in assessing the biomedical, socio-demographic, behavioral, physical and psychological factors which contribute to the health and disease of the general population, with a special focus on multimorbidity. In addition, the Lifelines project comprises a number of cross-sectional sub-studies, which investigate specific age-related conditions. These include investigations into metabolic and hormonal diseases, including obesity, cardiovascular and renal diseases, pulmonary diseases and allergy, cognitive function and depression, and musculoskeletal conditions. All survey participants are between 18 and 90 years old at the time of enrollment. Recruitment has been going on since the end of 2006, and over 130,000 participants had been included by April 2013. At the baseline examination, the participants in the study were asked to fill in a questionnaire (on paper or online) before the first visit. During the first and second visit, the first or second part of the questionnaire, respectively, are checked for completeness, a number of investigations are conducted, and blood and urine samples are taken. Lifelines is a facility that is open for all researchers. Information on application and data access procedure is summarized on www.lifelines.nl. (Scholtens S, Smidt N, Swertz MA, Bakker SJ, Dotinga A, Vonk JM, et al. Cohort Profile: LifeLines, a three-generation cohort study and biobank. Int J Epidemiol. 2014 Dec 14.)

- LifeLines group authors (genetics): Behrooz Z Alizadeh,[1] H Marike Boezen,[1] Lude Franke,[2] Pim van der Harst,[3] Gerjan Navis,[4] Marianne Rots,[5] Harold Snieder,[1] Morris Swertz,[2] Bruce HR Wolffenbuttel,[6] Cisca Wijmenga.[2]

- [1] Department of Epidemiology, University of Groningen, University Medical Center Groningen, The Netherlands; [2] Department of Genetics, University of Groningen, University Medical Center Groningen, The Netherlands; [3] Department of Cardiology, University of Groningen, University Medical Center Groningen, The Netherlands; [4] Department of Internal Medicine, Division of Nephrology, University of Groningen, University Medical Center Groningen, The Netherlands; [5] Department of Pathology and Medical Biology, University of Groningen, University Medical Center

Groningen, The Netherlands; [6] Department of Endocrinology, University of Groningen, University Medical Center Groningen, The Netherlands.

LLFS (The Long Life Family Study): LLFS is a family-based cohort study, including four clinical centers: Boston University Medical Center in Boston, MA, USA, Columbia University College of Physicians and Surgeons in New York City, NY, USA, the University of Pittsburgh in Pittsburgh PA, USA, and University of Southern Denmark, Denmark. The study characteristics, recruitment, eligibility and enrollment have been previously described (Pedersen et al., 2006, PMID: 17150149; Sebastiani et al., 2009, PMID: 19910380; Newman et al., 2011, PMID: 21258136). In brief, the LLFS was designed to determine genetic, behavioral, and environmental factors related to families of exceptionally healthy, elderly individuals. Phase 1 was conducted between 2006 and 2009 recruiting 4,953 individuals from 539 families. The probands were at least 79 years old in the USA centers, and 90 years old or above in Denmark. The families were selected to participate in the study based on The Family Longevity Selection Score (FLoSS) (Sebastiani et al., 2009, PMID: 19910380), a score generated according to birth-year cohort survival probabilities of the proband and siblings; probands and their families with FLoSS score of 7 or higher, at least one living sibling, and at least one living offspring (minimum family size of 3), who were able to give informed consent and willing to participate were recruited. The individuals were genotyped using ~2.3 million SNPs from the Illumina Omni chip, and then imputed on phased 1000 Genomes with Cosmopolitan data as a reference using MACH and MINIMAC. After excluding participants with 80 years and older, ~3,200 individuals have been included in the analyses for replication.

Loyola GxE (Kingston Gene-by-environment; subset of International Collaborative Study of Hypertension in Blacks (ICSHIB)): The Kingston GxE cohort was obtained from a survey conducted in Kingston, Jamaica as part of a larger project to examine gene by environment interactions in the determination of blood pressure among adults 25-74 years [PMID: 9103091]. The principal criterion for eligibility was a body mass index in either the top or bottom third of BMI for the Jamaican population. Participants were identified principally from the records of the Heart Foundation of Jamaica, a non-governmental organization based in Kingston, which provides low-cost screening services (height and weight, blood pressure, glucose, cholesterol) to the general public. Other participants were identified from among participants in family studies of blood pressure at the Tropical Metabolism Research Unit (TMRU) and from among staff members at the University of the West Indies, Mona.

Loyola SPT (Spanish Town; subset of International Collaborative Study of Hypertension in Blacks (ICSHIB)): Participants were recruited from Spanish Town, a stable, residential urban area neighboring the capital city of Kingston, Jamaica as part of the ICSHIB [PMID: 9103091]. A stratified random sampling scheme was used to recruit adult males and females aged 25–74 years from the general population. Spanish Town was chosen because its demographic make-up was broadly representative of Jamaica as a whole.

NESDA (Netherlands Study of Depression and Anxiety): NESDA is a multi-center study designed to examine the long-term course and consequences of depressive and anxiety disorders (http://www.nesda.nl)[1]. NESDA included both individuals with depressive and/or anxiety disorders and controls without psychiatric conditions. Inclusion criteria were age 18-65 years and self-reported western European ancestry while exclusion criteria were not being fluent in Dutch and having a primary diagnosis of another psychiatric condition (psychotic disorder, obsessive compulsive disorder, bipolar disorder, or severe substance use disorder).

1. Penninx, B.W. et al. The Netherlands Study of Depression and Anxiety (NESDA): rationale, objectives and methods. Int J Methods Psychiatr Res 17, 121-40 (2008).

PREVEND (The Prevention of REnal and Vascular ENd stage Disease study): The PREVEND study is an ongoing prospective study investigating the natural course of increased levels of urinary albumin excretion and its relation to renal and cardiovascular disease. Inhabitants 28 to 75 years of age (n=85,421) in the city of Groningen, The Netherlands, were asked to complete a short questionnaire, 47% responded, and individuals were then selected with a urinary albumin concentration of at least 10 mg/L (n = 7,768) and a randomly selected control group with a urinary albumin concentration less than 10 mg/L (n = 3,395). Details of the protocol have been described elsewhere (Hillege HL et al. Circulation 2002;106:1777-82).

RHS (Ragama Health Study): The Ragama Health Study (RHS) is a population-based study of South Asian men and women aged 35-64yrs living in the Ragama Medical Officer of Health (MOH) area, near Colombo, Sri Lanka.* Consenting adults attended a clinic after a 12-h fast with available health records, and were interviewed by trained personnel to obtain information on medical, sociodemographic, and lifestyle variables. A 10-mL sample of venous blood was obtained from each subject. The concurrent study was performed in two tea plantation estates in the Lindula MOH area, near Nuwara Eliya (180 km from Colombo), to investigate the gene-environment interaction in a community with differing lifestyles (e.g., physical activity and diet). The RHS is a collaborative effort between the Faculty of Medicine, University of Kelaniya and the National Center for Global Health and Medicine, Japan.

1. Dassanayake, A.S. et al. Prevalence and risk factors for non-alcoholic fatty liver disease among adults in an urban Sri Lankan population. J Gastroenterol Hepatol 24, 1284-8 (2009).

SHEEP (Stockholm Heart Epidemiology Project): The SHEEP is a population based case-control study of risk factors for first episode of acute myocardial infarction. The study base comprised all Swedish citizens resident in the Stockholm county 1992-1994 who were 45-70 years of age and were free of previous clinically diagnosed myocardial infarction.

Cases were identified using three different sources: 1) coronary units and internal medicine wards for acute care in all Stockholm hospitals; 2) the National Patient Register; and 3) death certificates. For the present study, only cases who survived at least 28 days were considered (n=1213).

First time incident myocardial infarction cases (n=1213) were identified during a 2-year period (1992-1993) for men and during a 3-year period (1992-1994) for women. Controls (n=1561) were randomly recruited from the study population continuously over time within 2 days of the case occurrence and matched to cases on age (5-years interval), sex and hospital catchment area using computerized registers of the population of Stockholm. Five control candidates were sampled simultaneously to be able to replace potential non-respondent controls. Occasionally, because of late response of the initial control, both the first and alternative controls were considered resulting in the inclusion of more controls than cases. Postal questionnaires covering a wide range of exposure areas including occupational exposures, life style factors, social factors and health related factors were distributed to the participants. Clinical investigations were performed at least three months after myocardial infarction of cases and their matched controls. The investigations included blood samplings under fasting conditions with collection of whole blood for DNA extraction, serum and plasma. A biobank was established containing DNA, serum and plasma.

Exposure information based on both the questionnaire and biological data from the health examination was available for 78% of the male and 67% of the female non-fatal cases; the corresponding figures for their controls were 68% and 64%.

SHIP (Study of Health in Pomerania): The Study of Health In Pomerania (SHIP) is a prospective longitudinal population-based cohort study in Mecklenburg-West Pomerania assessing the prevalence and incidence of common diseases and their risk factors (PMID: 20167617). SHIP encompasses the two independent cohorts SHIP and SHIP-TREND. Participants aged 20 to 79 with German citizenship and principal residency in the study area were recruited from a random sample of residents living in the three local cities, 12 towns as well as 17 randomly selected smaller towns. Individuals were randomly selected stratified by age and sex in proportion to population size of the city, town or small towns, respectively. A total of 4,308 participants were recruited between 1997 and 2001 in the SHIP cohort. Between 2008 and 2012 a total of 4,420 participants were recruited in the SHIP-TREND cohort. Individuals were invited to the SHIP study centre for a computer-assisted personal interviews and extensive physical examinations. The study protocol was approved by the medical ethics committee of the University of Greifswald. Oral and written informed consents was obtained from each of the study participants

Genome-wide SNP-typing was performed using the Affymetrix Genome-Wide Human SNP Array 6.0 or the Illumina Human Omni 2.5 array (SHIP-TREND samples). Array processing was carried out in accordance with the manufacturer's standard recommendations. Genotypes were determined using GenomeStudio Genotyping Module v1.0 (GenCall) for SHIP-TREND and the Birdseed2 clustering algorithm for SHIP. Imputation of genotypes in SHIP and SHIP-TREND was performed with the software IMPUTE v2.2.2 based on 1000 Genomes release March 2012.

SWHS/SMHS (Shanghai Women's Health Study/ Shanghai Men's Health Study): The Shanghai Women's Health Study (SWHS) is an ongoing population-based cohort study of approximately 75,000 women who were aged 40-70 years at study enrollment and resided in in urban Shanghai, China; 56,832 (75.8%) provided a blood samples. Recruitment for the SWHS was initiated in 1997 and completed in 2000. The self-administered questionnaire includes information on demographic characteristics, disease and surgery histories, personal habits (such as cigarette smoking, alcohol consumption, tea drinking, and ginseng use), menstrual history, residential history, occupational history, and family history of cancer.

The Shanghai Men's Health Study (SMHS) is an ongoing population-based cohort study of 61,480 Chinese men who were aged between 40 and 74 years, were free of cancer at enrollment, and lived in urban Shanghai, China; 45,766 (74.4%) provided a blood samples. Recruitment for the SMHS was initiated in 2002 and completed in 2006. The self-administered questionnaire includes information on demographic characteristics, disease and surgery histories, personal habits (such as cigarette smoking, alcohol consumption, tea drinking, and ginseng use), residential history, occupational history, and family history of cancer.

Genotyping and imputation: Genomic DNA was extracted from buffy coats by using a Qiagen DNA purification kit (Valencia, CA) or Puregene DNA purification kit (Minneapolis, MN) according to the manufacturers' instructions and then used for genotyping assays. The GWAS genotyping was performed using the Affymetrix Genome-Wide Human SNP Array 6.0 (Affy6.0) platform or Illumina 660, following manufacturers' protocols. After sample quality control, we exclude SNPs with 1) MAF <0.01; 2) call rate <95%; 2) bad genotyping cluster; and 3) concordance rate <95%

among duplicated QC samples. Genotypes were imputed using the program MACH (http://www.sph.umich.edu/csg/abecasis/MACH/download/), which determines the probable distribution of missing genotypes conditional on a set of known haplotypes, while simultaneously estimating the fine-scale recombination map. Phased autosome SNP data from HapMap Phase II Asians (release 22) were used as the reference. To test for associations between the imputed SNP data with BMI, linear regression (additive model) was used, in which SNPs were represented by the expected allele count, an approach that takes into account the degree of uncertainty of genotype imputation (http://www.sph.umich.edu/csg/abecasis/MACH/download/).

The lipid profiles were measured at Vanderbilt Lipid Laboratory. Total cholesterol, high-density lipoprotein (HDL) cholesterol, and triglycerides (TG) were measured using an ACE Clinical Chemistry System (Alfa Wassermann, Inc, West Caldwell, NJ). Low-density lipoprotein (LDL) cholesterol levels were calculated by using the Friedwald equation. The levels of LDL cholesterol were directly measured using an ACE Clinical Chemistry System for subjects with TG levels ≥ 400 mg/dL. Fasting status was defined as an interval between the last meal and blood draw of 8 hours or longer.

TWINGENE (TwinGene of the Swedish Twin Registry): The aim of the TwinGene project has been to systematically transform the oldest cohorts of the Swedish Twin Registry (STR) into a molecular-genetic resource. Beginning in 2004, about 200 twins were contacted each month until the data collection was completed in 2008. A total of 21 500 twins were contacted where of 12 600 participated. Invitations to the study contained information of the study and its purpose. Along with the invitations consent forms and health questionnaire were sent to the subjects. When the signed consent forms where returned, the subjects were sent blood sampling equipment and asked to contact a local health facility for blood sampling. The study population was recruited among twins participating in the Screening Across the Lifespan Twin Study (SALT) which was a telephone interview study conducted in 1998-2002. Other inclusion criteria were that both twins in the pair had to be alive and living in Sweden. Subjects were excluded from the study if they preciously declined participation in future studies or if they had been enrolled in other STR DNA sampling projects. The subjects were asked to make an appointment for a health check-up at their local health-care facility on the morning Monday to Thursday and not the day before a national holiday, this to ensure that the sample would reach the KI biobank the following morning by overnight mail. The subjects were instructed to fast from 20.00 the previous night. By venipuncture a total of 50 ml of blood was drawn from each subject. Tubes with serum and blood for biobanking as well as for clinical chemistry tests were sent to KI by overnight mail. One 7ml EDTA tube of whole blood is stored in -80°C while a second 7ml EDTA tube of blood is used for DNA extraction using Puregene extraction kit (Gentra systems, Minneapolis, USA). After excluding subjects in which the DNA concentration in the stock-solution was below 20ng/µl as well as subset of 302 female monozygous twin pairs participating in a previous genome wide effort DNA from 9896 individual subjects was sent to SNP&SEQ Technology Platform Uppsala, Sweden for genome wide genotyping with Illumina OmniExpress bead chip (all available dizygous twins + one twin from each available MZ twin pair).

YFS (The Cardiovascular Risk in Young Finns Study): The YFS is a population-based follow up-study started in 1980. The main aim of the YFS is to determine the contribution made by childhood lifestyle, biological and psychological measures to the risk of cardiovascular diseases in adulthood. In 1980, over 3,500 children and adolescents all around Finland participated in the baseline study. The follow-up studies have been conducted mainly with 3-year intervals. The latest 30-year follow-up study was conducted in 2010-11 (ages 33-49 years) with 2,063 participants. The study was approved by the local ethics committees (University Hospitals of Helsinki, Turku, Tampere, Kuopio

and Oulu) and was conducted following the guidelines of the Declaration of Helsinki. All participants gave their written informed consent.

STAGE 1 STUDY ACKNOWLEDGMENTS

Infrastructure for the CHARGE Consortium is supported in part by the National Heart, Lung, and Blood Institute grant R01HL105756. Infrastructure for the Gene-Lifestyle Working Group is supported by the National Heart, Lung, and Blood Institute grant R01HL118305.

AGES (Age Gene/Environment Susceptibility Reykjavik Study): This study has been funded by NIH contract N01-AG012100, HSSN271201200022C, the NIA Intramural Research Program, an Intramural Research Program Award (ZIAEY000401) from the National Eye Institute, an award from the National Institute on Deafness and Other Communication Disorders (NIDCD) Division of Scientific Programs (IAA Y2-DC_1004-02), Hjartavernd (the Icelandic Heart Association), and the Althingi (the Icelandic Parliament). The study is approved by the Icelandic National Bioethics Committee, VSN: 00-063. The researchers are indebted to the participants for their willingness to participate in the study.

ARIC (Atherosclerosis Risk in Communities) Study: The ARIC study has been funded in whole or in part with Federal funds from the National Heart, Lung, and Blood Institute, National Institutes of Health, Department of Health and Human Services, under Contract nos. (75N92022D00001, 75N92022D00002, 75N92022D00003, 75N92022D00004, 75N92022D00005). The authors thank the staff and participants of the ARIC study for their important contributions. Funding was also supported by R01HL087641 and R01HL086694; National Human Genome Research Institute contract U01HG004402; and National Institutes of Health contract HHSN268200625226C. Infrastructure was partly supported by Grant Number UL1RR025005, a component of the National Institutes of Health and NIH Roadmap for Medical Research.

Baependi Heart Study (Brazil): The Baependi Heart Study was supported by Fundação de Amparo a Pesquisa do Estado de São Paulo (FAPESP) (Grant 2013/17368-0), Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) and Hospital Samaritano Society (Grant 25000.180.664/2011-35), through Ministry of Health to Support Program Institutional Development of the Unified Health System (SUS-PROADI).

CARDIA (Coronary Artery Risk Development in Young Adults): The Coronary Artery Risk Development in Young Adults Study (CARDIA) is conducted and supported by the National Heart, Lung, and Blood Institute (NHLBI) in collaboration with the University of Alabama at Birmingham (HHSN268201800005I & HHSN268201800007I), Northwestern University (HHSN268201800003I), University of Minnesota (HHSN268201800006I), and Kaiser Foundation Research Institute (HHSN268201800004I). Genotyping was funded as part of the NHLBI Candidate-gene Association Resource (N01-HC-65226) and the NHGRI Gene Environment Association Studies (GENEVA) (U01-HG004729, U01-HG04424, and U01-HG004446). This manuscript has been reviewed by CARDIA for scientific content.

CHS (Cardiovascular Health Study): This Cardiovascular Health Study (CHS) research was supported by NHLBI contracts HHSN268201200036C, HHSN268200800007C, HHSN268200960009C, HHSN268201800001C N01HC55222, N01HC85079, N01HC85080,

N01HC85081, N01HC85082, N01HC85083, N01HC85086, 75N92021D00006; and NHLBI grants U01HL080295, U01HL130114, R01HL087652, R01HL105756, R01HL103612, R01HL085251, and R01HL120393 with additional contribution from the National Institute of Neurological Disorders and Stroke (NINDS). Additional support was provided through R01AG023629 from the National Institute on Aging (NIA). A full list of principal CHS investigators and institutions can be found at CHS-NHLBI.org. The provision of genotyping data was supported in part by the National Center for Advancing Translational Sciences, CTSI grant UL1TR001881, and the National Institute of Diabetes and Digestive and Kidney Disease Diabetes Research Center (DRC) grant DK063491 to the Southern California Diabetes Endocrinology Research Center. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

CROATIA-Korcula: We would like to acknowledge the staff of several institutions in Croatia that supported the field work, including but not limited to The University of Split and Zagreb Medical Schools and the Croatian Institute for Public Health. We would like to acknowledge the invaluable contributions of the recruitment team in Korcula, the administrative teams in Croatia and Edinburgh and the participants. The SNP genotyping for the CROATIA-Korcula cohort was performed in Helmholtz Zentrum München, Neuherberg, Germany. CROATIA-Korcula (CRKorcula) was funded by the Medical Research Council UK, The Croatian Ministry of Science, Education and Sports (grant 216-1080315-0302), the European Union framework program 6 EUROSPAN project (contract no. LSHG-CT-2006-018947), the Croatian National Centre of Research Excellence in Personalized Healthcare (grant number KK.01.1.1.01.0010), and the Centre of Competence in Molecular Diagnostics (KK.01.2.2.03.0006). JEH is now Director of Genomics Research, Center for Population Genomics. Boston VA, and JM is currently an employee of Genomics PLC, but both completed this work at the Medical Research Council Human Genetics Unit at the University of Edinburgh.

CROATIA-Vis: We would like to acknowledge the staff of several institutions in Croatia that supported the field work, including but not limited to The University of Split and Zagreb Medical Schools, the Institute for Anthropological Research in Zagreb and Croatian Institute for Public Health. The SNP genotyping for the CROATIA-Vis cohort was performed in the core genotyping laboratory of the Wellcome Trust Clinical Research Facility at the Western General Hospital, Edinburgh, Scotland. CROATIA-Vis (CR-Vis) was funded by the Medical Research Council UK, The Croatian Ministry of Science, Education and Sports (grant 216-1080315-0302), and the European Union framework program 6 EUROSPAN project (contract no. LSHG-CT-2006-018947).

ERF (Erasmus Rucphen Family study): The ERF study as a part of EUROSPAN (European Special Populations Research Network) was supported by European Commission FP6 STRP grant number 018947 (LSHG-CT-2006-01947) and also received funding from the European Community's Seventh Framework Programme (FP7/2007-2013)/grant agreement HEALTH-F4-2007-201413 by the European Commission under the programme "Quality of Life and Management of the Living Resources" of 5th Framework Programme (no. QLG2-CT-2002-01254). High-throughput analysis of the ERF data was supported by joint grant from Netherlands Organisation for Scientific Research and the Russian Foundation for Basic Research (NWO-RFBR 047.017.043). We are grateful to all study participants and their relatives, general practitioners and neurologists for their contributions and to P. Veraart for her help in genealogy, J. Vergeer for the supervision of the laboratory work, P. Snijders for his help in data collection and E.M. van Leeuwen for genetic imputation.

FamHS (Family Heart Study): The FamHS is funded by R01HL118305 and R01HL117078 NHLBI grants, and 5R01DK07568102 and 5R01DK089256 NIDDK grant.

FHS (Framingham Heart Study): This research was conducted in part using data and resources from the Framingham Heart Study of the National Heart Lung and Blood Institute of the National Institutes of Health and Boston University School of Medicine. The analyses reflect intellectual input and resource development from the Framingham Heart Study investigators participating in the SNP Health Association Resource (SHARe) project. This work was partially supported by the National Heart, Lung and Blood Institute's Framingham Heart Study (Contract Nos. N01-HC-25195 and HHSN268201500001I) and its contract with Affymetrix, Inc for genotyping services (Contract No. N02-HL-6-4278). A portion of this research utilized the Linux Cluster for Genetic Analysis (LinGA-II) funded by the Robert Dawson Evans Endowment of the Department of Medicine at Boston University School of Medicine and Boston Medical Center. This research was partially supported by grant R01-DK089256 from the National Institute of Diabetes and Digestive and Kidney Diseases (MPIs: Ingrid B. Borecki, L. Adrienne Cupples, Kari North).

GENOA (Genetic Epidemiology Network of Arteriopathy): Support for GENOA was provided by the National Heart, Lung and Blood Institute (HL119443, HL118305, HL054464, HL054457, HL054481, HL071917 and HL087660) of the National Institutes of Health. Genotyping was performed at the Mayo Clinic (Stephen T. Turner, MD, Mariza de Andrade PhD, Julie Cunningham, PhD). We thank Eric Boerwinkle, PhD and Megan L. Grove from the Human Genetics Center and Institute of Molecular Medicine and Division of Epidemiology, University of Texas Health Science Center, Houston, Texas, USA for their help with genotyping. We would also like to thank the families that participated in the GENOA study.

GenSalt (Genetic Epidemiology Network of Salt Sensitivity): The Genetic Epidemiology Network of Salt Sensitivity is supported by research grants (U01HL072507, R01HL087263, and R01HL090682) from the National Heart, Lung, and Blood Institute, National Institutes of Health, Bethesda, MD.

GS:SFHS: Generation Scotland received core support from the Chief Scientist Office of the Scottish Government Health Directorates [CZD/16/6] and the Scottish Funding Council [HR03006] and is currently supported by the Wellcome Trust [216767/Z/19/Z]. Genotyping of the GS:SFHS samples was carried out by the Genetics Core Laboratory at the Edinburgh Clinical Research Facility, University of Edinburgh, Scotland and was funded by the Medical Research Council UK and the Wellcome Trust (Wellcome Trust Strategic Award "STratifying Resilience and Depression Longitudinally" (STRADL) Reference 104036/Z/14/Z). Ethical approval for the GS:SFHS study was obtained from the Tayside Committee on Medical Research Ethics (on behalf of the National Health Service).

We are grateful to all the families who took part, the general practitioners and the Scottish School of Primary Care for their help in recruiting them, and the whole Generation Scotland team, which includes interviewers, computer and laboratory technicians, clerical workers, research scientists, volunteers, managers, receptionists, healthcare assistants and nurses.

CH was supported by an MRC Human Genetics Unit programme grant 'Quantitative traits in health and disease' (U. MC_UU_00007/10).

GOLDN (Genetics of Diet and Lipid Lowering Network): Support for the genome-wide association studies in GOLDN was provided by the National Heart, Lung, and Blood Institute

grant U01HL072524-04 and R01HL091357.

HANDLS (Healthy Aging in Neighborhoods of Diversity across the Life Span): The Healthy Aging in Neighborhoods of Diversity across the Life Span (HANDLS) study was supported by the Intramural Research Program of the NIH, National Institute on Aging and the National Center on Minority Health and Health Disparities (project # Z01-AG000513 and human subjects protocol number 09-AG-N248). Data analyses for the HANDLS study utilized the high-performance computational resources of the Biowulf Linux cluster at the National Institutes of Health, Bethesda, MD. (http://biowulf.nih.gov; http://hpc.nih.gov)).

Health ABC (Health, Aging, and Body Composition): Health ABC was funded by the National Institutes of Aging. This research was supported by NIA contracts N01AG62101, N01AG62103, and N01AG62106. The GWAS was funded by NIA grant 1R01AG032098-01A1 to Wake Forest University Health Sciences and genotyping services were provided by the Center for Inherited Disease Research (CIDR). CIDR is fully funded through a federal contract from the National Institutes of Health to The Johns Hopkins University, contract number HHSN268200782096C. This research was supported in part by the Intramural Research Program of the NIH, National Institute on Aging.

HERITAGE (Health, Risk Factors, Exercise Training and Genetics): The HERITAGE Family Study was supported by National Heart, Lung, and Blood Institute grant HL-45670.

HyperGEN (Hypertension Genetic Epidemiology Network): The hypertension network is funded by cooperative agreements (U10) with NHLBI: HL54471, HL54472, HL54473, HL54495, HL54496, HL54497, HL54509, HL54515, and 2 R01 HL55673-12. The study involves: University of Utah: (Network Coordinating Center, Field Center, and Molecular Genetics Lab); Univ. of Alabama at Birmingham: (Field Center and Echo Coordinating and Analysis Center); Medical College of Wisconsin: (Echo Genotyping Lab); Boston University: (Field Center); University of Minnesota: (Field Center and Biochemistry Lab); University of North Carolina: (Field Center); Washington University: (Data Coordinating Center); Weil Cornell Medical College: (Echo Reading Center); National Heart, Lung, & Blood Institute. For a complete list of HyperGEN Investigators: http://www.biostat.wustl.edu/hypergen/Acknowledge.html

JHS (Jackson Heart Study): The Jackson Heart Study is supported by contracts HSN268201300046C, HHSN268201300047C, HHSN268201300048C, HHSN268201300049C, HHSN268201300050C from the National Heart, Lung, and Blood Institute on Minority Health and Health Disparities. The authors acknowledge the Jackson Heart Study team institutions (University of Mississippi Medical Center, Jackson State University and Tougaloo College) and participants for their long-term commitment that continues to improve our understanding of the genetic epidemiology of cardiovascular and other chronic diseases among African Americans.

MESA (Multi-Ethic Study of Atherosclerosis) and the **MESA SHARe** projects are conducted and supported by the National Heart, Lung, and Blood Institute (NHLBI) in collaboration with MESA investigators. Support for MESA is provided by contracts 75N92020D00001, HHSN268201500003I, N01-HC-95159, 75N92020D00005, N01-HC-95160, 75N92020D00002, N01-HC-95161, 75N92020D00003, N01-HC-95162, 75N92020D00006, N01-HC-95163, 75N92020D00004, N01-HC-95164, 75N92020D00007, N01-HC-95165, N01-HC-95166, N01-HC-95167, N01-HC-95168, N01-HC-95169, UL1-TR-000040, UL1-TR-001079, UL1-TR-001420. Funding for SHARe genotyping was provided by NHLBI Contract N02-HL-64278. Genotyping was performed at Affymetrix (Santa Clara, California, USA) and the Broad Institute of Harvard and MIT (Boston, Massachusetts, USA) using the Affymetrix Genome-Wide Human SNP Array 6.0. Also supported in

part by the National Center for Advancing Translational Sciences, CTSI grant UL1TR001881, and the National Institute of Diabetes and Digestive and Kidney Disease Diabetes Research Center (DRC) grant DK063491 to the Southern California Diabetes Endocrinology Research Center. Infrastructure for the CHARGE Consortium is supported in part by the National Heart, Lung, and Blood Institute (NHLBI) grant R01HL105756.

NEO (The Netherlands Epidemiology of Obesity study): The authors of the NEO study thank all individuals who participated in the Netherlands Epidemiology in Obesity study, all participating general practitioners for inviting eligible participants and all research nurses for collection of the data. We thank the NEO study group, Petra Noordijk, Pat van Beelen and Ingeborg de Jonge for the coordination, lab and data management of the NEO study. The genotyping in the NEO study was supported by the Centre National de Génotypage (Paris, France), headed by Jean-Francois Deleuze. The NEO study is supported by the participating Departments, the Division and the Board of Directors of the Leiden University Medical Center, and by the Leiden University, Research Profile Area Vascular and Regenerative Medicine. Dennis Mook-Kanamori is supported by Dutch Science Organization (ZonMW-VENI Grant 916.14.023).

Pelotas Birth Cohort Study (The 1982 Pelotas Birth Cohort Study, Brazil): The 1982 Pelotas Birth Cohort Study is conducted by the Postgraduate Program in Epidemiology at Universidade Federal de Pelotas with the collaboration of the Brazilian Public Health Association (ABRASCO). From 2004 to 2013, the Wellcome Trust supported the study. The International Development Research Center, World Health Organization, Overseas Development Administration, European Union, National Support Program for Centers of Excellence (PRONEX), the Brazilian National Research Council (CNPq), and the Brazilian Ministry of Health supported previous phases of the study.

Genotyping of 1982 Pelotas Birth Cohort Study participants was supported by the Department of Science and Technology (DECIT, Ministry of Health) and National Fund for Scientific and Technological Development (FNDCT, Ministry of Science and Technology), Funding of Studies and Projects (FINEP, Ministry of Science and Technology, Brazil), Coordination of Improvement of Higher Education Personnel (CAPES, Ministry of Education, Brazil).

RS (Rotterdam Study): The Rotterdam Study is supported by the Erasmus MC University Medical Center and Erasmus University Rotterdam; The Netherlands Organisation for Scientific Research (NWO); The Netherlands Organisation for Health Research and Development (ZonMw); the Research Institute for Diseases in the Elderly (RIDE); The Netherlands Genomics Initiative (NGI); the Ministry of Education, Culture and Science; the Ministry of Health, Welfare and Sports; the European Commission (DG XII); and the Municipality of Rotterdam. The contribution of inhabitants, general practitioners and pharmacists of the Ommoord district to the Rotterdam Study is gratefully acknowledged. The contributions to the Management Team specifically and the Rotterdam Study at large of the following persons are pivotal to the daily operations of the study and highly appreciated: Jacobus Lubbe, Gabriëlle Bakker, Eric Neeleman, Jeannette Vergeer, Anneke Korving, Pauline van Wijngaarden, Jolande Verkroost – van Heemst, Silvan Licher, and Isabel van der Velpen.

The generation and management of GWAS genotype data for the Rotterdam Study was executed by the Human Genotyping Facility of the Genetic Laboratory of the Department of Internal Medicine, Erasmus MC, Rotterdam, The Netherlands. The GWAS datasets are supported by the Netherlands Organisation of Scientific Research NWO Investments (nr. 175.010.2005.011, 911-03-012), the

Genetic Laboratory of the Department of Internal Medicine, Erasmus MC, the Research Institute for Diseases in the Elderly (014-93-015; RIDE2), the Netherlands Genomics Initiative (NGI)/Netherlands Organisation for Scientific Research (NWO) Netherlands Consortium for Healthy Aging (NCHA), project nr. 050-060-810. We thank Pascal Arp, Mila Jhamai, Marijn Verkerk, Lizbeth Herrera, Marjolein Peters and Carolina Medina-Gomez for their help in creating the GWAS database, and Karol Estrada, Yurii Aulchenko and Carolina Medina-Gomez for the creation and analysis of imputed data.

SCES (Singapore Chinese Eye Study), SiMES (Singapore Malay Eye Study), SINDI (Singapore Indian Eye Study): The Singapore Malay Eye Study (SiMES), the Singapore Indian Eye Study (SINDI), and the Singapore Chinese Eye Study (SCES) are supported by the National Medical Research Council (NMRC), Singapore (grants 0796/2003, 1176/2008, 1149/2008, STaR/0003/2008, 1249/2010, CG/SERI/2010, CIRG/1371/2013, and CIRG/1417/2015), and Biomedical Research Council (BMRC), Singapore (08/1/35/19/550 and 09/1/35/19/616). Ching-Yu Cheng is supported by an award from NMRC (CSA/033/2012). The Singapore Tissue Network and the Genome Institute of Singapore, Agency for Science, Technology and Research, Singapore provided services for tissue archival and genotyping, respectively.

SCHS-CHD (Singapore Chinese Health Study - Coronary Heart Disease): The Singapore Chinese Health Study is supported by the National Institutes of Health, USA (RO1 CA144034 and UM1 CA182876), the nested case-control study of myocardial infarction by the Singapore National Medical Research Council (NMRC 1270/2010) and genotyping by the HUJ-CREATE Programme of the National Research Foundation, Singapore (Project Number 370062002).

SP2 (Singapore Prospective Study Program): The SP2 study is supported by individual research and clinical scientist award schemes from the National Medical Research Council (NMRC) and the Biomedical Research Council (BMRC) of Singapore, the Singapore Ministry of Health, National University of Singapore and National University Health System, Singapore. The Genome Institute of Singapore, Agency for Science, Technology and Research, Singapore provided services for genotyping.

WGHS (Women's Genome Health Study): The WGHS is supported by the National Heart, Lung, and Blood Institute (HL043851 and HL080467) and the National Cancer Institute (CA047988 and UM1CA182913), with collaborative scientific support and funding for genotyping provided by Amgen.

STAGE 2 STUDY ACKNOWLEDGMENTS

Infrastructure for the CHARGE Consortium is supported in part by the National Heart, Lung, and Blood Institute grant R01HL105756. Infrastructure for the Gene-Lifestyle Working Group is supported by the National Heart, Lung, and Blood Institute grant R01HL118305.

AA-DHS (African American Diabetes Heart Study): The investigators acknowledge the cooperation of our Diabetes Heart Study (DHS) and AA-DHS participants. This work was supported by NIH R01 DK071891, R01 HL092301, R01 AG058921, and the General Clinical Research Center of Wake Forest School of Medicine M01-RR-07122.

BES (Beijing Eye Study): The Beijing Eye Study has been funded by the National Natural Science Foundation of China (grant 81170890, 81570835) and the Beijing Municipal of Health Reform and Development Project (grant 2019-4).

BRIGHT (British Genetics of Hypertension): This work was supported by the Medical Research Council of Great Britain (grant number G9521010D) and the British Heart Foundation (grant number PG/02/128). The BRIGHT study is extremely grateful to all the patients who participated in the study and the BRIGHT nursing team. This work forms part of the research program of the National Institutes of Health Research (NIHR Cardiovascular Biomedical Research) Cardiovascular Biomedical Unit at Barts and The London, QMUL.

DESIR (Data from an Epidemiological Study on the Insulin Resistance): The DESIR study has been funded by INSERM contracts with Caisse nationale de l'assurance Maladie des Travailleurs Salariés (CNAMTS), Lilly, Novartis Pharma, and Sanofi-Aventis; INSERM (Réseaux en Santé Publique, Interactions entre les déterminants de la santé, Cohortes Santé TGIR 2008); the Association Diabète Risque Vasculaire; the Fédération Française de Cardiologie; La Fondation de France; Association de Langue Française pour l'Etude du Diabète et des Maladies Métaboliques (ALFEDIAM)/Société Francophone de Diabétologie (SFD); l'Office National Interprofessionnel des Vins (ONIVINS); Ardix Medical; Bayer Diagnostics; Becton Dickinson; Cardionics; Merck Santé; Novo Nordisk; Pierre Fabre; Roche; Topcon.

The DESIR Study Group. INSERM U1018: B. Balkau, P. Ducimetière, E. Eschwège; INSERM U367: F. Alhenc-Gelas; CHU D'Angers: Y Gallois, A. Girault; Centre de Recherche des Cordeliers, INSERM U1138, Bichat Hospital: F. Fumeron, M. Marre, R. Roussel; CHU de Rennes: F. Bonnet; CNRS UMR8090, Lille: A. Bonnefond, S. Cauchi, P. Froguel; Centres d'Examens de Santé: Alençon, Angers, Blois, Caen, Chateauroux, Chartres, Cholet, Le Mans, Orléans, Tours; Institut de Recherche Médecine Générale: J. Cogneau; General practitioners of the region; Institut inter-régional pour la Santé: C. Born, E. Caces, M. Cailleau, O Lantieri, J.G. Moreau.

DFTJ (Dongfeng-Tongji Cohort Study): This work was supported by grants from the Foundation of National Key Program of Research and Development of China (2016YFC0900800), the Programme of Introducing Talents of Discipline, the grants from the National Natural Science Foundation (grant NSFC-81473051, 81522040 and 81230069), and the Program for the New Century Excellent Talents in University (NCET-11-0169).

DHS (Diabetes Heart Study): The authors thank the investigators, staff, and participants of the DHS for their valuable contributions. This study was supported by the National Institutes of Health through R01 HL067348, R01 HL092301, and R01 AG058921.

DR's EXTRA (Dose-Responses to Exercise Training): The study was supported by grants from Ministry of Education and Culture of Finland (722 and 627; 2004-2010); Academy of Finland (102318, 104943, 123885, 211119); European Commission FP6 Integrated Project (EXGENESIS), LSHM-CT-2004-005272; City of Kuopio; Juho Vainio Foundation; Finnish Diabetes Association; Finnish Foundation for Cardiovascular Research; Kuopio University Hospital; Päivikki and Sakari Sohlberg Foundation; Social Insurance Institution of Finland 4/26/2010.

EGCUT (Estonian Genome Center - University of Tartu (Estonian Biobank)): This study was funded by EU H2020 grant 692145, 676550, 654248, Estonian Research Council Grant IUT20-60, IUT24-6, PUT1660, and European Union through the European Regional Development Fund (Project No. 2014-2020.4.01.15-0012), GENTRANSMED end Project No. 2014-2020.4.01.16-0125, NIH – GIANT and NIH-BMI Grant 2R01DK075787-06A1.

EPIC (European Prospective Investigation into Cancer and Nutrition)-Norfolk: The EPIC-Norfolk study (DOI 10.22025/2019.10.105.00004) has received funding from the Medical Research Council (MR/N003284/1, MC-PC_13048 and MC-UU_12015/1). We are grateful to all the participants who have been part of the project and to the many members of the study teams at the University of Cambridge who have enabled this research.

The EPIC-InterAct Case-Cohort Study: We thank all EPIC participants and staff for their contribution to the study. The InterAct study received funding from the European Union (Integrated Project LSHM-CT-2006-037197 in the Framework Programme 6 of the European Community). We thank staff from the Technical, Field Epidemiology and Data Functional Group Teams of the Medical Research Council Epidemiology Unit in Cambridge, UK, for carrying out sample preparation, DNA provision and quality control, genotyping and data handling work. We specifically thank Sarah Dawson for coordinating the sample provision for biomarker measurements, Abigail Britten for coordinating DNA sample provision and genotyping of candidate markers, Nicola Kerrison, Chris Gillson and Abigail Britten for data provision and genotyping quality control, Matt Sims for writing the technical laboratory specification for the intermediate pathway biomarker measurements and for overseeing the laboratory work.

GeneSTAR (Genetic Studies of Atherosclerosis Risk): GeneSTAR was supported by National Institutes of Health grants from the National Heart, Lung, and Blood Institute (HL49762, HL58625, HL071025, U01 HL72518, HL087698, HL092165, HL099747, and K23HL105897), National Institute of Nursing Research (NR0224103), National Institute of Neurological Disorders and Stroke (NS062059), and by grants from the National Center for Research Resources to the Johns Hopkins General Clinical Research Center (M01-RR000052) and the Johns Hopkins Institute for Clinical & Translational Research (UL1 RR 025005).

GLACIER (Gene x Lifestyle Interactions and Complex Traits Involved in Elevated Disease Risk): We thank the participants, health professionals and data managers involved in the Västerbotten Intervention Programme. We are also grateful to the staff of the Northern Sweden Biobank for preparing materials and to K Enqvist and T Johansson (Västerbottens County Council, Umeå, Sweden) for DNA preparation. The Västerbotten Intervention Programme is finance by Västerbotten County Council. The current study was supported by Novo Nordisk (PWF), the Swedish Research Council (PWF), the Swedish Heart Lung Foundation (PWF), the European Research Council (PWF), and the Skåne Health Authority (PWF). This study was also supported by the Swedish Research Council, Strategic Research Area Exodiab (Dnr 2009-1039), the Swedish

Foundation for Strategic Research (Dnr IRC15-0067) and the Swedish Research Council, Linnaeus grant (Dnr 349-2006-237).

GRAPHIC (Genetic Regulation of Arterial Pressure of Humans in the Community): The GRAPHIC Study was funded by the British Heart Foundation (BHF/RG/2000004). This work falls under the portfolio of research supported by the NIHR Leicester Cardiovascular Biomedical Research Unit. CPN and NJS are funded by the BHF and NJS is a NIHR Senior Investigator.

HCHS/SOL (Hispanic Community Health Study/ Study of Latinos): The baseline examination of HCHS/SOL was supported by contracts from the National Heart, Lung, and Blood Institute (NHLBI) to the University of North Carolina (N01-HC65233), University of Miami (N01-HC65234), Albert Einstein College of Medicine (N01-HC65235), Northwestern University (N01-HC65236), and San Diego State University (N01-HC65237). The National Institute on Minority Health and Health Disparities, National Institute on Deafness and Other Communication Disorders, National Institute of Dental and Craniofacial Research (NIDCR), National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK), National Institute of Neurological Disorders and Stroke, and NIH Office of Dietary Supplements additionally contributed funding to HCHS/SOL. The Genetic Analysis Center at the University of Washington was supported by NHLBI and NIDCR contracts (HHSN268201300005C AM03 and MOD03). Additional analysis support was provided by 1R01DK101855-01 and 13GRNT16490017. Genotyping was also supported by National Center for Advancing Translational Sciences UL1TR000124 and NIDDK DK063491 to the Southern California Diabetes Endocrinology Research Center. This research was also supported in part by the Intramural Research Program of the NIDDK, contract no. HHSB268201200054C, and Illumina.

HRS (Health & Retirement Study): HRS is supported by the National Institute on Aging (NIA U01AG009740 and R03 AG046389). Genotyping was funded separately by NIA (RC2 AG036495, RC4 AG039029). Our genotyping was conducted by the NIH Center for Inherited Disease Research (CIDR) at Johns Hopkins University. Genotyping quality control and final preparation of the data were performed by the Genetics Coordinating Center at the University of Washington.

HyperGEN-AXIOM (Hypertension Genetic Epidemiology Network): The study was support by the National Institutes of Health, the National Heart, Lung, Blood Institute grant HL086718.

INGI-CARL (Italian Network Genetic Isolates): This study was partially supported by Regione FVG (L.26.2008) and Italian Ministry of Health (GR-2011-02349604).

INGI-FVG (Italian Network Genetic Isolates): This study was partially supported by Regione FVG (L.26.2008) and Italian Ministry of Health (GR-2011-02349604).

IRAS (Insulin Resistance Atherosclerosis Study): The IRAS is supported by the National Heart Lung Institute (HL047887, HL047889, HL047890, and HL47902). Genotyping and analysis for this study was supported by the GUARDIAN Consortium with grant support from the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK; DK085175 and DK118062) and in part by UL1TR000124 (CTSI) and DK063491 (DRC). The authors thank study investigators, staff, and participants for their valuable contributions.

IRASFS (Insulin Resistance Atherosclerosis Family Study): The IRASFS is supported by the National Heart Lung and Blood Institute (HL060944, HL061019, and HL060919). Genotyping and analysis for this study was supported by the GUARDIAN Consortium with grant support from the

National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK; DK085175 and DK118062) and in part by UL1TR000124 (CTSI) and DK063491 (DRC). The authors thank study investigators, staff, and participants for their valuable contributions.

JUPITER (Justification for the Use of Statins in Primary Prevention: An Intervention Trial Evaluating Rosuvastatin): Support for genotype data collection and collaborative genetic analysis in JUPITER was provided by Astra-Zeneca.

KORA (Cooperative Health Research in the Augsburg Region): The KORA study was initiated and financed by the Helmholtz Zentrum München – German Research Center for Environmental Health, which is funded by the German Federal Ministry of Education and Research (BMBF) and by the State of Bavaria. Furthermore, KORA research was supported within the Munich Center of Health Sciences (MC-Health), Ludwig-Maximilians-Universität, as part of LMUinnovativ.

LBC1936 (Lothian Birth Cohort 1936): We thank the LBC1936 cohort participants and team members who contributed to these studies. Phenotype collection was supported by Age UK (The Disconnected Mind project). Genotyping was funded by the BBSRC (BB/F019394/1). The work was undertaken by The University of Edinburgh Centre for Cognitive Ageing and Cognitive Epidemiology, part of the cross council Lifelong Health and Wellbeing Initiative (MR/K026992/1). Funding from the BBSRC and Medical Research Council (MRC) is gratefully acknowledged. We gratefully acknowledge the contribution of Lothian Birth Cohort co-author Professor John M. Starr, who died prior to the publication of this manuscript.

LifeLines (Netherlands Biobank): The Lifelines Cohort Study, and generation and management of GWAS genotype data for the Lifelines Cohort Study is supported by the Netherlands Organization of Scientific Research NWO (grant 175.010.2007.006), the Economic Structure Enhancing Fund (FES) of the Dutch government, the Ministry of Economic Affairs, the Ministry of Education, Culture and Science, the Ministry for Health, Welfare and Sports, the Northern Netherlands Collaboration of Provinces (SNN), the Province of Groningen, University Medical Center Groningen, the University of Groningen, Dutch Kidney Foundation and Dutch Diabetes Research Foundation.

The authors wish to acknowledge the services of the Lifelines Cohort Study, the contributing research centers delivering data to Lifelines, and all the study participants.

LLFS (The Long Life Family Study): The study is supported by the National Institute on Aging (NIA) grant U01AG023746, U01AG023712, U01AG023749, U01AG023755, U01AG023744, and U19AG063893.

Loyola GxE (Kingston Gene-by-environment; subset of International Collaborative Study of Hypertension in Blacks (ICSHIB)): The Loyola GxE project was supported by NIH Grant R01HL53353.

Loyola SPT (Spanish Town; subset of International Collaborative Study of Hypertension in Blacks (ICSHIB)): The Loyola SPT project was supported by NIH Grant R01HL53353.

NESDA (Netherlands Study of Depression and Anxiety): Funding was obtained from the Netherlands Organization for Scientific Research (Geestkracht program grant 10-000-1002); the Center for Medical Systems Biology (CSMB, NWO Genomics), Biobanking and Biomolecular Resources Research Infrastructure (BBMRI-NL), VU University's Institutes for Health and Care Research (EMGO+) and Neuroscience Campus Amsterdam, University Medical Center Groningen,

Leiden University Medical Center, National Institutes of Health (NIH, R01D0042157-01A, MH081802, Grand Opportunity grants 1RC2 MH089951 and 1RC2 MH089995). Part of the genotyping and analyses were funded by the Genetic Association Information Network (GAIN) of the Foundation for the National Institutes of Health.Computing was supported by BiG Grid, the Dutch e-Science Grid, which is financially supported by NWO.

PREVEND (The Prevention of REnal and Vascular ENd stage Disease study): PREVEND genetics is supported by the Dutch Kidney Foundation (Grant E033), the EU project grant GENECURE (FP-6 LSHM CT 2006 037697), the National Institutes of Health (grant 2R01LM010098), The Netherlands organization for health research and development (NWO-Groot grant 175.010.2007.006, NWO VENI grant 916.761.70, ZonMw grant 90.700.441).

RHS (Ragama Health Study): The RHS was supported by the Grant of National Center for Global Health and Medicine (NCGM).

SHEEP (Stockholm Heart Epidemiology Project): This study was supported by grants from the Swedish Research Council for Health, Working Life and Welfare (http://www.forte.se/en/), the Stockholm County Council (http://www.sll.se/om-landstinget/Information-in-English1/), the Swedish Research Council (http://www.vr.se/inenglish.4.12fff4451215cbd83e4800015152.html), the Swedish Heart and Lung Foundation (https://www.hjart-lungfonden.se/HLF/Om-Hjart-lungfonden/About-HLF/), and the Cardiovascular Programme at Karolinska Institutet (http://ki.se/en/mmk/cardiovascular-research-networks).

SHIP (Study of Health in Pomerania): SHIP is part of the Community Medicine Research net of the University of Greifswald, Germany, which is funded by the Federal Ministry of Education and Research (grants no. 01ZZ9603, 01ZZ0103, and 01ZZ0403), the Ministry of Cultural Affairs as well as the Social Ministry of the Federal State of Mecklenburg-West Pomerania, and the network 'Greifswald Approach to Individualized Medicine (GANI_MED)' funded by the Federal Ministry of Education and Research (grant 03IS2061A). Genome-wide data were supported by the Federal Ministry of Education and Research (grant no. 03ZIK012) and a joint grant from Siemens Healthcare, Erlangen, Germany and the Federal State of Mecklenburg- West Pomerania. The University of Greifswald is a member of the 'Center of Knowledge Interchange' program of the Siemens AG.

SWHS/SMHS (Shanghai Women's Health Study/ Shanghai Men's Health Study): We thank all the individuals who took part in these studies and all the researchers who have enabled this work to be carried out. The Shanghai Women's Health Study and the Shanghai Men's Health Study are supported by research grants UM1CA182910 and UM1CA173640 from the U.S. National Cancer Institute, respectively.

TWINGENE (TwinGene of the Swedish Twin Registry): The Swedish Twin Registry is financially supported by Karolinska Institutet. TwinGene project received funding from the Swedish Research Council (M-2005-1112), GenomEUtwin (EU/QLRT-2001-01254; QLG2-CT-2002-01254), NIH DK U01-066134, The Swedish Foundation for Strategic Research (SSF) and the Heart and Lung foundation no. 20070481

YFS (The Cardiovascular Risk in Young Finns Study): The Young Finns Study has been financially supported by the Academy of Finland: grants 286284, 134309 (Eye), 126925, 121584, 124282, 129378 (Salve), 117787 (Gendi), and 41071 (Skidi); the Social Insurance Institution of

Finland; Kuopio, Tampere and Turku University Hospital Medical Funds (grant X51001); Juho Vainio Foundation; Paavo Nurmi Foundation; Finnish Foundation for Cardiovascular Research ; Finnish Cultural Foundation; Tampere Tuberculosis Foundation; Emil Aaltonen Foundation; Yrjö Jahnsson Foundation; Signe and Ane Gyllenberg Foundation; and Diabetes Research Foundation of Finnish Diabetes Association.

The expert technical assistance in the statistical analyses by Leo-Pekka Lyytikäinen and Irina Lisinen is gratefully acknowledged.

Figure S1: Flow Chart of Study Design

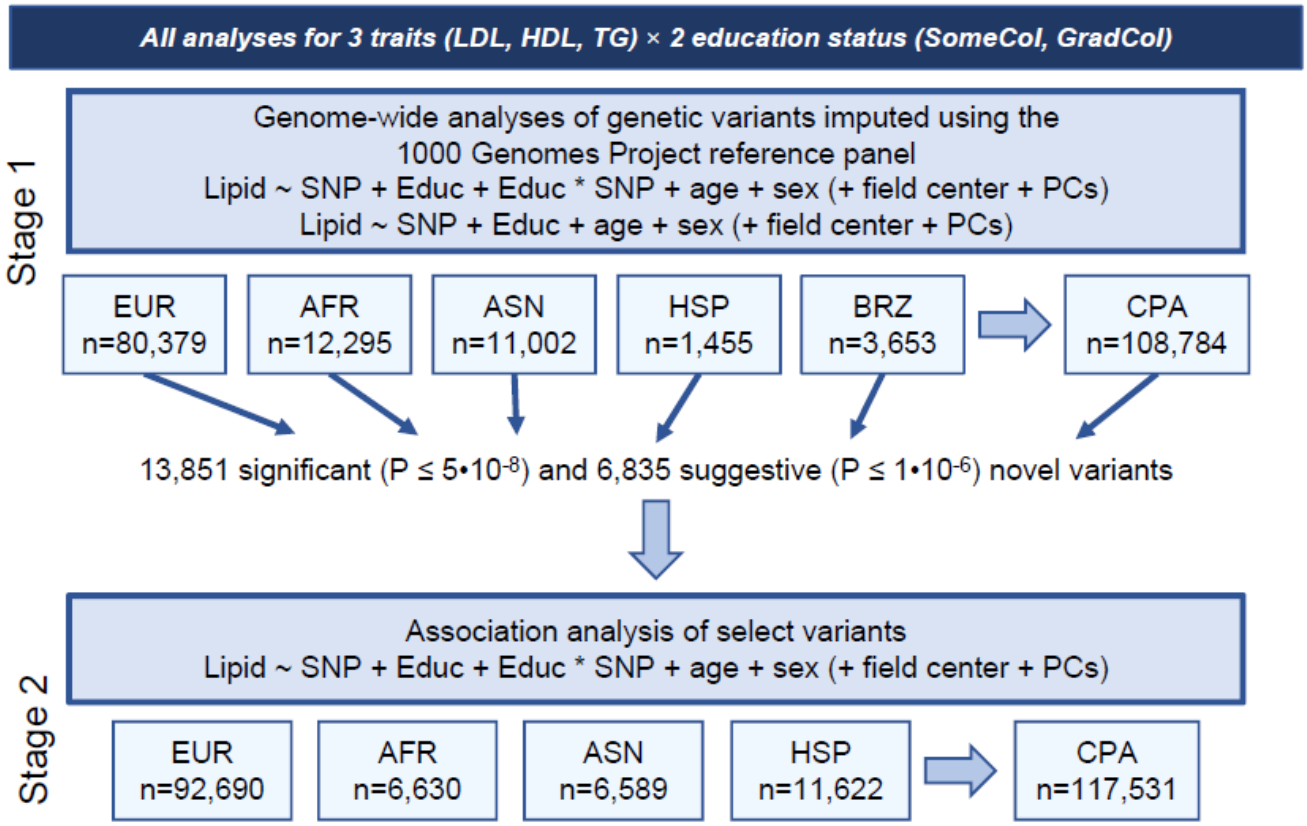
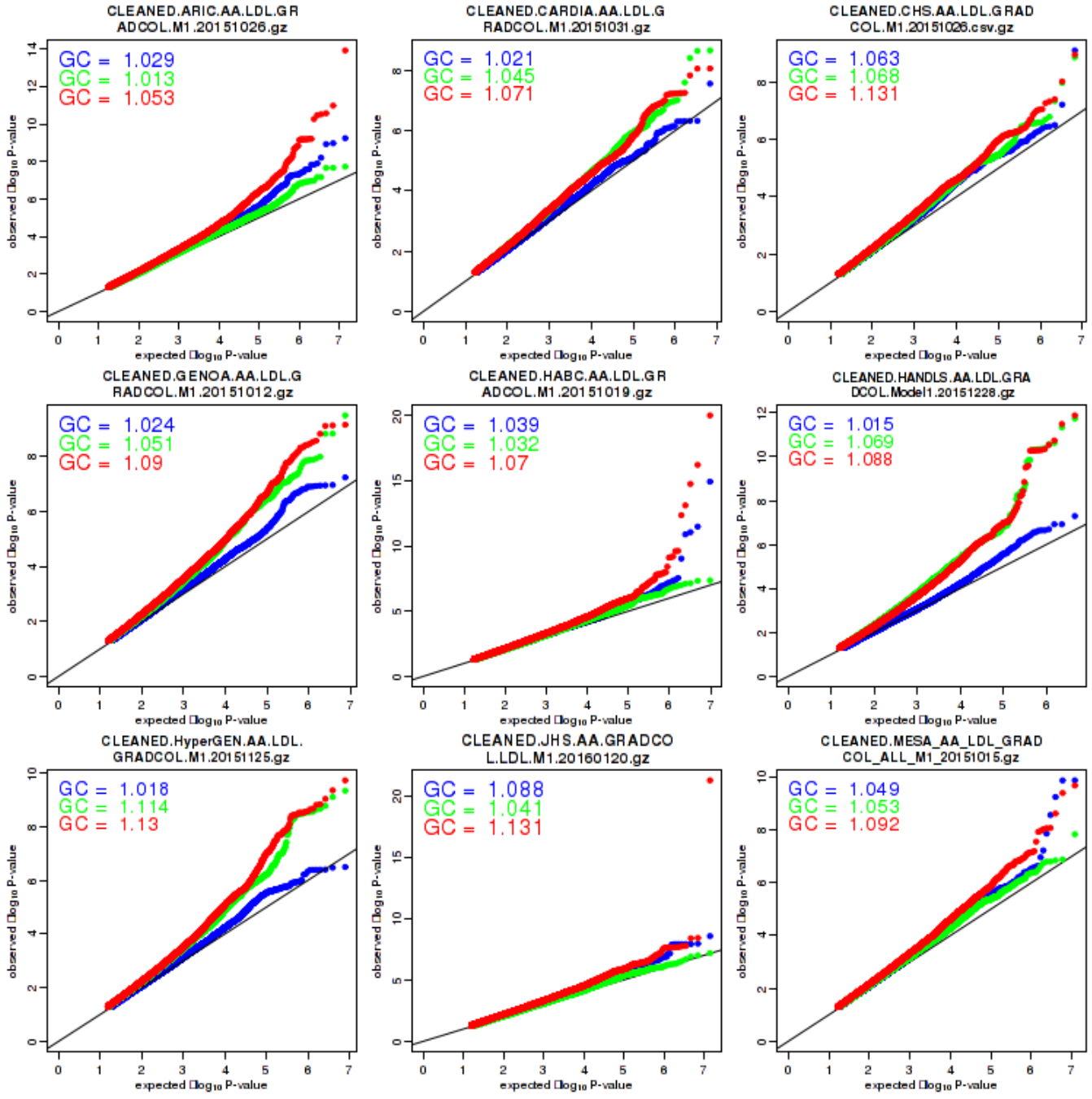
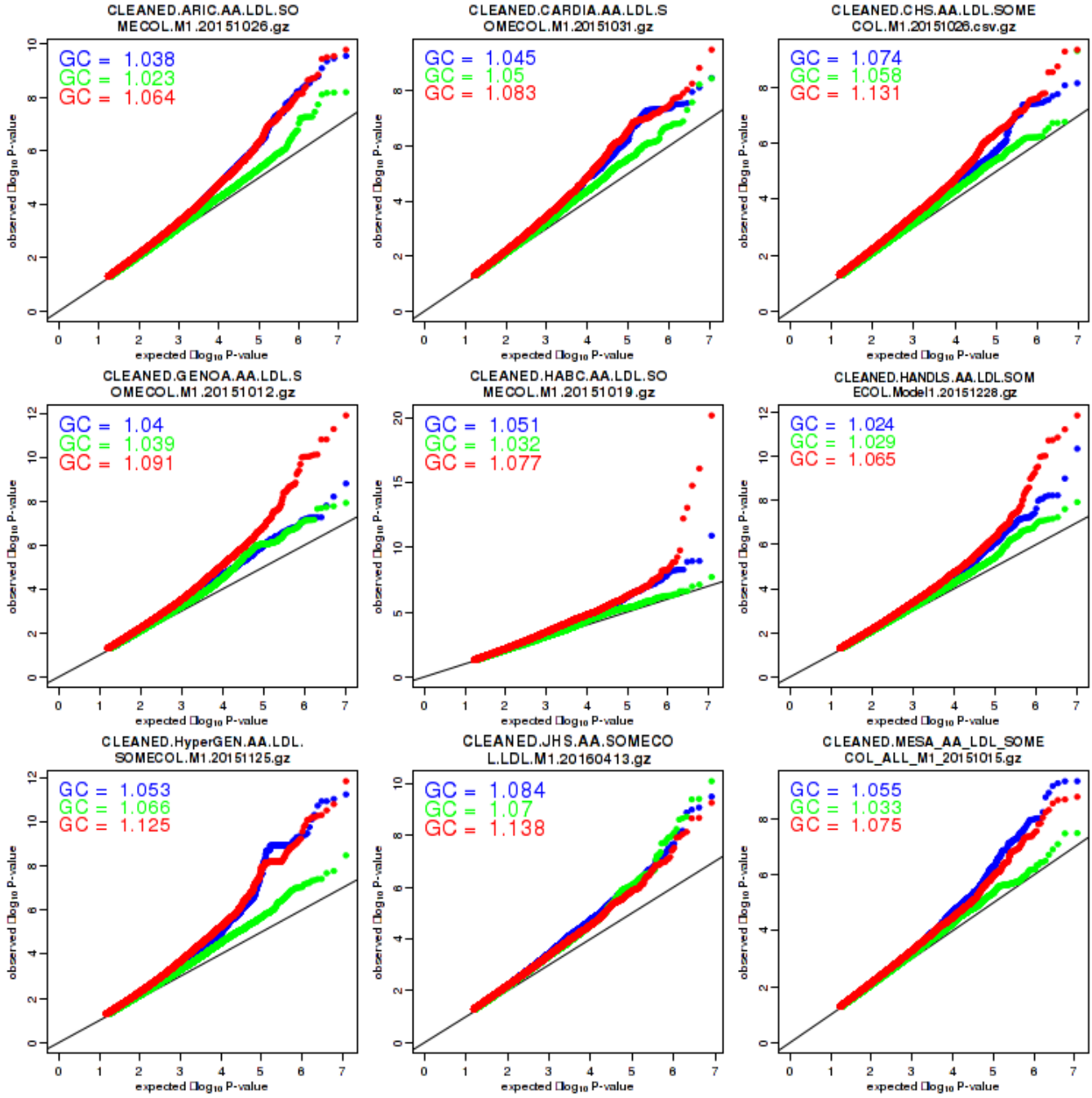


Figure S2A: QQ Plots of African-Ancestry Cohorts.

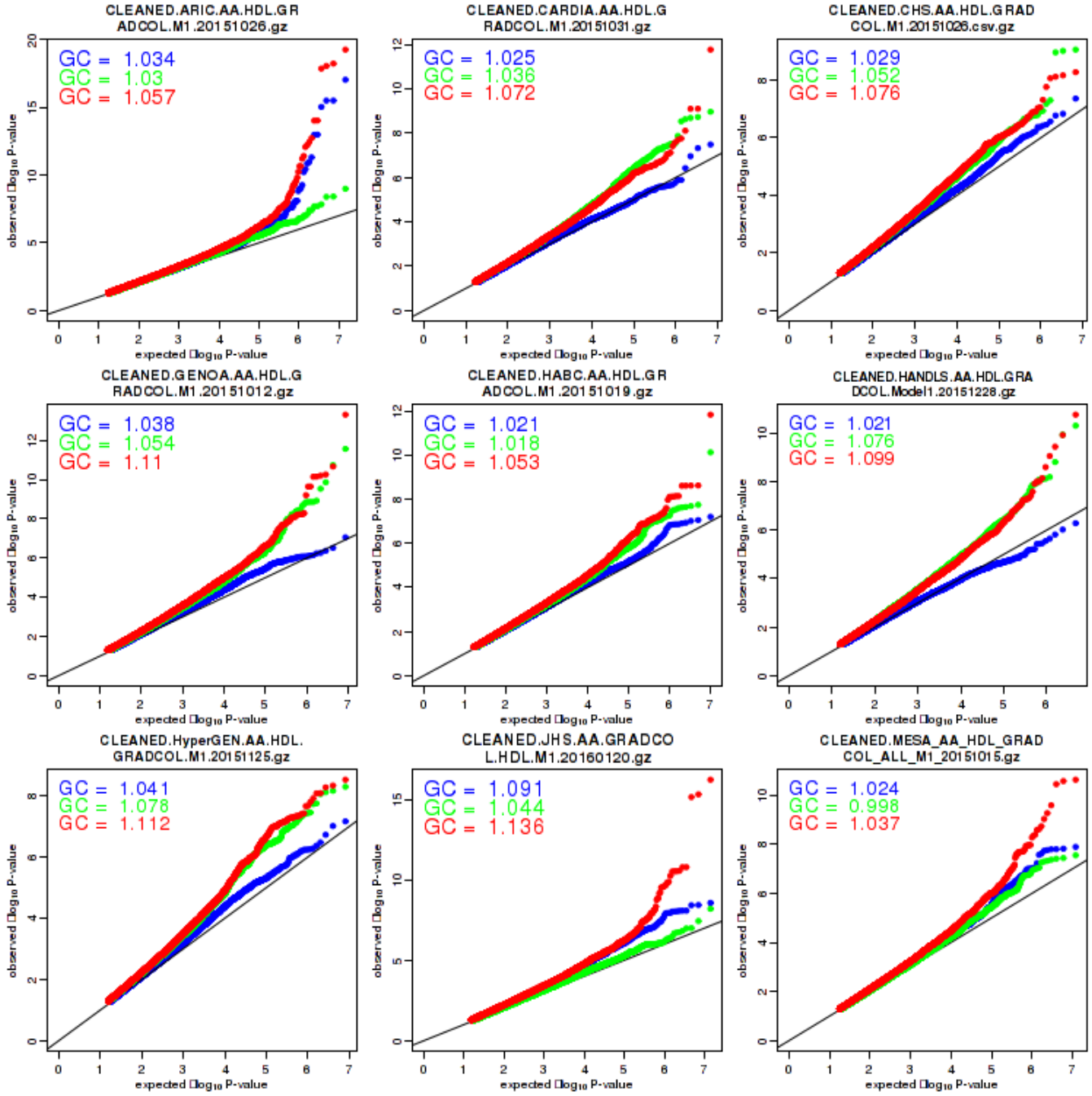
<u>AFR LDL GradCol M1: DF >20</u>



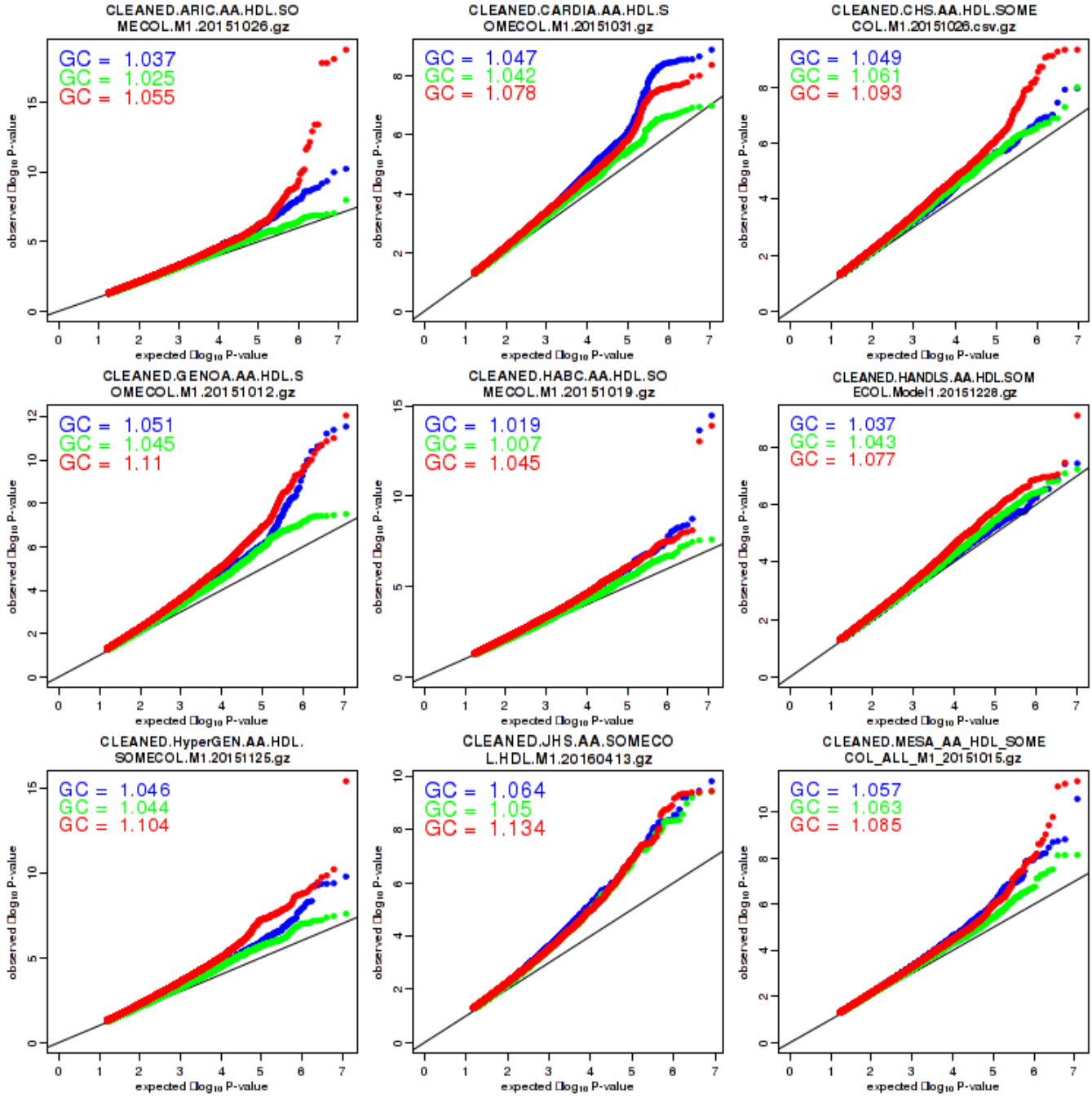
AFR LDL SomeCol M1: DF >20



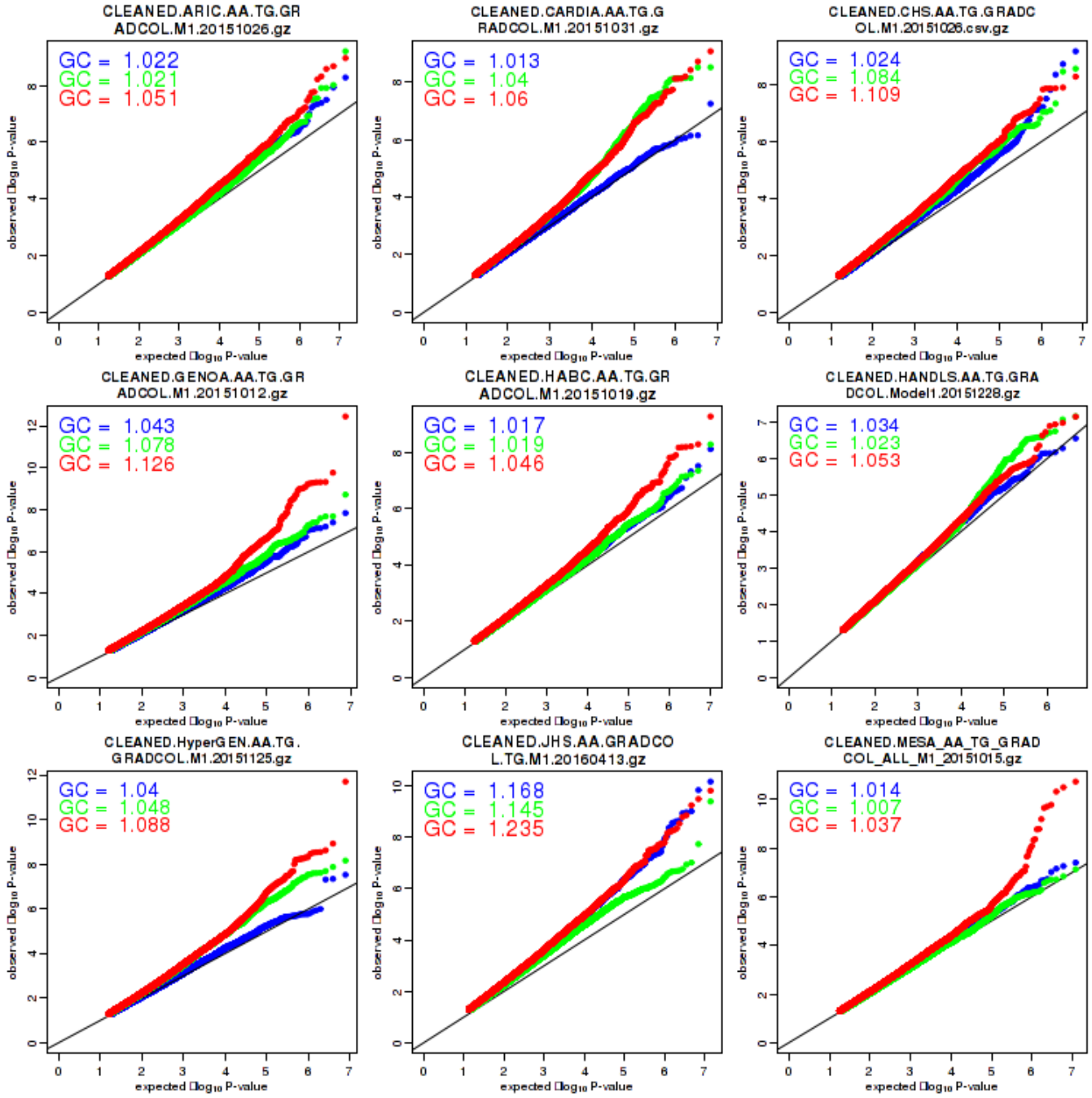
AFR HDL GradCol M1: DF >20



AFR HDL SomeCol M1: DF >20



AFR TG GradCol M1: DF >20



AFR TG SomeCol M1: DF >20

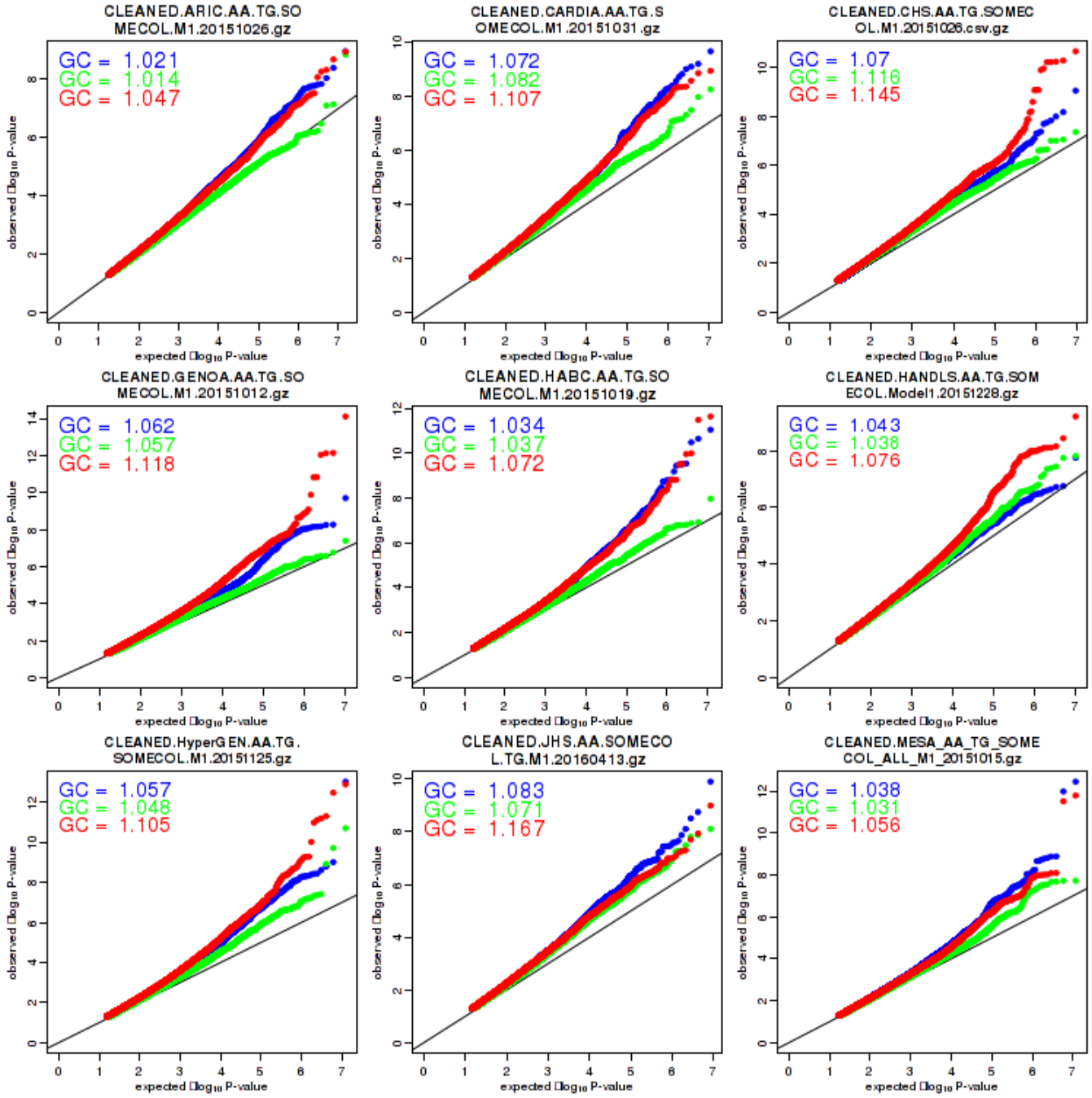
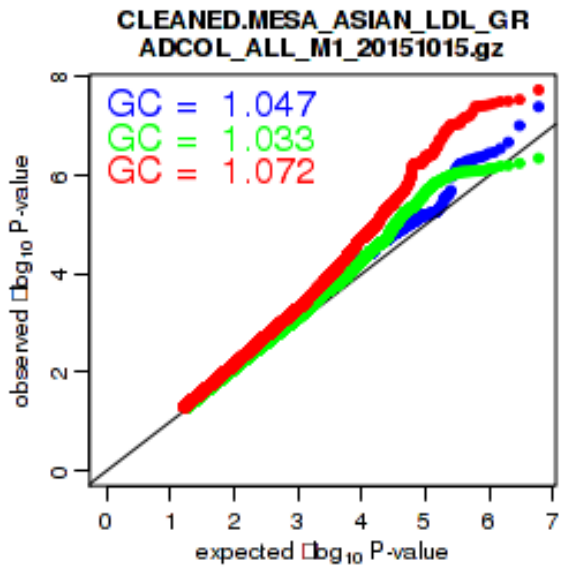
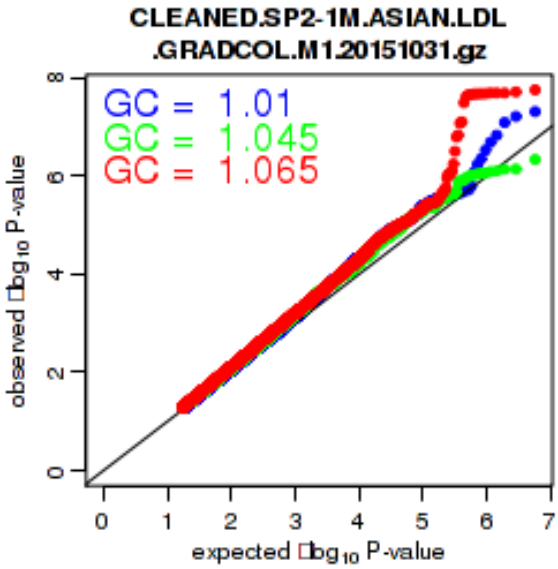
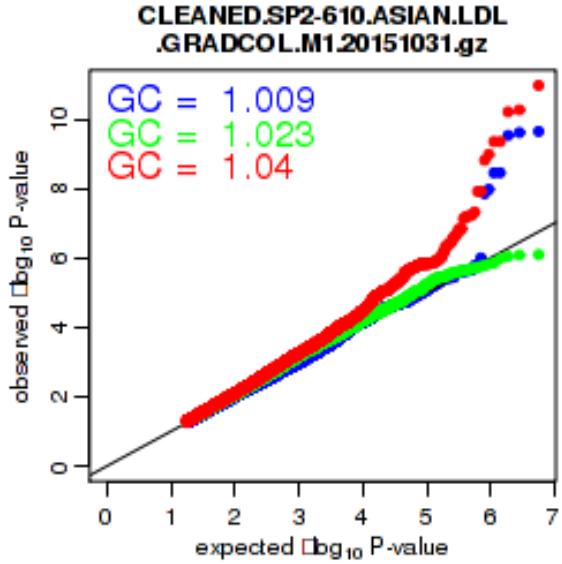


Figure S2B: QQ Plots of Asian-Ancestry Cohorts

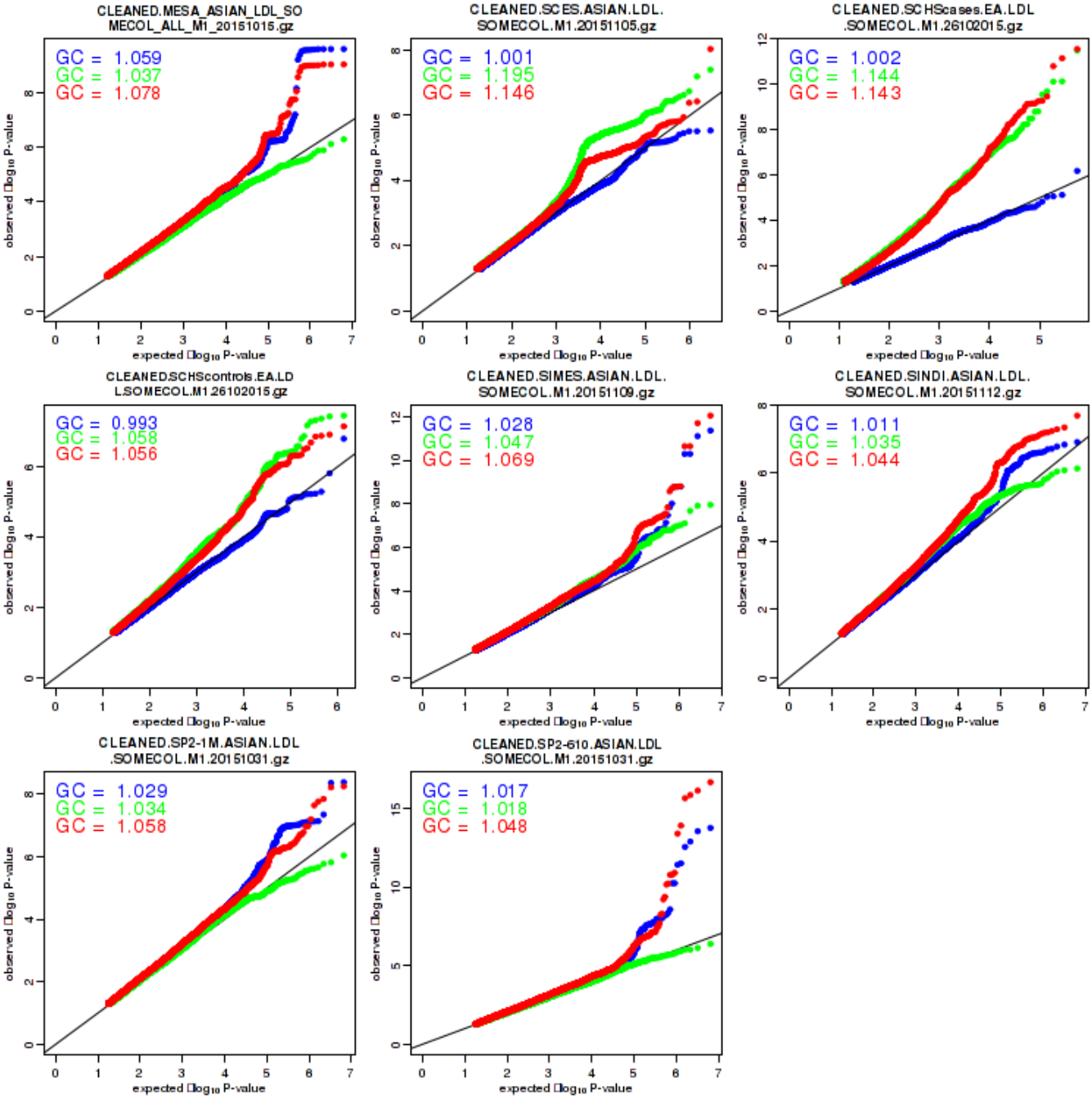
<u>EAS LDL GradCol M1: DF >20</u>



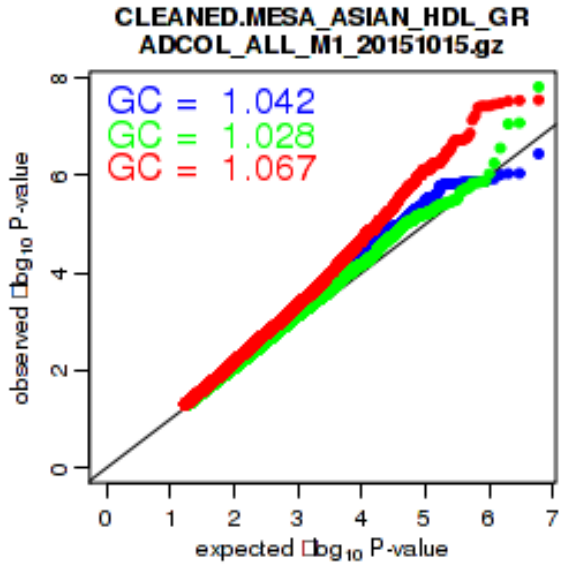


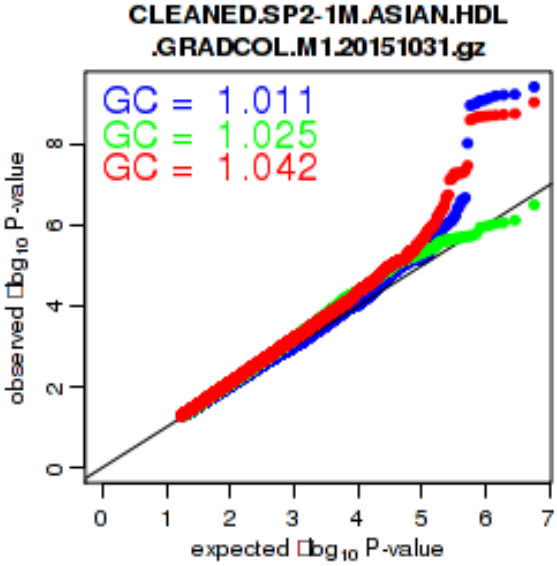


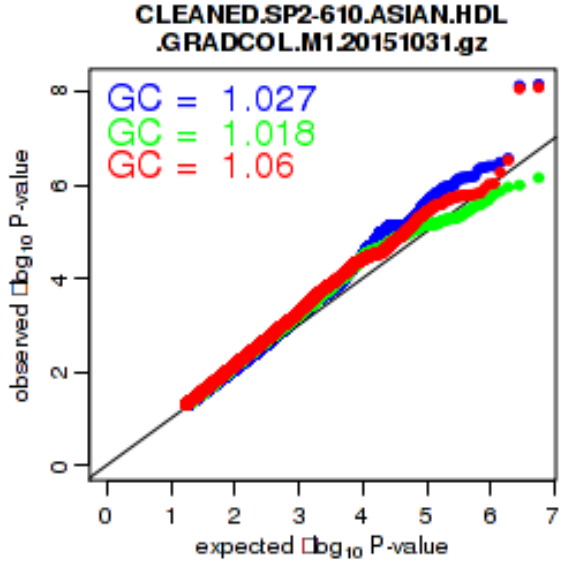
EAS LDL SomeCol M1: DF >20



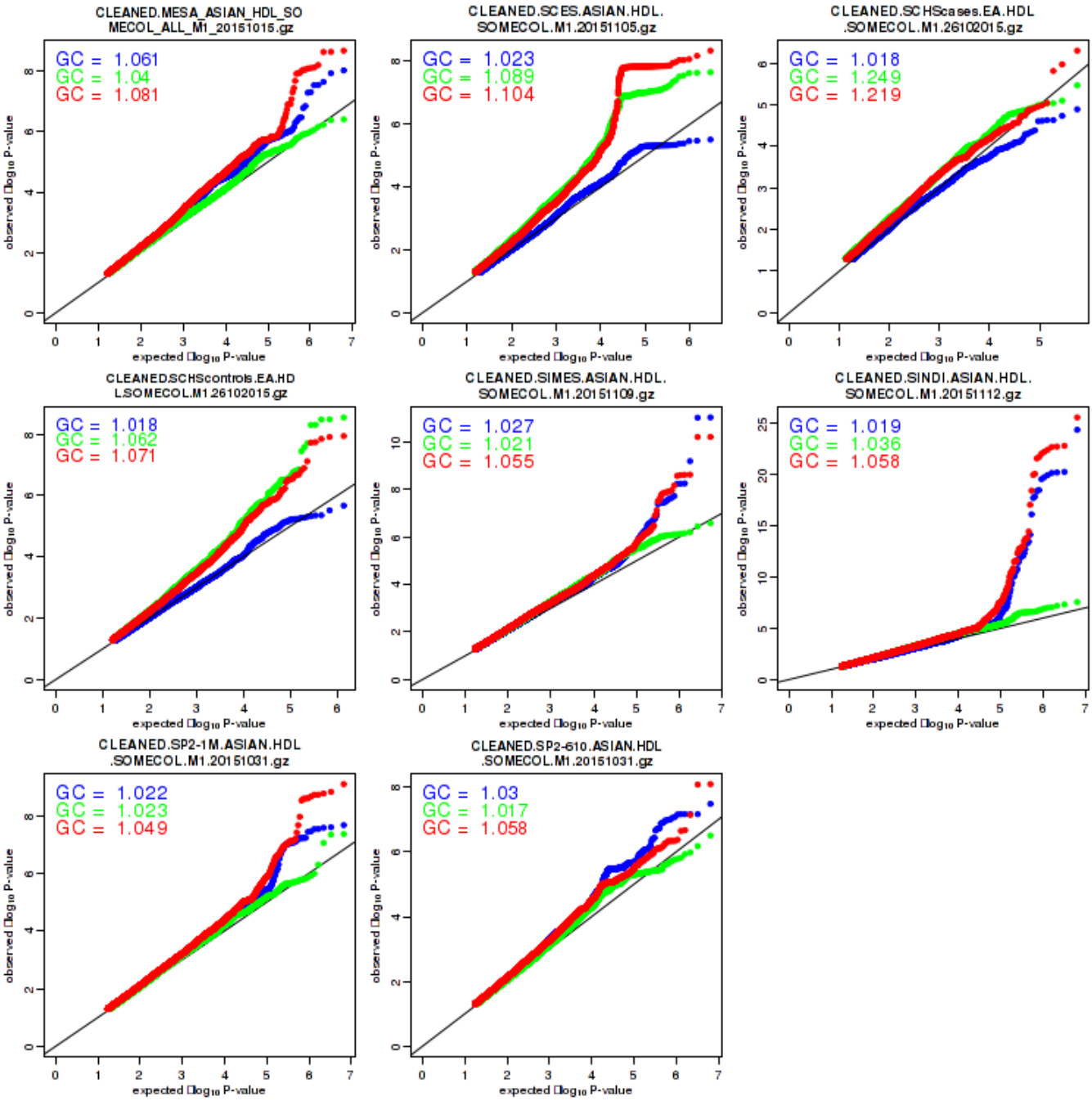
EAS HDL GradCol M1: DF >20



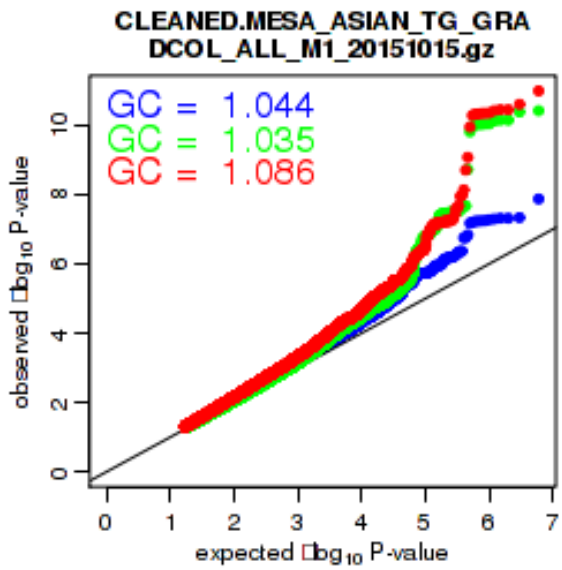


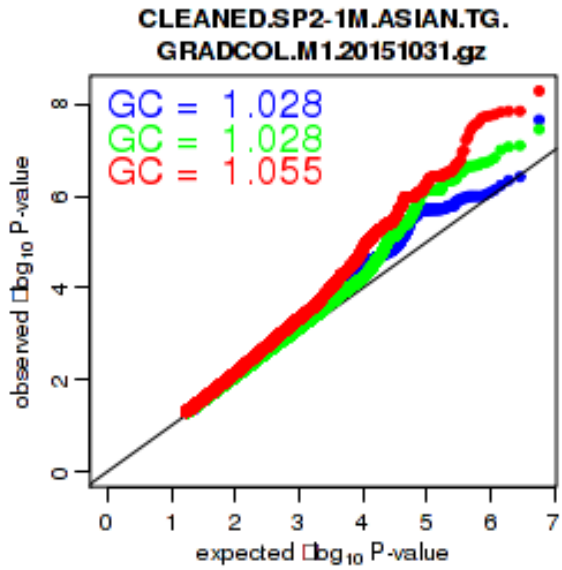


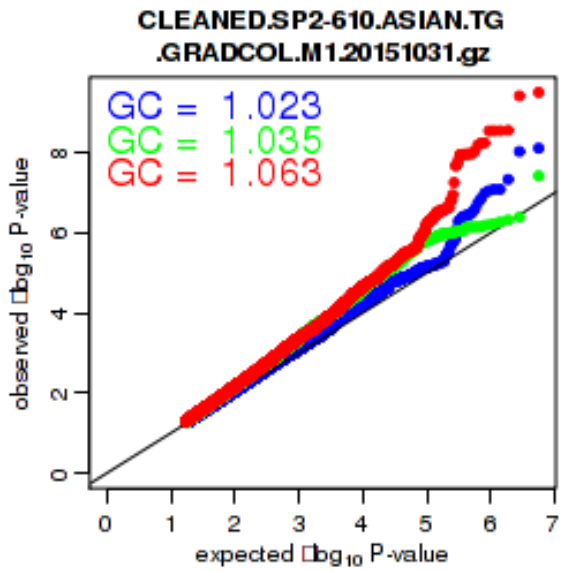
EAS HDL SomeCol M1: DF >20



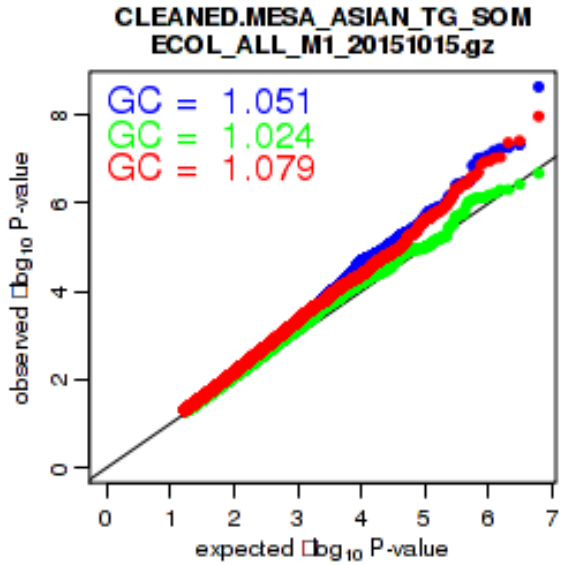
EAS TG GradCol M1: DF >20

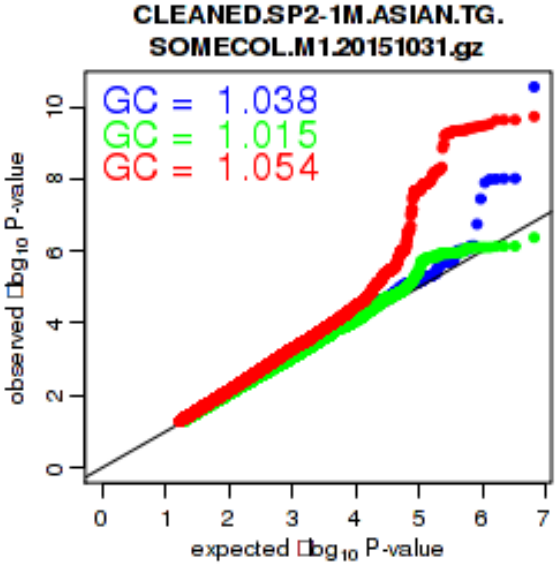






EAS TG SomeCol M1: DF >20





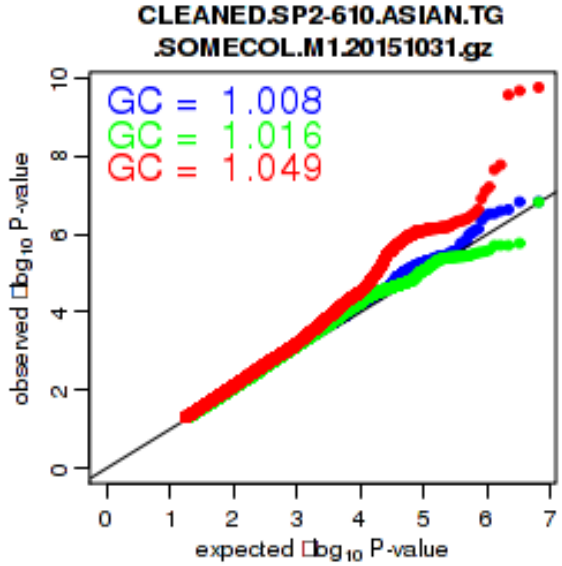
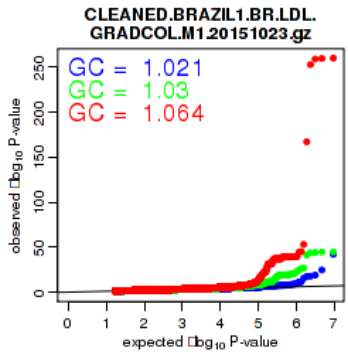
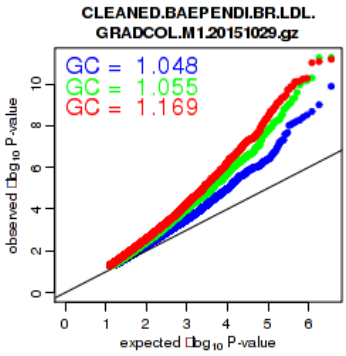


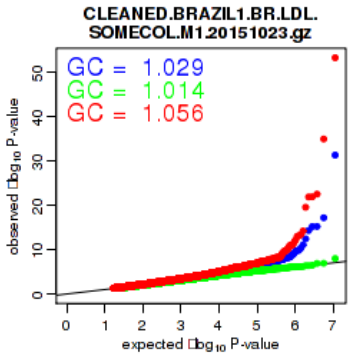
Figure S2C: QQ Plots of Brazilian-Ancestry Cohorts.

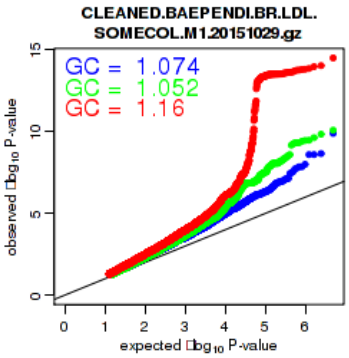
BRZ LDL GradCol M1: DF >20



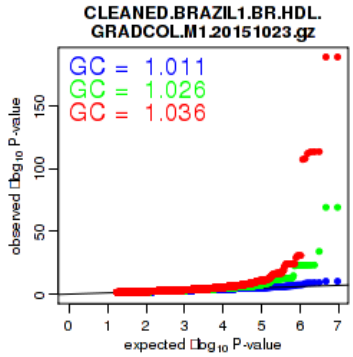


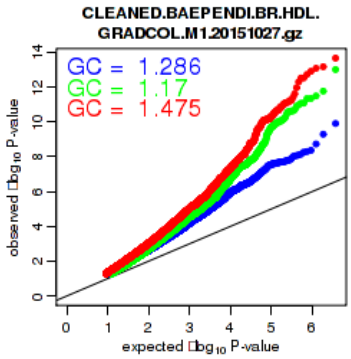
BRZ LDL SomeCol M1: DF >20



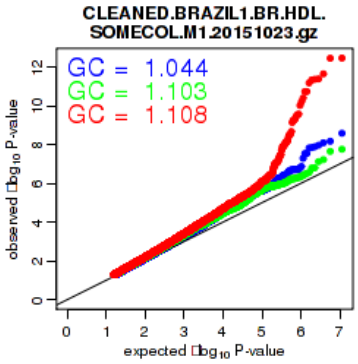


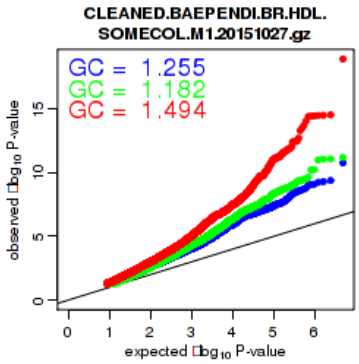
BRZ HDL GradCol M1: DF >20



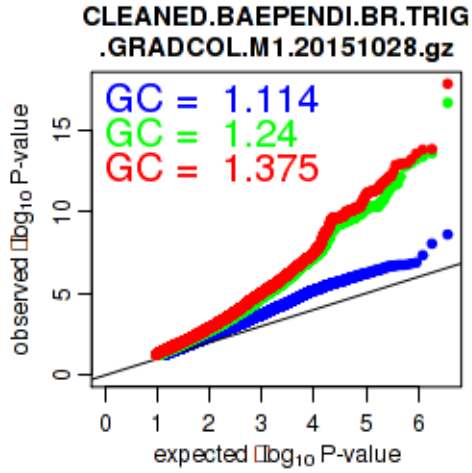


BRZ HDL SomeCol M1: DF >20





BRZ TG GradCol M1: DF >20



BRZ TG SomeCol M1: DF >20

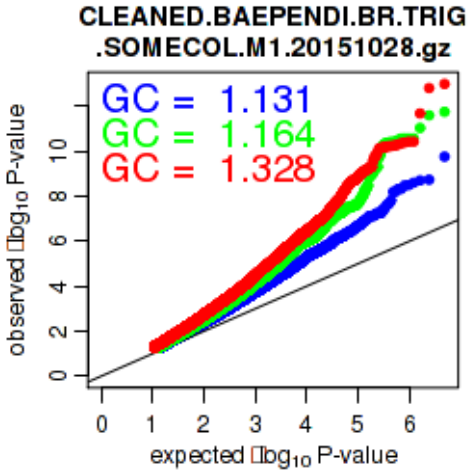
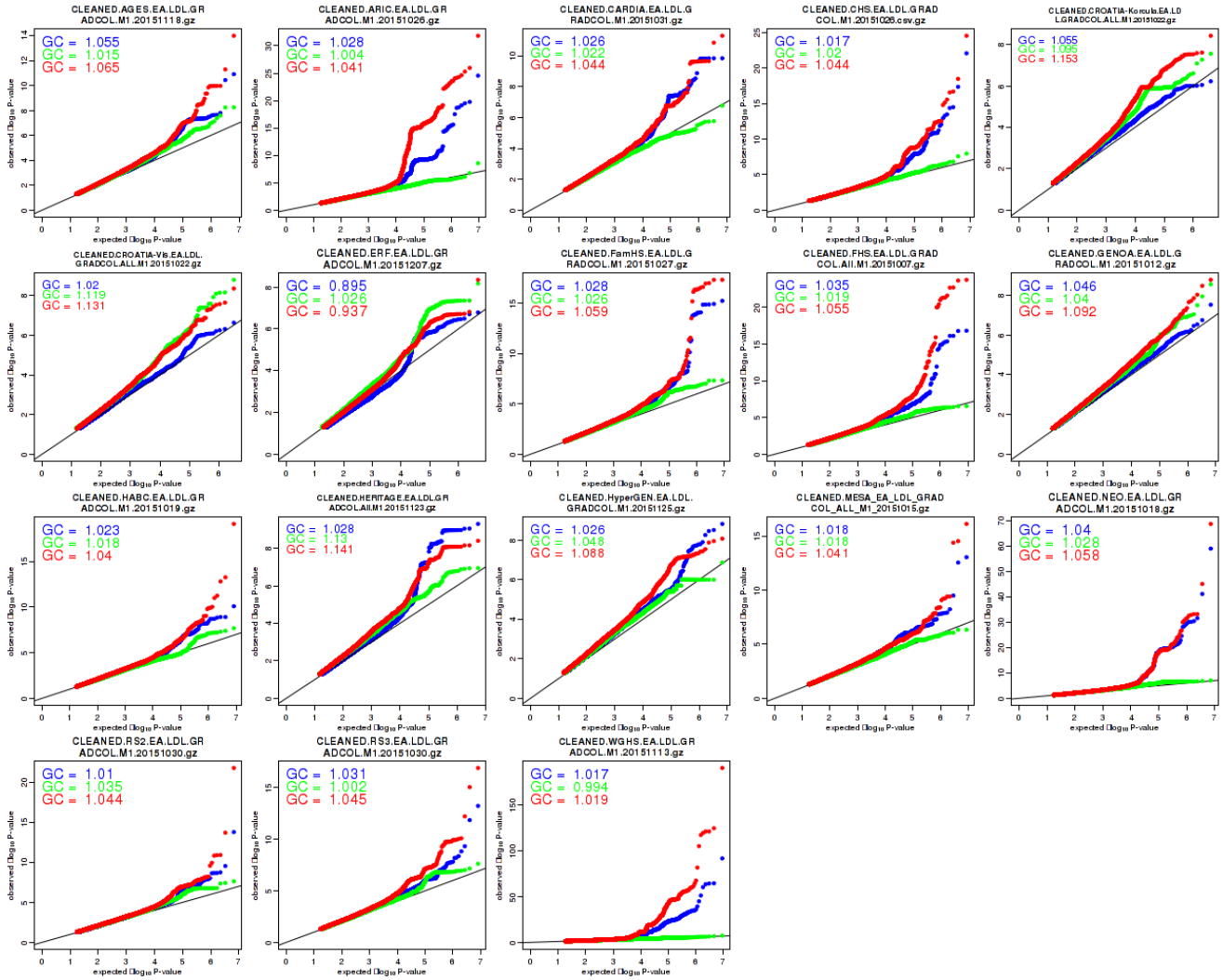
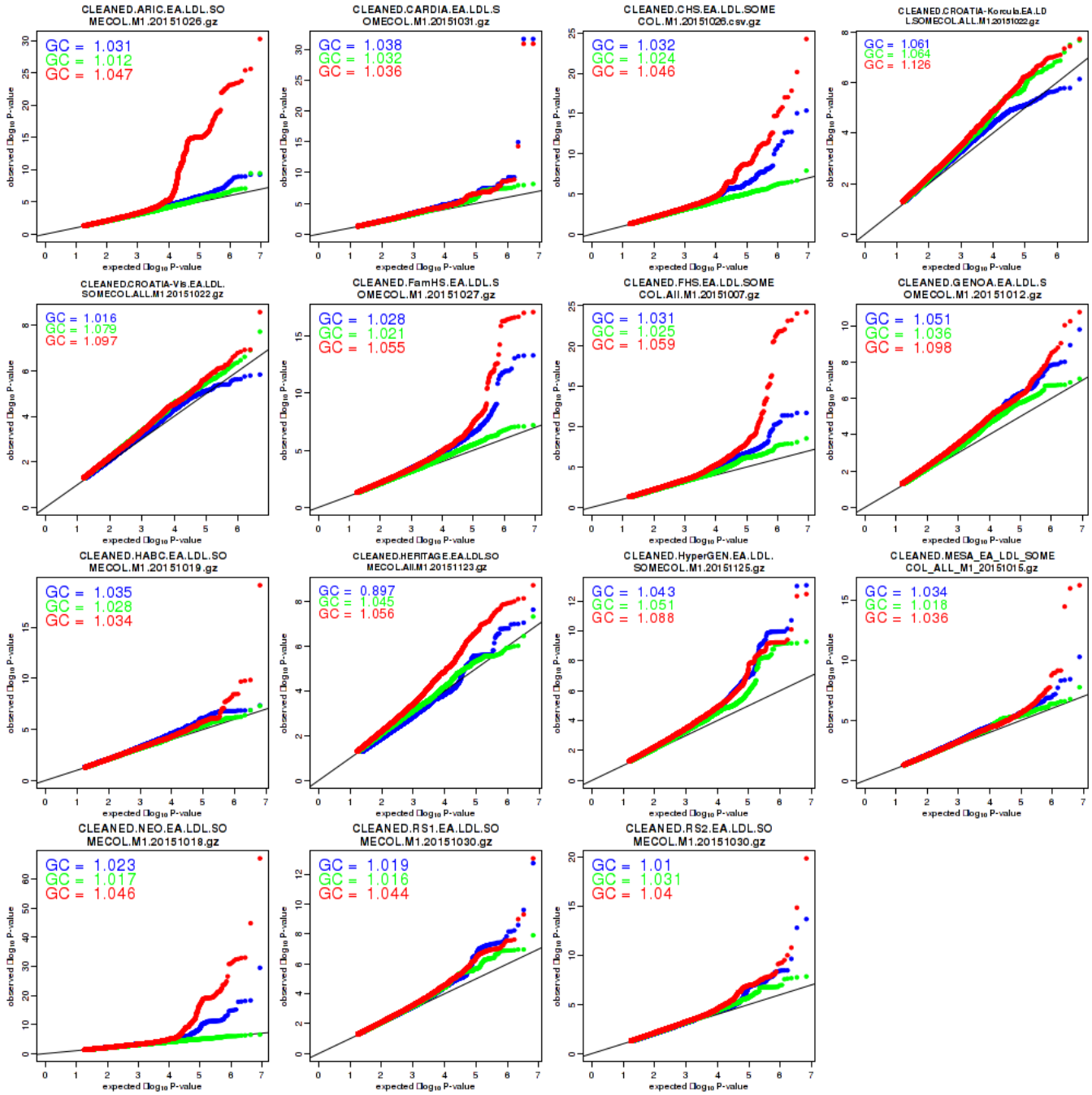


Figure S2D: QQ Plots of European-Ancestry Cohorts.

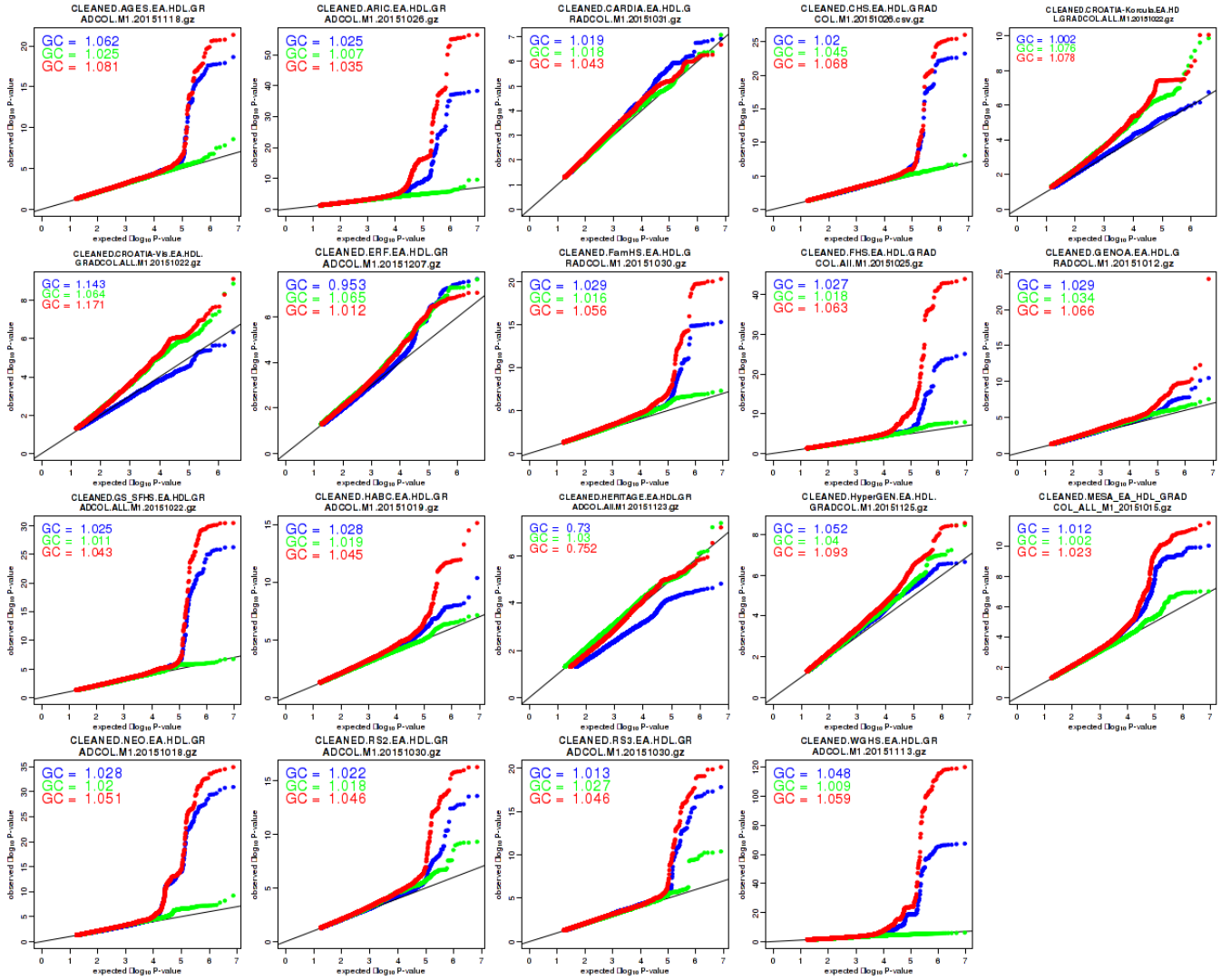
<u>EUR LDL GradCol M1: DF >20</u>



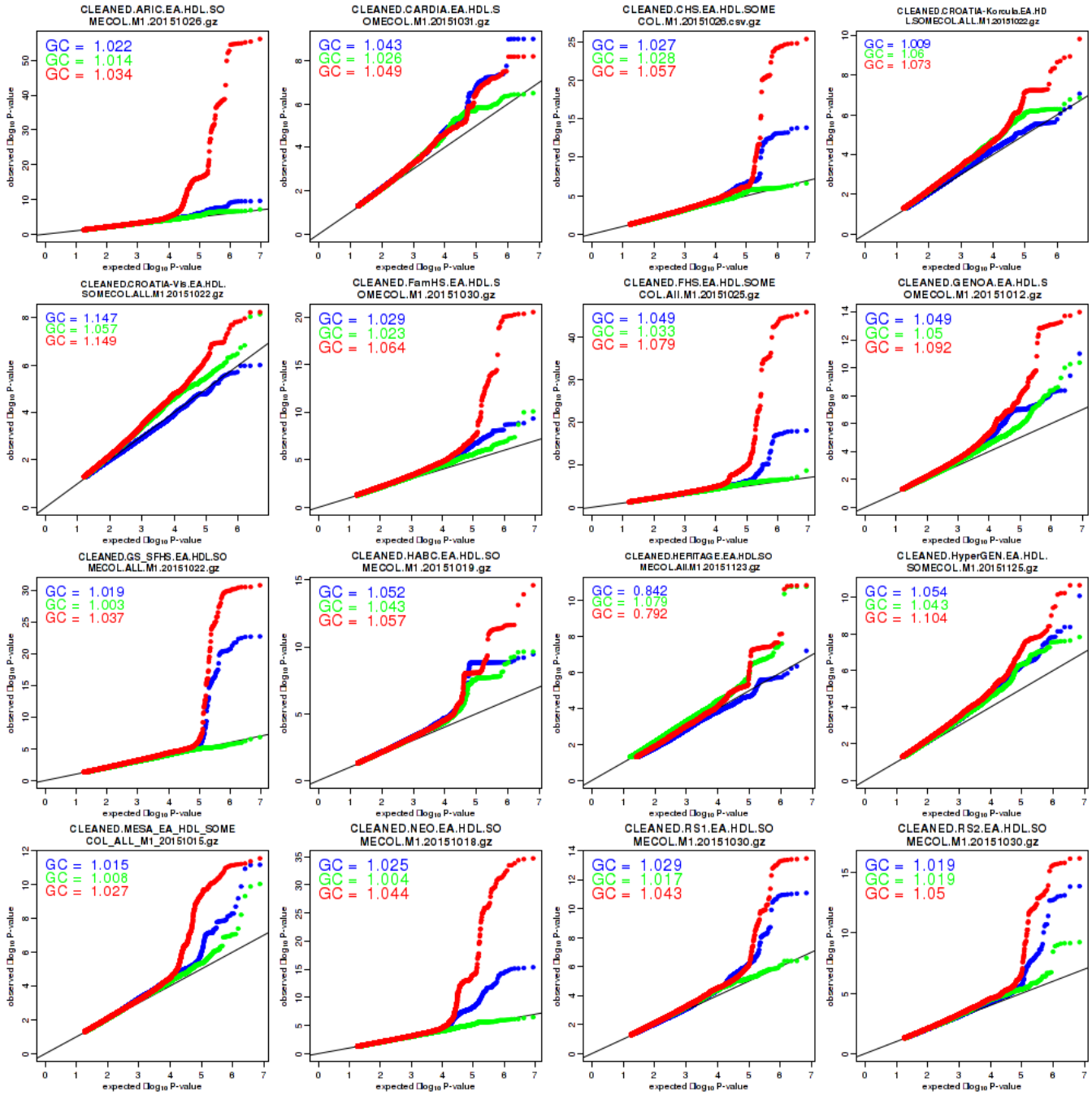
EUR LDL SomeCol M1: DF >20



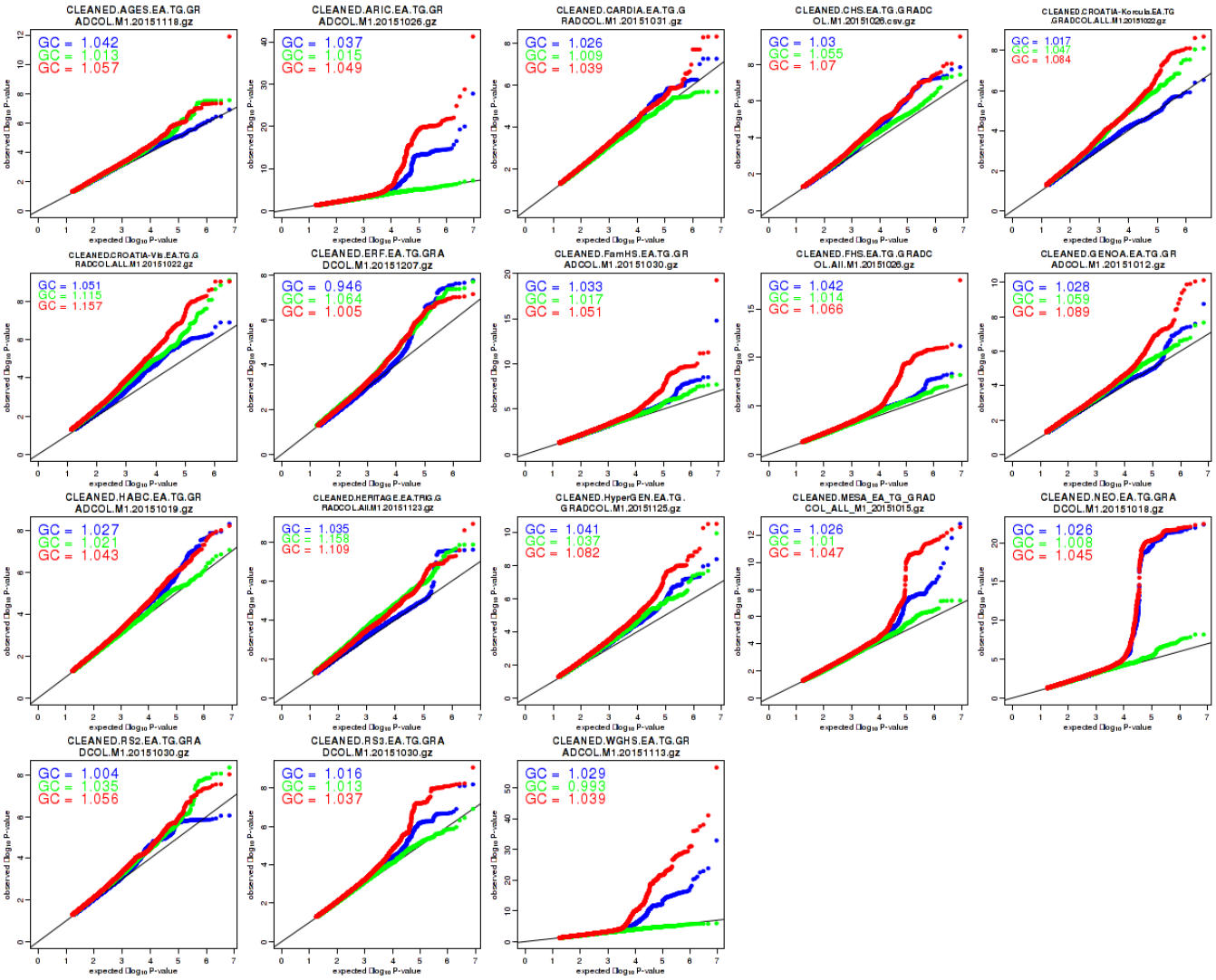
EUR HDL GradCol M1: DF >20



EUR HDL SomeCol M1: DF >20



EUR TG GradCol M1: DF >20



EUR TG SomeCol M1: DF >20

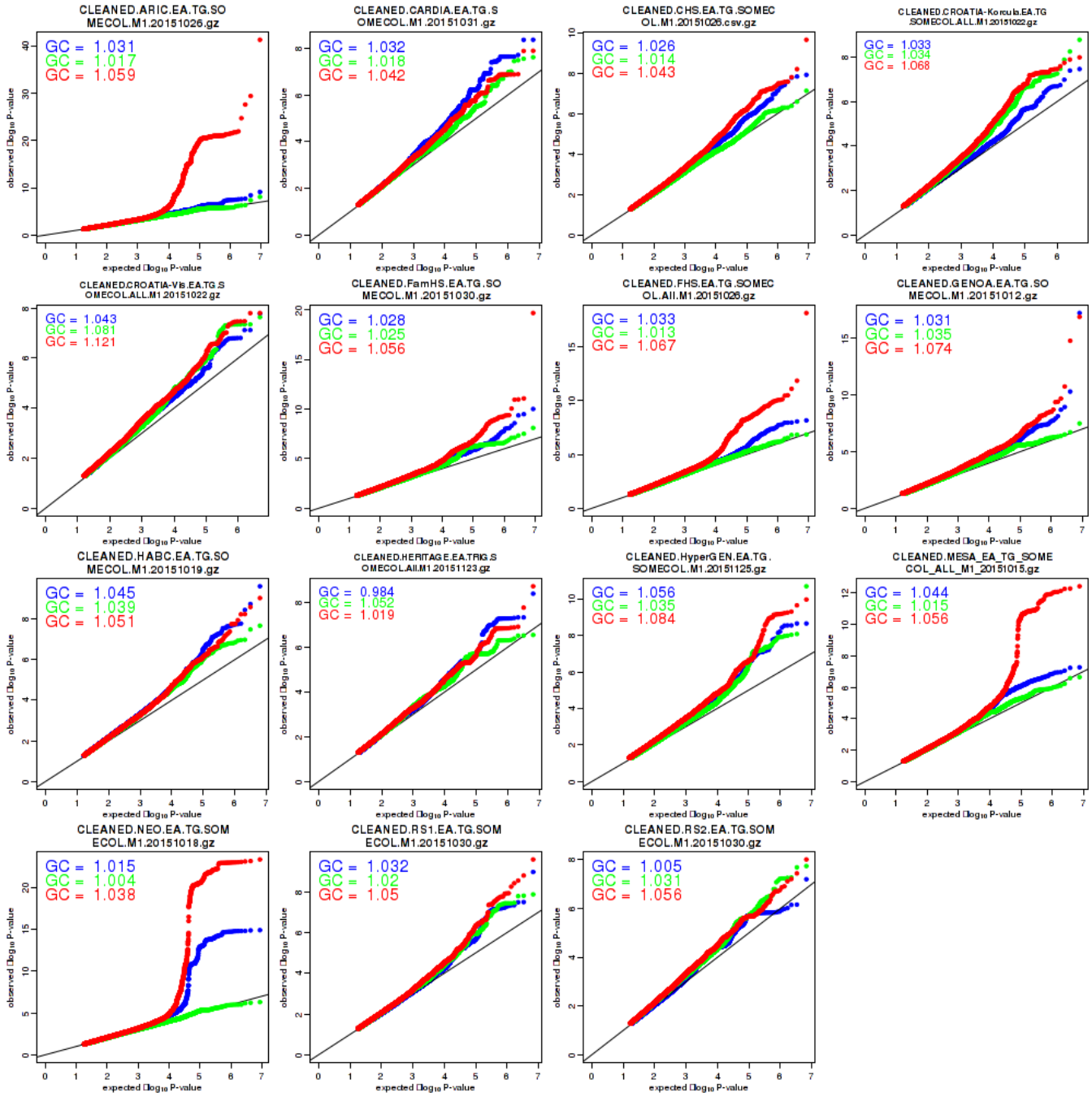
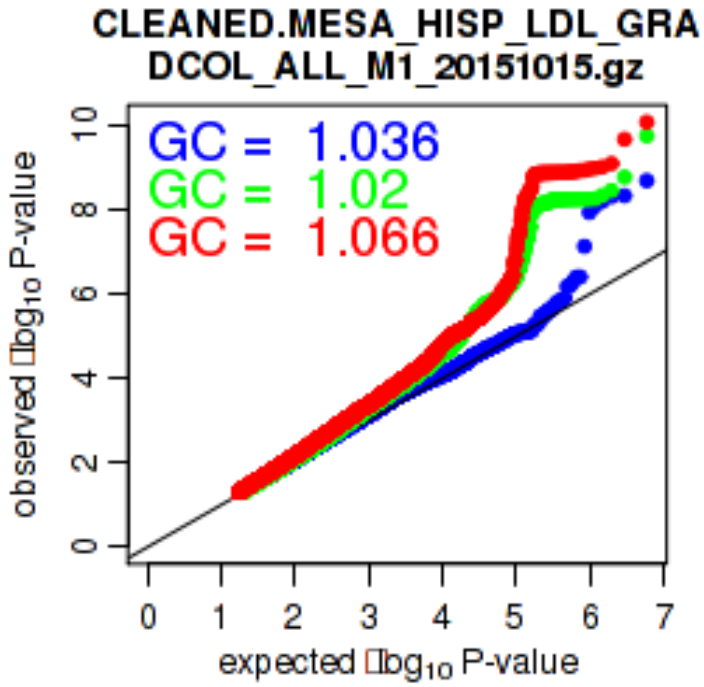
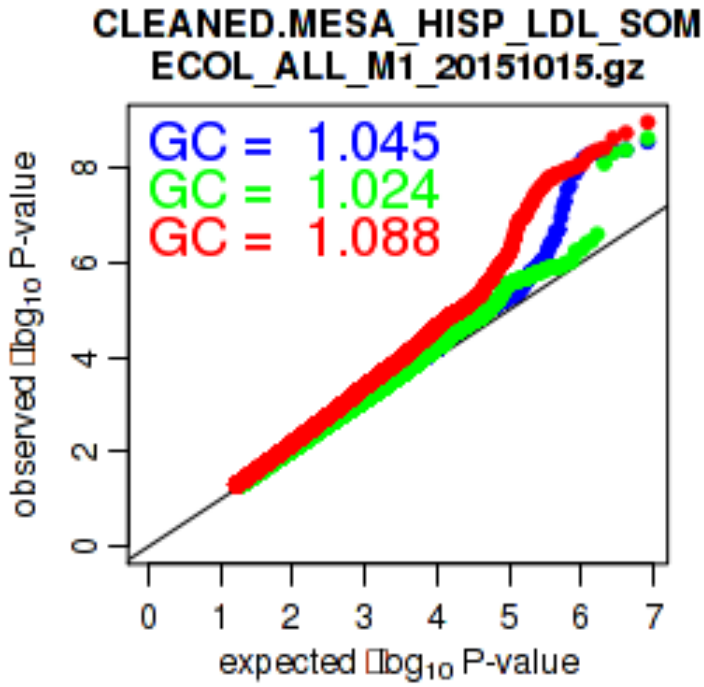


Figure S2E: QQ Plots of Hispanic-Ancestry Cohorts.

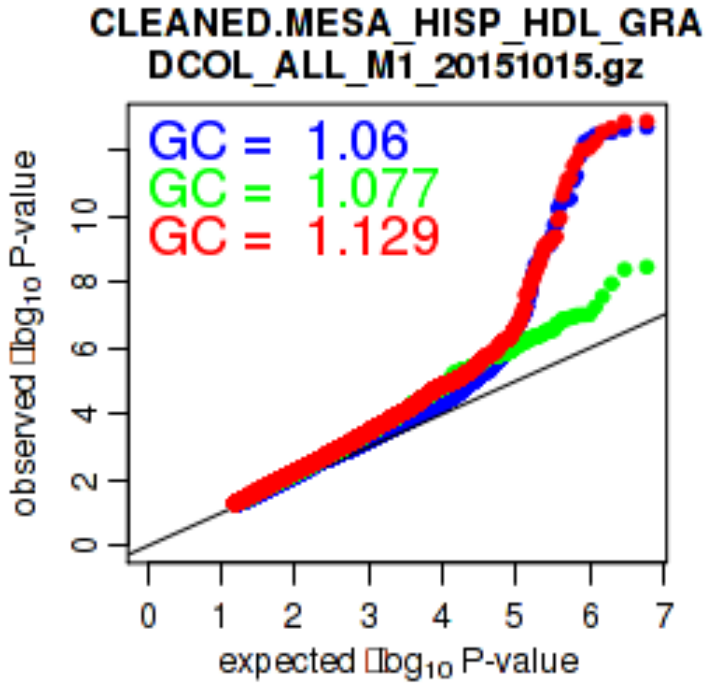
<u>HIS LDL GradCol M1: DF >20</u>



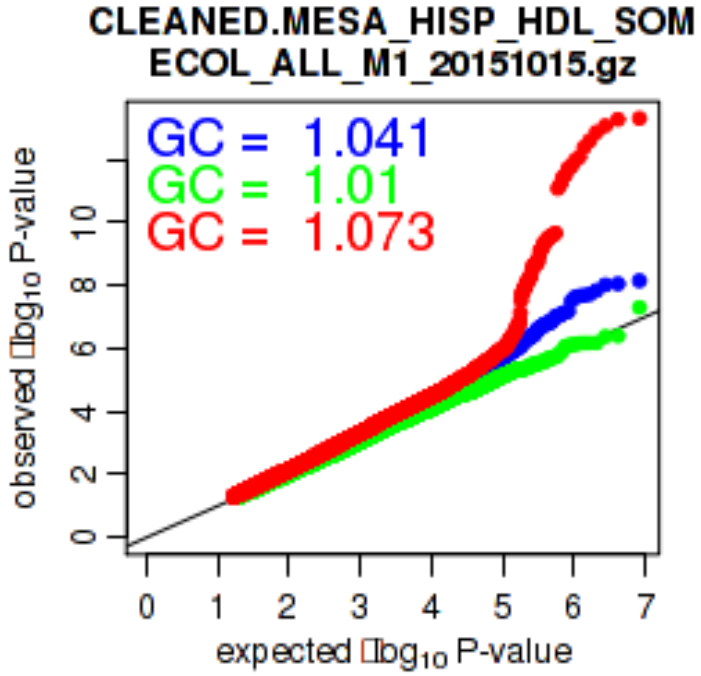
<u>HIS LDL SomeCol M1: DF >20</u>



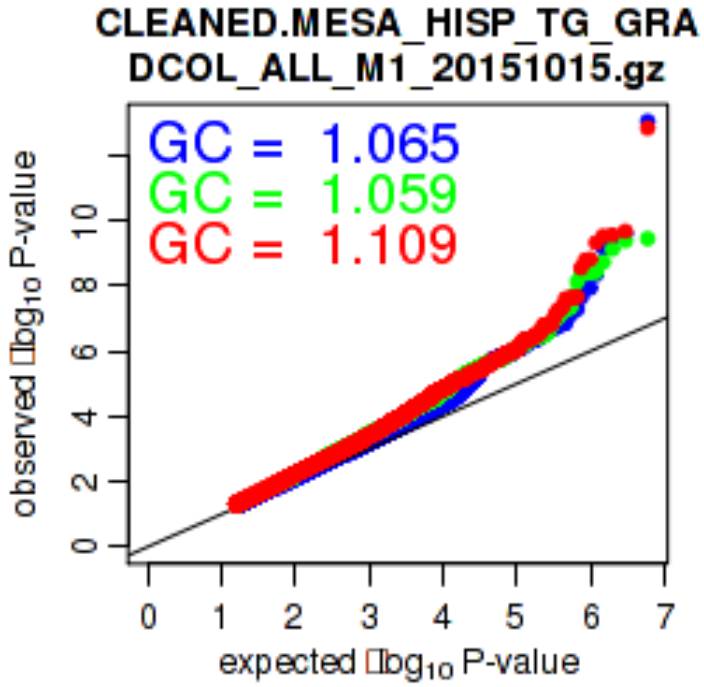
HIS HDL GradCol M1: DF >20



HIS HDL SomeCol M1: DF >20



HIS TG GradCol M1: DF >20



HIS TG SomeCol M1: DF >20

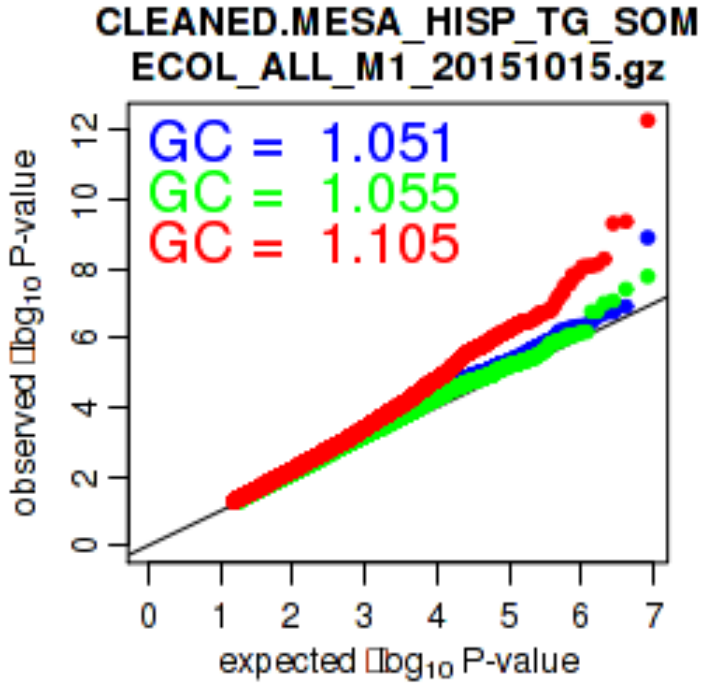
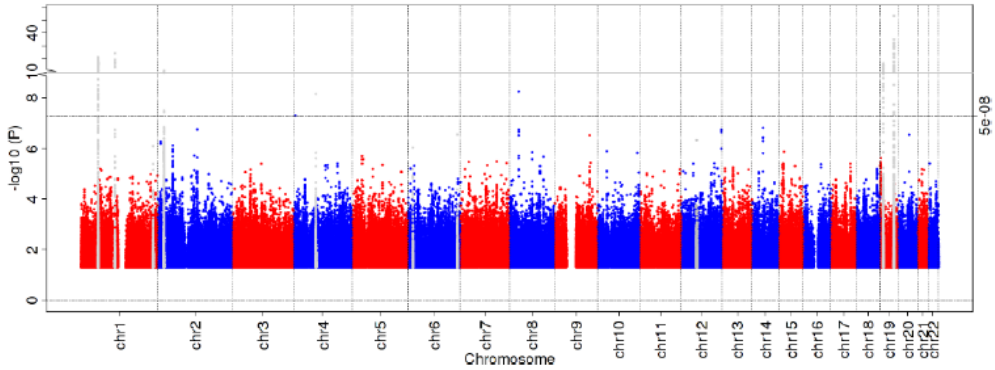


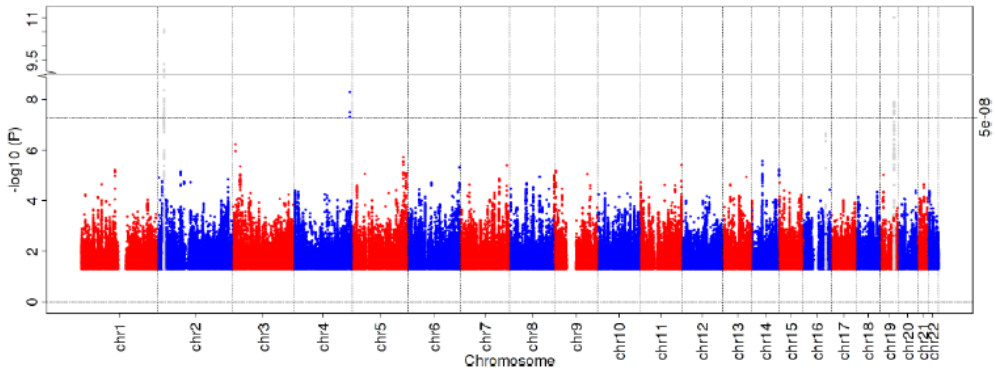
Figure S3A: Manhattan plots of the combined analyses of Stages 1 and 2 using the 2DF for LDL and GradCol by ancestry.

The -log10(P) of each SNV was plotted at the chromosomal location of each variant. P-values are based on the meta-analysis combining Stages 1 and 2 analysis and the Stage 1 analysis for the remaining 18.8 million variants.

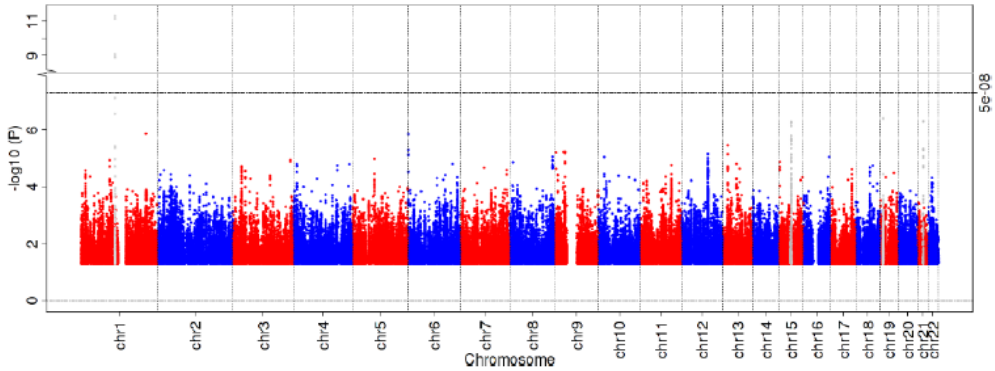
African Ancestry



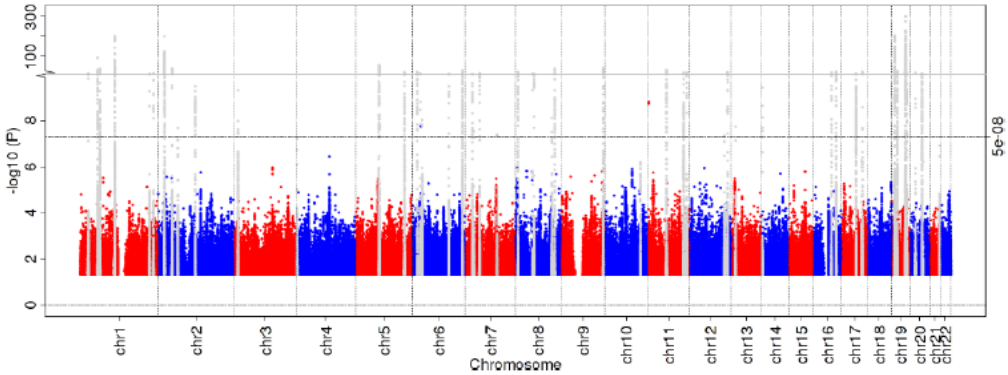
Asian Ancestry



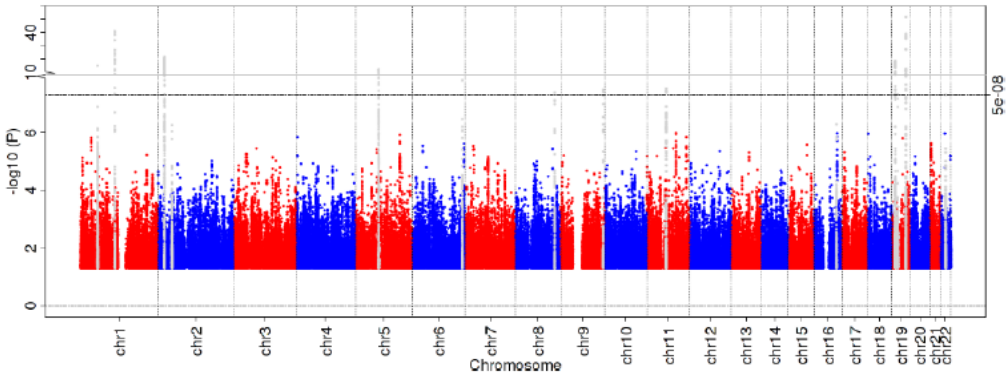
Brazilian Ancestry



European Ancestry



Hispanic Ancestry



Cross-Population Analysis

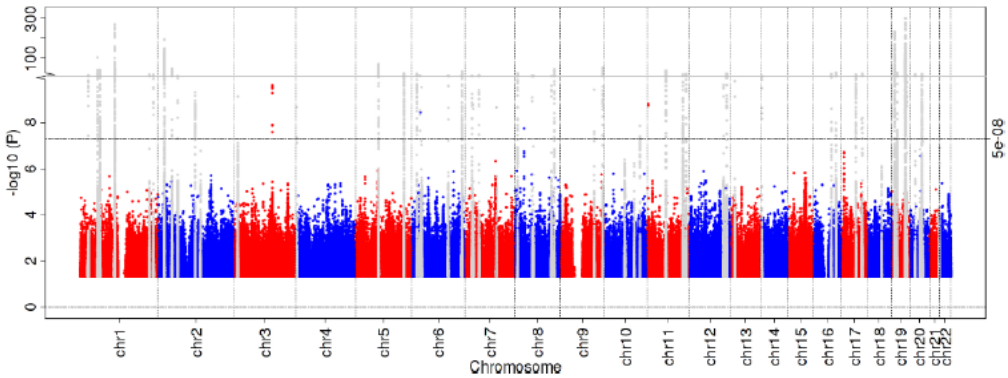
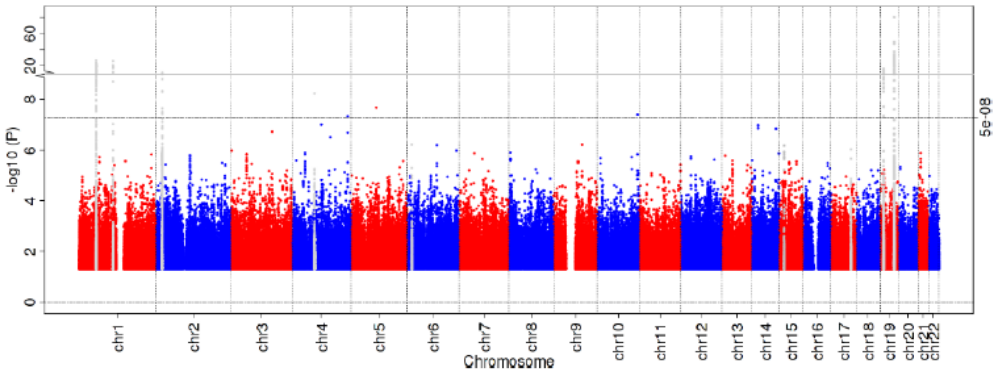


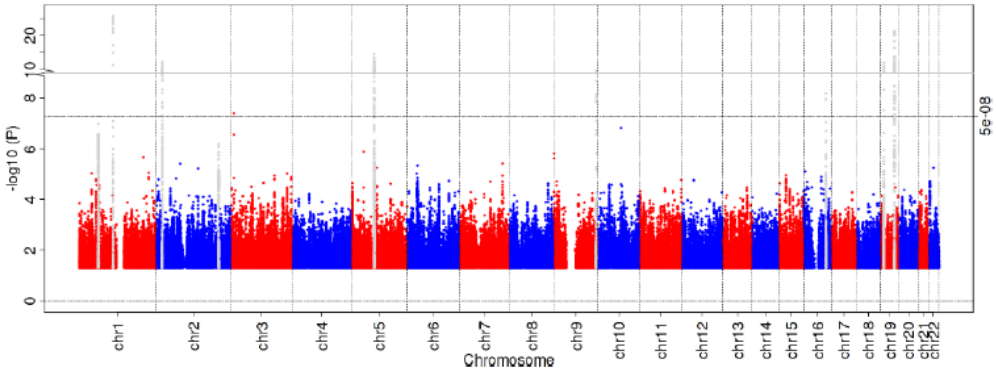
Figure S3B: Manhattan plots of the combined analyses of Stages 1 and 2 using the 2DF for LDL and SomeCol by ancestry.

The -log10(P) of each SNV was plotted at the chromosomal location of each variant. P-values are based on the meta-analysis combining Stages 1 and 2 analysis and the Stage 1 analysis for the remaining 18.8 million variants.

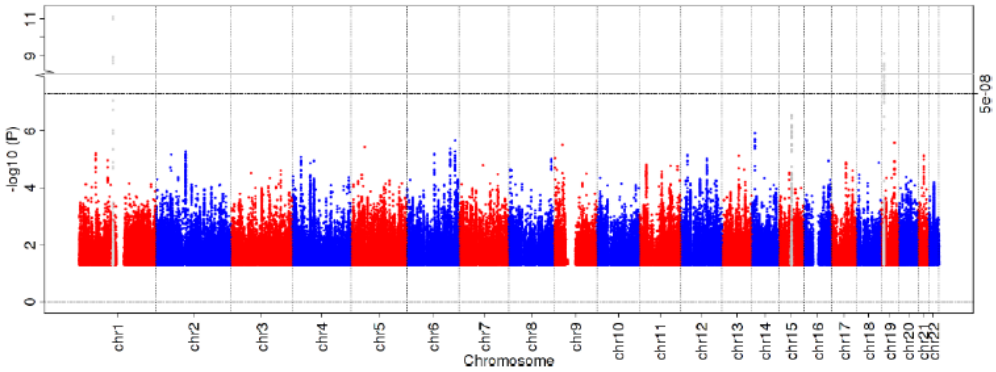
African Ancestry



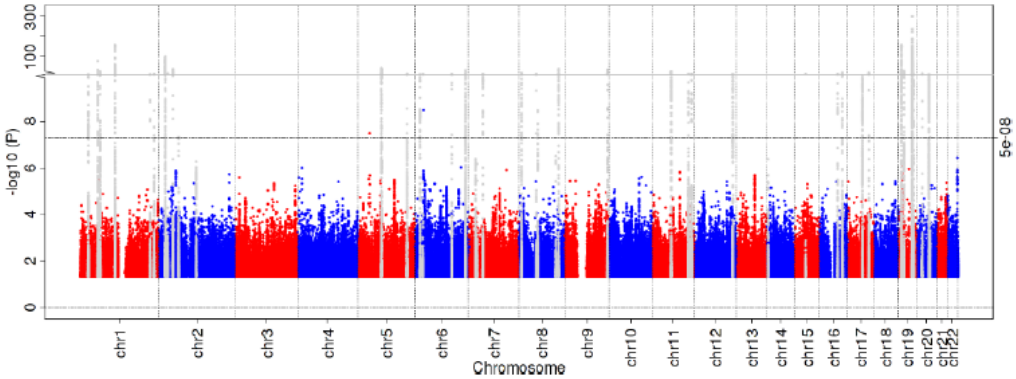
Asian Ancestry



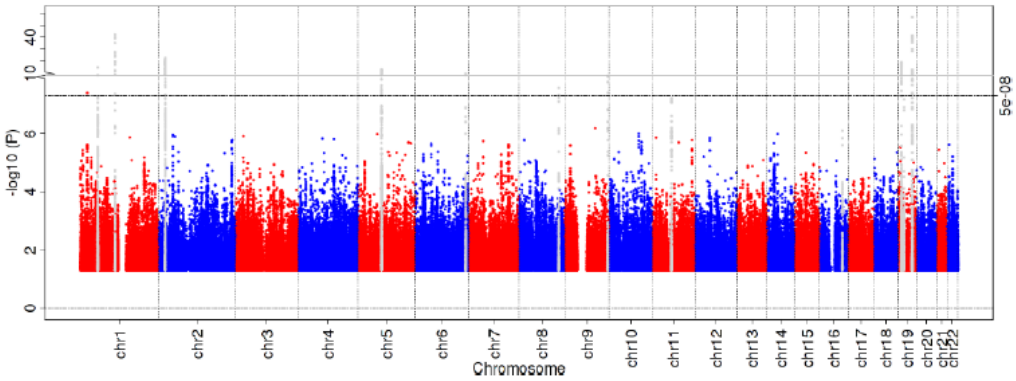
Brazilian Ancestry



European Ancestry



Hispanic Ancestry



Cross-Population Analysis

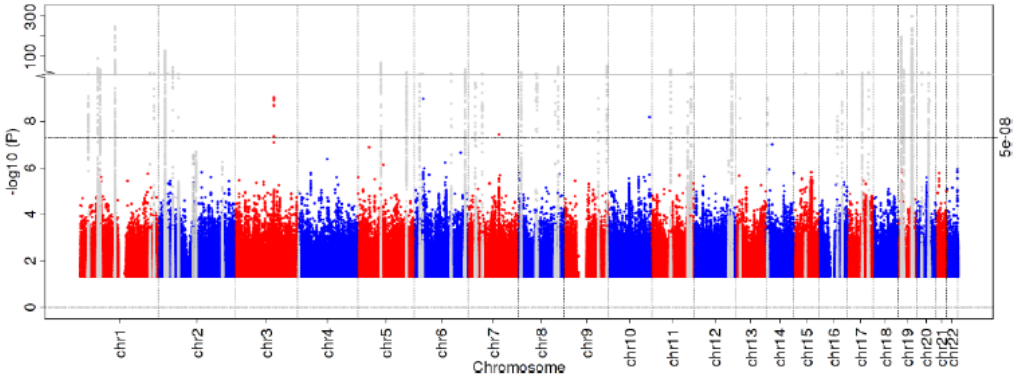
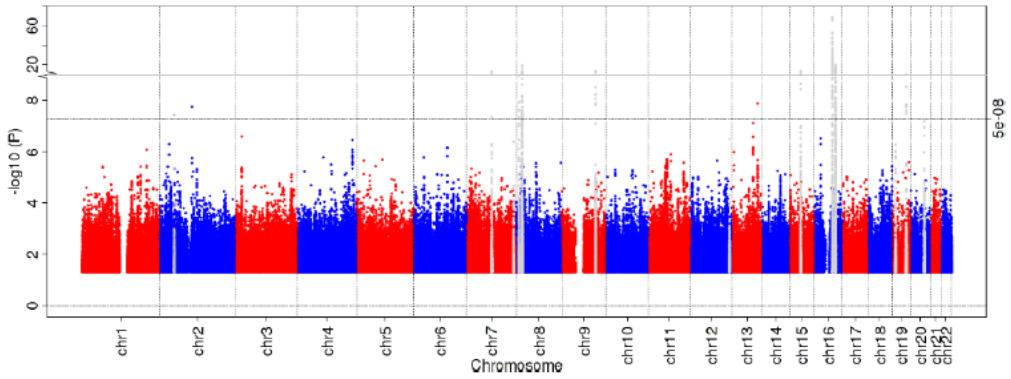


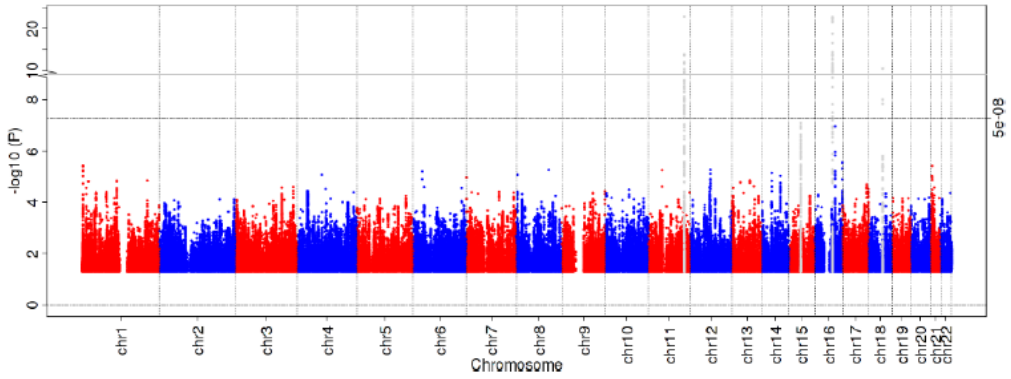
Figure S3C: Manhattan plots of the combined analyses of Stages 1 and 2 using the 2DF for HDL and GradCol by ancestry.

The -log10(P) of each SNV was plotted at the chromosomal location of each variant. P-values are based on the meta-analysis combining Stages 1 and 2 analysis and the Stage 1 analysis for the remaining 18.8 million variants.

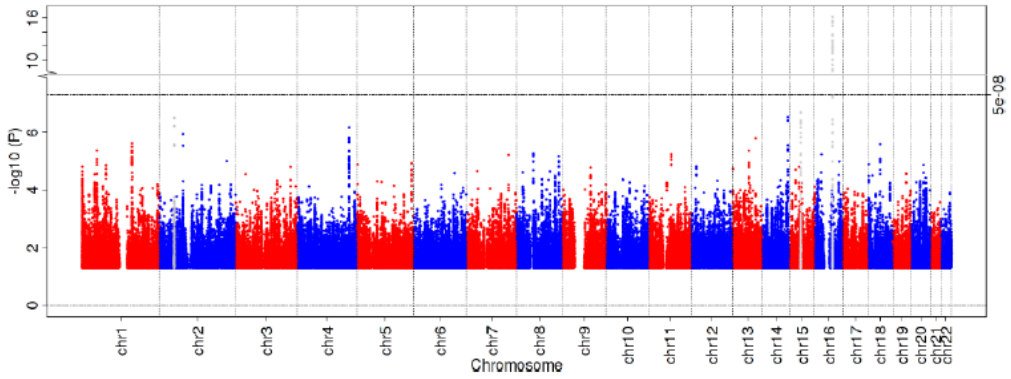
African Ancestry



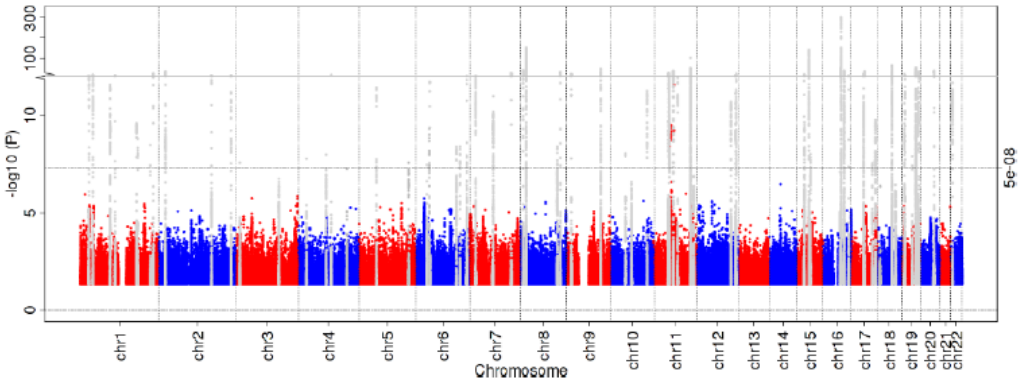
Asian Ancestry



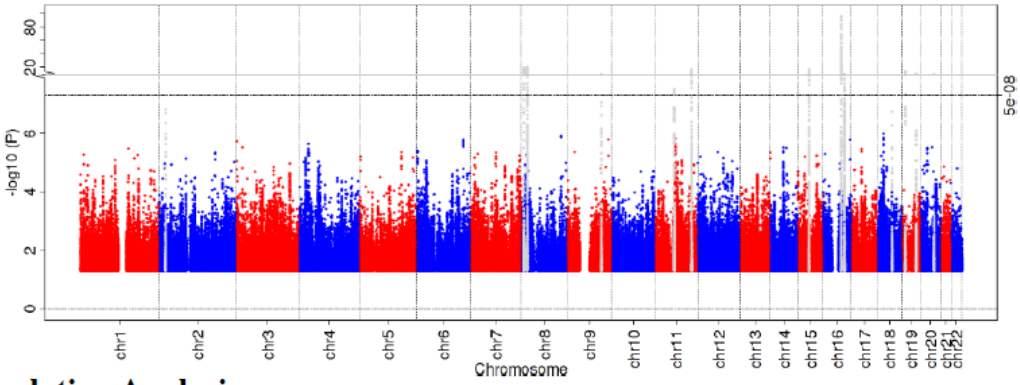
Brazilian Ancestry



European Ancestry



Hispanic Ancestry



Cross-Population Analysis

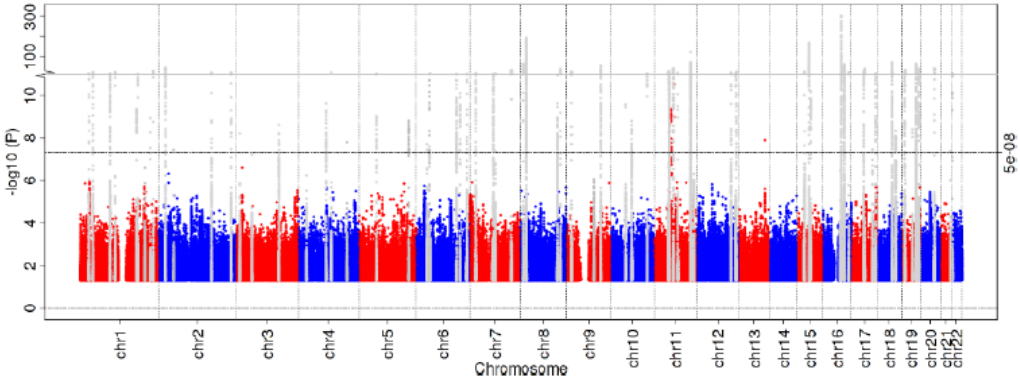
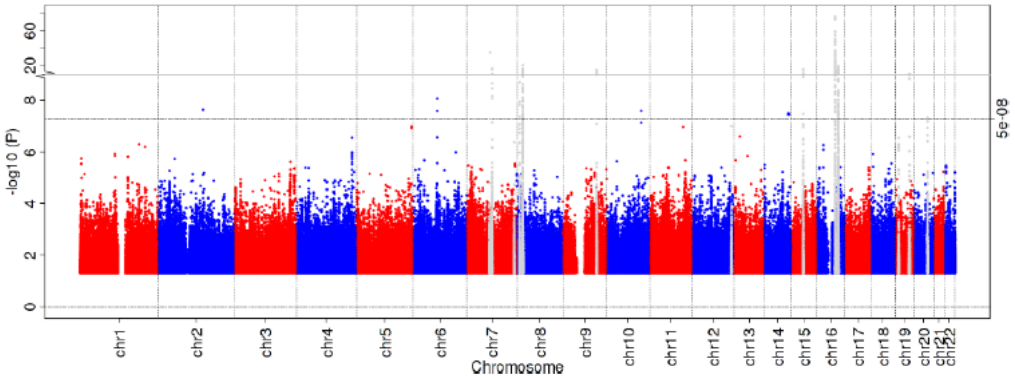


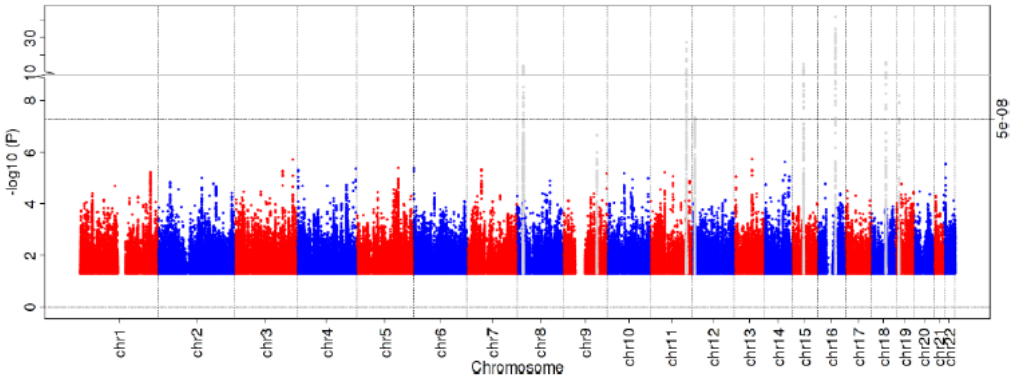
Figure S3D: Manhattan plots of the combined analyses of Stages 1 and 2 using the 2DF for HDL and SomeCol by ancestry.

The -log10(P) of each SNV was plotted at the chromosomal location of each variant. P-values are based on the meta-analysis combining Stages 1 and 2 analysis and the Stage 1 analysis for the remaining 18.8 million variants.

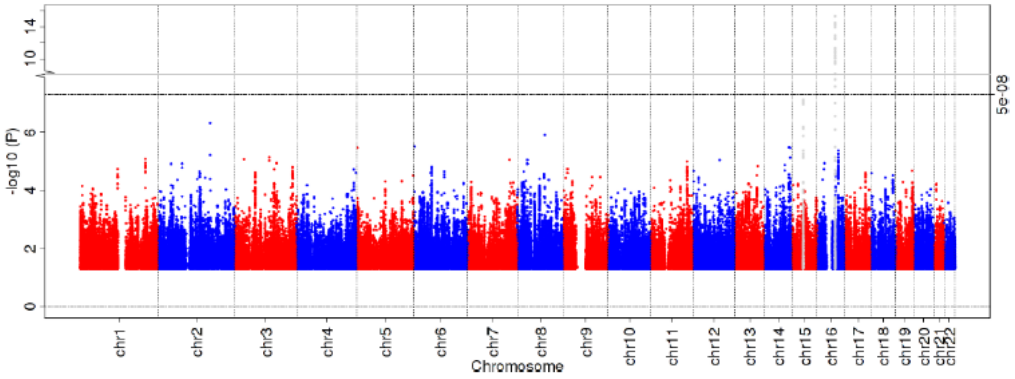
African Ancestry



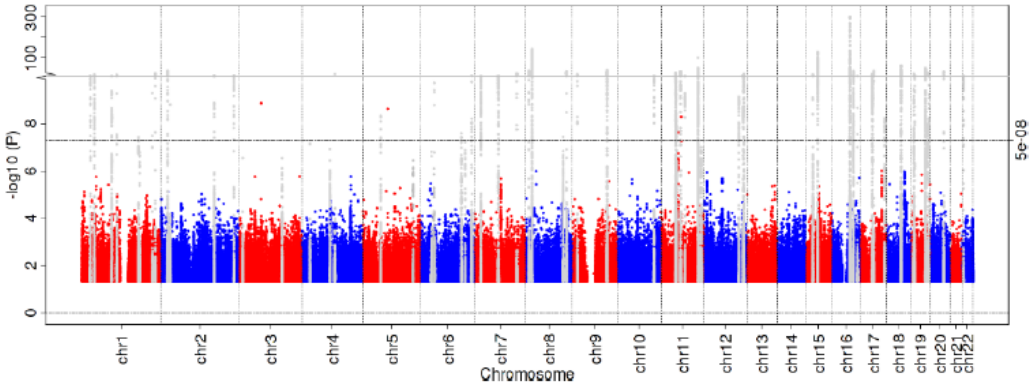
Asian Ancestry



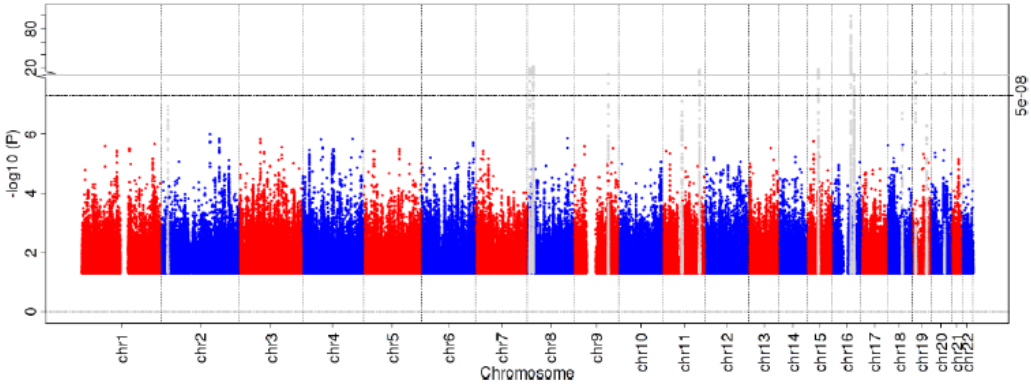
Brazilian Ancestry



European Ancestry



Hispanic Ancestry



Cross-Population Analysis

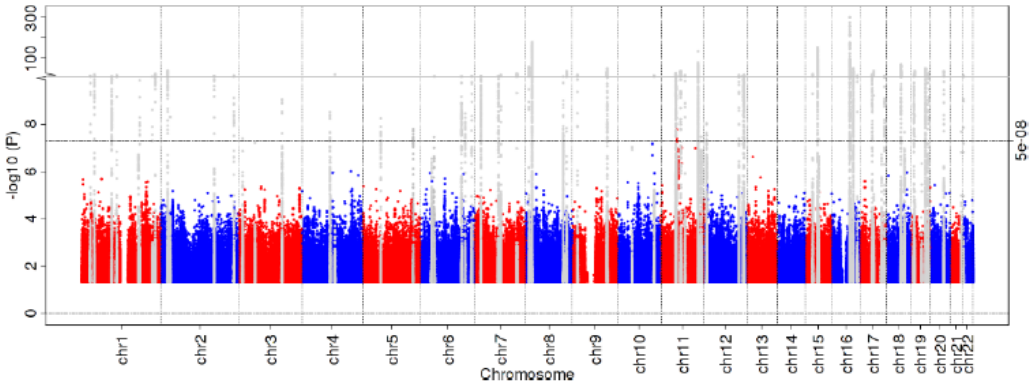
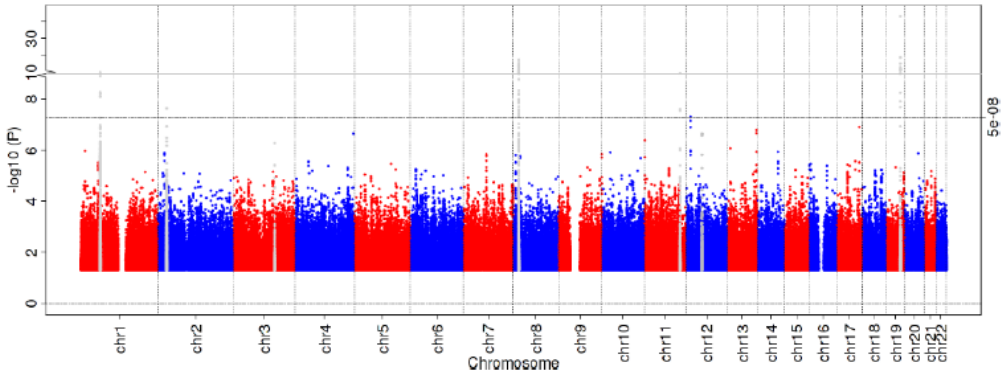


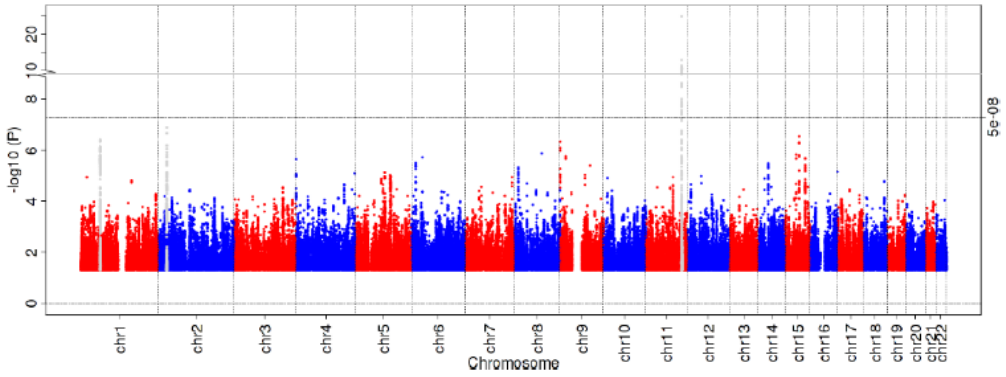
Figure S3E: Manhattan plots of the combined analyses of Stages 1 and 2 using the 2DF for TG and GradCol by ancestry.

The -log10(P) of each SNV was plotted at the chromosomal location of each variant. P-values are based on the meta-analysis combining Stages 1 and 2 analysis and the Stage 1 analysis for the remaining 18.8 million variants.

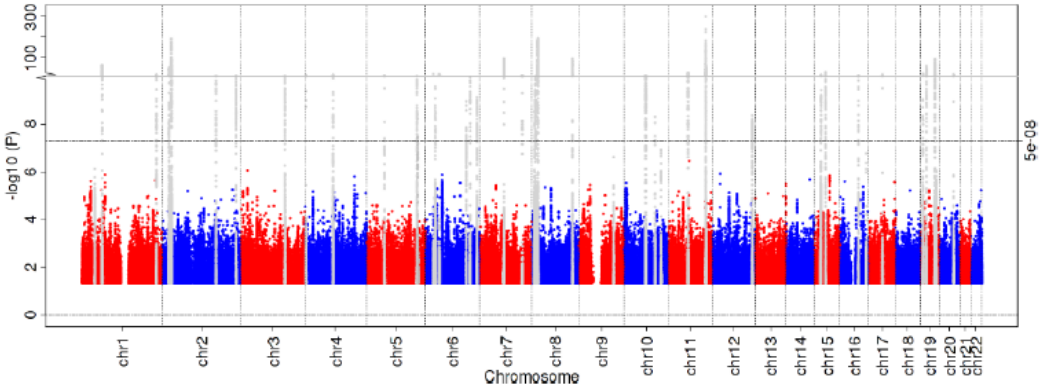
<u>African Ancestry</u>



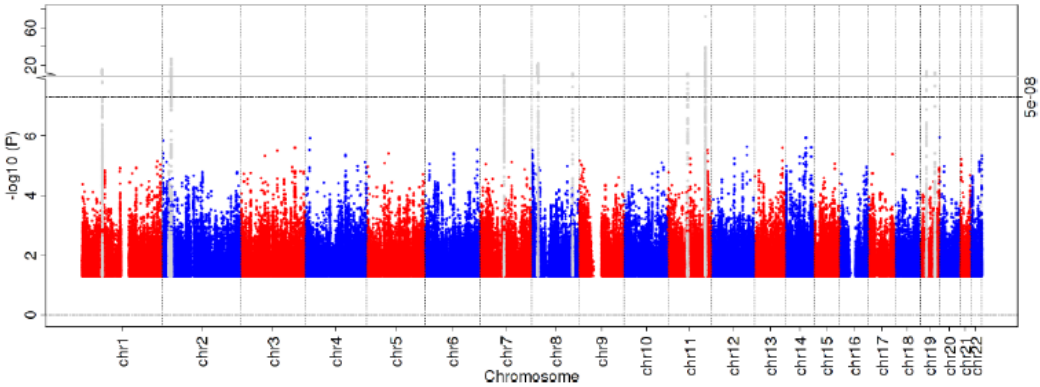
<u>Asian Ancestry</u>



European Ancestry



Hispanic Ancestry



Cross-Population Analysis

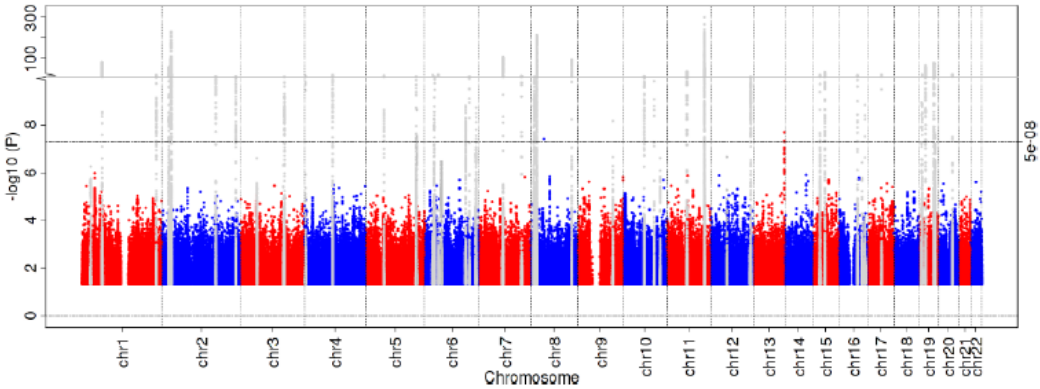
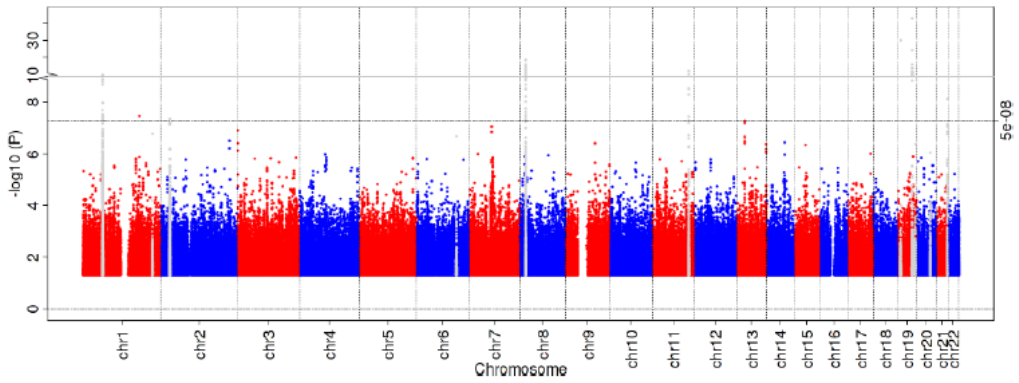


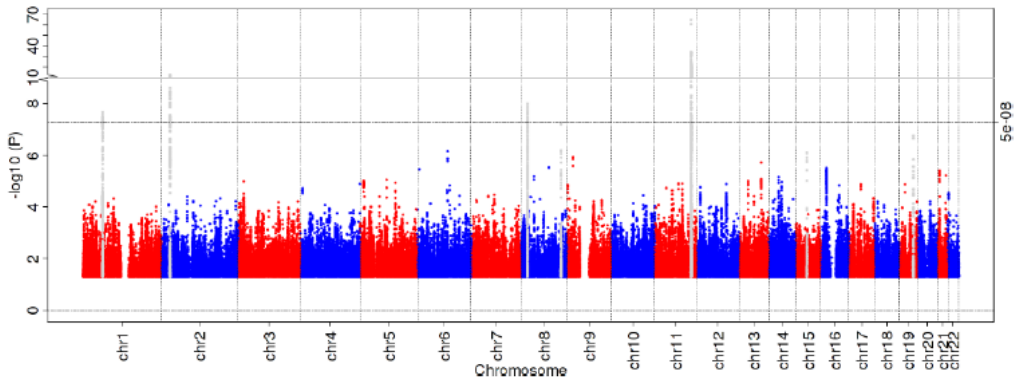
Figure S3F: Manhattan plots of the combined analyses of Stages 1 and 2 using the 2DF for TG and SomeCol by ancestry.

The -log10(P) of each SNV was plotted at the chromosomal location of each variant. P-values are based on the meta-analysis combining Stages 1 and 2 analysis and the Stage 1 analysis for the remaining 18.8 million variants.

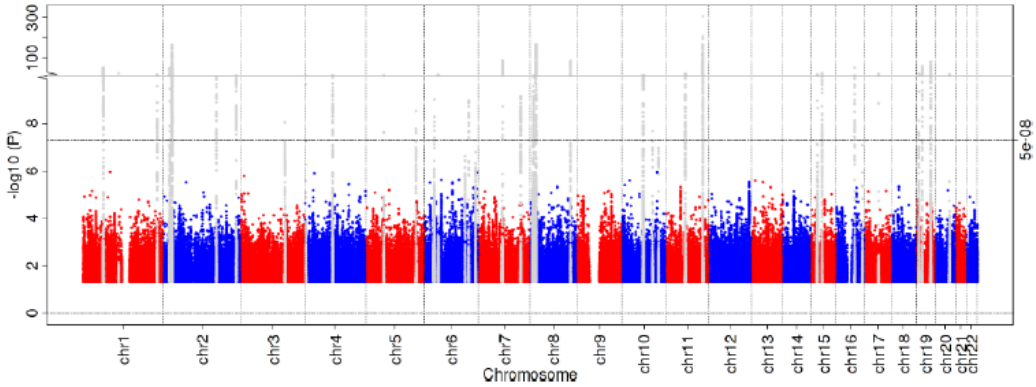
<u>African Ancestry</u>



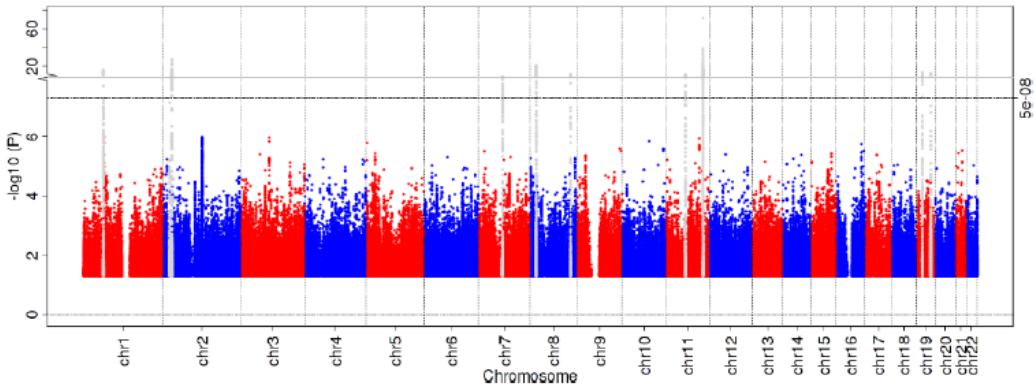
<u>Asian Ancestry</u>



European Ancestry



Hispanic Ancestry



Cross-Population Analysis

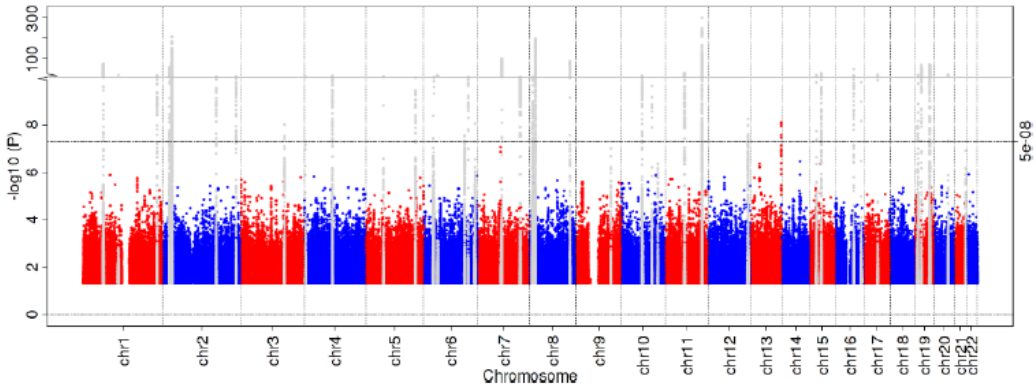
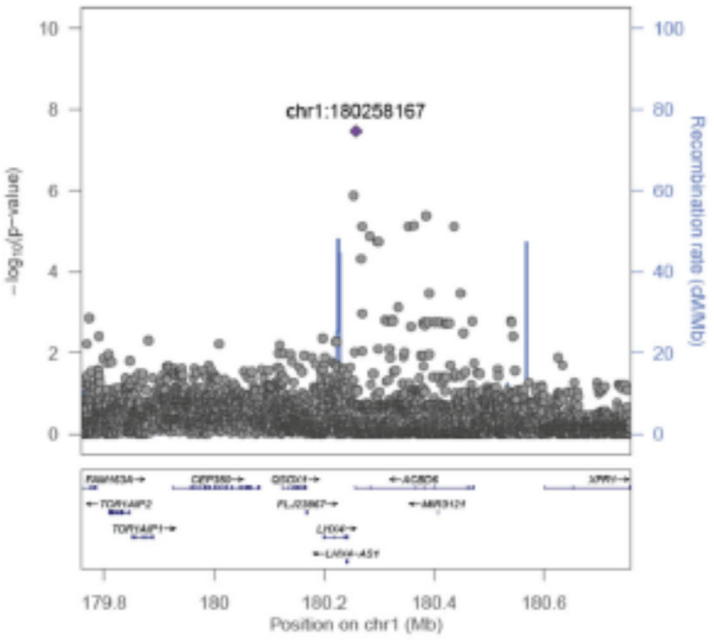


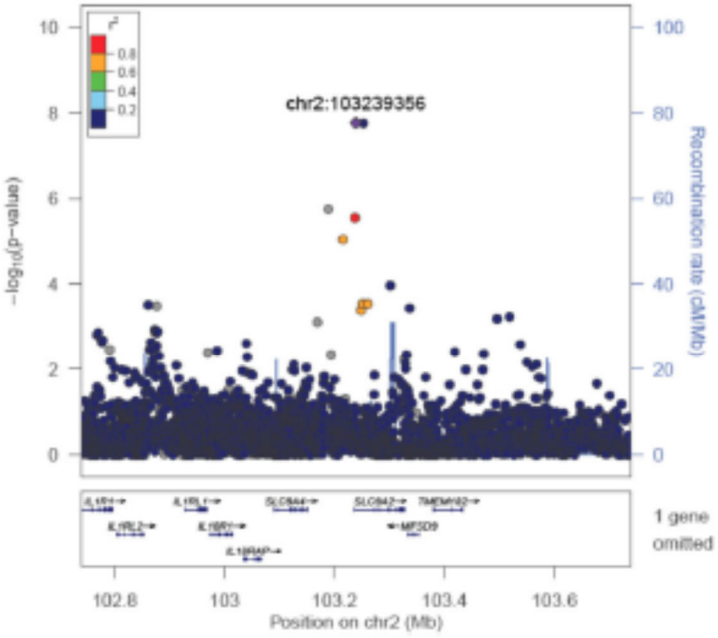
Figure S4: LocusZoom plots for 18 newly identified loci (Table 1).

Each locus has at least one lipid trait reaching genome-wide significance with at least one lipid trait in any ancestry group. LD substructure is not available for six insertion/deletion index variants.

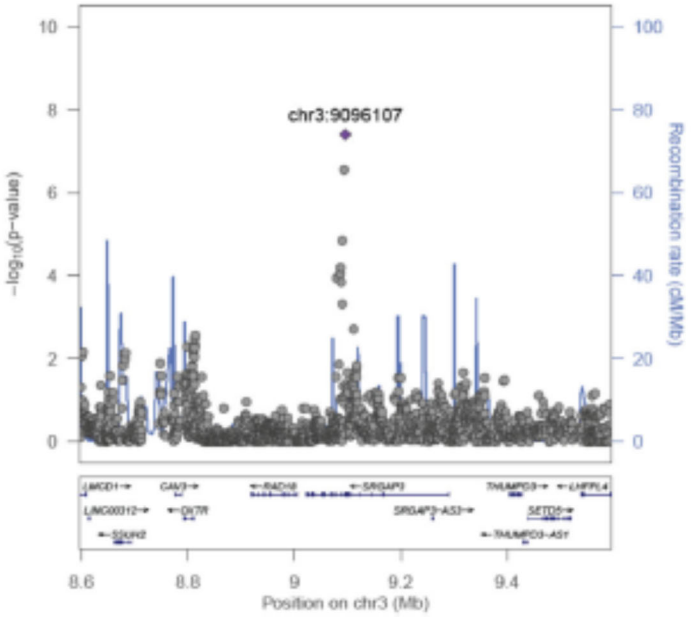
Locus 1: HDL-GradCol in African Ancestry (2DF).



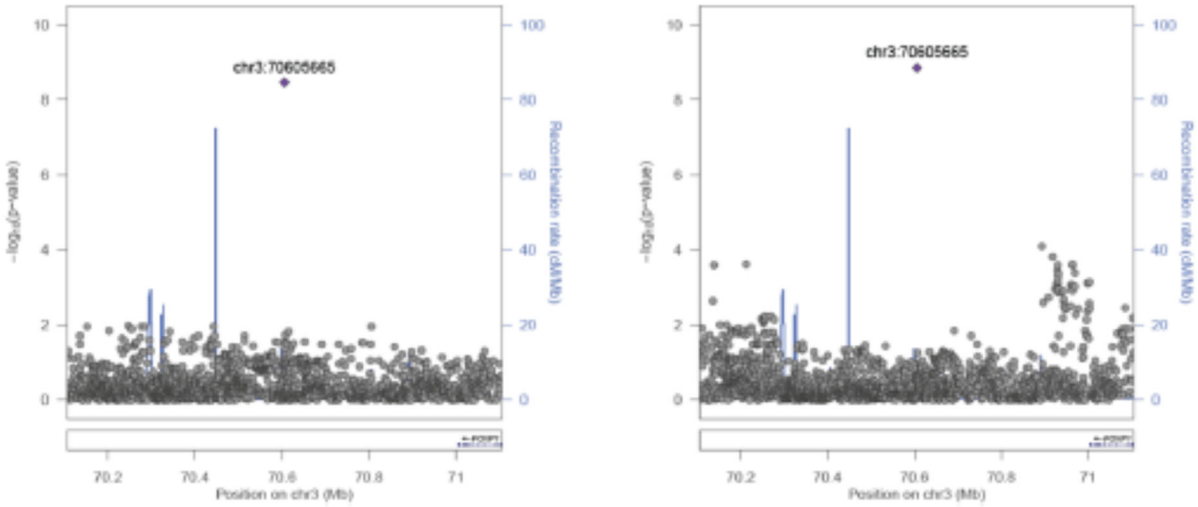
Locus 2: TG-SomeCol in African Ancestry (2DF).



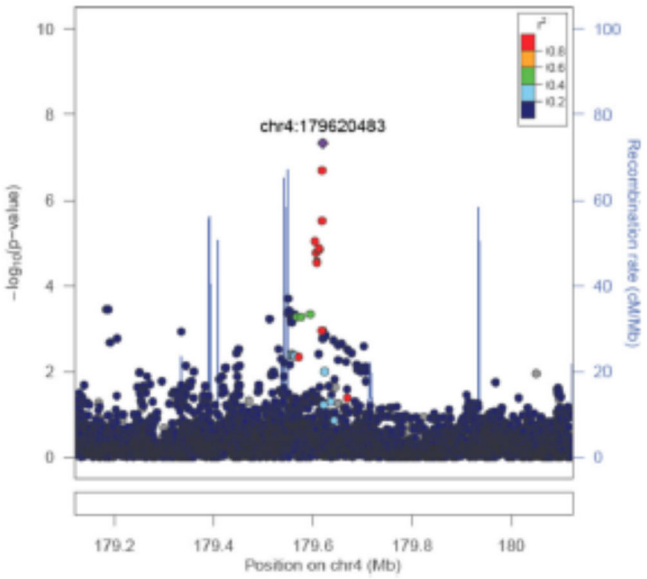
Locus 3: LDL-SomeCol in Asian Ancestry (2DF).



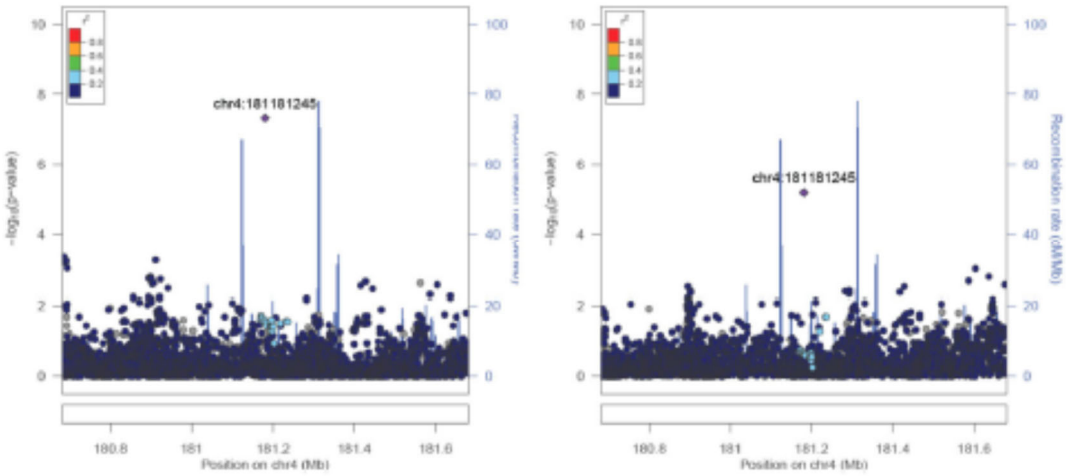
Locus 4: HDL-SomeCol in European Ancestry (left – interaction; right – 2DF).



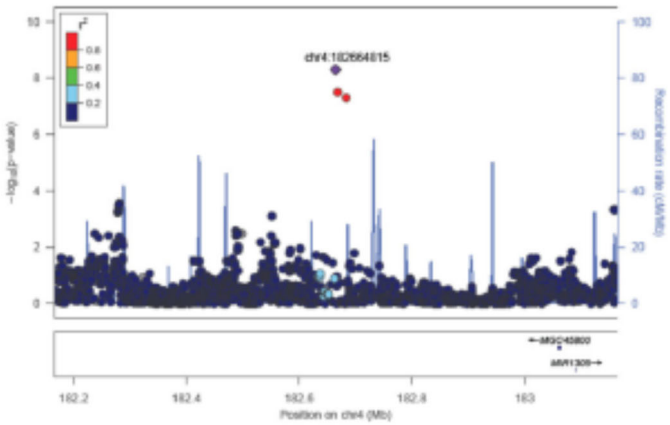
Locus 5: LDL-SomeCol in African Ancestry (2DF).



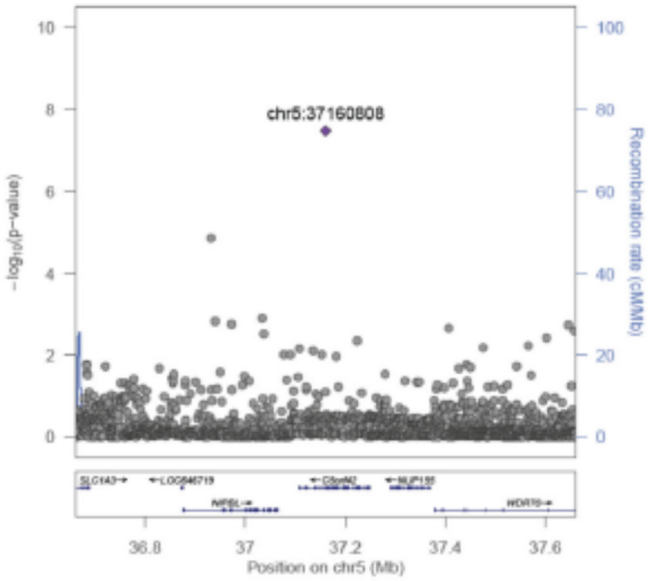
Locus 6: HDL-GradCol in Trans-Ancestry (left – interaction; right – 2DF).



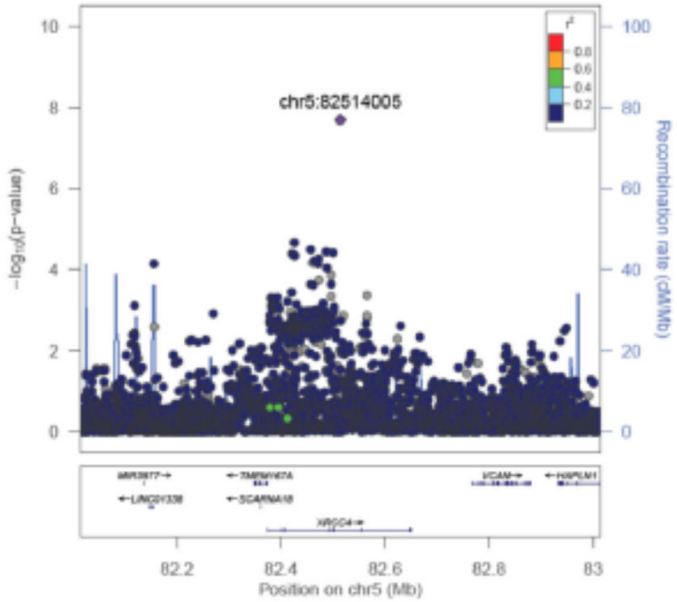
Locus 7: LDL-GradCol in Asian Ancestry (2DF).



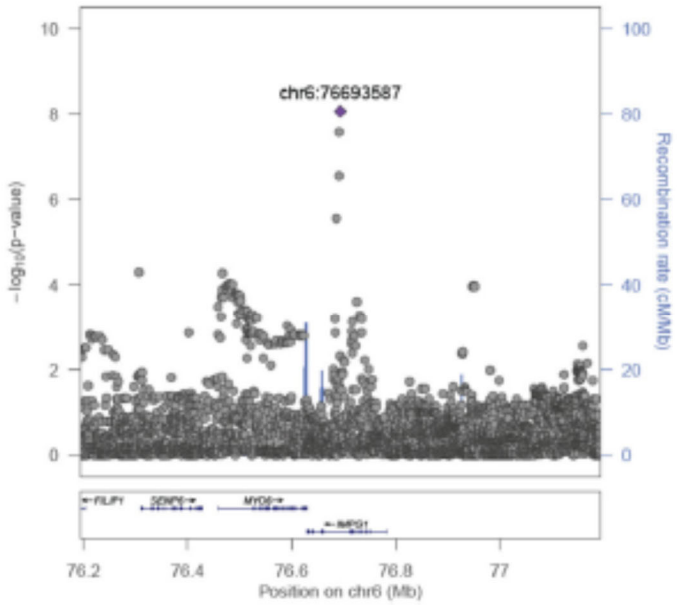
Locus 8: LDL-SomeCol in European Ancestry (2DF).



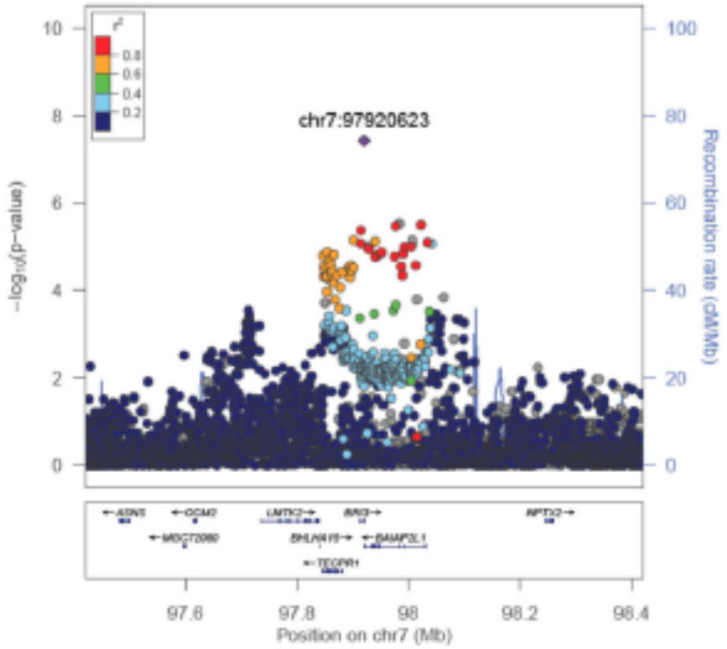
Locus 9: LDL-SomeCol in African Ancestry (2DF).



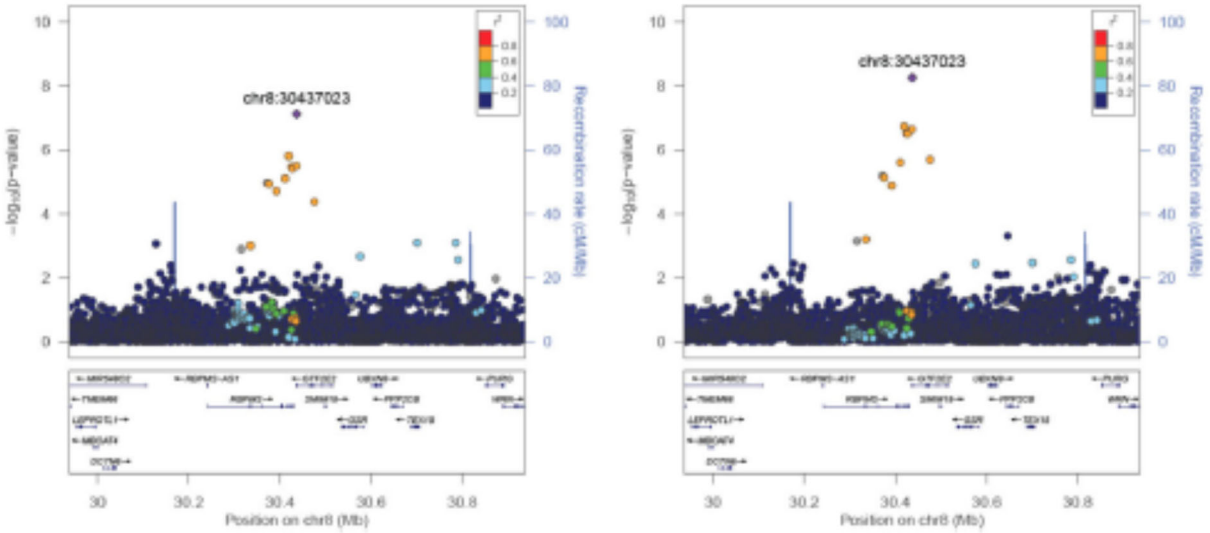
Locus 10: HDL-SomeCol in African Ancestry (2DF).



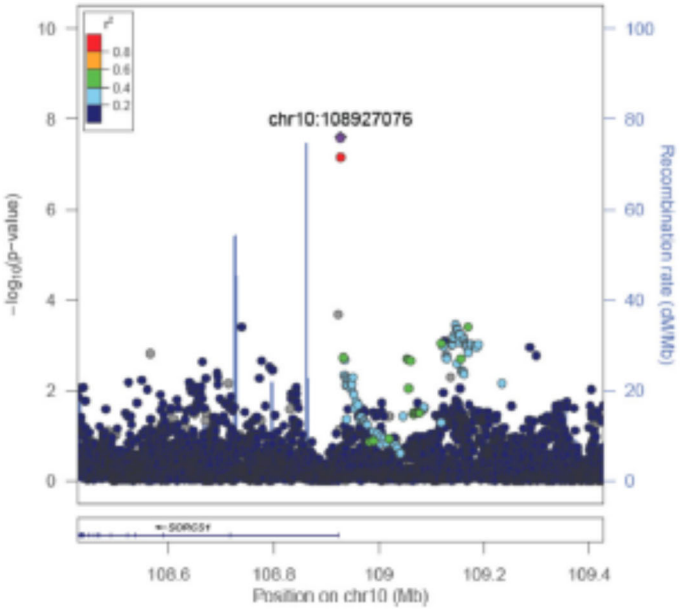
Locus 11: LDL-SomeCol in Trans-Ancestry (2DF).



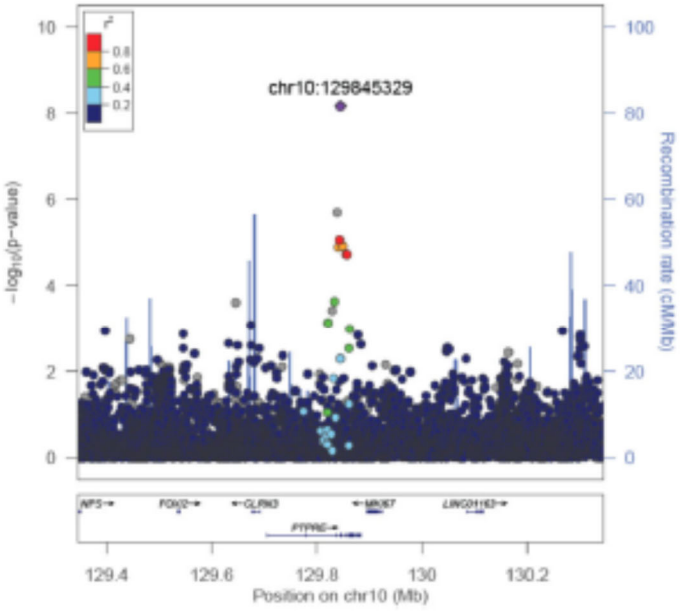
Locus 12: LDL-GradCol in African Ancestry (left – interaction; right – 2DF).



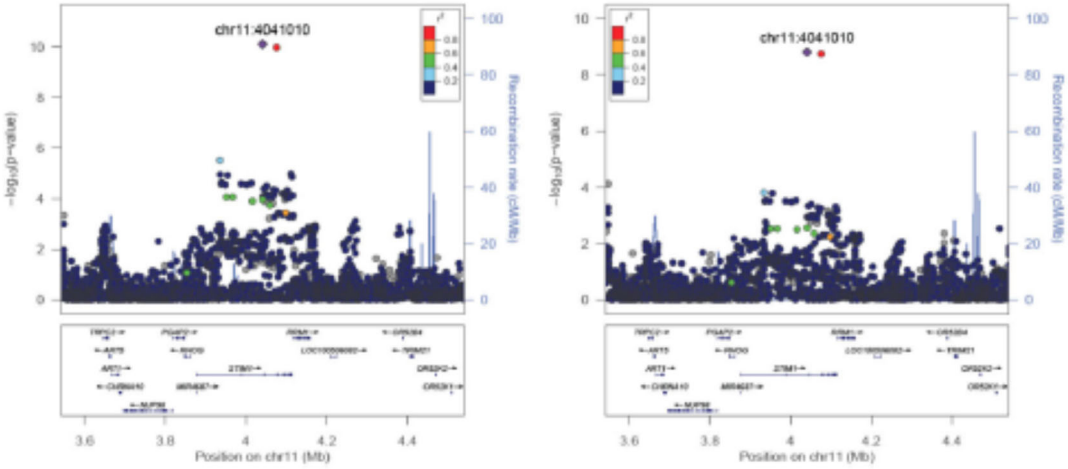
Locus 13: HDL-SomeCol in African Ancestry (2DF).



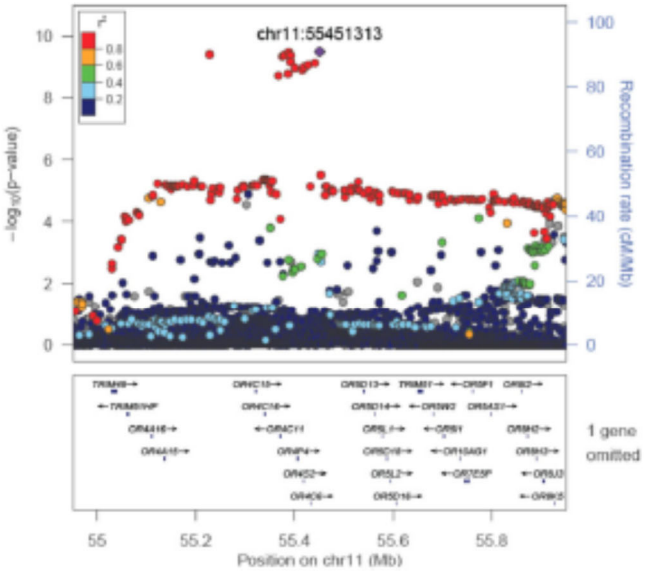
Locus 14: LDL-SomeCol in Trans-Ancestry (2DF).



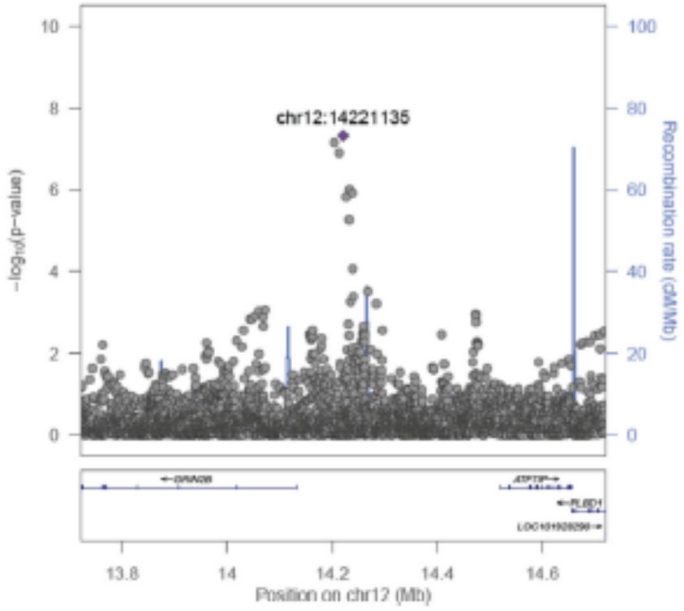
Locus 15: LDL-GradCol in Trans-Ancestry (left – interaction; right – 2DF).



Locus 16: HDL-GradCol in European Ancestry (2DF).



Locus 17: TG-GradCol in African Ancestry (2DF).



Locus 18: HDL-GradCol in African Ancestry (2DF).

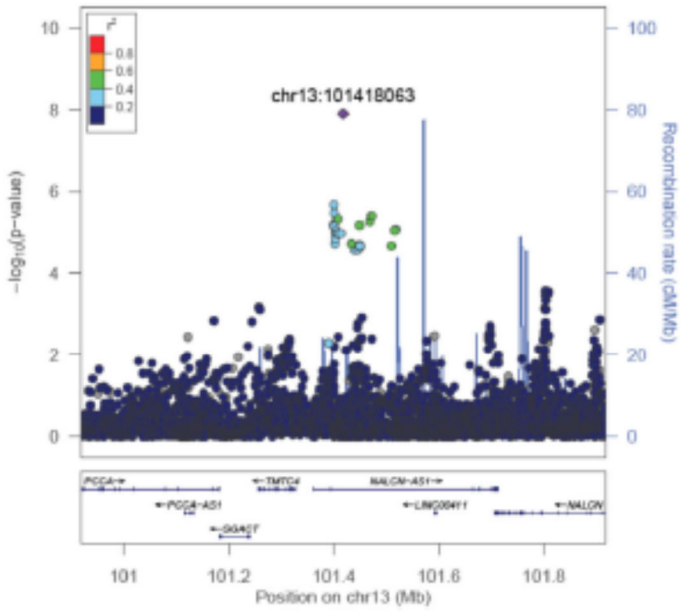


Table 1. Summary of novel loci.

Locus	rsID[1]	Chr	Pos (hg19)	Coded Allele	Coded Allele Freq[2]	$Beta_0$[3]	$Beta_1$[4]	2 DF P[5]	1 DF G×E P[6]	N	nFUMA SNPs6	Func[7]	Genes[8]	RefPop-Trait-Exp[9]
1	rs139845754	1	180,258,167	A	0.052	-0.101	-0.06	3.40E-08	2.00E-01	10,913	98	68 intronic; 18 intergenic; 2 upstream; 1 downstream; 3 ncRNA intronic; 2 ncRNA exonic; 4 NA	*LHX4; RP5-11 80C10.2;* ***ACBD6****; XPR1; KIAA1614; STX6*	AFR-TG-SomeCol
2*	rs7567133	2	103,239,356	C	0.031	0.102	0.021	1.72E-08	1.16E-03	6,361	11	6 intronic; 3 intergenic; 1 ncRNA exonic; 1 NA	*IL1RL1; IL18R1; IL18RAP; SLC9A4;* ***SLC9A2;*** *MFSD9; TMEM182*	AFR-HDL-GradCol
3	rs79367750	3	9,096,107	C	0.096	1.251	-10.279	3.99E-08	1.94E-04	8,855	18	17 intronic; 1 NA	***SRGAP3***	ASN-LDL-SomeCol
4*	rs147731578	3	70,605,665	ATTATT	0.018	-0.145	0.034	1.39E-09	3.51E-09	15,584	1	1 intergenic	*MITF; FOXP1*	EUR-HDL-SomeCol
5*	rs74620279	4	179,620,483	C	0.025	-12.24	11.59	4.67E-08	1.39E-06	5,669	49	47 intergenic; 1 downstream; 1 NA	*SNORD65*	AFR-LDL-SomeCol
6	rs77249395	4	181,181,245	G	0.035	0.006	-0.026	3.17E-06	2.41E-08	134,413	2	2 intergenic	*LINC00290*	TA-HDL-GradCol EUR-HDL-GradCol
7*	rs11132093	4	182,664,815	A	0.061	-1.65	15.84	5.04E-09	4.10E-07	2,809	53	53 intergenic	*RP11-540E16.2; TENM3*	ASN-LDL-GradCol
8	rs190502162	5	37,160,808	C	0.013	9.266	-0.715	3.35E-08	4.04E-07	48,467	1	1 intronic	*SKP2 ; NADK2; RANBP3L; SLC1A3; NIPBL;* ***CPLANE1****; NUP155; WDR70; GNDF*	EUR-LDL-SomeCol
9	rs192718305	5	82,514,005	G	0.025	3.481	19.901	2.05E-08	1.60E-03	5,664	51	49 intronic; 1 intergenic; 1 3'UTR	*RP11-343L5.2; TMEM167A;* ***XRCC4***	AFR-LDL-SomeCol
10*	rs147892694	6	76,693,587	G	0.034	-0.037	-0.087	8.61E-09	1.20E-01	8,599	56	44 intronic; 5 intergenic; 7 NA	*TMEM30A; FILIP1; SENP6; MYO6;* ***IMPG1***	AFR-HDL-SomeCol
11	rs7015	7	97,920,623	A	0.203	-1.142	-0.526	4.56E-08	6.18E-02	157,929	95	61 intronic; 24 intergenic; 3 exonic; 1 upstream; 1 ncRNA exonic; 2 3'UTR; 3 NA	*LMTK2; TECPR1;* ***BRI3****; BAIAP2L1; NPTX2*	TA-LDL-SomeCol
12†	rs77655002	8	30,437,023	C	0.042	-9.97	10.73	5.52E-09	7.65E-08	7,166	106	89 intronic; 5 intergenic; 2 exonic; 1 upstream; 6 downstream; 1 5'UTR; 1 3'UTR; 1 NA	*MBOAT4; RBPMS-AS1; RBPMS;* ***GTF2E2****; GSR; TEX15*	AFR-LDL-GradCol TA-LDL-GradCol
13	rs144190766	10	108,927,076	C	0.03	-0.015	-0.103	2.52E-08	5.64E-03	9,677	188	14 intronic; 173 intergenic; 1 NA		AFR-HDL-SomeCol
14	rs116562538	10	129,845,329	A	0.028	-4.573	-9.049	6.79E-09	4.82E-02	20,939	32	31 intronic; 1 NA	***PTPRE***	TA-LDL-SomeCol
15	rs35287906	11	4,041,010	C	0.014	3.405	-5.247	1.53E-09	7.71E-11	81,020	13	11 intronic; 2 intergenic	*PGAP2; RHOG;* ***STIM1****; RRM1; OR55B1P; RP11-23F23.3*	EUR-LDL-GradCol
16	rs11230661	11	55,451,313	A	0.185	0.01	0.007	3.15E-10	3.90E-01	151,320	925	3 intronic; 746 intergenic; 22 exonic; 54 upstream; 48 downstream; 1 upstream:downstream; 16 ncRNA intronic; 19 ncRNA exonic; 16 NA	*TRIM48; TRIM51; 87* Olfactory Receptor Genes; *APLNR*	EUR-HDL-GradCol EUR-HDL-SomeCol
17*	rs148063115	12	14,221,135	T	0.026	0.022	0.222	4.86E-08	4.35E-03	6,168	37	4 intronic; 33 intergenic	*GRIN2B; RP11-72J9.1; RP11-298E10.1; RN7SL676P; GUCY2C*	AFR-TG-GradCol
18*	rs190746034	13	101,418,063	G	0.022	-0.002	-0.158	1.26E-08	1.06E-05	6,361	35	1 intronic; 2 intergenic; 31 ncRNA intronic; 1 ncRNA exonic	*RP11-151A6.4; TMTC4;* ***NALCN-AS1****; ARF4P3*	AFR-HDL-GradCol

Notes: 1. rsID based on dbSNP build 146; 2. Coded allele frequency; 3. Effect size (beta) of the unexposed group; 4. Effect size (beta) of the exposed group; 5. GWAS 2 DF p-value of the significant lead SNP for this locus; 6. GWAS 1 DF genetic-educational attainment interaction p-value of the significant lead SNP for this locus; 6. Number of SNVs in LD (r2>0.2) with lead variant from FUMA; when N=1, the SNV was not found in FUMA; 7. Functional consequence of the SNP on the gene obtained from ANNOVAR (NA: not available); 8. Nearest gene of all SNVs in LD (r2>0.2) with lead variant; if SNV not in FUMA, gene or nearest flanking coding genes noted. Bolded genes reflect intragenic lead SNV. 9. The reference panel used in FUMA to obtain functional annotations-trait-exposure; if more than one RefPop-Trait-Exp listed, the data provided is for the more significant association which is listed first. * Only significant in Stage I analyses. † 2DF P = 1.83E-08, 1 DF G×E P = 4.14E-08 in TA-LDL-GradCol analyses.
AFR, African ancestry; ASN, Asian ancestry; EUR, European ancestry; GradCol, graduated college; HDL, high-density lipoprotein cholesterol; LDL, low-density lipoprotein cholesterol; NA, not available; SomeCol, some college; TA, trans-ancestry; TG, triglyceride.

Table 1. Summary of novel loci.

Locus	rsID[1]	Chr	Pos (hg19)	Coded Allele	Coded Allele Freq[2]	$Beta_0$[3]	$Beta_1$[4]	2 DF P[5]	1 DF G×E P[6]	N	Genes[7]	RefPop-Trait-Exp[9]
1	rs139845754	1	180,258,167	A	0.052	-0.101	-0.06	3.40E-08	2.00E-01	10,913	*LHX4; RP5-11 80C10.2;* ***ACBD6****; XPR1; KIAA1614; STX6*	AFR-TG-SomeCol
2*	rs7567133	2	103,239,356	C	0.031	0.102	0.021	1.72E-08	1.16E-03	6,361	*IL1RL1; IL18R1; IL18RAP; SLC9A4;* ***SLC9A2;*** *MFSD9; TMEM182*	AFR-HDL-GradCol
3	rs79367750	3	9,096,107	C	0.096	1.251	-10.279	3.99E-08	1.94E-04	8,855	***SRGAP3***	ASN-LDL-SomeCol
4*	rs147731578	3	70,605,665	ATTATT	0.018	-0.145	0.034	1.39E-09	3.51E-09	15,584	*MITF; FOXP1*	EUR-HDL-SomeCol
5*	rs74620279	4	179,620,483	C	0.025	-12.24	11.59	4.67E-08	1.39E-06	5,669	*SNORD65*	AFR-LDL-SomeCol
6	rs77249395	4	181,181,245	G	0.035	0.006	-0.026	3.17E-06	2.41E-08	134,413	*LINC00290*	TA-HDL-GradCol EUR-HDL-GradCol
7*	rs11132093	4	182,664,815	A	0.061	-1.65	15.84	5.04E-09	4.10E-07	2,809	*RP11-540E16.2; TENM3*	ASN-LDL-GradCol
8	rs190502162	5	37,160,808	C	0.013	9.266	-0.715	3.35E-08	4.04E-07	48,467	*SKP2 ; NADK2; RANBP3L; SLC1A3; NIPBL;* ***CPLANE1****; NUP155; WDR70; GNDF*	EUR-LDL-SomeCol
9	rs192718305	5	82,514,005	G	0.025	3.481	19.901	2.05E-08	1.60E-03	5,664	*RP11-343L5.2; TMEM167A;* ***XRCC4***	AFR-LDL-SomeCol
10*	rs147892694	6	76,693,587	G	0.034	-0.037	-0.087	8.61E-09	1.20E-01	8,599	*TMEM30A; FILIP1; SENP6; MYO6;* ***IMPG1***	AFR-HDL-SomeCol
11	rs7015	7	97,920,623	A	0.203	-1.142	-0.526	4.56E-08	6.18E-02	157,929	*LMTK2; TECPR1;* ***BRI3****; BAIAP2L1; NPTX2*	TA-LDL-SomeCol
12†	rs77655002	8	30,437,023	C	0.042	-9.97	10.73	5.52E-09	7.65E-08	7,166	*MBOAT4; RBPMS-AS1; RBPMS;* ***GTF2E2****; GSR; TEX15*	AFR-LDL-GradCol TA-LDL-GradCol
13	rs144190766	10	108,927,076	C	0.03	-0.015	-0.103	2.52E-08	5.64E-03	9,677	*SORCS1; RNA5SP326*	AFR-HDL-SomeCol
14	rs116562538	10	129,845,329	A	0.028	-4.573	-9.049	6.79E-09	4.82E-02	20,939	***PTPRE***	TA-LDL-SomeCol
15	rs35287906	11	4,041,010	C	0.014	3.405	-5.247	1.53E-09	7.71E-11	81,020	*PGAP2; RHOG;* ***STIM1****; RRM1; OR55B1P; RP11-23F23.3*	EUR-LDL-GradCol
16	rs11230661	11	55,451,313	A	0.185	0.01	0.007	3.15E-10	3.90E-01	151,320	*TRIM48; TRIM51; 87* Olfactory Receptor Genes; *APLNR*	EUR-HDL-GradCol EUR-HDL-SomeCol
17*	rs148063115	12	14,221,135	T	0.026	0.022	0.222	4.86E-08	4.35E-03	6,168	*GRIN2B; RP11-72J9.1; RP11-298E10.1; RN7SL676P; GUCY2C*	AFR-TG-GradCol
18*	rs190746034	13	101,418,063	G	0.022	-0.002	-0.158	1.26E-08	1.06E-05	6,361	*RP11-151A6.4; TMTC4;* ***NALCN-AS1****; ARF4P3*	AFR-HDL-GradCol

Notes: 1. rsID based on dbSNP build 146; 2. Coded allele frequency; 3. Effect size (beta) of the unexposed group; 4. Effect size (beta) of the exposed group; 5. GWAS 2 DF p-value of the significant lead SNP for this locus; 6. GWAS 1 DF genetic-educational attainment interaction p-value of the significant lead SNP for this locus; 7. Nearest gene of all SNVs in LD (r2>0.2) with lead variant; if SNV not in FUMA, gene or nearest flanking coding genes noted. Bolded genes reflect intragenic lead SNV; 8. The reference panel used in FUMA to obtain functional annotations-trait-exposure; if more than one RefPop-Trait-Exp listed, the data provided is for the more significant association which is listed first. * Only significant in Stage I analyses. † 2DF P = 1.83E-08, 1 DF G×E P = 4.14E-08 in TA-LDL-GradCol analyses.
AFR, African ancestry; ASN, Asian ancestry; EUR, European ancestry; GradCol, graduated college; HDL, high-density lipoprotein cholesterol; LDL, low-density lipoprotein cholesterol; TA, trans-ancestry; TG, triglyceride.

Table S1: Sample sizes used in Stage 1 (genome-wide discovery) analysis.

Ancestry	Cohort	Max N	SomeCol			GradCol		
			N	E1	E0	N	E1	E0
European	**AGES**	2,395	NA	NA	NA	2,395	287	2,108
	ARIC	9,459	9,459	7,940	1,519	9,459	3,609	5,850
	CARDIA	1,633	1,633	1,238	395	1,633	749	884
	CHS	2,966	2,966	1,347	1,619	2,966	638	2,328
	Croatia-Korcula	483	483	103	380	483	66	417
	Croatia-Vis	459	459	89	370	459	50	409
	ERF	2,575	NA	NA	NA	2,575	154	2,421
	FamHS	3,551	3,551	2,270	1,281	3,551	1,061	2,490
	FHS	7,036	7,036	5,320	1,716	7,036	3,144	3,892
	GENOA	1,429	1,429	741	688	1,429	255	1,174
	GENSCOT	6,125	6,125	2,703	3,422	6,125	1,685	4,440
	GOLDN	816	816	581	235	816	86	730
	Health ABC	1,633	1,633	1,436	197	1,633	870	763
	HERITAGE	500	500	262	238	500	146	354
	HyperGEN	1,246	1,246	753	493	1,246	366	880
	MESA	2,585	2,585	2,053	532	2,585	1,302	1,283
	NEO	5,671	5,671	3,350	2,321	5,671	531	5,140
	RS1	2,957	2,957	310	2,647	NA	NA	NA
	RS2	1,967	1,967	329	1,638	1,967	304	1,663
	RS3	2,971	NA	NA	NA	2,971	798	2,173
	WGHS	22,738	NA	NA	NA	22,738	9,936	12,802
European Total		*80,379*	*49,700*	*30,244*	*19,456*	*77,422*	*25,951*	*51,471*
African	**ARIC**	2,726	2,726	1,626	1,100	2,726	864	1,862
	CARDIA	908	908	481	427	908	130	778
	CHS	725	725	256	469	725	123	602
	GENOA	984	984	378	606	984	195	789
	HANDLS	985	985	325	660	985	70	915
	Health ABC	1,091	1,091	602	489	1,091	277	814
	HyperGEN	1,240	1,240	584	656	1,240	179	1,061
	JHS	1,984	1,984	1,792	192	1,984	864	1,120
	MESA	1,652	1,652	1,148	504	1,652	557	1,095
African Total		*12,295*	*12,295*	*7,192*	*5,103*	*12,295*	*3,259*	*9,036*
Asian	**MESA**	747	747	436	311	747	290	457
	SCES	1,853	1,853	137	1,716	NA	NA	NA
	SCHS-CHD-Cases	719	719	31	688	NA	NA	NA
	SCHS-CHD-Controls	1,283	1,283	89	1,194	NA	NA	NA
	SiMES	2,533	2,533	162	2,371	NA	NA	NA
	SINDI	2,493	2,493	254	2,239	NA	NA	NA
	SP2-1M	950	950	448	502	950	199	751
	SP2-610	1,143	1,143	438	705	1,143	208	935
Asian Total		*11,002*	*11,002*	*1,964*	*9,038*	*2,840*	*697*	*2,143*
Hispanic	**MESA**	1,455	1,455	512	943	1,455	142	1,313
Hispanic Total		*1,455*	*1,455*	*512*	*943*	*1,455*	*142*	*1,313*
Brazilian	**Baependi HS**	902	902	117	785	902	66	836
	Pelotas BCS	2,751	2,751	1,218	1,533	2,751	563	2,188
Brazilian Total		*3,653*	*3,653*	*1,335*	*2,318*	*3,653*	*629*	*3,024*
TOTAL		**108,784**	**78,105**	**41,247**	**36,858**	**97,665**	**30,678**	**66,987**

E0, unexposed group; E1, exposed group.
NOTE: Data from GOLDN and SCHS-CHD-Cases were omitted from meta-analysis due to QC issues

Table S2: Sample sizes used in Stage 2 (focused discovery/replication) analysis.

Ancestry	Cohort	Max N	SomeCol			GradCol		
			N	E1	E0	N	E1	E0
European	**BRIGHT**	1,725	1,725	695	1,030	1,725	158	1,567
	CARL	374	NA	NA	NA	374	27	347
	DESIR1	647	647	102	545	647	70	577
	DHS	1,166	1,166	319	847	1,166	144	1,022
	DR's EXTRA	1,230	1,230	471	759	1,230	307	923
	EGCUT_Human370CN	683	683	440	243	683	306	377
	EGCUT_OmniExpress	1,078	1,078	692	386	1,078	338	740
	EPIC	18,849	18,849	10,387	8,462	18,849	2,630	16,219
	FVG	1,039	NA	NA	NA	1,039	89	950
	GeneSTAR	1,677	1,677	904	773	1,677	512	1,165
	GLACIER	4,327	NA	NA	NA	4,327	793	3,534
	GRAPHIC	988	988	334	654	988	162	826
	HRS	6,904	6,904	2,206	4,698	6,904	1,798	5,106
	JUPITER	8,397	8,397	4,196	4,201	8,397	2,452	5,945
	KORA_S3	3,047	3,047	1,013	2,034	3,047	421	2,626
	KORA_S4	3,750	3,750	865	2,885	3,749	552	3,197
	LBC1936	908	908	228	680	908	124	784
	Lifelines	12,462	12,462	6,671	5,791	12,462	557	11,905
	LLFS	3,158	3,158	2,668	490	3,158	1,685	1,473
	NESDA	2,581	2,581	1,399	1,182	2,581	291	2,290
	PREVEND	3,282	3,028	2,393	635	3,282	1,046	2,236
	SHEEP-Case	1,109	1,109	245	864	NA	NA	NA
	SHEEP-Control	1,478	1,478	420	1,058	NA	NA	NA
	SHIP-0	3,521	3,521	868	2,653	NA	NA	NA
	SHIP-Trend	900	900	442	458	NA	NA	NA
	TWINGENE	5,572	5,572	1,882	3,690	5,572	1,271	4,301
	YFS	1,838	1,838	1,205	633	1,838	587	1,251
European Total		*92,690*	*86,696*	*41,045*	*45,651*	*85,681*	*16,320*	*69,361*
African	**AA-DHS**	577	577	201	376	577	84	493
	GeneSTAR	1,083	1,083	450	633	1,083	189	894
	HRS	1,646	1,646	312	1,334	1,646	220	1,426
	HyperGEN-AXIOM	424	424	212	212	424	85	339
	JUPITER	1,606	1,606	23	1,583	1,606	12	1,594
	LOYOLA-GXE	603	603	188	415	603	120	483
	LOYOLA-SPT	691	691	42	649	691	29	662
African Total		*6,630*	*6,630*	*1,428*	*5,202*	*6,630*	*739*	*5,891*
Asian	**BES-610**	487	487	40	447	NA	NA	NA
	BES-omniexpress	389	389	127	262	NA	NA	NA
	DFTJ	1,410	1,410	142	1,268	NA	NA	NA
	RHS	2,159	2,159	440	1,719	NA	NA	NA
	SMWHS	2,144	2,144	288	1,856	2,144	130	2,014
Asian Total		*6,589*	*6,589*	*1,037*	*5,552*	*2,144*	*130*	*2,014*
Hispanic	**IRAS-C**	184	184	69	115	184	31	153
	IRAS Family Study	984	984	393	591	984	133	851
	HCHS/SOL	10,454	10,454	3,679	6,775	10,454	1,551	8,903
Hispanic Total		*11,622*	*11,622*	*4,141*	*7,481*	*11,622*	*1,715*	*9,907*
TOTAL		**117,531**	**111,537**	**47,651**	**63,886**	**106,077**	**18,904**	**87,173**

Table S3: Demographic and descriptive statistics for Stage 1 (genome-wide discovery) analysis.

Ancestry	Cohort	Max N	% Male	% On Statins	Age					HDL (natural log-transformed)			TG (natural log-transformed)			LDL (adjusted for medications)		
					Mean	SD	Median	Min.	Max.	N	Mean	SD	N	Mean	SD	N	Mean	SD
European	AGES	2,395	42.00	24.38	73.95	3.64	74.00	66	80	2,395	4.07	0.28	2,395	4.58	0.45	2,385	164.20	48.11
	ARIC	9,459	53.07	0.59	54.32	5.69	54.00	44	66	9,459	3.87	0.32	9,255	4.78	0.51	9,058	137.95	38.20
	CARDIA	1,633	46.62	0.00	25.55	3.33	26.00	18	32	1,633	3.92	0.25	1,632	4.22	0.50	1,625	108.61	29.25
	CHS	2,966	38.00	1.48	71.21	4.03	70.00	65	80	2,966	3.97	0.28	2,944	4.85	0.43	2,903	131.73	36.20
	Croatia-Korcula	483	39.34	1.45	55.55	13.32	56.00	18	90	482	4.02	0.23	483	4.70	0.48	481	149.02	40.23
	Croatia-Vis	459	41.48	1.96	53.98	14.32	54.00	18	80	458	3.75	0.17	459	4.91	0.47	458	124.89	35.44
	ERF	2,575	44.56	11.69	48.24	14.13	48.64	18	80	2,575	3.86	0.28	2,552	4.65	0.51	2,550	150.48	38.85
	FamHS	3,551	47.62	6.05	51.76	13.29	53.66	25	80	3,551	3.86	0.29	3,551	4.85	0.57	3,549	128.21	37.67
	FHS	7,036	46.70	3.73	38.11	9.32	38.00	18	72	7,025	3.93	0.29	7,036	4.44	0.62	6,952	119.70	34.61
	GENOA	1,429	44.86	15.26	55.17	10.63	55.64	25	80	1,429	3.89	0.30	1,361	5.15	0.46	1,303	128.61	36.82
	GENSCOT	6,125	41.17	14.81	53.54	11.64	55.00	18	80	6,125	4.01	0.28	NA	NA	NA	NA	NA	NA
	GOLDN	816	49.57	0.00	48.34	15.87	48.00	18	83	816	3.80	0.27	816	4.76	0.59	816	123.04	31.84
	Health ABC	1,633	52.85	14.41	73.78	2.85	74.00	69	80	1,632	3.90	0.30	1,633	4.90	0.49	1,598	126.34	35.32
	HERITAGE	500	48.90	0.00	35.84	14.56	31.66	17	65	500	3.67	0.24	500	4.68	0.49	500	116.11	31.36
	HyperGEN	1,246	49.44	10.75	49.80	13.82	49.50	18	80	1,246	3.81	0.27	1,233	4.94	0.59	1,177	121.62	36.06
	MESA	2,585	47.86	17.10	62.09	9.65	63.00	44	80	2,583	3.92	0.29	2,585	4.74	0.53	2,551	125.75	32.11
	NEO	5,671	48.10	15.40	55.97	5.95	56.00	44	66	5,671	3.97	0.28	5,671	4.71	0.54	5,595	145.45	36.16
	RS1	2,957	43.79	13.97	70.20	5.10	70.00	61	80	2,957	3.94	0.27	2,785	4.80	0.42	2,585	153.22	34.95
	RS2	1,967	46.26	13.37	63.38	6.22	61.55	55	80	1,967	3.93	0.26	1,838	4.83	0.44	1,817	150.89	34.41
	RS3	2,971	43.92	22.05	56.72	6.09	56.69	45	80	2,971	3.97	0.31	2,940	4.76	0.48	2,917	144.94	40.65
	WGHS	22,738	0.00	3.20	54.67	7.05	52.88	39	81	22,738	3.95	0.28	16,392	4.78	0.54	22,738	124.28	34.42
African	ARIC	2,726	62.40	0.26	53.40	5.77	53.00	44	66	2,726	3.96	0.30	2,532	4.58	0.47	2,480	138.53	42.69
	CARDIA	908	38.68	0.00	24.44	3.80	25.00	18	34	908	3.97	0.24	908	4.08	0.44	908	111.38	30.88
	CHS	725	37.66	4.55	71.57	4.15	71.00	65	80	725	4.02	0.26	709	4.65	0.43	704	132.62	39.18
	GENOA	984	29.78	4.07	55.83	10.71	55.63	21	80	984	3.96	0.30	792	4.84	0.44	780	121.94	43.78
	HANDLS	985	44.56	12.38	48.31	8.80	49.00	30	64	985	3.96	0.31	939	4.54	0.48	936	112.59	41.60
	Health ABC	1,091	42.89	10.69	73.45	2.89	73.00	68	80	1,090	4.00	0.29	1,091	4.68	0.43	1,080	129.65	40.94
	HyperGEN	1,240	32.58	4.27	45.09	13.26	45.00	18	79	1,240	3.94	0.27	1,207	4.52	0.51	1,201	121.78	39.96
	JHS	1,984	39.64	9.24	49.49	11.15	49.06	21	80	1,984	3.88	0.27	1,984	4.50	0.54	1,962	122.31	42.96
	MESA	1,652	45.75	12.60	61.70	9.64	62.00	45	80	1,651	3.92	0.28	1,652	4.52	0.48	1,644	123.90	36.74
Asian	MESA	747	49.00	11.84	61.75	9.90	62.00	44	80	747	3.87	0.24	747	4.83	0.52	734	121.42	30.84
	SCES	1,853	51.48	23.96	58.04	9.05	56.92	44	80	1,853	3.87	0.30	NA	NA	NA	1,853	138.45	35.89
	SCHS-CHD-Cases	719	64.39	NA	66.78	7.82	66.80	50	79	719	3.88	0.22	NA	NA	NA	719	128.84	33.73
	SCHS-CHD-Controls	1,283	63.06	NA	66.48	7.78	66.56	50	79	1,282	3.92	0.23	NA	NA	NA	1,283	122.37	30.76
	SiMES	2,533	49.41	12.19	59.08	11.04	58.00	40	80	2,533	3.93	0.24	NA	NA	NA	2,532	142.90	40.79
	SINDI	2,493	50.86	30.16	57.62	9.59	56.00	43	80	2,493	3.67	0.29	NA	NA	NA	2,492	141.42	37.95
	SP2-1M	950	63.68	6.74	46.94	10.26	46.28	25	80	950	3.99	0.24	950	4.61	0.56	939	123.18	32.63
	SP2-610	1,143	23.45	9.19	48.63	11.29	48.85	23	80	1,143	4.05	0.23	1,142	4.55	0.53	1,136	124.94	35.46
Brazilian	Baependi HS	902	40.13	9.64	48.34	14.98	48.00	22	80	902	3.82	0.25	902	4.78	0.46	895	128.00	37.67
	Pelotas BCS	2,751	48.56	NA	30.18	0.34	30.17	29	31	2,751	4.04	0.24	NA	NA	NA	2,751	109.64	29.33
Hispanic	MESA	1,455	48.35	15.16	60.80	9.73	61.00	44	80	1,455	3.83	0.26	1,455	4.93	0.51	1,422	126.45	35.79

Table S4: Demographic and descriptive statistics for Stage 2 (focused discovery/replication) analysis.

Ancestry	Cohort	Max N	% Male	% On Statins	Age					HDL (natural log-transformed)			TG (natural log-transformed)			LDL (adjusted for medications)		
					Mean	SD	Median	Min.	Max.	N	Mean	SD	N	Mean	SD	N	Mean	SD
European	BRIGHT	1,725	39.71	NA	56.96	10.99	58	21	80	1,725	3.90	0.30	1,720	5.10	0.54	1,621	130.95	36.00
	CARL	374	38.23	3.72	48.64	16.21	51	18	79	NA	NA	NA	374	4.73	0.50	NA	NA	NA
	DESIR1	647	39.72	0.00	53.49	5.67	53	45	65	647	4.18	0.25	647	4.36	0.47	647	130.75	32.36
	DHS	1,166	46.83	40.19	61.69	9.07	62	34	80	1,166	3.72	0.28	1,166	5.14	0.56	1,091	121.00	35.91
	DR's EXTRA	1,230	47.20	34.60	66.48	5.38	66	57	79	1,230	4.15	0.29	1,230	4.66	0.44	1,230	139.96	34.54
	EGCUT_Human370CNV	683	49.60	1.00	38.63	15.88	35	18	80	683	4.04	0.29	89	4.67	0.53	683	133.77	43.34
	EGCUT_OmniExpress	1,078	52.00	5.10	49.18	15.98	51	18	80	1,078	4.07	0.28	884	4.65	0.50	1,078	135.50	39.52
	EPIC	18,849	47.00	1.02	59.18	9.22	59	39	79	18,222	3.96	0.29	18,849	4.94	0.53	18,223	155.00	43.13
	FVG	1,039	41.38	4.23	50.78	15.56	51	18	79	1,039	3.96	0.24	1,039	4.57	0.48	1,039	145.40	42.87
	GeneSTAR	1,677	46.57	8.47	43.23	10.71	43	18	79	1,676	3.91	0.30	1,677	4.81	0.55	1,641	137.17	42.91
	GLACIER	4,327	35.06	1.41	50.33	8.41	50	29	64	3,226	3.96	0.29	4,327	4.70	0.44	2,591	165.21	45.09
	GRAPHIC	988	50.10	3.85	52.89	4.41	53	40	61	988	4.00	0.26	988	4.98	0.48	988	127.73	29.39
	HRS	6,904	42.32	48.15	64.51	8.82	65	33	80	6,904	3.96	0.30	NA	NA	NA	NA	NA	NA
	JUPITER	8,397	68.50	0.00	65.37	7.03	65	50	80	8,397	3.91	0.28	8,397	4.79	0.49	8,397	106.37	17.35
	KORA_S3	3,047	48.50	3.30	47.32	12.96	47	24	75	3,047	3.96	0.29	246	4.70	0.60	3,041	141.21	45.45
	KORA_S4	3,750	48.94	5.97	49.17	13.86	49	24	75	3,750	4.02	0.29	1,277	4.77	0.49	3,740	140.60	45.13
	LBC1936	908	50.22	28.74	69.68	0.76	70	68	71	908	4.03	0.28	NA	NA	NA	NA	NA	NA
	Lifelines	12,462	41.71	15.34	48.29	11.28	47	18	80	12,461	3.99	0.27	12,462	4.58	0.51	12,342	135.53	41.17
	LLFS	3,158	44.65	23.56	61.26	8.71	61	24	80	3,158	4.05	0.30	2,924	4.59	0.52	3,125	132.34	34.86
	NESDA	2,581	33.40	6.50	41.79	13.17	43	18	65	2,575	4.11	0.27	2,581	4.60	0.51	2,561	123.94	39.23
	PREVEND	3,282	52.41	3.47	48.97	12.36	48	28	75	3,282	3.88	0.30	3,181	4.68	0.53	3,117	156.91	43.17
	SHEEP-Case	1,109	71.00	2.80	59.19	7.17	60	45	70	1,095	3.69	0.28	1,109	5.04	0.51	1,053	165.26	39.14
	SHEEP-Control	1,478	67.70	1.42	59.84	7.11	61	45	70	1,470	3.87	0.28	1,478	4.76	0.49	1,453	154.32	37.97
	SHIP-0	3,521	49.59	7.53	53.28	14.08	54	29	80	3,521	3.98	0.30	NA	NA	NA	3,502	146.14	46.75
	SHIP-Trend	900	42.78	7.90	52.35	11.97	52	30	80	900	4.02	0.25	900	4.71	0.51	900	138.29	36.16
	TWINGENE	5,572	47.97	11.84	63.68	7.53	64	47	80	5,572	3.95	0.29	5,572	4.66	0.49	5,507	150.17	37.24
	YFS	1,838	44.45	1.90	37.70	5.01	39	30	45	1,838	3.92	0.24	1,838	4.65	0.49	1,838	122.01	32.18
African	AADHS	577	43.15	46.97	56.18	9.45	56	33	79	577	3.83	0.26	577	4.67	0.52	563	128.28	40.20
	GeneSTAR	1,083	37.58	5.17	42.58	10.46	43	20	75	1,083	3.94	0.29	1,083	4.56	0.52	1,075	130.86	44.57
	HRS	1,646	36.63	64.34	61.76	8.36	61	28	80	1,646	3.98	0.29	NA	NA	NA	NA	NA	NA
	HyperGEN-AXIOM	424	46.22	4.01	48.29	12.23	47	18	76	424	3.93	0.27	412	4.53	0.54	405	121.30	36.54
	JUPITER	1,606	50.93	0.00	65.47	7.08	66	50	80	1,606	3.91	0.29	1,606	4.70	0.47	1,606	94.22	23.59
	LOYOLA-GXE	603	23.05	0.00	40.18	8.29	40	25	78	603	3.92	0.22	603	4.32	0.50	601	135.61	39.06
	LOYOLA-SPT	691	39.07	0.00	46.53	13.84	45	24	76	691	3.85	0.25	691	4.39	0.50	689	123.43	39.32
Asian	BES-610	487	34.09	7.19	61.81	7.80	61	50	80	486	4.04	0.31	482	4.78	0.59	487	136.52	41.08
	BES-omniexpress	389	38.82	15.94	64.03	8.77	65	50	80	389	3.93	0.28	389	4.87	0.55	389	137.93	41.43
	DF-TJ	1,410	77.94	12.89	63.48	7.70	63	35	80	1,410	3.96	0.31	1,410	4.66	0.55	1,410	121.33	39.60
	RHS	2,159	46.64	0.00	52.05	7.97	52	30	77	2,159	3.90	0.09	2,158	4.75	0.46	2,159	136.90	36.16
	SMWHS	2,144	13.81	NA	57.29	9.26	59	40	75	2,144	3.88	0.25	288	4.92	0.54	2,144	105.16	32.60
Hispanic	IRAS-C	184	40.77	5.95	53.92	8.20	54	40	69	184	3.71	0.33	184	4.90	0.55	184	144.07	42.72
	IRAS Family Study	984	41.97	1.65	39.24	12.84	38	18	80	984	3.74	0.27	984	4.60	0.59	984	110.13	30.50
	HCHS/SOL	10,454	41.04	10.29	45.69	13.76	47	18	76	10,453	3.86	0.26	10,454	4.76	0.55	10,272	128.41	39.70

Table S5: Genotyping, imputation, and statistical analysis for Stage 1 (genome-wide discovery) analysis.

Ancestry	Cohort	Genotyping Platform	Imputation Software	Association Test
European	**AGES**	Illumina Hu370CNV	MaCH/minimac	ProbABEL
	ARIC	Affymetrix 6.0	IMPUTE2	MMAP
	CARDIA	Affymetrix 6.0	BEAGLE	ProbABEL
	CHS	Illumina 370CNV / ITMAT-Broad-CARe Illumina iSELECT chip	MaCH/minimac	R
	CROATIA-Korcula	Illumina HumanHap 370CNV DUO/QUAD Phase 1	ShapeIt/IMPUTE2	ProbABEL
	CROATIA-Vis	Illumina HumanHap300v1	ShapeIt/IMPUTE2	ProbABEL
	ERF	Illumina 6k/318K/250K/350K/610K	MaCH/minimac	ProbABEL
	FamHS	Illumina HumMap550K/610quad/Human1MDuov3	MaCH/minimac	MMAP
	FHS	Affymetrix Nsp, Sty and 50K gene centric	MaCH/minimac	MMAP
	GENOA	Affymetrix 6.0 / Illumina 1M-Duo	IMPUTE2	MMAP
	GENSCOT	HumanOmniExpressExome8v1-2_A	ShapeIt/IMPUTE2	ProbABEL
	GOLDN	Affymetrix 6.0	MaCH	MMAP
	Health ABC	Illumina Human 1M-Duov3	MaCH/minimac	R
	HERITAGE	Illumina 370CNV	minimac	ProbABEL
	HyperGEN	Affymetrix 5.0	MaCH/minimac	MMAP
	MESA	Affymetrix 6.0	IMPUTE2	ProbABEL
	NEO	Illumina HumanCoreExome-24v1_A	IMPUTE2	ProbABEL
	RS1	Illumina 550 (+duo) / Illumina 610 quad	MaCH/minimac	ProbABEL
	RS2	Illumina 550 duo	MaCH/minimac	ProbABEL
	RS3	Illumina 610 quad	MaCH/minimac	ProbABEL
	WGHS	Illumina HumanHap 300 DuoPlus	MaCH/minimac	ProbABEL
African	**ARIC**	Affymetrix 6.0	IMPUTE2	MMAP
	CARDIA	Affymetrix 6.0	minimac	ProbABEL
	CHS	Illumina HumanOmni1-Quad_v1	IMPUTE2	R
	GENOA	Affymetrix 6.0 / Illumina 1M-Duo	IMPUTE2	MMAP
	HANDLS	Illumina 1M/1M-Duo/550K/370K/510S/240S	MaCH/minimac	ProbABEL
	Health ABC	Illumina Human 1M-Duov3	MaCH/minimac	R
	HyperGEN	Affymetrix 6.0	MaCH/minimac	MMAP
	JHS	Affymetrix 5.0	IMPUTE2	ProbABEL
	MESA	Affymetrix 6.0	IMPUTE2	ProbABEL
Asian	**MESA**	Affymetrix 6.0	IMPUTE2	ProbABEL
	SCES	Illumina610Quad	MaCH (UM imputation server)	ProbABEL
	SCHS-CHD-Cases	Illumina Illumina Omni Zhonghua-8	IMPUTE2	ProbABEL
	SCHS-CHD-Controls	Illumina Illumina Omni Zhonghua-8	IMPUTE2	ProbABEL
	SiMES	Illumina610Quad	MaCH (UM imputation server)	ProbABEL
	SINDI	Illumina610Quad	MaCH (UM imputation server)	ProbABEL
	SP2-1M	Illumina1Mduov3	MaCH	ProbABEL
	SP2-610	Illumina610Quad	MaCH	ProbABEL
Brazilian	**Baependi HS**	Affymetrix 6.0	ShapeIt/IMPUTE2	MMAP
	Pelotas BCS	Illumina HumanOmni2.5-8v1	IMPUTE2	ProbABEL
Hispanic	**MESA**	Affymetrix 6.0	IMPUTE2	ProbABEL

Table S6: Genotyping, imputation, and statistical analysis for Stage 2 (focused discovery/replication) analysis.

Ancestry	Cohort	Genotyping Platform	Imputation Software	Association Test
European	**BRIGHT**	Affymetrix GeneChip 500k array	MaCH/minimac	ProbABEL
	CARL	Illumina 370K	IMPUTE2	ProbABEL
	DESIR-1	Illumina	ShapeIt/IMPUTE2	R
	DHS	Affymetrix 5.0	IMPUTE2	GWAF
	DR's EXTRA	Cardio-Metabo_Chip_11395247_A / HumanExome-12v1-1_A	MaCH/minimac	ProbABEL
	EGCUT_Human370CNV	Illumina (OmniExpress; Infinium CoreExome-24; HumanCNV370)	IMPUTE2	ProbABEL/R
	EGCUT_OmniExpress	Illumina (OmniExpress; Infinium CoreExome-24; HumanCNV370)	IMPUTE2	ProbABEL/R
	EPIC	UKBioBank Axiom	ShapeIt/IMPUTE	QUICKTEST
	FVG	Illumina 370K	IMPUTE2	ProbABEL
	GeneSTAR	Illumina 1M_v1C	IMPUTE2	MMAP
	GLACIER	Illumina CardioMetaboCHip	NA	QUICKTEST
	GRAPHIC	HumanOmniExpress-12v1 array	IMPUTE2	SNPTEST
	HRS	Illumina Omni 2.5	ShapeIt/IMPUTE	ProbABEL
	JUPITER	Illumina Omni 1M Quad platform	MaCH/minimac	ProbABEL
	KORA_S3	Illumina Omni 2.5/Illumina Omni Express	IMPUTE2	SNPTEST
	KORA_S4	Affymetrix Axiom	IMPUTE2	SNPTEST
	LBC1936	Illumina 610-Quadv1	minimac	ProbABEL
	LifeLines	Illumina Cyto SNP12 v2	minimac	QUICKTEST
	LLFS	Illumina Omni 2.5	MaCH/minimac	MMAP
	NESDA	Affymetrix 5.0; Affymetrix 6.0 907K	MaCH/minimac	QUICKTEST
	PREVEND	Illumina Cyto SNP12 v2	BEAGLE	SNPTEST
	SHEEP-Case	Illumina Cardiometabochip	NA	PLINK
	SHEEP-Control	Illumina Cardiometabochip	NA	PLINK
	SHIP-0	Affymetrix Genome-wide Human SNP Array 6.0	IMPUTE2	ProbABEL
	SHIP-Trend	Illumina HumanOmni 2.5	IMPUTE2	ProbABEL
	TWINGENE	Illumina OmniExpress	MaCH/minimac	ProbABEL
	YFS	Illumina 670k custom	IMPUTE2	R/sandwich
African	**AADHS**	Illumina Omni5 array	IMPUTE2	SAS PROC REG
	GeneSTAR	Illumina 1M_v1C	IMPUTE2	MMAP
	HRS	Illumina Omni 2.5	ShapeIt/IMPUTE	ProbABEL
	HyperGEN-AXIOM	Affymetrix Axiom	MaCH-Admix	ProbABEL
	JUPITER	Illumina Omni 1M Quad	MaCH/minimac	ProbABEL
	LOYOLA-GXE	Illumina Metabochip	MaCH/minimac	ProbABEL
	LOYOLA-SPT	Illumina Metabochip	MaCH/minimac	ProbABEL
Asian	**BES-610**	Illumina Human610-Quad	MaCH (UM imputation server)	ProbABEL
	BES-omniexpress	Illumina OmniExpress	MaCH (UM imputation server)	ProbABEL
	DF-TJ	Affymetrix 6.0	MaCH/minimac	ProbABEL
	RHS	Illumina 550K / Omni 2.5M	BEAGLE	R/sandwich
	SMWHS	Affymetrix 6.0, Illumina Omni Express, Illumina 550, Illumina 1M	MaCH/minimac	R
Hispanic	**IRAS-C**	Illumina OmniExpress+1S	IMPUTE2	GWAF
	IRAS Family Study	Illumina OmniExpress+1S	IMPUTE2	GWAF
	HCHS/SOL	Illumina SOL HCHS Custom 15041502 B3 array (Illumina Omni 2.5M-8v1-1 + Custom SNPs)	IMPUTE2	SUGEN

Table S7: Summary of known loci.

Notes: 1. rsID based on dbSNP build 146, selected as representative of the locus based on published literature; 2.References for published literature located in table footer.

Known Locus	Chr	Pos	rsID[1]	Locus	Refs[2]	LDL						HDL						TG					
						AFR	ASN	BRZ	EUR	HIS	TA	AFR	ASN	BRZ	EUR	HIS	TA	AFR	ASN	BRZ	EUR	HIS	TA
1	1	25,768,937	rs10903129	TMEM57/LDLRAP1	6,10,17				1.75E-15		2.20E-16												
2	1	27,138,393	rs12748152	PIGV/NR0B2	17										3.95E-14		2.48E-15						
3	1	40,028,180	rs4660293	PABPC4	10,17										9.14E-18		9.84E-21						
4	1	55,496,039	rs11206510	PCSK9	3,5,7,10,14,17,19,26	1.01E-27			2.05E-93	1.10E-15	3.57E-104												
5	1	63,049,593	rs1748195	DOCK7/ANGPTL3	3,5,6,7,10,15,17,26													9.96E-11	2.04E-08		9.56E-66	1.17E-16	1.26E-84
6	1	93,009,438	rs7515577	EVI5	10,17										2.76E-12		9.79E-15						
7	1	109,818,530	rs646776	CELSR/SORT1/CELSR2	3,4,5,6,7,9,10,17,26	1.63E-26	2.02E-26	5.51E-12	1.43E-201	4.19E-43	3.15E-272												
8	1	178,515,312	rs4650994	ANGPTL1	17										2.68E-10		4.30E-10						
9	1	182,168,885	rs1689800	ZNF648	10,17										1.67E-09		3.32E-12						
10	1	220,973,563	rs2642442	MOSC1	10,17				5.06E-13		1.48E-18												
11	1	230,295,691	rs4846914	GALNT2	3,7,10,17										1.82E-25		4.41E-28						
12	1	234,858,597	rs514230	IRF2BP2	10,17				3.96E-16		7.42E-17												
13	2	21,232,195	rs693	APOB	3,5,6,7,9,10,14,15,17,19,26	1.03E-11	4.93E-13		2.62E-200	2.72E-23	9.96E-195												
14	2	27,741,237	rs780094	GCKR	1,3,5,6,7,9,10,14,15,17													2.26E-08	9.89E-13		2.98E-191	1.66E-27	6.65E-230
15	2	44,073,881	rs6544713	ABCG5/ABCG8	6,7,10,14,17,23				1.90E-39		3.11E-47												
16	2	46,166,321	rs12712955	PRKCE	25							3.66E-08					3.66E-08						
17	2	63,149,557	rs2710642	EHBP1	17				2.13E-08		5.11E-13												
18	2	109,232,388	rs826682	LIMS1	25						4.32E-08												
19	2	118,835,841	rs10490626	INSIG2	17				3.36E-10		5.05E-10												
20	2	165,548,569	rs10490694	COBLL1	8,10,17										3.48E-16		4.91E-18						
21	2	227,100,698	rs2972146	IRS1	10,17										1.17E-16		4.13E-19						
22	3	12,628,920	rs2290159	RAF1	10,17				4.99E-10		7.74E-10												
23	3	50,129,399	rs2013208	RBM5	17																		
24	3	135,926,622	rs645040	MSL2L1	10,17																3.93E-12		7.36E-12
25	4	3,473,139	rs6831256	LRPAP1	17																3.63E-16		5.22E-16
26	4	26,062,990	rs10019888	SMIM20	17										1.73E-08								
27	4	73,696,709	rs117087731	ADAMTS3	21	5.69E-09																	
28	4	88,030,261	rs442177	KLHL8	10																1.00E-16		6.75E-19
29	4	103,188,709	rs13107325	SLC39A8	10,11,17										6.38E-20		1.03E-20						
30	4	154,224,048	rs17369400	TRIM2	25												1.57E-08						
31	5	55,861,786	rs9686661	MAP3K1	10,17																8.25E-14		6.70E-16
32	5	74,648,603	rs12654264	HMGCR	3,6,7,10,17		3.22E-15		1.50E-57	6.95E-14	5.84E-73												
33	5	156,398,169	rs1501908	TIMD4/HAVCR1	5,10,15,17,26				1.48E-19		1.24E-24												
34	5	158,007,165	rs2914228	EBF1	25												1.57E-09				6.37E-09		
35	6	16,127,407	rs3757354	MYLIP	10,11,17				5.14E-22		2.77E-22												
36	6	26,093,141	rs1800562	HFE	10,17																5.13E-20		2.39E-19
37	6	31,265,490	rs2247056	HLA	10,15,17						2.97E-10										2.66E-10		3.52E-11
38	6	34,546,560	rs2814982	ILRUN	10,17										3.58E-08		1.92E-08						
39	6	43,757,896	rs998584	VEGFA	17																3.12E-19		5.90E-22
40	6	116,312,893	rs9488822	FRK	10,17						1.70E-14												
41	6	127,436,064	rs1936800	RSPO3	17												3.25E-12				1.15E-09		
42	6	139,829,666	rs605066	CITED2	10,17																1.16E-10		3.20E-13
43	6	160,578,860	rs1564348	LPA	10,17,26				9.10E-32	3.42E-10	5.59E-37												
44	7	17,919,258	rs4142995	SNX13	17										5.42E-14		8.73E-17						
45	7	21,607,352	rs12670798	DNAH11	6,10,17				8.47E-11		1.16E-12												
46	7	25,991,826	rs4722551	MIR148A	16,17				2.94E-08		3.39E-10												
47	7	44,579,180	rs2072183	NPC1L1	10,17				1.32E-16		6.77E-16												
48	7	72,982,874	rs17145738	MLXIPL	3,5,6,7,10,17							2.76E-35									1.31E-94	9.27E-10	1.49E-106
49	7	80,340,622	rs2366858	CD36	14							4.15E-17					6.31E-19						
50	7	100,285,974	rs221797	GIGYF1	25				4.42E-08		2.20E-09												
51	7	130,433,384	rs4731702	KLF14	10,17										8.01E-27		2.54E-31						
52	8	9,183,358	rs9987289	PPP1R3B	10,11,14,17,26							1.05E-09			1.34E-41	1.95E-19	1.45E-62						
53	8	11,045,161	rs7819412	XKR6	7																7.28E-12		4.55E-12
54	8	18,272,881	rs1495741	NAT2	10,17																5.13E-15		8.82E-17
55	8	19,819,724	rs328	LPL	2,3,5,6,7,10,13,14,17,19,23,26							1.13E-20	5.23E-15								1.06E-194	6.25E-23	1.41E-214
56	8	55,421,614	rs10102164	SOX17	17				2.21E-13		3.55E-13												
57	8	59,388,565	rs2081687	CYP7A1	10,17				5.54E-12		1.23E-12												
58	8	116,599,199	rs2293889	TRPS1	10,17										4.10E-09		1.12E-11						
59	8	126,486,409	rs17321515	TRIB1	3,5,6,7,10,15,17																1.42E-94	6.10E-12	3.70E-94
60	9	15,289,578	rs471364	TTC39B	7,10,11,17										2.77E-23		1.08E-23						
61	9	107,647,655	rs3890182	ABCA1	3,5,6,7,1,15,17,26										5.63E-49	3.73E-10	9.89E-57						
62	9	136,155,000	rs635634	ABO	10,17		2.74E-10		5.22E-43	1.05E-08	1.54E-55												
63	10	46,013,277	rs970548	MARCH8/ALOX5	17										9.05E-09		2.58E-10						
64	10	65,027,610	rs10761731	JMJD1C	10,17																3.57E-14		3.91E-17
65	10	94,839,642	rs2068888	CYP26A1	10,17																5.01E-09		1.53E-10
66	10	113,933,886	rs2255141	GPAM	10,15,17										5.72E-12		3.73E-16						
67	11	46,743,247	rs3136441	LRP4	10,17										7.36E-28		6.86E-28						
68	11	61,569,830	rs174546	FADS1/FADS2/FADS3	6,7,9.10,11,15,17										1.13E-36		5.29E-40					7.45E-12	
69	11	75,455,021	rs499974	MOGAT2/DGAT2	17										1.19E-12		1.23E-17						
70	11	116,652,207	rs12286037	APO/APOA1/APOA5/APOA/APOC/ZNF259	3,5,6,7,10,13,14,15,17,19,26													5.50E-13	3.21E-65		2.05E-304	7.67E-73	5.00E-300
71	11	126,243,952	rs11220462	ST3GAL4	10,17				6.50E-19		1.73E-21												
72	12	7,651,138	rs10743940	CD163/APOBEC1	24								4.52E-08				9.39E-09						
73	12	109,895,168	rs2338104	MVK	5,7,10,17										2.11E-11		3.23E-17						
74	12	111,414,461	rs12229654	MYL2	12				4.52E-08		2.32E-10												
75	12	121,388,962	rs2650000	HNF1A	7,12,17				1.51E-19		3.22E-22												
76	12	123,796,238	rs4759375	SBNO1	10,17										2.58E-14		4.24E-15						
77	12	125,261,593	rs838880	SCARB1	10,17										5.20E-25		1.83E-25						
78	13	32,953,388	rs4942486	BRCA2	17				1.88E-08		1.65E-10												
79	14	24,883,887	rs8017377	NYNRIN	10.17				3.77E-10		3.32E-12												
80	15	44,245,931	rs2929282	FRMD5	10,17										4.27E-20		3.04E-23						
81	15	58,674,695	rs4775041	LIPC	3,5,6,7,9,10,17,26							7.09E-16	5.58E-16		5.86E-142	6.41E-18	1.28E-168						
82	15	63,396,867	rs2652834	LACTB	10,17										1.00E-08		6.30E-11						
83	16	56,993,324	rs3764261	CETP	2,3,5,6,7,9,10,14,15,17,19,26							1.18E-76	3.57E-42	7.39E-17	1.11E-300		7.29E-303						
84	16	67,902,070	rs2271293	CTCF/PRMT8/LCAT	6,7,26							4.17E-20			1.41E-37	6.05E-10	3.97E-60						
85	16	72,048,632	rs8062895	DHODH/HPR	23,26		6.60E-09		1.41E-22		1.60E-28												
86	16	81,534,790	rs2925979		10,17										7.23E-15		4.51E-21						
87	17	37,813,856	rs11869286	STARD3	10,17										1.94E-23		2.07E-25						
88	17	41,878,166	rs8077889	MPP3	17,23										1.47E-36		9.54E-40						
89	17	45,425,115	rs7206971	OSBPL7	10,17				1.47E-13		5.03E-18												
90	17	64,210,580	rs1801689	APOH/PRXCA	17						7.64E-09												
91	17	66,875,294	rs4148008	ABCA8	10,17				2.09E-23		2.05E-23												
92	17	76,403,984	rs4129767	PGS1	10,17										1.83E-10		3.75E-15						
93	18	47,181,668	rs2156552	LIPG	3,5,6,7,10,17,19,26								6.88E-17		1.73E-65		9.82E-73						
94	18	57,849,023	rs12967135	MC4R	10,17												4.65E-09						
95	19	7,224,431	rs7248104	INSR	17																9.70E-10		1.18E-09
96	19	8,469,738	rs2967605	ANGPTL4	7,10,17,19										3.30E-23		2.41E-24	1.30E-30					
97	19	11,202,306	rs6511720	LDLR	3,5,6,7,10,14,17,19,26	4.44E-17	9.38E-13	7.75E-10	1.70E-203	8.11E-20	1.48E-233												
98	19	19,658,472	rs16996148	TM6SF2/CILP2/ZNF101	3,5,6,7,10,15,17,18																3.38E-60	3.47E-14	4.06E-72
99	19	33,899,065	rs731839	PEPD	17										1.30E-08		4.02E-09						
100	19	45,422,946	rs4420638	APOE	3,5,6,7,10,14,15,19,26	6.32E-81	5.52E-22		8.11E-13	3.02E-57	5.00E-300				<5.00E-305								
101	19	49,206,417	rs492602	MAMSTR	10,17				6.39E-11		8.00E-14												
102	19	52,324,216	rs17695224	HAS1	17										3.88E-12		7.24E-12						
103	19	54,792,761	rs386000	LILRA3	10,17										5.04E-34		3.53E-38						
104	20	17,845,921	rs2328223	SNX5	17				4.77E-18		1.08E-19												
105	20	39,228,784	rs6102059	MAFB	7,10,17				1.01E-14		4.20E-20												
106	20	43,042,364	rs1800961	HNF4A	7,10,17										1.89E-32	8.42E-11	1.51E-40						
107	20	44,576,502	rs7679	PLTP	7,10,17							4.67E-08			5.54E-38		1.30E-42						
108	21	46,875,775	rs114139997	COL18A1	19,26												2.07E-08	7.38E-09					
109	22	21,932,068	rs181362	UBE2L3	10,17										2.23E-12		3.75E-16						
110	22	35,711,098	rs138777	TOM1	17					4.96E-08													

References:

1. Saxena R, Voight BF, Lyssenko V, Burtt NP, de Bakker PI, Chen H, Roix JJ, Kathiresan S, Hirschhorn JN, Daly MJ, et al. (2007) Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels. Science 31:1331-1336.
2. Heid IM, Boes E, Muller M, Kollerits B, Lamina C, Coassin S, Gieger C, Doring A, Klopp N, Frikke-Schmidt R, et al. (2008) Genome-wide association analysis of high-density lipoprotein cholesterol in the population-based KORA study sheds new light on intergenic regions. Circ Cardiovasc Genet 1:10-20.
3. Kathiresan S, Melander O, Guiducci C, Surti A, Burtt NP, Rieder MJ, Cooper GM, Roos C, Voight BF, Havulinna AS, et al. (2008) Six new loci associated with blood low-density lipoprotein cholesterol, high-density lipoprotein cholesterol or triglycerides in humans. Nat Genet 40:189-197.
4. Sandhu MS, Waterworth DM, Debenham SL, Wheeler E, Papadakis K, Zhao JH, Song K, Yuan X, Johnson T, Ashford S, et al (2008) LDL-cholesterol concentrations: a genome-wide association study. Lancet 371:483-491.
5. Willer CJ, Sanna S, Jackson AU, Scuteri A, Bonnycastle LL, Clarke R, Heath SC, Timpson NJ, Najjar SS, Stringham HM, et al. (2008) Newly identified loci that influence lipid concentrations and risk of coronary artery disease. Nat Genet 40:161-169.
6. Aulchenko YS, Ripatti S, Lindqvist I, Boomsma D, Heid IM, Pramstaller PP, Penninx BW, Janssens AC, Wilson JF, Spector T, et al. (2009) Loci influencing lipid levels and coronary heart disease risk in 16 European population cohorts. Nat Genet 41:47-55.
7. Kathiresan, S Willer, CJ Peloso, GM Demissie, S Musunuru, K Schadt, EE Kaplan L, Bennett D, Li Y, Tanaka T, et al. (2009) Common variants at 30 loci contribute to polygenic dyslipidemia. Nat Genet 41:56-65.
8. Ridker PM, Pare G, Parker AN, Zee RY, Miletich JP, Chasman, DI. (2009) Polymorphism in the CETP gene region HDL cholesterol, and risk of future myocardial infarction: Genomewide analysis among 18,245 initially healthy women from the Women's Genome Health Study. Circ Cardiovasc Genet 2:26-33.
9. Sabatti C, Service SK, Hartikainen AL, Pouta A, Ripatti S, Brodsky J, Jones CG, Zaitlen NA, Varilo T, Kaakinen M, et al. (2009) Genome-wide association analysis of metabolic traits in a birth cohort from a founder population. Nat Genet 41:35-46.
10. Teslovich TM, Musunuru K, Smith AV, Edmondson AC, Stylianou IM, Koseki M, Pirruccello JP, Ripatti S, Chasman DI, Willer CJ, et al. (2010) Biological, clinical and population relevance of 95 loci for blood lipids. Nature 466:707-713.
11. Waterworth DM, Ricketts SL, Song K, Chen L, Zhao JH, Ripatti S, Aulchenko YS, Zhang W, Yuan X, Lim N, et al. (2010) Genetic variants influencing circulating lipid levels and risk of coronary artery disease. Arterioscler Thromb Vasc Biol 30:2264-2276.
12. Kim YJ, Go MJ, Hu C, Hong CB, Kim YK, Lee JY, Hwang JY, Oh JH, Kim DJ, Kim NH, et al. (2011) Large-scale genome-wide association studies in East Asians identify new genetic loci influencing metabolic traits. Nat Genet 43:990-995.
13. Tan A, Sun J, Xia N, Qin X, Hu Y, Zhang S, Tao S, Gao Y, Yang X, Zhang H, et al (2012) A genome-wide association and gene-environment interaction study for serum triglycerides levels in a healthy Chinese male population. Hum Mol Genet 21:1658-1664.
14. Coram MA, Duan Q, Hoffmann TJ, Thornton T, Knowles JW, Johnson NA, Ochs-Balcom HM, Donlon TA, Martin LW, Eaton CB, et al. (2013) Genome-wide characterization of shared and distinct genetic components that influence blood lipid levels in ethnically diverse human populations. Am J Hum Genet 92:904-916.
15. Do R, Willer CJ, Schmidt EM, Sengupta S, Gao C, Peloso GM, 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.
16. Weissglas-Volkov D, Aguilar-Salinas CA, Nikkola E, Deere KA, Cruz-Bautista I, Arellano-Campos O, Munoz-Hernandez LL, Gomez-Munguia L, Ordonez-Sanchez ML, Reddy PM, et al. (2013) Genomic study in Mexicans identifies a new locus for triglycerides and refines European lipid loci. J Med Genet 50:298-308.
17. Willer CJ, Schmidt EM, Sengupta S, Peloso GM, Gustafsson S, Kanoni S, Ganna A, Chen J, Buchkovich ML, Mora S, et al. (2013) Discovery and refinement of loci associated with lipid levels. Nat Genet 45:1274-1283.
18. Holmen OL, Zhang H, Fan Y, Hovelson DH, Schmidt EM, Zhou W, Guo Y, Zhang J, Langhammer A, Lochen ML, et al. (2014) Systematic evaluation of coding variation identifies a candidate causal variant in TM6SF2 influencing total cholesterol and myocardial infarction risk. Nat Genet 46:345-351.
19. Peloso GM, Auer PL, Bis JC, Voorman A, Morrison AC, Stitziel NO, Brody JA, Khetarpal SA, Crosby JR, 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.
20. Lanktree MB, Elbers CC, Li Y, Zhang G, Duan Q, Karczewski KJ, Guo Y, Tragante V, North KE, Cushman M, et al. (2015) Genetic meta-analysis of 15,901 African Americans identifies variation in EXOC3L1 is associated with HDL concentration. J Lipid Res 56:1781-1786.
21. Surakka I, Horikoshi M, Magi R, Sarin AP, Mahajan A, Lagou V, Marullo L, Ferreira T, Miraglio B, Timonen S, et al. (2015) The impact of low-frequency and rare variants on lipid levels. Nat Genet 47:589-597.
22. van Leeuwen EM, Karssen LC, Deelen J, Isaacs A, Medina-Gomez C, Mbarek H, Kanterakis A, Trompet S, Postmus I, Verweij N, et al. (2015) Genome of The Netherlands population-specific imputations identify an ABCA6 variant associated with cholesterol levels. Nat Commun 6:6065.
23. Zubair N, Graff M, Luis Ambite J, Bush WS, Kichaev G, Lu Y, Manichaikul A, Sheu WH, Absher D, Assimes TL, et al. (2016) Fine-mapping of lipid regions in global populations discovers ethnic-specific signals and refines previously identified lipid loci. Hum Mol Genet 25:5500-5512.
24. Spracklen CN, Chen P, Kim YJ, Wang X, Cai H, Li S, Long J, Wu Y, Wang YX, Takeuchi F, et al. (2017) Association analyses of East Asian individuals and trans-ancestry analyses with European individuals reveal new loci associated with cholesterol and triglyceride levels. Hum Mol Genet 26:1770-1784.
25. Klarin D, Damrauer SM, Cho K, Sun YV, Teslovich TM, Honerlaw J, et al. (2018) Genetics of blood lipids among ~300,000 multi-ethnic participants of the Million Veteran Program. Nat Genet 50:1514-1523.
26. Choudhury A, Brandenburg JT, Chikowore T, Sengupta D, Boua PR, Crowther NJ, Agongo G, Asiki G, et al. (2022) Meta-analysis of sub-Saharan African studies provides insights into genetic architecture of lipid traits. Nat Commun 13:2578.

Table S8. Functional annotation of novel loci.

Locus	Chr	Pos (hg19)	rsID[1]	Coded Allele	Coded Allele Freq[2]	RefPop-Trait[3]	2 DF P[4]		nFUMA SNPs[5]	Func[6]	Genes[7]
							SomeCol	GradCol			
1	1	180,258,167	rs139845754	A	0.052	AFR-TG	3.40E-08	9.03E-03	98	68 intronic; 18 intergenic; 2 upstream; 1 downstream; 3 ncRNA intronic; 2 ncRNA exonic; 4 NA	*LHX4; RP5-11 80C10.2;* ***ACBD6****; XPR1; KIAA1614; STX6*
2*	2	103,239,356	rs7567133	C	0.031	AFR-HDL	2.22E-02	1.72E-08	11	6 intronic; 3 intergenic; 1 ncRNA exonic; 1 NA	*IL1RL1; IL18R1; IL18RAP; SLC9A4;* ***SLC9A2;*** *MFSD9; TMEM182*
3	3	9,096,107	rs79367750	C	0.096	ASN-LDL	3.99E-08	1.09E-06	18	17 intronic; 1 NA	***SRGAP3***
4*	3	70,605,665	rs147731578	ATTATT	0.018	EUR-HDL	1.39E-09	7.23E-04	1	1 intergenic	*MITF; FOXP1*
5*	4	179,620,483	rs74620279	C	0.025	AFR-LDL	4.67E-08	NA	49	47 intergenic; 1 downstream; 1 NA	*SNORD65*
6†	4	181,181,245	rs77249395	G	0.035	TA-HDL	4.97E-01	3.17E-06	2	2 intergenic	*LINC00290*
						EUR-HDL	3.96E-01	6.33E-06			
7*	4	182,664,815	rs11132093	A	0.061	ASN-LDL	1.99E-02	5.04E-09	53	53 intergenic	*RP11-540E16.2; TENM3*
8	5	37,160,808	rs190502162	C	0.013	EUR-LDL	3.35E-08	4.22E-03	1	1 intronic	*SKP2 ; NADK2; RANBP3L; SLC1A3; NIPBL;* ***CPLANE1****; NUP155; WDR70; GNDF*
9	5	82,514,005	rs192718305	G	0.025	AFR-LDL	2.05E-08	NA	51	49 intronic; 1 intergenic; 1 3'UTR	*RP11-343L5.2; TMEM167A;* ***XRCC4***
10*	6	76,693,587	rs147892694	G	0.034	AFR-HDL	8.61E-09	5.87E-02	56	44 intronic; 5 intergenic; 7 NA	*TMEM30A; FILIP1; SENP6; MYO6;* ***IMPG1***
11	7	97,920,623	rs7015	A	0.203	TA-LDL	4.56E-08	6.01E-07	95	61 intronic; 24 intergenic; 3 exonic; 1 upstream; 1 ncRNA exonic; 2 3'UTR; 3 NA	*LMTK2; TECPR1;* ***BRI3****; BAIAP2L1; NPTX2*
12	8	30,437,023	rs77655002	C	0.042	AFR-LDL	6.87E-02	5.52E-09	106	89 intronic; 5 intergenic; 2 exonic; 1 upstream; 6 downstream; 1 5'UTR; 1 3'UTR; 1 NA	*MBOAT4; RBPMS-AS1; RBPMS;* ***GTF2E2****; GSR; TEX15*
						TA-LDL	7.50E-02	1.83E-08			
13	10	108,927,076	rs144190766	C	0.03	AFR-HDL	2.52E-08	NA	188	14 intronic; 173 intergenic; 1 NA	*SORCS1; RNA5SP326*
14	10	129,845,329	rs116562538	A	0.028	TA-LDL	6.79E-09	1.46E-02	32	31 intronic; 1 NA	***PTPRE***
15	11	4,041,010	rs35287906	C	0.014	EUR-LDL	3.51E-01	1.53E-09	13	11 intronic; 2 intergenic	*PGAP2; RHOG;* ***STIM1****; RRM1; OR55B1P; RP11-23F23.3*
16	11	55,451,313	rs11230661	A	0.185	EUR-HDL	0.00000002	3.15E-10	925	3 intronic; 746 intergenic; 22 exonic; 54 upstream; 48 downstream; 1 upstream:downstream; 16 ncRNA intronic; 19 ncRNA exonic; 16 NA	*TRIM48; TRIM51; 87* Olfactory Receptor Genes; *APLNR*
17*	12	14,221,135	rs148063115	T	0.026	AFR-TG	2.20E-04	4.86E-08	37	4 intronic; 33 intergenic	*GRIN2B; RP11-72J9.1; RP11-298E10.1; RN7SL676P; GUCY2C*
18*	13	101,418,063	rs190746034	G	0.022	AFR-HDL	3.53E-01	1.26E-08	35	1 intronic; 2 intergenic; 31 ncRNA intronic; 1 ncRNA exonic	*RP11-151A6.4; TMTC4;* ***NALCN-AS1****; ARF4P3*

Notes: 1. rsID based on dbSNP build 146; 2. Coded allele frequency; 3. The reference panel used in FUMA to obtain functional annotations-trait (NA: not available for that educational exposure); 4. GWAS 2 DF p-value of the significant lead SNP for this locus; 5. Number of SNVs in LD (r2>0.2) with lead variant from FUMA; when N=1, the SNV was not found in FUMA; 6. Functional consequence of the SNP on the gene obtained from ANNOVAR (NA: not available); 7. Nearest gene of all SNVs in LD (r2>0.2) with lead variant; if SNV not in FUMA, gene or nearest flanking coding genes noted. Bolded genes reflect intragenic lead SNV. * Only significant in Stage I analyses. † Only significant by 1 DF G×E analyses.
AFR, African ancestry; ASN, Asian ancestry; EUR, European ancestry; GradCol, graduated college; HDL, high-density lipoprotein cholesterol; LDL, low-density lipoprotein cholesterol; NA, not available; SomeCol, some college; TA, trans-ancestry; TG, triglyceride.

Supplementary Table 10. Drug-gene interaction analysis and summary of approved drugs and active ligands for high and medium priority candidate gene targets.

Priority	Locus	Chr	Genes	HGNC ID	UniProt	KEGG Pathways	Complexes	Target Category	IDG Target Level	DGIdb #Interactions	DGIdb Interactions	Interaction Type	Top Interaction Name	ChEMBL ID	Sources	PMID	Interaction Score	Approved Drugs	Active Ligands
High	1	1	STX6	HGNC:11441	O43752	SNARE interactions in vesicular transport - Homo sapiens (human), Nicotine Pathway (Dopaminergic Neuron), Pharmacodynamics, Retrograde transport at the Trans-Golgi-Network, Intra-Golgi traffic, Intra-Golgi and retrograde Golgi-to-ER traffic, Membrane Trafficking, Vesicle-mediated transport	STX7; VTI1B; STX6, STX6; SNAP29, SNAP23; STX6, VAMP8; STX7; VTI1B; VAMP3; VAMP7; STX4; STX6, STX6; VAMP4; VAMP3; VAMP8, STX6; VAMP3; VTI1A, STX16; STX6; VAMP4; VTI1A; VTI1B, STX16; VAMP4; VTI1B; STX6, STX16; VAMP4; STX6; VTI1A	N/A	Tbio	0	N/A								
High	3	3	SRGAP3	HGNC:19744	O43295	Axon guidance - Homo sapiens (human), Developmental Biology, Signal Transduction, Rho GTPase cycle, Signaling by Rho GTPases, Signaling by Robo receptor, Axon guidance, Inactivation of Cdc42 and Rac	N/A	CLINICALLY ACTIONABLE	Tbio	0	N/A								
High	4	3	FOXP1	HGNC:3823	Q9H334	MicroRNAs in cancer - Homo sapiens (human), Developmental Biology, Transcriptional regulation of pluripotent stem cells	N/A	CLINICALLY ACTIONABLE, TRANSCRIPTION FACTOR	Tbio	0	N/A								
High	8	5	SKP2	HGNC:10901	Q13309	mTOR signaling pathway - Homo sapiens (human), Cell cycle - Homo sapiens (human), Small cell lung cancer - Homo sapiens (human), Ubiquitin mediated proteolysis - Homo sapiens (human), FoxO signaling pathway - Homo sapiens (human), Pathways in cancer - Homo sapiens (human), Viral carcinogenesis - Homo sapiens (human), Epstein-Barr virus infection - Homo sapiens (human), Herpes simplex infection - Homo sapiens (human), Retinoblastoma (RB) in Cancer, Adaptive Immune System, C-MYC pathway, Regulation of retinoblastoma protein, FOXM1 transcription factor network, FoxO family signaling, p53 pathway, Cyclin D associated events in G1, G1 Phase, regulation of p27 phosphorylation during cell cycle progression, SCF(Skp2)-mediated degradation of p27/p21, Cyclin E associated events during G1/S transition , G1/S Transition, Mitotic G1-G1/S phases, Cyclin A:Cdk2-associated events at S phase entry, Immune System, S Phase, e2f1 destruction pathway, Antigen processing: Ubiquitination & Proteasome degradation, Class I MHC mediated antigen processing & presentation, Ub-specific processing proteases, Post-translational protein modification, Deubiquitination, Metabolism of proteins, Notch, Cell Cycle	RBX1; SKP1; SKP2; CUL1, SKP1; SKP2; CUL1, URI1; POLR2E; SKP1; SKP2; CUL1; PFDN2; STAP1; RUVBL2; RUVBL1, CCNT1; CDK9; SKP2, RBX1; SKP1; GPS1; SKP2; CUL1; CUL2; CUL3; COPS8, WEE1; SKP2; BTRC, CCNE1; CDK2; CDKN1B; CKS1B; RBX1; SKP1; SKP2; CUL1, CKS1B; RBX1; SKP1; SKP2; CUL1, CDH1; SKP2	TUMOR SUPPRESSOR, KINASE	Tbio	0	N/A								
High	8	5	LMBRD2	HGNC:25287	Q68DH5	N/A	N/A	N/A	Tdark	0	N/A								
High	9	5	SLC1A3	HGNC:10941	P43003	Glutamatergic synapse - Homo sapiens (human), Synaptic Vesicle Pathway, Sudden Infant Death Syndrome (SIDS) Susceptibility Pathways, Glutamate Neurotransmitter Release Cycle, Neurotransmitter uptake and Metabolism In Glial Cells, Neurotransmitter Release Cycle, Astrocytic Glutamate-Glutamine Uptake And Metabolism, Transmission across Chemical Synapses, Amino acid and oligopeptide SLC transporters, Transport of inorganic cations/anions and amino acids/oligopeptides, SLC-mediated transmembrane transport, Transmembrane transport of small molecules, Neuronal System, Histidine metabolism, Purine metabolism	N/A	DRUGGABLE GENOME, TRANSPORTER, CELL SURFACE	Tchem	0	N/A								32
High	9	5	TMEM167A	HGNC:28330	Q8TBQ9	N/A	N/A	N/A	Tdark	1	fluorouracil		FLUOROURACIL	chembl:CHEMBL185	PharmGKB	28796378	1.21	1	
High	10	6	MYO6	HGNC:7605	Q9UM54	VEGFA-VEGFR2 Signaling Pathway, Stabilization and expansion of the E-cadherin adherens junction, Trafficking of AMPA receptors, Glutamate Binding, Activation of AMPA Receptors and Synaptic Plasticity, Neurotransmitter Receptor Binding And Downstream Transmission In The Postsynaptic Cell, Transmission across Chemical Synapses, Membrane Trafficking, Neuronal System, Gap junction degradation, Gap junction trafficking, Gap junction trafficking and regulation, EGFR1, Vesicle-mediated transport	N/A	N/A	Tbio	0	N/A								
High	10	6	TMEM30A	HGNC:16667	Q9NV96	Innate Immune System, Immune System, Neutrophil degranulation	N/A	CLINICALLY ACTIONABLE	Tbio	0	N/A								
High	10	6	FILIP1	HGNC:21015	Q4L180	N/A	FLNA; FILIP1, RHOD; FILIP1	N/A	Tdark	0	N/A								
High	10	6	SENP6	HGNC:20944	Q9GZR1	N/A	N/A	DRUGGABLE GENOME, PROTEASE, ENZYME	Tchem	15	2-(4-NITROBENZYLSULFANYL)BENZOXAZOLE, 4-PHENYLAMINO-[1,2]NAPHTHOQUINONE, 6,4'-DIHYDROXYFLAVONE, BAS-0338868, CID 4101591, DIHYDROXANTHOHUMOL, DNDI1417156, GNF-PF-1702, GNF-PF-4068, GNF-PF-5151, GNF-PF-5159, GNF-PF-992, IPRIFLAVONE, TCMDC-125543, XANTHOANGELOL B		CHEMBL378903	chembl:CHEMBL378903	DTC		1.41		
High	11	7	BRI3	HGNC:1109	O95415	Innate Immune System, Immune System, Neutrophil degranulation	N/A	N/A	Tbio	0	N/A								
High	12	8	RBPMS	HGNC:19097	Q6ZRY4	Exercise-induced Circadian Regulation	N/A	N/A	Tbio	0	N/A								
High	12	8	GTF2E2	HGNC:4651	P29084	Basal transcription factors - Homo sapiens (human), Viral carcinogenesis - Homo sapiens (human), Epstein-Barr virus infection - Homo sapiens (human), Eukaryotic Transcription Initiation, Disease, RNA Polymerase II Promoter Escape, HIV Transcription Initiation, RNA Polymerase II Transcription Pre-Initiation And Promoter Opening, RNA Polymerase II HIV Promoter Escape, Transcription of the HIV genome, RNA Polymerase II Pre-transcription Events, RNA Polymerase II Transcription Initiation, RNA Polymerase II Transcription Initiation And Promoter Clearance, Late Phase of HIV Life Cycle, HIV Life Cycle, HIV Infection, RNA Polymerase II Transcription, Infectious disease, Gene Expression, RNA polymerase II transcribes snRNA genes	POLR2D; GTF2F2; ERCC2; POLR2C; POLR2E; ERCC3; TBP; POLR2A; GTF2E1; GTF2E2; POLR2B; GTF2H1; GTF2F1; POLR2I; POLR2H; POLR2J; POLR2K; POLR2F; POLR2G; POLR2L; GTF2B; GTF2H2; GTF2H3; GTF2H4, GTF2E2; POLR2D; GTF2F2; ERCC2; POLR2C; POLR2E; ERCC3; TBP; POLR2A; GTF2E1; GTF2E2; POLR2B; GTF2H1; GTF2F1; POLR2I; BRCA1; POLR2H; POLR2J; POLR2K; POLR2F; POLR2G; POLR2L; GTF2B; MED21; GTF2H2; GTF2H3; GTF2H4	KINASE, TRANSCRIPTION FACTOR	Tbio	0	N/A								
High	12	8	MBOAT4	HGNC:32311	Q96T53	Peptide hormone metabolism, Metabolism of proteins, Synthesis, secretion, and deacylation of Ghrelin	N/A	DRUGGABLE GENOME	Tbio	0	N/A								
High	15	11	STIM1	HGNC:11386	Q13586	Platelet activation - Homo sapiens (human), Calcium signaling pathway - Homo sapiens (human), Adaptive Immune System, Elevation of cytosolic Ca2+ levels, Platelet calcium homeostasis, Platelet homeostasis, TCR signaling in naï ve CD8+ T cells, TCR signaling in naï ve CD4+ T cells, Ion homeostasis, Immune System, Hemostasis, Antigen activates B Cell Receptor (BCR) leading to generation of second messengers, Muscle contraction, Signaling by the B Cell Receptor (BCR), Cardiac conduction	STIM1; ORAI1; TRPC1, STIM1; TRPV4	TRANSPORTER	Tbio	0	N/A								
High	15	11	RRM1	HGNC:10451	P23921	Glutathione metabolism - Homo sapiens (human), Fluoropyrimidine Pathway, Pharmacokinetics, Gemcitabine Pathway, Pharmacodynamics, Purine Nucleoside Phosphorylase Deficiency, Mercaptopurine Action Pathway, Azathioprine Action Pathway, Xanthine Dehydrogenase Deficiency (Xanthinuria), Adenylosuccinate Lyase Deficiency, AICA-Ribosiduria, Thioguanine Action Pathway, Adenine phosphoribosyltransferase deficiency (APRT), Mitochondrial DNA depletion syndrome, Myoadenylate deaminase deficiency, Purine Metabolism, Gemcitabine Action Pathway, Molybdenum Cofactor Deficiency, Adenosine Deaminase Deficiency, Gemcitabine Metabolism Pathway, Gout or Kelley-Seegmiller Syndrome, Lesch-Nyhan Syndrome (LNS), Xanthinuria type I, Xanthinuria type II, Fluoropyrimidine Activity, Integrated Pancreatic Cancer Pathway, Retinoblastoma (RB) in Cancer, Nucleotide Metabolism, Purine nucleotides nucleosides metabolism, Pyrimidine nucleotides nucleosides metabolism, guanosine deoxyribonucleotides de novo biosynthesis, guanosine nucleotides de novo biosynthesis, adenosine deoxyribonucleotides de novo biosynthesis, Metabolism of nucleotides, E2F transcription factor network, pyrimidine deoxyribonucleotides de novo biosynthesis, Metabolism, purine nucleotides de novo biosynthesis, superpathway of pyrimidine deoxyribonucleotides de novo biosynthesis, pyrimidine deoxyribonucleotides biosynthesis from CTP, Synthesis and interconversion of nucleotide di- and triphosphates, superpathway of purine nucleotide salvage, Pyrimidine metabolism, Purine metabolism	N/A	DRUGGABLE GENOME, DNA REPAIR, ENZYME	Tclin	15	CISPLATIN, CLOFARABINE, DOCETAXEL, FLUDARABINE, FLUDARABINE PHOSPHATE, GALLIUM NITRATE, GEMCITABINE, GEMCITABINE HYDROCHLORIDE, HYDROXYUREA, PACLITAXEL, TEZACITABINE, VINORELBINE, cisplatin, cladribine, cytarabine	inhibitor	TEZACITABINE	chembl:CHEMBL2105467	ChemblInteractions	10547615	3.03	3	1
High	16	11	APLNR	HGNC:339	P35414	Neuroactive ligand-receptor interaction - Homo sapiens (human), GPCRs, Class A Rhodopsin-like, G alpha (i) signalling events, Signaling by GPCR, GPCR downstream signaling, Signal Transduction, Peptide ligand-binding receptors, Class A/1 (Rhodopsin-like receptors), GPCR ligand binding	N/A	CLINICALLY ACTIONABLE, DRUGGABLE GENOME, G PROTEIN COUPLED RECEPTOR	Tchem	0	N/A								102
Medium	1	1	ACBD6	HGNC:23339	Q9BR61	Fatty acid, triacylglycerol, and ketone body metabolism, Metabolism of lipids and lipoproteins, TCR, Metabolism, Fatty Acyl-CoA Biosynthesis, Triglyceride Biosynthesis	N/A	N/A	Tbio	0	N/A								
Medium	2	2	SLC9A4	HGNC:11077	Q6AI14	Gastric acid secretion - Homo sapiens (human), Beta-agonist/Beta-blocker Pathway, Pharmacodynamics, Sodium/Proton exchangers, Transport of inorganic cations/anions and amino acids/oligopeptides, SLC-mediated transmembrane transport, Transmembrane transport of small molecules	N/A	DRUGGABLE GENOME, EXCHANGER, TRANSPORTER	Tdark	0	N/A								
Medium	2	2	SLC9A2	HGNC:11072	Q9UBY0	Beta-agonist/Beta-blocker Pathway, Pharmacodynamics, Sodium/Proton exchangers, Transport of inorganic cations/anions and amino acids/oligopeptides, SLC-mediated transmembrane transport, Transmembrane transport of small molecules	N/A	DRUGGABLE GENOME, TRANSPORTER	Tchem	0	N/A								5
Medium	2	2	IL18R1	HGNC:5988	Q13478	Inflammatory bowel disease (IBD) - Homo sapiens (human), TNF signaling pathway - Homo sapiens (human), Cytokine-cytokine receptor interaction - Homo sapiens (human), JAK STAT MolecularVariation 1, JAK STAT MolecularVariation 2, JAK STAT pathway and regulation, IL23-mediated signaling events, IL12 signaling mediated by STAT4, IL12-mediated signaling events, Immune System, il12 and stat4 dependent signaling pathway in th1 development, Signaling by Interleukins, Cytokine Signaling in Immune system	N/A	DRUGGABLE GENOME	Tbio	0	N/A								
Medium	4	3	MITF	HGNC:7105	O75030	Melanoma - Homo sapiens (human), Pathways in cancer - Homo sapiens (human), Osteoclast differentiation - Homo sapiens (human), Melanogenesis - Homo sapiens (human), RANKL-RANK (Receptor activator of NFKB (ligand)) Signaling Pathway, Neural Crest Differentiation, Kit receptor signaling pathway, KitReceptor, Regulation of retinoblastoma protein, Regulation of nuclear beta catenin signaling and target gene transcription, Signaling events mediated by Stem cell factor receptor (c-Kit), IL6-mediated signaling events, melanocyte development and pigmentation pathway, SUMOylation of transcription factors, SUMO E3 ligases SUMOylate target proteins, Signaling mediated by p38-alpha and p38-beta, SUMOylation, Post-translational protein modification, Metabolism of proteins	MITF; HINT1; KARS,	CLINICALLY ACTIONABLE, DRUGGABLE GENOME, DRUG RESISTANCE, ENZYME, TRANSCRIPTION FACTOR	Tchem	53	2-PHENYLAMINO-[1,4]NAPHTHOQUINONE, 8-AZAGUANINE, 8-FLUOROTRYPTANTHRIN, ALPHA,BETA-DEHYDROCURVULARIN, ARTEANNUIN B, AURIN, BAS03450373, CID 1540545, CID 649849, CID 650707, DABRAFENIB, DNDI1416934, DNDI1416969, DNDI1417055, DNDI1417346, DNDI1417413, DNDI1417560, DNDI1417809, GNF-PF-107, GNF-PF-186, GNF-PF-2260, GNF-PF-2298, GNF-PF-2335, GNF-PF-2812, GNF-PF-2906, GNF-PF-2954, GNF-PF-3407, GNF-PF-3716, GNF-PF-3951, GNF-PF-4867, HOMIDIUM BROMIDE, METAPHIT METHANESULFONATE, N-(BENZYLCARBAMOTHIOYL)BENZAMIDE, NIFUROXAZIDE, NSC-645827, NSC-7936, PERHEXILINE MALEATE, PIPLARTINE, PLX4720, SELUMETINIB, SJ000001592, SJ000012773, SJ000114741, SJ000128325, SJ000154295, TCMDC-124032, TCMDC-124063, TCMDC-124356, TCMDC-124676, TCMDC-125150, TCMDC-125814, TDR78274, VEMURAFENIB		CHEMBL588965	chembl:CHEMBL588965	DTC		1.2		34
Medium	7	4	TENM3	HGNC:29944	Q9P273	N/A	N/A	N/A	Tbio	0	N/A								
Medium	9	5	XRCC4	HGNC:12831	Q13426	Non-homologous end-joining - Homo sapiens (human), Non-homologous end joining, Nonhomologous End-Joining (NHEJ), DNA Double-Strand Break Repair, Disease, ATM pathway, DNA-PK pathway in nonhomologous end joining, DNA Repair, Integration of provirus, Early Phase of HIV Life Cycle, SUMOylation of DNA damage response and repair proteins, SUMO E3 ligases SUMOylate target proteins, 2-LTR circle formation, HIV Life Cycle, HIV Infection, SUMOylation, Infectious disease, Post-translational protein modification, Metabolism of proteins	LIG4; XRCC4, LIG4; AHNAK; XRCC4, LIG4; XRCC4; PNKP, PRKDC; LIG4; XRCC4, LIG4; XRCC4; NHEJ1, NBN; XRCC6; XRCC5; LIG4; MRE11A; XRCC4; RAD50	KINASE	Tbio	2	docetaxel, fluorouracil		DOCETAXEL	chembl:CHEMBL92	PharmGKB	20157331	0.81		
Medium	10	6	IMPG1	HGNC:6055	Q17R60	N/A	N/A	DRUGGABLE GENOME	Tbio	0	N/A								
Medium	11	7	TECPR1	HGNC:22214	Q7Z6L1	N/A	N/A	N/A	Tbio	0	N/A								
Medium	11	7	BAIAP2L1	HGNC:21649	Q9UHR4	EGFR1	N/A	KINASE	Tbio	0	N/A								
Medium	12	8	GSR	HGNC:4623	P00390	Thyroid hormone synthesis - Homo sapiens (human), Glutathione metabolism - Homo sapiens (human), Pathway_PA165980337, Oxidative Stress Regulatory Pathway (Erythrocyte), 2-Hydroxyglutric Aciduria (D And L Form), Gamma-glutamyl-transpeptidase deficiency, 5-oxoprolinase deficiency, Gamma-Glutamyltransferase Deficiency, Glutathione Metabolism, Homocarnosinosis, Hyperinsulinism-Hyperammonemia Syndrome, Glutathione Synthetase Deficiency, Succinic semialdehyde dehydrogenase deficiency, 4-Hydroxybutyric Aciduria/Succinic Semialdehyde Dehydrogenase Deficiency, 5-Oxoprolinuria, Glutamate Metabolism, Glutathione metabolism, Selenium Micronutrient Network, Nuclear Receptors Meta-Pathway, NRF2 pathway, One carbon metabolism and related pathways, Amino Acid metabolism, Oxidative Stress, Sulfation Biotransformation Reaction, Metapathway biotransformation, Metabolism of ingested H2SeO4 and H2SeO3 into H2Se, glutathione redox reactions II, Detoxification of Reactive Oxygen Species, Cellular responses to stress, Metabolism of nucleotides, Metabolism of amino acids and derivatives, glutathione redox reactions I, Metabolism, Transcriptional Regulation by TP53, Synthesis and interconversion of nucleotide di- and triphosphates, Generic Transcription Pathway, Gene Expression, Glutamate Glutamine metabolism, Urea cycle and metabolism of arginine, proline, glutamate, aspartate and asparagine, Oxidative Stress Pathway (Erythrocyte), TP53 Regulates Metabolic Genes, Selenoamino acid metabolism	N/A	DRUGGABLE GENOME, EXTERNAL SIDE OF PLASMA MEMBRANE, ENZYME	Tclin	15	AMIFOSTINE, CARMUSTINE, CARVEDILOL, CEFOTAXIME, ENALAPRIL, L-alanyl-l-glutamine, LANSOPRAZOLE, LEVOFLOXACIN, MISONIDAZOLE, OFLOXACIN, ONDANSETRON, OXIGLUTATIONE, QUERCETIN, THYROXINE, carboplatin	inhibitor, substrate	OXIGLUTATIONE	chembl:CHEMBL1372	ChemblInteractions, GuideToPharmacology	8651901	6.37	1	7
Medium	12	8	TEX15	HGNC:11738	Q9BXT5	N/A	N/A	DNA REPAIR	Tbio	0	N/A								
Medium	13	10	SORCS1	HGNC:16697	Q8WY21	Ectoderm Differentiation	N/A	N/A	Tbio	0	N/A								
Medium	14	10	PTPRE	HGNC:9669	P23469	IL6-mediated signaling events, EGFR1	N/A	DRUGGABLE GENOME, PROTEIN PHOSPHATASE, KINASE	Tbio	0	N/A								
Medium	17	12	GRIN2B	HGNC:4586	Q13224	Long-term potentiation - Homo sapiens (human), Dopaminergic synapse - Homo sapiens (human), Glutamatergic synapse - Homo sapiens (human), Systemic lupus erythematosus - Homo sapiens (human), Alzheimer,s disease - Homo sapiens (human), Amyotrophic lateral sclerosis (ALS) - Homo sapiens (human), Huntington,s disease - Homo sapiens (human), Circadian entrainment - Homo sapiens (human), cAMP signaling pathway - Homo sapiens (human), Rap1 signaling pathway - Homo sapiens (human), Ras signaling pathway - Homo sapiens (human), Amphetamine addiction - Homo sapiens (human), Neuroactive ligand-receptor interaction - Homo sapiens (human), Nicotine addiction - Homo sapiens (human), Alcoholism - Homo sapiens (human), Cocaine addiction - Homo sapiens (human), Alzheimers Disease, Brain-Derived Neurotrophic Factor (BDNF) signaling pathway, Interactions of neurexins and neuroligins at synapses, Protein-protein interactions at synapses, Signaling by EGFR, IRS-related events triggered by IGF1R, IGF1R signaling cascade, Signaling by Type 1 Insulin-like Growth Factor 1 Receptor (IGF1R), Signalling to RAS, Signalling to p38 via RIT and RIN, Gastrin-CREB signalling pathway via PKC and MAPK, Signaling by GPCR, Frs2-mediated activation, ARMS-mediated activation, Prolonged ERK activation events, Signalling to ERKs, Signalling by NGF, ErbB4 signaling events, Ras activation uopn Ca2+ infux through NMDA receptor, CREB phosphorylation through the activation of Ras, Unblocking of NMDA receptor, glutamate binding and activation, CREB phosphorylation through the activation of CaMKII, Post NMDA receptor activation events, Activation of NMDA receptor upon glutamate binding and postsynaptic events, DAP12 signaling, DAP12 interactions, Developmental Biology, Reelin signaling pathway, FCERI mediated MAPK activation, VEGFR2 mediated cell proliferation, Fc epsilon receptor (FCERI) signaling, RAF/MAP kinase cascade, SHC1 events in EGFR signaling, Neurotransmitter Receptor Binding And Downstream Transmission In The Postsynaptic Cell, Transmission across Chemical Synapses, Innate Immune System, Immune System, VEGFA-VEGFR2 Pathway, nitric oxide signaling pathway, Interleukin receptor SHC signaling, Interleukin-2 signaling, Signal Transduction, MAPK1/MAPK3 signaling, NGF signalling via TRKA from the plasma membrane, Signaling by VEGF, Neuronal System, SOS-mediated signalling, IRS-mediated signalling, Insulin receptor signalling cascade, Signaling by Insulin receptor, EPHB-mediated forward signaling, MAPK family signaling cascades, Signaling by Leptin, EPH-Ephrin signaling, Interleukin-3, 5 and GM-CSF signaling, Downstream signal transduction, Signaling by PDGF, Signaling by Interleukins, NCAM signaling for neurite out-growth, Cytokine Signaling in Immune system, GPCR Dopamine D1like receptor, RET signaling, Axon guidance, Signaling by SCF-KIT, SALM protein interactions at the synapses, GRB2 events in EGFR signaling	NOS1; DLG4; GRIN2B, CAMK2A; DLG4; GRIN1; GRIN2B; SYNGAP1, CAMK2A; DLG4; GRIN1; GRIN2A; GRIN2B; SYNGAP1, CAMK2A; LRRC7; GRIN2B; ACTN2; APBA1; CASK; GRIN2B; LIN7A; KIF17, DLG4; GRIN1; GRIN2A; GRIN2B, GRIN2B; LRP8	DRUGGABLE GENOME, CELL SURFACE, ION CHANNEL	Tclin	50	ACAMPROSATE CALCIUM, AM-101, AMANTADINE, AMANTADINE HYDROCHLORIDE, AV-101, AZD-6765, AZD8108, CERC-301, CNS-5161, D-SERINE, DELUCEMINE, DEXANABINOL, DIMIRACETAM, DIZOCILPINE, ELIPRODIL, ESKETAMINE, ESKETAMINE HYDROCHLORIDE, EVT-101, FELBAMATE, GLYCINE, GW468816, Indantadol, KETAMINE, KETAMINE HYDROCHLORIDE, LANICEMINE, LATREPIRDINE, MEMANTINE, MEMANTINE HYDROCHLORIDE, NBQX, NEBOGLAMINE, NERAMEXANE, NERAMEXANE MESYLATE, NMDA, NRX-1074, ORPHENADRINE (CHLORIDE), ORPHENADRINE CITRATE, PHENCYCLIDINE, RALFINAMIDE, RAPASTINEL, RGH-896, Radiprodil, SELFOTEL, TRAXOPRODIL, V3381, acamprosate, clozapine, haloperidol, methylphenidate, quetiapine, risperidone	antagonist	MEMANTINE	chembl:CHEMBL807	TdgClinicalTrial, TEND	9120573, 11854433, 16009352, 15452358, 14530799	1.94	5	201
Medium	17	12	GUCY2C	HGNC:4688	P25092	Purine nucleotides nucleosides metabolism, Miscellaneous digestion events, Disease, actions of nitric oxide in the heart, Metabolism, Intestinal infectious diseases, Uptake and actions of bacterial toxins, Infectious disease, Purine metabolism	N/A	DRUGGABLE GENOME, KINASE	Tclin	8	DOLCANATIDE, LINACLOTIDE, Linaclotide, MD-1100, MLN0264, PLECANATIDE, Plecanatide, SP-333	agonist	LINACLOTIDE	chembl:CHEMBL3301675	TdgClinicalTrial, ChemblInteractions, GuideToPharmacology, TTD	23090647	79.56	2	
Medium	18	13	TMTC4	HGNC:25904	Q5T4D3	N/A	N/A	N/A	Tdark	0	N/A								

Notes: Used DGIdb, Pharos, DrugBank and ChEMBL databases. Target genes with approved drugs are highlighted in grey.

Supplementary Table 11. Drug repurposing analysis for high and medium priority candidate drug-gene targ

Priority	Locus	Chr	Target Gene	Pharmacologically Active	Action	Drug Generic	Drug Groups	First Approval	Summary	ChEMBL_ID	DrugBank_ID	Trade Names	Type	Route	Late Phase Trials	Latest Phase	MESH Indications	MESH Description	DrugBank Indications	Recent Clinical Trials	Status	Purpose	Title
High	9	5	TMEM167A			Fluorouracil	Approved	1962	Fluorouracil is a pyrimidine analog used to treat basal cell carcinomas, and as an injection in palliative cancer treatment.	CHEMBL185	DB00544	Aztikerall, Carac, Efudex, Fluoroplex, Tolak	Small molecule	Intravenous, topical	241	4	D002280; D009369; D055623; D007642	Carcinoma, Basal Cell; Neoplasms; Keratosis, Actinic; Keratosis	DBCOND0046770 (Open Angle Glaucoma (OAG)); DBCOND0042053 (Stage IV Colorectal Cancer); DBCOND0046917 (Malignant Neoplasm of Stomach); DBCOND0028460 (Esophageal Cancer) DBCOND0029426 (Skin Neoplasms); DBCOND0043025 (Neoplasms, Basal Cell); DBCOND0028161 (Skin Diseases); DBCOND0043488 (Squamous Cell Carcinoma (SCC)); DBCOND0033322 (Neoplasms, Squamous Cell); DBCOND0022280 (Carcinoma); DBCOND0079791 (Carcinoma, Basal Cell (BCC)); DBCOND0065881 (Actinic Keratosis (AK))	NCT03904381; NCT01271582; NCT00911092; NCT00847912; NCT00696488; NCT00377273; NCT00377273	Active Not Recruiting; Active Not Recruiting; Completed; Completed; Completed; Completed	Health Services Research; Treatment; Not Available; Prevention; Treatment	Efficacy and Safety of XEN® Gel Stent and Post-operative Management in Patients With Open Angle Glaucoma; Investigation of Association Between UGT1A1 Polymorphisms and Irinotecan Toxicity in Korean Patients; Predictive Proteomic Factors of the Response to Concomitant Radiochemotherapy in Esophageal Cancer; CSP #562 - The VA Keratinocyte Carcinoma Chemoprevention Trial; Measuring Adherence To Topical 5-Fluorouracil in a Clinic Population; An Open-Label 18-Month Safety Study of Fluorouracil Cream 0.5% for the Treatment of Actinic Keratoses; Measuring Adherence To Topical 5-Fluorouracil in a Clinic Population; An Open-Label 18-Month Safety Study of Fluorouracil Cream 0.5% for the Treatment of Actinic Keratoses
High	15	11	RRM1	yes	inhibitor	Gemcitabine	Approved	1996	Gemcitabine is a nucleoside metabolic inhibitor used as adjunct therapy in the treatment of certain types of ovarian cancer, non-small cell lung carcinoma, metastatic breast cancer, and as a single agent for pancreatic cancer.	CHEMBL888	DB00441	Gemzar, Infugem	Small molecule	Intravenous, parenteral	240	4	D009369; D002289; D010051; D010190	Neoplasms; Carcinoma, Non-Small-Cell Lung; Ovarian Neoplasms; Pancreatic Neoplasms	DBCOND0117097 (Non Small Cell Lung Cancer Squamous (NSCLC-Sq)); DBCOND0084936 (Adenocarcinomas of the Pancreas); DBCOND0034623 (Non-Small Cell Lung Carcinoma (NSCLC)); DBCOND0043377 (Metastatic Non-Small Cell Lung Cancer)	NCT02996214; NCT02812992; NCT02316327; NCT02000531; NCT04331626	Active Not Recruiting; Completed; Completed; Not Yet Recruiting	Treatment; Treatment; Treatment; Treatment	Paclitaxel Liposome for Squamous Non-Small-cell Lung Cancer Study(LIPUSU); Geriatric Assessment Directed Trial to Evaluate Gemcitabine +/- Nab-paclitaxel in Elderly Pancreatic Cancer Patients; Open-label Study in Patients With Metastatic NSLC Treated With Cisplatin, Gemcitabine and Bevacizumab; Progression Free Survival (PFS) Using Erlotinib for Non-Small-Cell Lung Cancer (NSCLC) in Chinese Population; Low-dose Gemcitabine Combined With Nivolumab for Second-line and Above Line Treatment of NSCLC
High	15	11	RRM1	yes	inhibitor	Hydroxyurea	Approved	1967	Hydroxyurea is an antimetabolite used to treat sickle cell anemia crisis	CHEMBL467	DB01005	Droxia, Hydrea, Siklos	Small molecule	Oral	36	4	D015464; D002294; D008545; D009369; 000755	Leukemia, Myelogenous, Chronic, BCR-ABL Positive; Carcinoma, Squamous Cell; Melanoma; Neoplasms; Anemia, Sickle Cell	DBCOND0128712 (Sickle Cell Disease (SCD)); DBCOND0028117 (Sickle Cell Anemia); DBCOND0041412 (Renal Function Disorder); DBCOND0058019 (Drepanocytic Men Treated by Hydroxyurea for the First Time); DBCOND0128712 (Sickle Cell Disease (SCD))	NCT02149537; NCT02522104; NCT01609192; NCT05081349	Active Not Recruiting; Completed; Completed; Completed	Treatment; Other; Treatment; Treatment	Risk Clinical Stratification of Sickle Cell Disease in Nigeria, Assessment of Efficacy/Safety of Hydroxyurea Treatment; Evaluation of the Impact of Renal Function on the Pharmacokinetics of SIKLOS ® (DARH); Future of Spermatogenesis in Men With Sickle Cell Disease Medically Treated; Hydroxycarbamide and L-Carnitine Therapy in Sickle Cell Anemia
High	15	11	RRM1	yes	inhibitor	Clofarabine	Approved, Investigational	2004	Clofarabine is a purine nucleoside used to treat relapsed or refractory acute lymphoblastic leukemia in patients 1 to 21 years old.	CHEMBL1750	DB00631	Clolar, Evoltra	Small molecule	Intravenous	7	3	D054198; D009369; D015470; D007938; D006646	Precursor Cell Lymphoblastic Leukemia-Lymphoma; Neoplasms; Leukemia, Myeloid, Acute; Leukemia; Histiocytosis, Langerhans-Cell	DBCOND0031335 (Acute Lymphoblastic Leukemia (ALL)); DBCOND0028530 (Adult Erythroleukemia (M6a)); DBCOND0028523 (Adult Acute Myeloblastic Leukemia Without Maturation (M1)); DBCOND0036072 (Adult Acute Myeloid Leukemia With T(16;16)(p13;q22)); DBCOND0041078 (Secondary Acute Myeloid Leukemia (Secondary AML, sAML)); DBCOND0028533 (Adult Acute Megakaryoblastic Leukemia (M7)); DBCOND0036073 (Adult Acute Myeloid Leukemia With T(8;21)(q22;q22)); DBCOND0028524 (Adult Acute Myeloid Leukemia With 11q23 (MLL) Abnormalities); DBCOND0028534 (Adult Acute Minimally Differentiated Myeloid Leukemia (M0)); DBCOND0028525 (Adult Acute Myeloid Leukemia With Del(5q)); DBCOND0028522 (Adult Acute Myeloblastic Leukemia With Maturation (M2)); DBCOND0028531 (Adult Pure Erythroid Leukemia (M6b)); DBCOND0028536 (Adult Acute Monocytic Leukemia (M5b)); DBCOND0028535 (Adult Acute Monoblastic Leukemia (M5a)); DBCOND0028529 (Adult Acute Myelomonocytic Leukemia (M4)); DBCOND0029645 (Acute Myeloid Leukemia With Multilineage Dysplasia Following Myelodysplastic Syndrome); DBCOND0028532 (Untreated Adult Acute Myeloid Leukemia); DBCOND0028526 (Adult Acute Myeloid Leukemia With Inv(16)(p13;q22)); DBCOND0053647 (Central Nervous System Leukemia); DBCOND0090211 (Ph-Like Acute Lymphoblastic Leukemia); DBCOND0053648 (Testicular Leukemia); DBCOND0053645 (B Acute Lymphoblastic Leukemia)	NCT00549848; NCT02085408; NCT02883049	Active Not Recruiting; Active Not Recruiting; Active Not Recruiting	Treatment; Treatment; Treatment	Total Therapy Study XVI for Newly Diagnosed Patients With Acute Lymphoblastic Leukemia; Clofarabine or Daunorubicin Hydrochloride and Cytarabine Followed By Decitabine or Observation in Treating Older Patients With Newly Diagnosed Acute Myeloid Leukemia; Combination Chemotherapy in Treating Young Patients With Newly Diagnosed High-Risk B Acute Lymphoblastic Leukemia and Ph-Like TKI Sensitive Mutations
High	15	11	RRM1	yes	inhibitor	Fludarabine	Approved	1991	Fludarabine is a purine analog antimetabolite that inhibits DNA synthesis.	CHEMBL1096882	DB01073	Fludara	Small molecule	Oral, Intravenous	75	4	D015451; D009369; D008258; D012595; D015477	Leukemia, Lymphocytic, Chronic, B-Cell; Neoplasms; Waldenstrom Macroglobulinemia; Scleroderma, Systemic; Leukemia, Myelomonocytic, Chronic	DBCOND0030268 (Acute Myeloid Leukemia (AML)); DBCOND0034535 (B-Cell Chronic Lymphocytic Leukemia (B-CLL)); DBCOND0030902 (Leukemia, Lymphocytic, Acute, Adult); DBCOND0046362 (Malignant Lymphomas); DBCOND0032773 (Myeloproliferative Disorders (MPD)); DBCOND0032544 (Myelodysplastic Syndromes (MDS)); DBCOND0030421 (Plasma Cell Myeloma); DBCOND0029684 (Leukemia, Lymphocytic)	NCT00488709; NCT00220311; NCT00199056; NCT00361140	Completed; Completed; Completed; Completed	Treatment; Treatment; Treatment; Treatment	Fludarabine, Cytarabine, Topotecan in Treating Patients With Relapsed or Refractory Acute Myeloid Leukemia; A Study to Confirm the Efficacy and Safety of Fludarabine Phosphate Administered in Untreated Chronic Lymphocytic Leukemia Patients With Anemia and/or Thrombocytopenia; German Multicenter Trial for Treatment of Newly Diagnosed Acute Lymphoblastic Leukemia in Adults (06/99); Busulfan Safety/Efficacy as Conditioning Prior to Hematopoietic Cell Transplantation (HCT)
High	15	11	RRM1	unknown		Imexon	Investigational		Imexon is currently being studied for the treatment of pancreatic, lung, breast, prostate, melanoma, and multiple myeloma cancers. It belongs to the family of drugs called cyanoazidine derivatives. Also called Amplimexon. Imexon is a cyanoazidine derivative. Imexon is a thiol-binding small molecule which induces mitochondrial oxidation, a loss of membrane potential and cytochrome C, leading to apoptosis.	CHEMBL146428	DB05003		Small molecule			2	D008258; D015451; D002289; D008224; D010190	Waldenstrom Macroglobulinemia; Leukemia, Lymphocytic, Chronic, B-Cell; Carcinoma, Non-Small-Cell Lung; Lymphoma, Follicular; Pancreatic Neoplasms	DBCOND0110322 (Marginal Zone Lymphoma (MZL)); DBCOND0031072 (Lymphoplasmacytic Lymphoma); DBCOND0071979 (Mantle Cell Lymphoma (MCL)); DBCOND0028221 (Diffuse Large B-Cell Lymphoma (DLBCL)); DBCOND0083518 (Burkitt Lymphoma (BL)); DBCOND0023538 (Small Lymphocytic Lymphoma); DBCOND0137658 (Follicular Lymphoma (FL)); DBCOND0054129 (Neoplasms, Pancreatic); DBCOND0034623 (Non-Small Cell Lung Carcinoma (NSCLC))	NCT01314014; NCT00637247; NCT00697060	Completed; Completed; Withdrawn	Treatment; Treatment; Treatment	Imexon for Relapsed Follicular and Aggressive Lymphomas; A Randomized Study of Amplimexon (Imexon) With Gemcitabine in Pancreatic Cancer; A Safety and Efficacy Trial of Amplimexon Plus Taxotere in Metastatic Non-Small Cell Lung Cancer
High	15	11	RRM1	unknown		Tezacitabine	Investigational		A synthetic purine nucleoside analogue with potential antineoplastic activity.	CHEMBL2105467	DB06433		Small molecule			2	D000230; D013274; D004938; D015179	Adenocarcinoma; Stomach Neoplasms; Esophageal Neoplasms; Colorectal Neoplasms	DBCOND0060431 (Adenocarcinomas); DBCOND0055130 (Neoplasms, Esophageal); DBCOND0000827 (Stomach Neoplasms); DBCOND0036298 (Neoplasms, Colorectal)	NCT00054873; NCT00051688	Completed; Terminated	Treatment; Treatment	Tezacitabine With or Without 5-Fluorouracil (5-FU) for Advanced Esophageal Cancer or Gastric Cancer; Tezacitabine and Oxaliplatin for the Treatment of Patients With Metastatic Colorectal Cancer
High	15	11	RRM1	yes	inhibitor	Cladribine	Approved, Investigational	1993	Cladribine is a purine antimetabolite used for the management of relapsing forms of Multiple Sclerosis (MS), used in patients who have not responded to or who were unable to tolerate alternative MS drugs.	CHEMBL1619	DB00242	Litak, Mavenclad	Small molecule	Oral, Intravenous, Subcutaneous	19	4	D007154; D007943; D009103; D009369; D000740	Immune System Diseases; Leukemia, Hairy Cell; Multiple Sclerosis; Neoplasms; Anemia	DBCOND0027885 (Multiple Sclerosis); DBCOND0032744 (Lymphoma, Lymphoblastic); DBCOND0027885 (Multiple Sclerosis); DBCOND0036413 (Relapsing Remitting Multiple Sclerosis (RRMS))	NCT03961204; NCT00199017; NCT03364036; NCT03366965; NCT03963375	Completed; Completed; Completed; Recruiting	Other; Treatment; Treatment; Basic Science	Long-Term Outcomes and Durability of Effect Following Treatment With Cladribine Tablets for MS (CLASSIC-MS); German Multicenter Trial for the Treatment of Newly Diagnosed T-lymphoblastic Lymphoma in Adults; Evaluation of the Onset of Action in Highly Active MS (MAGNIFY); Evaluation of Quality of Life (QoL) in Subjects With Highly Active Relapsing Multiple Sclerosis (MS) (CLARIFY MS); Cladribine Tablets: Collaborative Study to Evaluate Impact On Central Nervous System Biomarkers in Multiple Sclerosis
Medium	12	8	GSR			Glutathione disulfide	Approved, Experimental, Investigational	2008	Glutathione disulfide is an ingredient of ophthalmic irrigation solutions used during ophthalmological procedures.	CHEMBL1372	DB03310	Bss Ophthalmic Solution	Small molecule	Intraocular, Ophthalmic	1	3	D002289; D001943; D007938; D010051	Carcinoma, Non-Small-Cell Lung; Breast Neoplasms; Leukemia; Ovarian Neoplasms	DBCOND0028228 (Cataracts)	NCT00333268; NCT00346528	Completed	Treatment	Safety and Efficacy of Next Generation Ophthalmic Irrigating Solution Compared to BSS Plus; Safety of NGOIS Compared to BSS Plus in Pediatric Patients
Medium	17	12	GRIN2B			Memantine	Approved, Investigational	2003	Memantine is an NMDA receptor antagonist used to treat moderate to severe dementia in Alzheimer's.	CHEMBL807	DB01043	Axura, Ebixa, Marixino, Namenda, Namenda 49 Titration Pack, Namzaric	Small molecule	Oral	83	4	D003704; D000544; D012559; D005356	Dementia; Alzheimer Disease; Schizophrenia; Fibromyalgia	DBCOND0049114 (Alzheimer's Disease (AD))	NCT00120874; NCT00933608	Completed; Completed	Not Available; Basic Science	Memantine and Comprehensive, Individualized Management of Alzheimer's Disease and Caregiver Training; Effects of Memantine on Magnetic Resonance (MR) Spectroscopy in Subjects at Risk for Alzheimer's Disease
Medium	17	12	GUCY2C			Linaclotide	Approved	2012	Linaclotide is a guanylate cyclase-C agonist used to treat constipation associated with irritable bowel syndrome or that is idiopathic in nature.	CHEMBL3301675	DB08890	Constella, Linzess	Small molecule	Oral	13	4	D043183; D003248; D000236; D015179	Irritable Bowel Syndrome; Constipation; Adenoma; Colorectal Neoplasms	DBCOND0031256 (Chronic Constipation); DBCOND0027886 (Diabetes Mellitus); DBCOND0060222 (Constipation-predominant Irritable Bowel Syndrome (IBS-C)); DBCOND0102972 (chronic idiopathic constipation (CIC)); DBCOND0052193 (Crohn's Disease (CD)); DBCOND0029012 (Ulcerative Colitis); DBCOND0031498 (Gastrointestinal Hemorrhage)	NCT03119584; NCT02590432; NCT02465385	Active Not Recruiting; Completed; Completed	Treatment; Treatment; Treatment	Efficacy of Linaclotide in Type II Diabetics With Chronic Constipation; An Open-Label, Long-term Study to Assess the Immunogenicity of LINZESS® (Linaclotide) Administered Orally to Adult Participants With Irritable Bowel Syndrome With Constipation or Chronic Idiopathic Constipation; Single-dose Linaclotide for Capsule Endoscopy Preparation

Notes: Used ChEMBL, DrugBank and ClinicalTrials.gov databases to highlight top approved or investigational drugs fo

Supplementary Table 1. Characteristics of cases and controls with ancestry data.

Case	Control	Sex	Age_Dx	Age_Ref	Ancestry	Ancestry_Source	Ancestry_Method	Ancestry_Date	Ancestry_Version	Ancestry_Comments	Ancestry_Confidence	Ancestry_Flag	Ancestry_Exclusion	Ancestry_Exclusion_Reason	Ancestry_Exclusion_Date
1	1	M	78.0	78.0	Ashkenazi Jew	1	1	2.00E+00	2.00E+00	1000G_PHASE3	0.97	Ashkenazi Jew	0	NA	NA
2	2	M	78.0	78.0	Ashkenazi Jew	1	1	2.00E+00	2.00E+00	1000G_PHASE3	0.97	Ashkenazi Jew	0	NA	NA
3	3	M	78.0	78.0	Ashkenazi Jew	1	1	2.00E+00	2.00E+00	1000G_PHASE3	0.97	Ashkenazi Jew	0	NA	NA
4	4	M	78.0	78.0	Ashkenazi Jew	1	1	2.00E+00	2.00E+00	1000G_PHASE3	0.97	Ashkenazi Jew	0	NA	NA
5	5	M	78.0	78.0	Ashkenazi Jew	1	1	2.00E+00	2.00E+00	1000G_PHASE3	0.97	Ashkenazi Jew	0	NA	NA
6	6	M	78.0	78.0	Ashkenazi Jew	1	1	2.00E+00	2.00E+00	1000G_PHASE3	0.97	Ashkenazi Jew	0	NA	NA
7	7	M	78.0	78.0	Ashkenazi Jew	1	1	2.00E+00	2.00E+00	1000G_PHASE3	0.97	Ashkenazi Jew	0	NA	NA
8	8	M	78.0	78.0	Ashkenazi Jew	1	1	2.00E+00	2.00E+00	1000G_PHASE3	0.97	Ashkenazi Jew	0	NA	NA
9	9	M	78.0	78.0	Ashkenazi Jew	1	1	2.00E+00	2.00E+00	1000G_PHASE3	0.97	Ashkenazi Jew	0	NA	NA
10	10	M	78.0	78.0	Ashkenazi Jew	1	1	2.00E+00	2.00E+00	1000G_PHASE3	0.97	Ashkenazi Jew	0	NA	NA
11	11	M	78.0	78.0	Ashkenazi Jew	1	1	2.00E+00	2.00E+00	1000G_PHASE3	0.97	Ashkenazi Jew	0	NA	NA
12	12	M	78.0	78.0	Ashkenazi Jew	1	1	2.00E+00	2.00E+00	1000G_PHASE3	0.97	Ashkenazi Jew	0	NA	NA
13	13	M	78.0	78.0	Ashkenazi Jew	1	1	2.00E+00	2.00E+00	1000G_PHASE3	0.97	Ashkenazi Jew	0	NA	NA
14	14	M	78.0	78.0	Ashkenazi Jew	1	1	2.00E+00	2.00E+00	1000G_PHASE3	0.97	Ashkenazi Jew	0	NA	NA
15	15	M	78.0	78.0	Ashkenazi Jew	1	1	2.00E+00	2.00E+00	1000G_PHASE3	0.97	Ashkenazi Jew	0	NA	NA
16	16	M	78.0	78.0	Ashkenazi Jew	1	1	2.00E+00	2.00E+00	1000G_PHASE3	0.97	Ashkenazi Jew	0	NA	NA
17	17	M	78.0	78.0	Ashkenazi Jew	1	1	2.00E+00	2.00E+00	1000G_PHASE3	0.97	Ashkenazi Jew	0	NA	NA
18	18	M	78.0	78.0	Ashkenazi Jew	1	1	2.00E+00	2.00E+00	1000G_PHASE3	0.97	Ashkenazi Jew	0	NA	NA
19	19	M	78.0	78.0	Ashkenazi Jew	1	1	2.00E+00	2.00E+00	1000G_PHASE3	0.97	Ashkenazi Jew	0	NA	NA
20	20	M	78.0	78.0	Ashkenazi Jew	1	1	2.00E+00	2.00E+00	1000G_PHASE3	0.97	Ashkenazi Jew	0	NA	NA
21	21	M	78.0	78.0	Ashkenazi Jew	1	1	2.00E+00	2.00E+00	1000G_PHASE3	0.97	Ashkenazi Jew	0	NA	NA
22	22	M	78.0	78.0	Ashkenazi Jew	1	1	2.00E+00	2.00E+00	1000G_PHASE3	0.97	Ashkenazi Jew	0	NA	NA
23	23	M	78.0	78.0	Ashkenazi Jew	1	1	2.00E+00	2.00E+00	1000G_PHASE3	0.97	Ashkenazi Jew	0	NA	NA
24	24	M	78.0	78.0	Ashkenazi Jew	1	1	2.00E+00	2.00E+00	1000G_PHASE3	0.97	Ashkenazi Jew	0	NA	NA
25	25	M	78.0	78.0	Ashkenazi Jew	1	1	2.00E+00	2.00E+00	1000G_PHASE3	0.97	Ashkenazi Jew	0	NA	NA
26	26	M	78.0	78.0	Ashkenazi Jew	1	1	2.00E+00	2.00E+00	1000G_PHASE3	0.97	Ashkenazi Jew	0	NA	NA
27	27	M	78.0	78.0	Ashkenazi Jew	1	1	2.00E+00	2.00E+00	1000G_PHASE3	0.97	Ashkenazi Jew	0	NA	NA
28	28	M	78.0	78.0	Ashkenazi Jew	1	1	2.00E+00	2.00E+00	1000G_PHASE3	0.97	Ashkenazi Jew	0	NA	NA
29	29	M	78.0	78.0	Ashkenazi Jew	1	1	2.00E+00	2.00E+00	1000G_PHASE3	0.97	Ashkenazi Jew	0	NA	NA
30	30	M	78.0	78.0	Ashkenazi Jew	1	1	2.00E+00	2.00E+00	1000G_PHASE3	0.97	Ashkenazi Jew	0	NA	NA
31	31	M	78.0	78.0	Ashkenazi Jew	1	1	2.00E+00	2.00E+00	1000G_PHASE3	0.97	Ashkenazi Jew	0	NA	NA
32	32	M	78.0	78.0	Ashkenazi Jew	1	1	2.00E+00	2.00E+00	1000G_PHASE3	0.97	Ashkenazi Jew	0	NA	NA
33	33	M	78.0	78.0	Ashkenazi Jew	1	1	2.00E+00	2.00E+00	1000G_PHASE3	0.97	Ashkenazi Jew	0	NA	NA
34	34	M	78.0	78.0	Ashkenazi Jew	1	1	2.00E+00	2.00E+00	1000G_PHASE3	0.97	Ashkenazi Jew	0	NA	NA
35	35	M	78.0	78.0	Ashkenazi Jew	1	1	2.00E+00	2.00E+00	1000G_PHASE3	0.97	Ashkenazi Jew	0	NA	NA
36	36	M	78.0	78.0	Ashkenazi Jew	1	1	2.00E+00	2.00E+00	1000G_PHASE3	0.97	Ashkenazi Jew	0	NA	NA
37	37	M	78.0	78.0	Ashkenazi Jew	1	1	2.00E+00	2.00E+00	1000G_PHASE3	0.97	Ashkenazi Jew	0	NA	NA
38	38	M	78.0	78.0	Ashkenazi Jew	1	1	2.00E+00	2.00E+00	1000G_PHASE3	0.97	Ashkenazi Jew	0	NA	NA
39	39	M	78.0	78.0	Ashkenazi Jew	1	1	2.00E+00	2.00E+00	1000G_PHASE3	0.97	Ashkenazi Jew	0	NA	NA
40	40	M	78.0	78.0	Ashkenazi Jew	1	1	2.00E+00	2.00E+00	1000G_PHASE3	0.97	Ashkenazi Jew	0	NA	NA
41	41	M	78.0	78.0	Ashkenazi Jew	1	1	2.00E+00	2.00E+00	1000G_PHASE3	0.97	Ashkenazi Jew	0	NA	NA
42	42	M	78.0	78.0	Ashkenazi Jew	1	1	2.00E+00	2.00E+00	1000G_PHASE3	0.97	Ashkenazi Jew	0	NA	NA
43	43	M	78.0	78.0	Ashkenazi Jew	1	1	2.00E+00	2.00E+00	1000G_PHASE3	0.97	Ashkenazi Jew	0	NA	NA
44	44	M	78.0	78.0	Ashkenazi Jew	1	1	2.00E+00	2.00E+00	1000G_PHASE3	0.97	Ashkenazi Jew	0	NA	NA
45	45	M	78.0	78.0	Ashkenazi Jew	1	1	2.00E+00	2.00E+00	1000G_PHASE3	0.97	Ashkenazi Jew	0	NA	NA
46	46	M	78.0	78.0	Ashkenazi Jew	1	1	2.00E+00	2.00E+00	1000G_PHASE3	0.97	Ashkenazi Jew	0	NA	NA
47	47	M	78.0	78.0	Ashkenazi Jew	1	1	2.00E+00	2.00E+00	1000G_PHASE3	0.97	Ashkenazi Jew	0	NA	NA
48	48	M	78.0	78.0	Ashkenazi Jew	1	1	2.00E+00	2.00E+00	1000G_PHASE3	0.97	Ashkenazi Jew	0	NA	NA
49	49	M	78.0	78.0	Ashkenazi Jew	1	1	2.00E+00	2.00E+00	1000G_PHASE3	0.97	Ashkenazi Jew	0	NA	NA
50	50	M	78.0	78.0	Ashkenazi Jew	1	1	2.00E+00	2.00E+00	1000G_PHASE3	0.97	Ashkenazi Jew	0	NA	NA
51	51	M	78.0	78.0	Ashkenazi Jew	1	1	2.00E+00	2.00E+00	1000G_PHASE3	0.97	Ashkenazi Jew	0	NA	NA
52	52	M	78.0	78.0	Ashkenazi Jew	1	1	2.00E+00	2.00E+00	1000G_PHASE3	0.97	Ashkenazi Jew	0	NA	NA
53	53	M	78.0	78.0	Ashkenazi Jew	1	1	2.00E+00	2.00E+00	1000G_PHASE3	0.97	Ashkenazi Jew	0	NA	NA
54	54	M	78.0	78.0	Ashkenazi Jew	1	1	2.00E+00	2.00E+00	1000G_PHASE3	0.97	Ashkenazi Jew	0	NA	NA
55	55	M	78.0	78.0	Ashkenazi Jew	1	1	2.00E+00	2.00E+00	1000G_PHASE3	0.97	Ashkenazi Jew	0	NA	NA
56	56	M	78.0	78.0	Ashkenazi Jew	1	1	2.00E+00	2.00E+00	1000G_PHASE3	0.97	Ashkenazi Jew	0	NA	NA
57	57	M	78.0	78.0	Ashkenazi Jew	1	1	2.00E+00	2.00E+00	1000G_PHASE3	0.97	Ashkenazi Jew	0	NA	NA
58	58	M	78.0	78.0	Ashkenazi Jew	1	1	2.00E+00	2.00E+00	1000G_PHASE3	0.97	Ashkenazi Jew	0	NA	NA
59	59	M	78.0	78.0	Ashkenazi Jew	1	1	2.00E+00	2.00E+00	1000G_PHASE3	0.97	Ashkenazi Jew	0	NA	NA
60	60	M	78.0	78.0	Ashkenazi Jew	1	1	2.00E+00	2.00E+00	1000G_PHASE3	0.97	Ashkenazi Jew	0	NA	NA
61	61	M	78.0	78.0	Ashkenazi Jew	1	1	2.00E+00	2.00E+00	1000G_PHASE3	0.97	Ashkenazi Jew	0	NA	NA
62	62	M	78.0	78.0	Ashkenazi Jew	1	1	2.00E+00	2.00E+00	1000G_PHASE3	0.97	Ashkenazi Jew	0	NA	NA
63	63	M	78.0	78.0	Ashkenazi Jew	1	1	2.00E+00	2.00E+00	1000G_PHASE3	0.97	Ashkenazi Jew	0	NA	NA
64	64	M	78.0	78.0	Ashkenazi Jew	1	1	2.00E+00	2.00E+00	1000G_PHASE3	0.97	Ashkenazi Jew	0	NA	NA
65	65	M	78.0	78.0	Ashkenazi Jew	1	1	2.00E+00	2.00E+00	1000G_PHASE3	0.97	Ashkenazi Jew	0	NA	NA
66	66	M	78.0	78.0	Ashkenazi Jew	1	1	2.00E+00	2.00E+00	1000G_PHASE3	0.97	Ashkenazi Jew	0	NA	NA
67	67	M	78.0	78.0	Ashkenazi Jew	1	1	2.00E+00	2.00E+00	1000G_PHASE3	0.97	Ashkenazi Jew	0	NA	NA
68	68	M	78.0	78.0	Ashkenazi Jew	1	1	2.00E+00	2.00E+00	1000G_PHASE3	0.97	Ashkenazi Jew	0	NA	NA
69	69	M	78.0	78.0	Ashkenazi Jew	1	1	2.00E+00	2.00E+00	1000G_PHASE3	0.97	Ashkenazi Jew	0	NA	NA
70	70	M	78.0	78.0	Ashkenazi Jew	1	1	2.00E+00	2.00E+00	1000G_PHASE3	0.97	Ashkenazi Jew	0	NA	NA
71	71	M	78.0	78.0	Ashkenazi Jew	1	1	2.00E+00	2.00E+00	1000G_PHASE3	0.97	Ashkenazi Jew	0	NA	NA
72	72	M	78.0	78.0	Ashkenazi Jew	1	1	2.00E+00	2.00E+00	1000G_PHASE3	0.97	Ashkenazi Jew	0	NA	NA
73	73	M	78.0	78.0	Ashkenazi Jew	1	1	2.00E+00	2.00E+00	1000G_PHASE3	0.97	Ashkenazi Jew	0	NA	NA
74	74	M	78.0	78.0	Ashkenazi Jew	1	1	2.00E+00	2.00E+00	1000G_PHASE3	0.97	Ashkenazi Jew	0	NA	NA
75	75	M	78.0	78.0	Ashkenazi Jew	1	1	2.00E+00	2.00E+00	1000G_PHASE3	0.97	Ashkenazi Jew	0	NA	NA
76	76	M	78.0	78.0	Ashkenazi Jew	1	1	2.00E+00	2.00E+00	1000G_PHASE3	0.97	Ashkenazi Jew	0	NA	NA
77	77	M	78.0	78.0	Ashkenazi Jew	1	1	2.00E+00	2.00E+00	1000G_PHASE3	0.97	Ashkenazi Jew	0	NA	NA
78	78	M	78.0	78.0	Ashkenazi Jew	1	1	2.00E+00	2.00E+00	1000G_PHASE3	0.97	Ashkenazi Jew	0	NA	NA
79	79	M	78.0	78.0	Ashkenazi Jew	1	1	2.00E+00	2.00E+00	1000G_PHASE3	0.97	Ashkenazi Jew	0	NA	NA
80	80	M	78.0	78.0	Ashkenazi Jew	1	1	2.00E+00	2.00E+00	1000G_PHASE3	0.97	Ashkenazi Jew	0	NA	NA
81	81	M	78.0	78.0	Ashkenazi Jew	1	1	2.00E+00	2.00E+00	1000G_PHASE3	0.97	Ashkenazi Jew	0	NA	NA
82	82	M	78.0	78.0	Ashkenazi Jew	1	1	2.00E+00	2.00E+00	1000G_PHASE3	0.97	Ashkenazi Jew	0	NA	NA
83	83	M	78.0	78.0	Ashkenazi Jew	1	1	2.00E+00	2.00E+00	1000G_PHASE3	0.97	Ashkenazi Jew	0	NA	NA
84	84	M	78.0	78.0	Ashkenazi Jew	1	1	2.00E+00	2.00E+00	1000G_PHASE3	0.97	Ashkenazi Jew	0	NA	NA
85	85	M	78.0	78.0	Ashkenazi Jew	1	1	2.00E+00	2.00E+00	1000G_PHASE3	0.97	Ashkenazi Jew	0	NA	NA
86	86	M	78.0	78.0	Ashkenazi Jew	1	1	2.00E+00	2.00E+00	1000G_PHASE3	0.97	Ashkenazi Jew	0	NA	NA
87	87	M	78.0	78.0	Ashkenazi Jew	1	1	2.00E+00	2.00E+00	1000G_PHASE3	0.97	Ashkenazi Jew	0	NA	NA
88	88	M	78.0	78.0	Ashkenazi Jew	1	1	2.00E+00	2.00E+00	1000G_PHASE3	0.97	Ashkenazi Jew	0	NA	NA
89	89	M	78.0	78.0	Ashkenazi Jew	1	1	2.00E+00	2.00E+00	1000G_PHASE3	0.97	Ashkenazi Jew	0	NA	NA
90	90	M	78.0	78.0	Ashkenazi Jew	1	1	2.00E+00	2.00E+00	1000G_PHASE3	0.97	Ashkenazi Jew	0	NA	NA
91	91	M	78.0	78.0	Ashkenazi Jew	1	1	2.00E+00	2.00E+00	1000G_PHASE3	0.97	Ashkenazi Jew	0	NA	NA
92	92	M	78.0	78.0	Ashkenazi Jew	1	1	2.00E+00	2.00E+00	1000G_PHASE3	0.97	Ashkenazi Jew	0	NA	NA
93	93	M	78.0	78.0	Ashkenazi Jew	1	1	2.00E+00	2.00E+00	1000G_PHASE3	0.97	Ashkenazi Jew	0	NA	NA
94	94	M	78.0	78.0	Ashkenazi Jew	1	1	2.00E+00	2.00E+00	1000G_PHASE3	0.97	Ashkenazi Jew	0	NA	NA
95	95	M	78.0	78.0	Ashkenazi Jew	1	1	2.00E+00	2.00E+00	1000G_PHASE3	0.97	Ashkenazi Jew	0	NA	NA
96	96	M	78.0	78.0	Ashkenazi Jew	1	1	2.00E+00	2.00E+00	1000G_PHASE3	0.97	Ashkenazi Jew	0	NA	NA
97	97	M	78.0	78.0	Ashkenazi Jew	1	1	2.00E+00	2.00E+00	1000G_PHASE3	0.97	Ashkenazi Jew	0	NA	NA
98	98	M	78.0	78.0	Ashkenazi Jew	1	1	2.00E+00	2.00E+00	1000G_PHASE3	0.97	Ashkenazi Jew	0	NA	NA
99	99	M	78.0	78.0	Ashkenazi Jew	1	1	2.00E+00	2.00E+00	1000G_PHASE3	0.97	Ashkenazi Jew	0	NA	NA
100	100	M	78.0	78.0	Ashkenazi Jew	1	1	2.00E+00	2.00E+00	1000G_PHASE3	0.97	Ashkenazi Jew	0	NA	NA
101	101	M	78.0	78.0	Ashkenazi Jew	1	1	2.00E+00	2.00E+00	1000G_PHASE3	0.97	Ashkenazi Jew	0	NA	NA
102	102	M	78.0	78.0	Ashkenazi Jew	1	1	2.00E+00	2.00E+00	1000G_PHASE3	0.97	Ashkenazi Jew	0	NA	NA
103	103	M	78.0	78.0	Ashkenazi Jew	1	1	2.00E+00	2.00E+00	1000G_PHASE3	0.97	Ashkenazi Jew	0	NA	NA
104	104	M	78.0	78.0	Ashkenazi Jew	1	1	2.00E+00	2.00E+00	1000G_PHASE3	0.97	Ashkenazi Jew	0	NA	NA
105	105	M	78.0	78.0	Ashkenazi Jew	1	1	2.00E+00	2.00E+00	1000G_PHASE3	0.97	Ashkenazi Jew	0	NA	NA
106	106	M	78.0	78.0	Ashkenazi Jew	1	1	2.00E+00	2.00E+00	1000G_PHASE3	0.97	Ashkenazi Jew	0	NA	NA
107	107	M	78.0	78.0	Ashkenazi Jew	1	1	2.00E+00	2.00E+00	1000G_PHASE3	0.97	Ashkenazi Jew	0	NA	NA
108	108	M	78.0	78.0	Ashkenazi Jew	1	1	2.00E+00	2.00E+00	1000G_PHASE3	0.97	Ashkenazi Jew	0	NA	NA
109	109	M	78.0	78.0	Ashkenazi Jew	1	1	2.00E+00	2.00E+00	1000G_PHASE3	0.97	Ashkenazi Jew	0	NA	NA
110	110	M	78.0	78.0	Ashkenazi Jew	1	1	2.00E+00	2.00E+00	1000G_PHASE3	0.97	Ashkenazi Jew	0	NA	NA
111	111	M	78.0	78.0	Ashkenazi Jew	1	1	2.00E+00	2.00E+00	1000G_PHASE3	0.97	Ashkenazi Jew	0	NA	NA
112	112	M	78.0	78.0	Ashkenazi Jew	1	1	2.00E+00	2.00E+00	1000G_PHASE3	0.97	Ashkenazi Jew	0	NA	NA
113	113	M	78.0	78.0	Ashkenazi Jew	1	1	2.00E+00	2.00E+00	1000G_PHASE3	0.97	Ashkenazi Jew	0	NA	NA
114	114	M	78.0	78.0	Ashkenazi Jew	1	1	2.00E+00	2.00E+00	1000G_PHASE3	0.97	Ashkenazi Jew	0	NA	NA
115	115	M	78.0	78.0	Ashkenazi Jew	1	1	2.00E+00	2.00E+00	1000G_PHASE3	0.97	Ashkenazi Jew	0	NA	NA
116	116	M	78.0	78.0	Ashkenazi Jew	1	1	2.00E+00	2.00E+00	1000G_PHASE3	0.97	Ashkenazi Jew	0	NA	NA
117	117	M	78.0	78.0	Ashkenazi Jew	1	1	2.00E+00	2.00E+00	1000G_PHASE3	0.97	Ashkenazi Jew	0	NA	NA
118	118	M	78.0	78.0	Ashkenazi Jew	1	1	2.00E+00	2.00E+00	1000G_PHASE3	0.97	Ashkenazi Jew	0	NA	NA
119	119	M	78.0	78.0	Ashkenazi Jew	1	1	2.00E+00	2.00E+00	1000G_PHASE3	0.97	Ashkenazi Jew	0	NA	NA
120	120	M	78.0	78.0	Ashkenazi Jew	1	1	2.00E+00	2.00E+00	1000G_PHASE3	0.97	Ashkenazi Jew	0	NA	NA
121	121	M	78.0	78.0	Ashkenazi Jew	1	1	2.00E+00	2.00E+00	1000G_PHASE3	0.97	Ashkenazi Jew	0	NA	NA
122	122	M	78.0	78.0	Ashkenazi Jew	1	1	2.00E+00	2.00E+00	1000G_PHASE3	0.97	Ashkenazi Jew	0	NA	NA
123	123	M	78.0	78.0	Ashkenazi Jew	1	1	2.00E+00	2.00E+00	1000G_PHASE3	0.97	Ashkenazi Jew	0	NA	NA
124	124	M	78.0	78.0	Ashkenazi Jew	1	1	2.00E+00	2.00E+00	1000G_PHASE3	0.97	Ashkenazi Jew	0	NA	NA
125	125	M	78.0	78.0	Ashkenazi Jew	1	1	2.00E+00	2.00E+00	1000G_PHASE3	0.97	Ashkenazi Jew	0	NA	NA
126	126	M	78.0	78.0	Ashkenazi Jew	1	1	2.00E+00	2.00E+00	1000G_PHASE3	0.97	Ashkenazi Jew	0	NA	NA
127	127	M	78.0	78.0	Ashkenazi Jew	1	1	2.00E+00	2.00E+00	1000G_PHASE3	0.97	Ashkenazi Jew	0	NA	NA
128	128	M	78.0	78.0	Ashkenazi Jew	1	1	2.00E+00	2.00E+00	1000G_PHASE3	0.97	Ashkenazi Jew	0	NA	NA
129	129	M	78.0	78.0	Ashkenazi Jew	1	1	2.00E+00	2.00E+00	1000G_PHASE3	0.97	Ashkenazi Jew	0	NA	NA
130	130	M	78.0	78.0	Ashkenazi Jew	1	1	2.00E+00	2.00E+00	1000G_PHASE3	0.97	Ashkenazi Jew	0	NA	NA
131	131	M	78.0	78.0	Ashkenazi Jew	1	1	2.00E+00	2.00E+00	1000G_PHASE3	0.97	Ashkenazi Jew	0	NA	NA
132	132	M	78.0	78.0	Ashkenazi Jew	1	1	2.00E+00	2.00E+00	1000G_PHASE3	0.97	Ashkenazi Jew	0	NA	NA
133	133	M	78.0	78.0	Ashkenazi Jew	1	1	2.00E+00	2.00E+00	1000G_PHASE3	0.97	Ashkenazi Jew	0	NA	NA
134	134	M	78.0	78.0	Ashkenazi Jew	1	1	2.00E+00	2.00E+00	1000G_PHASE3	0.97	Ashkenazi Jew	0	NA	NA
135	135	M	78.0	78.0	Ashkenazi Jew	1	1	2.00E+00	2.00E+00	1000G_PHASE3	0.97	Ashkenazi Jew	0	NA	NA
136	136	M	78.0	78.0	Ashkenazi Jew	1	1	2.00E+00	2.00E+00	1000G_PHASE3	0.97	Ashkenazi Jew	0	NA	NA
137	137	M	78.0	78.0	Ashkenazi Jew	1	1	2.00E+00	2.00E+00	1000G_PHASE3	0.97	Ashkenazi Jew	0	NA	NA
138	138	M	78.0	78.0	Ashkenazi Jew	1	1	2.00E+00	2.00E+00	1000G_PHASE3	0.97	Ashkenazi Jew	0	NA	NA
139	139	M	78.0	78.0	Ashkenazi Jew	1	1	2.00E+00	2.00E+00	1000G_PHASE3	0.97	Ashkenazi Jew	0	NA	NA
140	140	M	78.0	78.0	Ashkenazi Jew	1	1	2.00E+00	2.00E+00	1000G_PHASE3	0.97	Ashkenazi Jew	0	NA	NA
141	141	M	78.0	78.0	Ashkenazi Jew	1	1	2.00E+00	2.00E+00	1000G_PHASE3	0.97	Ashkenazi Jew	0	NA	NA
142	142	M	78.0	78.0	Ashkenazi Jew	1	1	2.00E+00	2.00E+00	1000G_PHASE3	0.97	Ashkenazi Jew	0	NA	NA
143	143	M	78.0	78.0	Ashkenazi Jew	1	1	2.00E+00	2.00E+00	1000G_PHASE3	0.97	Ashkenazi Jew	0	NA	NA
144	144	M	78.0	78.0	Ashkenazi Jew	1	1	2.00E+00	2.00E+00	1000G_PHASE3	0.97	Ashkenazi Jew	0	NA	NA
145	145	M	78.0	78.0	Ashkenazi Jew	1	1	2.00E+00	2.00E+00	1000G_PHASE3	0.97	Ashkenazi Jew	0	NA	NA
146	146	M	78.0	78.0	Ashkenazi Jew	1	1	2.00E+00	2.00E+00	1000G_PHASE3	0.97	Ashkenazi Jew	0	NA	NA
147	147	M	78.0	78.0	Ashkenazi Jew	1	1	2.00E+00	2.00E+00	1000G_PHASE3	0.97	Ashkenazi Jew	0	NA	NA
148	148	M	78.0	78.0	Ashkenazi Jew	1	1	2.00E+00	2.00E+00	1000G_PHASE3	0.97	Ashkenazi Jew	0	NA	NA
149	149	M	78.0	78.0	Ashkenazi Jew	1	1	2.00E+00	2.00E+00	1000G_PHASE3	0.97	Ashkenazi Jew	0	NA	NA
150	150	M	78.0	78.0	Ashkenazi Jew	1	1	2.00E+00	2.00E+00	1000G_PHASE3	0.97	Ashkenazi Jew	0	NA	NA
151	151	M	78.0	78.0	Ashkenazi Jew	1	1	2.00E+00	2.00E+00	1000G_PHASE3	0.97	Ashkenazi Jew	0	NA	NA
152	152	M	78.0	78.0	Ashkenazi Jew	1	1	2.00E+00	2.00E+00	1000G_PHASE3	0.97	Ashkenazi Jew	0	NA	NA
153	153	M	78.0	78.0	Ashkenazi Jew	1	1	2.00E+00	2.00E+00	1000G_PHASE3	0.97	Ashkenazi Jew	0	NA	NA
154	154	M	78.0	78.0	Ashkenazi Jew	1	1	2.00E+00	2.00E+00	1000G_PHASE3	0.97	Ashkenazi Jew	0	NA	NA
155	155	M	78.0	78.0	Ashkenazi Jew	1	1	2.00E+00	2.00E+00	1000G_PHASE3	0.97	Ashkenazi Jew	0	NA	NA
156	156	M	78.0	78.0	Ashkenazi Jew	1	1	2.00E+00	2.00E+00	1000G_PHASE3	0.97	Ashkenazi Jew	0	NA	NA
157	157	M	78.0	78.0	Ashkenazi Jew	1	1	2.00E+00	2.00E+00	1000G_PHASE3	0.97	Ashkenazi Jew	0	NA	NA
158	158	M	78.0	78.0	Ashkenazi Jew	1	1	2.00E+00	2.00E+00	1000G_PHASE3	0.97	Ashken

Supplemental Table 1. Clinical characteristics of cases (Day 28) and controls (n=1,000,000)

Notes: LOF variants (ENSVEP 2018) or CNVs (ACMG Guidelines 2019) are in red, or if ClinVar stated CNV pathogenic or LOF variant in blue. Gene and Variant (ACMG HGVS nomenclature) or Functional consequence (ACMG Gene and Variant) are in bold if ACMG Gene and Variant or Functional consequence is different than ClinVar. Gene and Variant (ACMG HGVS nomenclature) or Functional consequence (ACMG Gene and Variant) are in italics if ACMG Gene and Variant or Functional consequence is different than ClinVar. Gene and Variant (ACMG HGVS nomenclature) or Functional consequence (ACMG Gene and Variant) are in bold and italics if ACMG Gene and Variant or Functional consequence is different than ClinVar. Gene and Variant (ACMG HGVS nomenclature) or Functional consequence (ACMG Gene and Variant) are in yellow if ACMG Gene and Variant or Functional consequence is different than ClinVar. Gene and Variant (ACMG HGVS nomenclature) or Functional consequence (ACMG Gene and Variant) are in green if ACMG Gene and Variant or Functional consequence is different than ClinVar.

Patient	ACMG	ClinVar	Gene and Variant (ACMG HGVS nomenclature)	Functional consequence (ACMG Gene and Variant)	Allele1	Allele2	Sex	Ethnicity	Age	Cancer diagnosis	Cancer	Sex	Ethnicity	Age	Cancer diagnosis	Cancer	Sex	Ethnicity	Age	Cancer diagnosis	Cancer
1-1-1000000			1-1-1000000-C-T		C	T	1-1-1000000	Male	0-5	1-1-1000000	No										
1-1-1000000			1-1-1000000-C-T		C	T	1-1-1000000	Male	0-5	1-1-1000000	No										
1-1-1000000			1-1-1000000-C-T		C	T	1-1-1000000	Male	0-5	1-1-1000000	No										
1-1-1000000			1-1-1000000-C-T		C	T	1-1-1000000	Male	0-5	1-1-1000000	No	1-1-1000000	Female	0-5	1-1-1000000	No

Supplementary Table 8: Functional annotation of novel index SNVs and SNVs in LD (r2=0.2)

Notes: 1. ra10 based on dbSNP build 146, 2. ra10 of the significant index SNP for this locus, 3. GWAS p-value of the significant index SNP for this locus, 4. The nearest Gene of the SNP based on ANNOVAR annotations, 5. Functional consequence of the SNP on the gene obtained from ANNOVAR, 6. Combined Annotation Dependent Depletion (CADD) score, based on 63 annotations. The higher the score, the more deleterious the SNP is, 12.37 is the suggested threshold, 7. Regulome DB categorical score (from 1a to 7). The higher the score (1a is the highest), the more biological evidence of the SNP to be regulatory element, 8. The minimum 15-core chromatin state across 127 tissue/cell type; 9. The most common 15-core chromatin state across 127 tissue/cell type; 10. The reference panel used in FUMA to obtain functional annotations.

Locus	rsID[1]	Chr	Pos (hg19)	unrsID	Allele1	Allele2	MAF	IndSigNP[2]	gwasP[3]	r2	nearestGene[4]	dist	func[5]	CADD[6]	RDB[7]	minChrState[8]	commonChrState[9]	Reference Genome[10]
1	rs11046971	3	9079291	3:9079291:A:G	G	A	0.0853	rs79367750	NA	0.7012	SRGAP3	0	intronic	2.71	5	2	7	ASN
1	rs2271205	3	9079449	3:9079449:C:G	G	C	0.1270	rs79367750	NA	0.4617	SRGAP3	0	intronic	1.34	7	4	15	ASN
1	rs2670007	3	9083526	3:9083526:C:T	C	T	0.1200	rs79367750	NA	0.5068	SRGAP3	0	intronic	2.83	NA	4	15	ASN
1	rs53046157	3	9084122	3:9084122:GA:GAA	GAA	GA	0.1577	rs79367750	NA	0.2800	SRGAP3	0	intronic	NA	NA	4	15	ASN
1	rs73562704	3	9084122	3:9084122:G:GAA	GAA	G	0.0427	rs79367750	NA	0.2800	SRGAP3	0	NA	2.75	NA	4	15	ASN
1	rs71314322	3	9086470	3:9086470:A:G	A	G	0.0883	rs79367750	NA	0.7121	SRGAP3	0	intronic	4.65	7	4	15	ASN
1	rs35253822	3	9086531	3:9086531:A:G	A	G	0.0883	rs79367750	NA	0.7121	SRGAP3	0	intronic	0.57	7	4	15	ASN
1	rs71314323	3	9086945	3:9086945:C:T	C	T	0.0843	rs79367750	NA	0.7483	SRGAP3	0	intronic	5.22	7	4	15	ASN
1	rs71305888	3	9088129	3:9088129:A:G	G	A	0.0893	rs79367750	NA	0.7481	SRGAP3	0	intronic	1.25	7	4	15	ASN
1	rs546854673	3	9088520	3:9088520:C:CT	C	CT	0.2440	rs79367750	NA	0.2320	SRGAP3	0	intronic	0.92	NA	4	15	ASN
1	rs2292250	3	9089178	3:9089178:A:G	A	G	0.0933	rs79367750	NA	0.7143	SRGAP3	0	intronic	8.51	NA	4	15	ASN
1	rs28741676	3	9090523	3:9090523:A:C	A	C	0.0933	rs79367750	NA	0.8030	SRGAP3	0	intronic	4.08	6	4	15	ASN
1	rs16809415	3	9093985	3:9093985:C:T	C	T	0.0823	rs79367750	NA	0.9755	SRGAP3	0	intronic	0.17	5	4	15	ASN
1	rs79367750	3	9096107	3:9096107:C:CT	CT	C	0.0804	rs79367750	1.0000000039	1.0000	SRGAP3	0	intronic	2.79	NA	4	15	ASN
1	rs144331599	3	9104063	3:9104063:G:GTAGTCCTATCATTCCCTCAGTCCTATCATTCCCT	G	GTAGTCCTATCATTCCCT	0.0883	rs79367750	NA	0.4227	SRGAP3	0	intronic	9.52	NA	4	15	ASN
1	rs139033001	3	9106545	3:9106545:T:TAATAAATA	TAATAAATA	T	0.1260	rs79367750	NA	0.2689	SRGAP3	0	intronic	7.68	NA	4	15	ASN
1	rs7632256	3	9107475	3:9107475:C:T	T	C	0.0903	rs79367750	NA	0.3665	SRGAP3	0	intronic	4.82	6	5	15	ASN
1	rs13088083	3	9111847	3:9111847:G:T	T	G	0.0714	rs79367750	NA	0.2067	SRGAP3	0	intronic	5.81	4	4	15	ASN
2	rs62340575	4	182595262	4:182595262:C:T	C	T	0.0675	rs111132093	NA	0.4493	RP11-540E1.6	145909	intergenic	5.48	7	5	15	ASN
2	rs12641856	4	182609865	4:182609865:A:G	G	A	0.0685	rs111132093	NA	0.4930	RP11-540E1.6	131306	intergenic	8.71	4	2	15	ASN
2	rs76799830	4	182622718	4:182622718:C:T	C	T	0.1885	rs111132093	NA	0.2368	RP11-540E1.6	118453	intergenic	10.51	4	7	15	ASN
2	rs2871272	4	182622966	4:182622966:A:G	G	A	0.0665	rs111132093	NA	0.9329	RP11-540E1.6	118205	intergenic	2.82	5	5	15	ASN
2	rs13147126	4	182626194	4:182626194:A:C	A	C	0.0605	rs111132093	NA	0.9001	RP11-540E1.6	114977	intergenic	1.19	6	9	15	ASN
2	rs55812346	4	182626204	4:182626204:C:CA	C	CA	0.0605	rs111132093	NA	0.9001	RP11-540E1.6	114967	intergenic	0.56	NA	9	15	ASN
2	rs10520505	4	182626599	4:182626599:A:G	G	A	0.0605	rs111132093	NA	0.9329	RP11-540E1.6	114572	intergenic	0.23	7	9	15	ASN
2	rs12502297	4	182626945	4:182626945:C:T	C	T	0.0815	rs111132093	NA	0.9168	RP11-540E1.6	114226	intergenic	3.33	6	9	15	ASN
2	rs4861491	4	182627579	4:182627579:A:G	G	A	0.0605	rs111132093	NA	0.9329	RP11-540E1.6	113592	intergenic	5.63	7	9	15	ASN
2	rs35350204	4	182629727	4:182629727:T:TA	TA	T	0.0536	rs111132093	NA	0.7232	RP11-540E1.6	111444	intergenic	2.01	NA	9	15	ASN
2	rs2309534	4	182631262	4:182631262:A:G	G	A	0.0625	rs111132093	NA	0.9665	RP11-540E1.6	109909	intergenic	0.66	5	9	15	ASN
2	rs12505685	4	182632409	4:182632409:A:G	G	A	0.1825	rs111132093	NA	0.2843	RP11-540E1.6	108762	intergenic	1.54	7	9	15	ASN
2	rs6552519	4	182634163	4:182634163:A:C	C	A	0.1825	rs111132093	NA	0.2843	RP11-540E1.6	107008	intergenic	4.78	6	9	15	ASN
2	rs6857076	4	182634531	4:182634531:A:C	C	A	0.0635	rs111132093	NA	0.9853	RP11-540E1.6	106640	intergenic	0.71	7	9	15	ASN
2	rs12505248	4	182634917	4:182634917:C:T	C	T	0.1825	rs111132093	NA	0.2843	RP11-540E1.6	106254	intergenic	0.09	6	9	15	ASN
2	rs2309637	4	182635303	4:182635303:G:T	G	T	0.1825	rs111132093	NA	0.2843	RP11-540E1.6	105868	intergenic	3.49	7	9	15	ASN
2	rs34350068	4	182636346	4:182636346:C:T	T	C	0.1835	rs111132093	NA	0.2823	RP11-540E1.6	104825	intergenic	0.38	7	9	15	ASN
2	rs76705985	4	182642551	4:182642551:G:T	T	G	0.1905	rs111132093	NA	0.2878	RP11-540E1.6	98620	intergenic	2.40	7	5	15	ASN
2	rs11405960	4	182642593	4:182642593:C:CA	CA	C	0.2093	rs111132093	NA	0.2480	RP11-540E1.6	98578	intergenic	0.66	NA	5	15	ASN
2	rs2309541	4	182643207	4:182643207:A:G	A	G	0.1915	rs111132093	NA	0.2748	RP11-540E1.6	97964	intergenic	1.67	7	5	15	ASN
2	rs2309542	4	182643912	4:182643912:A:G	A	G	0.1895	rs111132093	NA	0.2821	RP11-540E1.6	97259	intergenic	5.64	7	5	15	ASN
2	rs13126991	4	182648081	4:182648081:A:C	C	A	0.1806	rs111132093	NA	0.2883	RP11-540E1.6	93090	intergenic	4.71	7	5	15	ASN
2	rs6840636	4	182648457	4:182648457:G:T	T	G	0.0625	rs111132093	NA	1.0000	RP11-540E1.6	92714	intergenic	0.28	7	5	15	ASN
2	rs2309544	4	182648949	4:182648949:C:T	C	T	0.1825	rs111132093	NA	0.2804	RP11-540E1.6	92222	intergenic	4.27	5	5	15	ASN
2	rs2309545	4	182649354	4:182649354:C:T	T	C	0.1925	rs111132093	NA	0.2730	RP11-540E1.6	91817	intergenic	8.73	7	5	15	ASN
2	rs2309546	4	182649515	4:182649515:A:G	A	G	0.1925	rs111132093	NA	0.2730	RP11-540E1.6	91656	intergenic	0.17	7	5	15	ASN
2	rs7682313	4	182649811	4:182649811:G:T	T	G	0.1925	rs111132093	NA	0.2730	RP11-540E1.6	91360	intergenic	0.95	5	5	15	ASN
2	rs2309547	4	182650724	4:182650724:C:T	C	T	0.0625	rs111132093	NA	1.0000	RP11-540E1.6	90447	intergenic	1.01	7	5	15	ASN
2	rs11132089	4	182651374	4:182651374:A:C	A	C	0.1925	rs111132093	NA	0.2730	RP11-540E1.6	89797	intergenic	0.93	5	5	15	ASN
2	rs7662761	4	182654376	4:182654376:C:G	G	C	0.0635	rs111132093	NA	0.9833	RP11-540E1.6	86795	intergenic	3.28	7	9	15	ASN
2	rs11132090	4	182659787	4:182659787:C:T	T	C	0.0625	rs111132093	NA	1.0000	RP11-540E1.6	81384	intergenic	1.95	6	9	15	ASN
2	rs6850344	4	182660440	4:182660440:G:T	G	T	0.0625	rs111132093	NA	1.0000	RP11-540E1.6	80731	intergenic	3.84	6	7	15	ASN
2	rs6824363	4	182660458	4:182660458:A:C	C	A	0.0625	rs111132093	NA	1.0000	RP11-540E1.6	80713	intergenic	1.18	6	7	15	ASN
2	rs6824376	4	182660465	4:182660465:A:G	G	A	0.0625	rs111132093	NA	1.0000	RP11-540E1.6	80706	intergenic	1.14	6	7	15	ASN
2	rs4422450	4	182661635	4:182661635:C:G	G	C	0.1796	rs111132093	NA	0.2904	RP11-540E1.6	79536	intergenic	1.49	6	9	15	ASN
2	rs11132091	4	182663664	4:182663664:C:T	C	T	0.0625	rs111132093	NA	1.0000	RP11-540E1.6	77507	intergenic	0.54	7	9	15	ASN
2	rs2309538	4	182664749	4:182664749:C:G	G	C	0.0625	rs111132093	NA	1.0000	RP11-540E1.6	76422	intergenic	0.26	7	9	15	ASN
2	rs11132093	4	182664815	4:182664815:A:G	G	A	0.0625	rs111132093	.0000000050	1.0000	RP11-540E1.6	76356	intergenic	0.65	6	9	15	ASN
2	rs12512009	4	182667807	4:182667807:A:T	T	A	0.0645	rs111132093	NA	0.9026	RP11-540E1.6	73364	intergenic	0.93	6	5	15	ASN
2	rs6552521	4	182669892	4:182669892:A:G	A	G	0.0665	rs111132093	NA	0.8734	RP11-540E1.6	71279	intergenic	17.76	6	5	15	ASN
2	rs2309548	4	182671878	4:182671878:A:C	C	A	0.0655	rs111132093	NA	0.8878	RP11-540E1.6	69293	intergenic	2.62	6	5	15	ASN
2	rs35249909	4	182674683	4:182674683:GA:GAA	GA	GAA	0.2242	rs111132093	NA	0.6610	RP11-540E1.6	66488	intergenic	NA	NA	5	15	ASN
2	rs53143010	4	182674683	4:182674683:G:GA	GA	G	0.0833	rs111132093	NA	0.6610	RP11-540E1.6	66488	intergenic	2.85	NA	5	15	ASN
2	rs4449752	4	182675595	4:182675595:A:G	G	A	0.0655	rs111132093	NA	0.8878	RP11-540E1.6	65576	intergenic	2.14	5	5	15	ASN
2	rs6845685	4	182680490	4:182680490:C:T	T	C	0.0655	rs111132093	NA	0.8878	RP11-540E1.6	60681	intergenic	1.64	7	5	15	ASN
2	rs4241736	4	182680794	4:182680794:A:G	A	G	0.0655	rs111132093	NA	0.8878	RP11-540E1.6	60377	intergenic	1.17	7	5	15	ASN
2	rs4861492	4	182682024	4:182682024:A:T	A	T	0.0655	rs111132093	NA	0.8878	RP11-540E1.6	59147	intergenic	4.23	5	5	15	ASN
2	rs4241738	4	182682227	4:182682227:C:T	C	T	0.0655	rs111132093	NA	0.8878	RP11-540E1.6	58944	intergenic	5.72	5	5	15	ASN
2	rs4582191	4	182682724	4:182682724:C:T	T	C	0.0655	rs111132093	NA	0.8878	RP11-540E1.6	58447	intergenic	10.97	6	5	15	ASN
2	rs4861493	4	182683156	4:182683156:C:T	C	T	0.0655	rs111132093	NA	0.8878	RP11-540E1.6	58015	intergenic	1.50	7	5	15	ASN
2	rs4861494	4	182683289	4:182683289:A:G	G	A	0.0655	rs111132093	NA	0.8878	RP11-540E1.6	57882	intergenic	0.11	7	5	15	ASN
2	rs7663028	4	182684387	4:182684387:C:G	C	G	0.0655	rs111132093	NA	0.8878	RP11-540E1.6	56784	intergenic	0.10	6	5	15	ASN
2	rs71895142	4	182685406	4:182685406:A:AAG	A	AAG	0.0655	rs111132093	NA	0.8878	RP11-540E1.6	55765	intergenic	9.00	NA	5	15	ASN

Supplementary Table 8: Functional annotation of novel Indel variants in LD (r2=0.2)

Notes: 1. rs10 based on dbSNP build 146, 2. rs10 of the significant index SNP for this locus, 3. GWAS p-value of the significant index SNP for this locus, 4. The nearest Gene of the SNP based on ANNOVAR annotations, 5. Functional consequence of the SNP on the gene obtained from ANNOVAR, 6. Combined Annotation Dependent Depletion (CADD) score, based on 63 annotations. The higher the score, the more deleterious the SNP is. 12.37 is the suggested threshold, 7. RegulomeDB categorical score (from 1a to 7). The higher the score (1a is the highest), the more biological evidence of the SNP to be regulatory element, 8. The minimum 15-core chromatin state across 127 tissue/cell type; 9. The most common 15-core chromatin state across 127 tissue/cell type; 10. The reference panel used in PLINK to obtain functional annotations.

Locus	rs10[1]	Chr	Pos (hg19)	uniqID	Allele1	Allele2	MAF	IndSigSNP[2]	gwasP[3]	r2	nearestGene[4]	dist	func[5]	CADD[6]	RDB[7]	minChrState[8]	commonChrState[9]	Reference Genome[10]
1	rs200166881	4	181145020	4:181145020:A:AT	AT	A	0.0040	rs77249395	NA	0.2311	RP11-751A18;	349596	intergenic	1.75	NA	5	15	ALL
1	rs77249395	4	181181245	4:181181245:G:T	G	T	0.0114	rs77249395	1.0000000024	1.0000	RP11-751A18;	313317	intergenic	1.38	7	5	15	ALL
2	rs74672857	7	97864435	7:97864435:A:G	G	A	0.3117	rs7015	NA	0.2396	TEC1	0	UTR3	1.52	5	4	5	ALL
2	rs6951913	7	97847522	7:97847522:A:G	G	A	0.3011	rs7015	NA	0.2796	TEC1	0	intronic	0.03	4	2	5	ALL
2	rs73151174	7	97947738	7:97947738:A:G	A	G	0.2929	rs7015	NA	0.2916	TEC1	0	intronic	11.77	2a	1	5	ALL
2	rs112068728	7	97849652	7:97849652:A:G	G	A	0.2991	rs7015	NA	0.2857	TEC1	0	intronic	1.03	6	4	5	ALL
2	rs55738918	7	97852698	7:97852698:A:C	C	A	0.2991	rs7015	NA	0.2857	TEC1	0	intronic	3.28	4	4	5	ALL
2	rs60997612	7	97853489	7:97853489:C:T	C	T	0.2991	rs7015	NA	0.2857	TEC1	0	intronic	0.72	5	4	5	ALL
2	rs72291746	7	97853845	7:97853845:A:G	A	G	0.2991	rs7015	NA	0.2857	TEC1	0	intronic	1.77	NA	4	4	ALL
2	rs72291745	7	97854393	7:97854393:A:G	G	A	0.2885	rs7015	NA	0.3042	TEC1	0	exonic	8.06	5	4	4	ALL
2	rs114286157	7	97858973	7:97858973:TTC	TC	T	0.2967	rs7015	NA	0.3159	TEC1	0	intronic	2.73	NA	2	4	ALL
2	rs7476630	7	97859287	7:97859287:G:T	T	G	0.2967	rs7015	NA	0.3159	TEC1	0	intronic	0.56	NA	4	4	ALL
2	rs72276001	7	97860182	7:97860182:A:G	G	A	0.2967	rs7015	NA	0.3159	TEC1	0	intronic	1.14	NA	4	4	ALL
2	rs72291750	7	97863145	7:97863145:A:G	G	A	0.2987	rs7015	NA	0.3110	TEC1	0	exonic	0.08	NA	4	4	ALL
2	rs76754817	7	97864792	7:97864792:A:G	G	A	0.2256	rs7015	NA	0.4350	TEC1	0	intronic	0.16	7	4	5	ALL
2	rs80292432	7	97864868	7:97864868:A:G	G	A	0.2119	rs7015	NA	0.4898	TEC1	0	intronic	0.43	7	4	5	ALL
2	rs7777561	7	97867157	7:97867157:A:G	A	G	0.2137	rs7015	NA	0.4710	TEC1	0	intronic	4.87	5	4	4	ALL
2	rs77968419	7	97866480	7:97866480:A:G	G	A	0.2053	rs7015	NA	0.5164	TEC1	0	intronic	0.53	5	4	4	ALL
2	rs71693479	7	97866033	7:97866033:C:T	C	T	0.2057	rs7015	NA	0.4836	TEC1	0	intronic	3.83	4	4	4	ALL
2	rs75070716	7	97871985	7:97871985:A:G	A	G	0.2173	rs7015	NA	0.4773	TEC1	0	intronic	5.55	5	4	5	ALL
2	rs58989054	7	97876704	7:97876704:C:T	C	T	0.2250	rs7015	NA	0.4074	TEC1	0	intronic	6.66	7	4	5	ALL
2	rs72145133	7	97877962	7:97877962:G:GA	GA	G	0.2278	rs7015	NA	0.4516	TEC1	0	intronic	0.72	NA	5	5	ALL
2	rs80195155	7	97878153	7:97878153:A:G	G	A	0.2103	rs7015	NA	0.5112	TEC1	0	intronic	2.40	6	2	5	ALL
2	rs111962131	7	97882561	7:97882561:A:G	G	A	0.1917	rs7015	NA	0.6001	TEC1	0	upstream	5.30	1	7	7	ALL
2	7:97883069:G:T	7	97883069	7:97883069:G:T	G	T	0.4906	rs7015	NA	0.2002	BR3	0	NA	1.79	NA	2	5	ALL
2	7:97883069:A:G	7	97883069	7:97883069:A:G	G	A	0.4383	rs7015	NA	0.2002	BR3	0	intergenic	2.41	6	2	5	ALL
2	rs57008444[7]	7	97888924	7:97888924:TTG:TTGTG	TTG	TTGTG	0.3636	rs7015	NA	0.2250	BR3	0	intergenic	NA	NA	4	5	ALL
2	rs142861252	7	97888924	7:97888924:T:TTGTG	T	TTGTG	0.0925	rs7015	NA	0.2250	BR3	0	intergenic	3.22	NA	5	5	ALL
2	rs54657083	7	97888924	7:97888924:T:TTG	TTG	T	0.2107	rs7015	NA	0.2250	BR3	0	intergenic	3.09	NA	4	5	ALL
2	7:97890297:A:C	7	97890297	7:97890297:A:C	C	A	0.4419	rs7015	NA	0.2047	BR3	0	intergenic	0.52	5	4	5	ALL
2	7:97890297:C:T	7	97890297	7:97890297:C:T	C	T	0.4906	rs7015	NA	0.2047	BR3	0	NA	0.38	NA	4	5	ALL
2	rs74890108	7	97894397	7:97894397:A:G	G	A	0.1915	rs7015	NA	0.6086	BR3	0	intergenic	5.97	4	5	5	ALL
2	rs76976120	7	97894172	7:97894172:C:G	C	G	0.1915	rs7015	NA	0.6086	BR3	0	intergenic	2.33	7	4	5	ALL
2	rs73700840	7	97894254	7:97894254:C:T	C	T	0.1913	rs7015	NA	0.6136	BR3	0	intergenic	0.43	4	4	5	ALL
2	rs56694862	7	97897120	7:97897120:G:T	G	T	0.1913	rs7015	NA	0.6094	BR3	0	intergenic	1.98	7	4	5	ALL
2	rs56997713	7	97897780	7:97897780:A:G	G	A	0.1911	rs7015	NA	0.6081	BR3	0	intergenic	2.18	6	5	15	ALL
2	rs52825417[8]	7	97897827	7:97897827:T:TAA	T	TAA	0.3387	rs7015	NA	0.4904	BR3	0	intergenic	0.41	NA	5	15	ALL
2	rs12704861	7	97899777	7:97899777:C:T	C	T	0.4447	rs7015	NA	0.2070	BR3	0	intergenic	0.93	5	5	15	ALL
2	rs73873309	7	97900826	7:97900826:A:T	A	T	0.4447	rs7015	NA	0.2070	BR3	0	intergenic	0.63	5	5	15	ALL
2	rs12666622	7	97901565	7:97901565:C:T	C	T	0.4447	rs7015	NA	0.2070	BR3	0	intergenic	0.32	4	5	15	ALL
2	rs56477839	7	97901622	7:97901622:C:T	C	T	0.1919	rs7015	NA	0.6084	BR3	0	intergenic	2.63	5	5	15	ALL
2	rs10249366	7	97901860	7:97901860:C:G	C	G	0.4447	rs7015	NA	0.2070	BR3	0	intergenic	5.76	7	5	15	ALL
2	rs75006350	7	97902142	7:97902142:C:T	C	T	0.1879	rs7015	NA	0.6226	BR3	0	intergenic	2.65	7	5	15	ALL
2	rs199773685	7	97904898	7:97904898:A:ATT	ATT	A	0.4904	rs7015	NA	0.2011	BR3	0	intergenic	1.58	NA	5	15	ALL
2	rs574686505	7	97904898	7:97904898:A:ATATT	ATTT	AT	0.4417	rs7015	NA	0.2011	BR3	0	NA	NA	5	15	15	ALL
2	rs4506131	7	97913196	7:97913196:C:G	C	G	0.4183	rs7015	NA	0.3877	BR3	0	intronic	8.20	5	1	2	ALL
2	rs6950023	7	97915635	7:97915635:G:T	G	T	0.2352	rs7015	NA	0.9851	BR3	0	intronic	1.47	4	7	7	ALL
2	rs6967728	7	97915637	7:97915637:A:G	G	A	0.2352	rs7015	NA	0.9851	BR3	0	intronic	6.22	4	1	7	ALL
2	rs70115	7	97920623	7:97920623:A:G	A	G	0.2342	rs7015	1.0000000049	1.0000	BR3	0	UTR3	21.40	3a	4	4	ALL
2	rs6975519	7	97925797	7:97925797:C:T	C	T	0.1460	rs7015	NA	0.4086	BR3:BAIAP2L	0	intronic	0.26	6	4	5	ALL
2	rs13232861	7	97926091	7:97926091:A:G	G	A	0.2554	rs7015	NA	0.2069	BR3:BAIAP2L	0	intronic	0.54	7	4	5	ALL
2	rs11290747	7	97926573	7:97926573:T:TC	T	TC	0.2500	rs7015	NA	0.8930	BR3:BAIAP2L	0	intronic	0.43	NA	4	5	ALL
2	rs10450916	7	97928561	7:97928561:C:T	T	C	0.4443	rs7015	NA	0.2501	BR3:BAIAP2L	0	exonic	12.56	6	2	5	ALL
2	rs559504076	7	97934633	7:97934633:TAA:TAAA	TAA	TAAA	0.4986	rs7015	NA	0.2453	BR3:BAIAP2L	0	intronic	NA	NA	4	5	ALL
2	rs559432104	7	97934633	7:97934633:T:TAAA	TAAA	T	0.0266	rs7015	NA	0.2453	BR3:BAIAP2L	0	intronic	2.49	NA	4	5	ALL
2	rs74729431	7	97937981	7:97937981:A:T	A	T	0.4453	rs7015	NA	0.2162	BAIAP2L1:RP1	0	intronic	0.38	6	4	4	ALL
2	rs6951286	7	97939222	7:97939222:C:T	C	T	0.4060	rs7015	NA	0.3689	BAIAP2L1:RP1	0	ncRNA_exonic	0.01	7	4	5	ALL
2	rs2906184	7	97939987	7:97939987:A:T	T	A	0.2322	rs7015	NA	0.8224	BAIAP2L1	0	intronic	3.43	6	4	5	ALL
2	rs75139799	7	97940814	7:97940814:C:T	T	C	0.0990	rs7015	NA	0.3283	BAIAP2L1	0	intronic	0.01	6	4	5	ALL
2	rs11270106	7	97942436	7:97942436:CCCACCCCCACCCCCCCAACT	C	CCCACCCCCAC	0.2817	rs7015	NA	0.6612	BAIAP2L1	0	intronic	11.34	NA	4	5	ALL
2	rs34785619	7	97946299	7:97946299:C:CT	CT	C	0.2436	rs7015	NA	0.8331	BAIAP2L1	0	intronic	1.15	NA	4	5	ALL
2	rs955134283	7	97948257	7:97948257:C:CA	CA	C	0.2089	rs7015	NA	0.4436	BAIAP2L1	0	intronic	1.79	NA	4	5	ALL
2	rs1635609	7	97948814	7:97948814:C:T	T	C	0.2414	rs7015	NA	0.8391	BAIAP2L1	0	intronic	1.42	7	4	5	ALL
2	rs1688807	7	97951321	7:97951321:C:T	T	C	0.2420	rs7015	NA	0.8365	BAIAP2L1	0	intronic	3.03	6	4	5	ALL
2	rs76222115	7	97956184	7:97956184:C:CA	CA	C	0.2099	rs7015	NA	0.8797	BAIAP2L1	0	intronic	0.01	NA	5	15	ALL
2	rs539330515	7	97960777	7:97960777:C:CT	CT	C	0.4129	rs7015	NA	0.2637	BAIAP2L1	0	intronic	1.70	NA	2	15	ALL
2	rs10953258	7	97972639	7:97972639:A:G	G	A	0.4018	rs7015	NA	0.3779	BAIAP2L1	0	intronic	1.68	5	5	15	ALL
2	rs11688656	7	97974851	7:97974851:A:G	A	G	0.2446	rs7015	NA	0.8280	BAIAP2L1	0	intronic	2.31	NA	5	15	ALL
2	rs11275833[7]	7	97977268	7:97977268:A:G	G	A	0.1983	rs7015	NA	0.7400	BAIAP2L1	0	intronic	1.84	7	5	15	ALL
2	rs4727392	7	97977670	7:97977670:A:G	G	A	0.4014	rs7015	NA	0.3804	BAIAP2L1	0	intronic	3.20	6	5	15	ALL
2	rs77032872	7	97984321	7:97984321:T:TG	TG	T	0.1997	rs7015	NA	0.7343	BAIAP2L1	0	intronic	0.83	NA	2	15	ALL
2	rs12754698	7	97984430	7:97984430:C:T	C	T	0.2446	rs7015	NA	0.8240	BAIAP2L1	0	intronic	3.53	7	5	15	ALL
2	rs3779196	7	97990106	7:97990106:C:T	C	T	0.2446	rs7015	NA	0.8288	BAIAP2L1	0	intronic	9.77	NA	2	15	ALL
2	rs6965424	7	97991125	7:97991125:A:G	G	A	0.2446	rs7015	NA	0.8288	BAIAP2L1	0	intronic	2.54	6	4	15	ALL
2	rs3779195	7	97993362	7:97993362:A:T	T	A	0.1991	rs7015	NA	0.7355	BAIAP2L1	0	intronic	4.78	6	5	15	ALL
2	rs76325400	7	97999870	7:97999870:C:G	G	C	0.0831	rs7015	NA	0.3245	BAIAP2L1	0	intronic	1.84	6	5	15	ALL
2	rs201576562	7	98001109	7:98001109:C:CT	CT	C	0.1941	rs7015	NA	0.2192	BAIAP2L1	0	intronic	1.27	NA	1	15	ALL
2	rs34330041	7	98002045	7:98002045:G:GT	G	GT	0.3884	rs7015	NA	0.3735	BAIAP2L1	0	intronic	1.18	NA	4	15	ALL
2	rs12107717	7	98005358	7:98005358:C:T	T	C	0.4229	rs7015	NA	0.2087	BAIAP2L1	0	intronic	5.23	6	5	15	ALL
2	rs4268041	7	98005538	7:98005538:A:G	G	A	0.2376	rs7015	NA	0.7601	BAIAP2L1	0	intronic	3.93	6	5	15	ALL
2	rs6465679	7	98005617	7:98005617:A:T	T	A	0.4245	rs7015	NA	0.3806	BAIAP2L1	0	intronic	0.48	5	5	15	ALL
2	rs5886063	7	98007541	7:98007541:C:CAT	CAT	C	0.2340	rs7015	NA	0.8337	BAIAP2L1	0	intronic	1.35	NA	5	15	ALL
2	rs35903783	7	98012040	7:98012040:G:GT	GT	G	0.3836	rs7015	NA	0.3983	BAIAP2L1	0	intronic	0.15	NA	4	15	ALL
2	rs10953259	7	98013107	7:98013107:A:C	C	A	0.2324	rs7015	NA	0.8396	BAIAP2L1	0	intronic	2.94	6	4	15	ALL
2	rs1635612	7	98015047	7:98015047:T:G	T	G	0.2748	rs7015	NA	0.6640	BAIAP2L1	0	intronic	0.32	5	5	15	ALL
2	rs9649213	7	98021211	7:98021211:A:G	A	G	0.4970	rs7015	NA	0.2092	BAIAP2L1	0	intronic	0.43	7	4	15	ALL
2	rs531117461	7	98021479	7:98021479:T:TA	T	TA	0.4752	rs7015	NA	0.3012	BAIAP2L1	0	intronic	0.24	NA	4	15	ALL
2	rs59854442	7	98022388	7:98022388:C:T	C	T	0.2606	rs7015	NA	0.6379	BAIAP2L1	0	intronic	0.75	6	5	15	ALL
2	rs113310668	7	98022553	7:98022553:C:T	C	T	0.2352	rs7015	NA	0.8398	BAIAP2L1	0	intronic	0.21	7	5	15	ALL
2	rs139654368	7	98023554	7:98023554:AAAAACAAACAAACAAAC	A	ACAAACAAACA	0.4968	rs7015	NA	0.3302	BAIAP2L1	0	intronic	0.48	NA	5	15	ALL
2	rs80244318	7	98025543	7:98025543:A:G	G	A	0.0843	rs7015	NA	0.2387	BAIAP2L1	0	intronic	0.38	7	4	15	ALL
2	rs10953260	7	98033460	7:98033460:A:G	G	A	0.2356	rs7015	NA	0.8398	BAIAP2L1	3111	intergenic	1.53	4	5	14	ALL
2	rs75312394	7	98034660	7:98034660:G:T	T	G	0.3187	rs7015	NA	0.5241	BAIAP2L1	4279	intergenic	0.03	6	5	15	ALL
2	rs73775841[8]	7	98034664	7:98034664:T:TATCTG	ATCTG	T	0.1845	rs7015	NA	0.5887	BAIAP2L1	4283	intergenic	0.03	5	5	15	ALL
2	rs369336243	7	98034668	7:98034668:G:T	T	G	0.1845	rs7015	NA	0.5887	BAIAP2L1	4287	intergenic	0.03	6	5	15	ALL
2	rs6943025	7	98037557	7:98037557:A:T	A	T	0.2125	rs7015	NA	0.5457	BAIAP2L1	7176	intergenic	1.02	5	5	15	ALL
2	rs1688605	7	98038846	7:98038846:A:G	A	G	0.2344	rs7015	NA	0.3457	BAIAP2L1	9465	intergenic	1.27	6	5	15	ALL
3	rs73378106	10	12977854	10:12977854:A:C	C	A	0.0132	rs116562538	NA	0.2735	PTPRE	0	intronic	8.31	4	1	7	ALL
3	rs111833969	10	12978234	10:12978234:C:T	C	T	0.0094	rs116562538	NA	0.2141	PTPRE	0	intronic	0.01	7	4	15	ALL
3	rs11818165	10	12979064	10:12979064:A:CT	T	C	0.0104	rs116562538	NA	0.2377	PTPRE	0	intronic	3.96	5	4	15	ALL
3	rs116033357	10	12979167	10:12979167:A:T	A	T	0.0098	rs116562538	NA	0.2534	PTPRE	0	intronic	0.04	6	5	15	ALL
3	rs115075866	10	12979354	10:12979354:A:C	A	C	0.0098	rs116562538	NA	0.2534	PTPRE	0	intronic	0.73	5	5	15	ALL
3	rs116516150	10	12979469	10:12979469:C:T	C	T	0.0098	rs116562538	NA	0.2534	PTPRE	0	intronic	0.01	5	5	15	ALL
3	rs115071510	10	12979556	10:12979556:A:G	A	G	0.0098	rs116562538	NA	0.2534	PTPRE	0	intronic	3.18	5	5	15	ALL
3	rs115354595	10	12980794	10:12980794:I:AG	G	A	0.0334	rs116562538	NA	0.2494	PTPRE	0	intronic	0.63	5	5	15	ALL
3	rs141381883	10	12980855	10:12980855:I:ATCTG	ATCTG	A	0.0148	rs116562538	NA	0.2168	PTPRE	0	intronic	4.78	NA	1	7	ALL
3	rs7077701	10	12980968	10:12980968:ATCTG:CTCTG	ATCTG	CTCTG	0.0403	rs116562538	NA	0.2168	PTPRE	0	NA	NA	NA	1	7	ALL
3	rs4290142	10	12981265	10:12981265:C:T	C	T	0.0395	rs116562538	NA	0.2313	PTPRE	0	intronic	12.30	7	5	15	ALL
3	rs9733299	10	12981304	10:12981304:A:G	G	A	0.0353	rs116562538	NA	0.2583	PTPRE	0	intronic	1.60	5	5	15	ALL
3	rs114093176	10	12981391	10:12981391:T:C	C	T	0.0349	rs116562538	NA	0.2613	PTPRE	0	intronic	0.92	7	5	15	ALL
3	rs115221107	10	12981991	10:12981991:A:G	G	A	0.0405	rs116562538	NA	0.2460	PTPRE	0	intronic	0.15	4	5	15	ALL
3	rs116251518	10	12982073	10:12982073:C:T	T	C	0.0194	rs116562538	NA	0.5118	PTPRE	0	intronic	0.26	7	4	15	ALL
3	rs11818614	10	12982657	10:12982657:A:G	G	A	0.0357	rs116562538	NA	0.2614	PTPRE	0	intronic	0.20	5	5	15	ALL
3	rs11815719	10	12982833	10:12982833:C:T	C	T	0.0407	rs116562538	NA	0.2367	PTPRE	0	intronic	7.48	5	5	15	ALL
3	rs116335520	10	12982163	10:12982163:A:G	A	G	0.0152	rs116562538	NA	0.5929	PTPRE	0	intronic	1.65	6	5	15	ALL
3	rs114346735	10	12982793	10:12982793:C:T	C	T	0.0359	rs116562538	NA	0.2720	PTPRE	0	intronic	1.72	5	4	15	ALL
3	rs130263348	10	12983890	10:12983890:G:GT	G	GT	0.0148	rs116562538	NA	0.6002	PTPRE	0	intronic	1.17	NA	4	15	ALL
3	rs13328740	10	12983180	10:12983180:I:AG	A	G	0.0314	rs116562538	NA	0.3726	PTPRE	0	intronic	1.14	7	5	15	ALL
3	rs114070054	10	12983376	10:12983376:C:T	C	T	0.0194	rs116562538	NA	0.6086	PTPRE	0	intronic	0.17	4	5	15	ALL
3	rs142793837	10	12983062	10:12983062:G:GA	GA	G	0.0176	rs116562538	NA	0.7360	PTPRE	0	intronic	0.06	NA	4	15	ALL
3	rs115148459	10	12984015	10:12984015:G:T	G	T	0.0176	rs116562538	NA	0.3983	PTPRE	0	intronic	0.01	4	4	15	ALL
3	rs114468842	10	12984316	10:12984316:A:G	G	A	0.0176	rs116562538	NA	0.8032	PTPRE	0	intronic	1.62	5	2	15	ALL
3	rs116562538	10	12984925	10:12984925:A:C	C	A	0.0144	rs116562538	0.0000000067	1.0000	PTPRE	0	intronic	6.22	5	2	7	ALL
3	rs114805563	10	12984925	10:12984925:A:C	C	A	0.0178	rs116562538	NA	0.7944	PTPRE	0	intronic	0.83	4	3	5	ALL
3	rs149669680	10	12985621	10:12985621:C:T	C	T	0.0182	rs116562538	NA	0.7068	PTPRE	0	intronic	0.05	5	4	5	ALL
3	rs61732920	10	12985785	10:12985785:G:A	G	A	0.0164	rs116562538	NA	0.8565	PTPRE	0	intronic	0.09	5	5	15	ALL
3	rs113730842	10	12986162	10:12986162:A:G	A	G	0.0192	rs116562538	NA	0.4472	PTPRE	0	intronic	1.41	4	1	7	ALL
3	rs112978015	10	12986180	10:12986180:I:A	G	A	0.0204	rs116562538	NA	0.4472	PTPRE	0	intronic	8.84	4	7	7	ALL
3	rs11819356	10	12986251	10:12986251:C:T	C	T	0.0144	rs116562538	NA	0.5268	PTPRE	0	intronic	2.72	2b	2	7	ALL
4	rs34566763	11	3948455	11:3948455:T:G	T	G	0.0076	rs35287906	NA	0.4211	STIM1	0	intronic	5.39	3a	5	15	ALL
4	rs12801358	11	3963502	11:3963502:A:G	G	A	0.0076	rs35287906	NA	0.4211	STIM1	0	intronic	5.32	4	2	7	ALL
4	rs111235854	11	3988117	11:3988117:G:T	G	T	0.0016	rs35287906	NA	0.5277	STIM1	0	intronic	2.18	5	5	15	ALL
4	rs35573107	11	4012916	11:4012916:C:T	C	T	0.0076	rs35287906	NA	0.4729	STIM1	0	intronic	0.49	7	5	15	ALL
4	rs35570285	11	4030606	11:4030606:A:G	G	A	0.0076	rs35287906	NA	0.4612	STIM1	0	intronic	1.45	4	1	15	ALL
4	rs35287906	11	4041010	11:4041010:C:T	T	C	0.0034	rs35287906	X0000000007	1.0000	STIM1	0	intronic	2.51	5	5	15	ALL
4	rs143465062	11	4055391	11:4055391:A:G	G	A	0.0080	rs35287906	NA	0.4500	STIM1	0	intronic	0.89	7	5	15	ALL
4	rs201529537	11	4057449	11:4057449:G:GTA	G	GTA	0.0124	rs35287906	NA	0.2930	STIM1	0	intronic	7.46	NA	5	15	ALL
4	rs34241665	11	4072133	11:4072133:G:T	T	G	0.0034	rs35287906	NA	1.0000	STIM1	0	intronic	4.10	7	4	5	ALL
4	rs34900519	11	4094869	11:4094869:A:G	G	A	0.0056	rs35287906	NA	0.6045	STIM1	0	intronic	2.01	6	4	5	ALL
4	rs76519659	11	4100600	11:4100600:A:G	A	G	0.0076	rs35287906	NA	0.4689	STIM1	0	intronic	0.78	7	4	5	ALL
4	rs35094565	11	4170405	11:4170405:A:AT	A	AT	0.0034	rs35287906	NA	1.0000	OR5B17P	1879	intergenic	7.89	NA	5	15	ALL
4	rs71488439	11	4227801	11:4227801:A:G	A	G	0.0014	rs35287906	NA	0.4101	RP11-23F23.3	3782	intergenic	0.08	6	5	15	ALL

Supplementary Table 1. Functional annotation of novel indels SNVs and SNVs (V.2.0.0.0).

Variant is the variant in the form of chr:pos:ref:alt. Gene is the gene where the variant is located. Transcript is the transcript of the canonical isoform (NP-ID) for the gene if it is the known gene or the NM-ID transcript if it is a novel gene. Exon is the exon where the variant is located. c.DNA is the gene DNA change in the transcript. p.Protein is the protein change in the transcript. Type is the type of the variant. Novel indicates if the variant is novel or known. Dbsnp is the dbSNP ID if the variant is known. Frequency is the frequency of the variant in the Kaplan, the Intra-Kaplan and the Inter-Kaplan cohorts. Zygosity is the zygosity of the variant in the Kaplan, the Intra-Kaplan and the Inter-Kaplan cohorts. Genotype is the genotype of the variant in the Kaplan, the Intra-Kaplan and the Inter-Kaplan cohorts. Genotype_Quality is the genotype quality of the variant in the Kaplan, the Intra-Kaplan and the Inter-Kaplan cohorts. Pathogenicity is the pathogenicity of the variant in the Kaplan, the Intra-Kaplan and the Inter-Kaplan cohorts. Pathogenicity_Source is the source of the variant pathogenicity.

Variant	Gene	Transcript	Exon	c.DNA	p.Protein	Type	Novel	Dbsnp	Frequency	Zygosity	Genotype	Genotype_Quality	Pathogenicity	Pathogenicity_Source
chr9:96448:C:CCCC

Table 1: Summary of novel loci.

Locus	rsID[1]	Chr	Pos (hg19)	uniqID	gwasP[2]	MAF[3]	nGWAS SNPs[4]	nFUMA SNPs[5]	func[6]	genes[7]	RefPop-Trait-Exp[8]	Stage[9]	
1	rs139845754	1	180,258,167	1:180258167:A:AG	3.40E-08	0.052	1	98	68 intronic; 18 intergenic; 2 upstream; 1 downstream; 3 ncRNA intronic; 2 ncRNA exonic; 4 NA	LHX4;RP5-11 80C10.2; **ACBD6**; Y_RNA; AL358354.1; OVAAL	AFR-TG-SC	D+R	
2	rs75218113	2	103,239,356	2:103239356:C:G	1.72E-08	0.031	2	11	6 intronic; 3 intergenic; 1 ncRNA exonic; 1 NA	SLC9A4; **SLC9A2**	AFR-HDL-GC	D	
3	rs79367750	3	9,096,107	3:9096107:C:CT	3.99E-08	0.096	1	18	17 intronic; 1 NA	**SRGAP3**	ASN-LDL-SC	D+R	
4	rs147731578	3	70,605,665	3:70605665:ID	1.39E-09	0.018	1	N/A	1 intergenic	MITF; FOXP1	EUR-HDL-SC	D	Missing in FUMA - InDel in 1-SNP Locus
5	rs74620279	4	179,620,483	4:179620483:C:T	4.67E-08	0.025	1	49	47 intergenic; 1 downstream; 1 NA	SNORD65	AFR-LDL-SC	D	
6	rs77249395	4	181,181,245	4:181181245:G:T	4.74E-08	0.035	1	2	2 intergenic	RP11-751A18.1	TA-HDL-GC	D+R	
7	rs11132093	4	182,664,815	4:182664815:A:G	5.04E-09	0.061	1	53	53 intergenic	RP11-540E16.2; TENM3	ASN-LDL-GC	D	Missing in FUMA - triallelic 1-SNP locus
8	rs190502162	5	37,160,808	5:37160808:C:G	3.35E-08	0.013	1	N/A	1 intronic	**C5orf42**	EUR-LDL-SC	D+R	
9	rs192718305	5	82,514,005	5:82514005:G:T	2.05E-08	0.025	1	51	49 intronic; 1 intergenic; 1 3'UTR	RP11-343L5.2; TMEM167A; **XRCC4**	AFR-LDL-SC	D+R	
10	rs147892694	6	76,693,587	6:76693587:G:GC	8.61E-09	0.034	1	56	44 intronic; 5 intergenic; 7 NA	**IMPG1**; RNU6-248P	AFR-HDL-SC	D	
11	rs7015	7	97,920,623	7:97920623:A:G	4.56E-08	0.203	1	95	61 intronic; 24 intergenic; 3 exonic; 1 upstream; 1 ncRNA exonic; 2 3'UTR; 3 NA	TECPR1; **BRI3**; BAIAP2L1; RP4-607J23.2	TA-LDL-SC	D+R	BRI3 highly expressed in liver (2.27E-24) & adipose (2.17E-12)
12	rs77655002	8	30,437,023	8:30437023:C:T	5.52E-09	0.042	1	106	89 intronic; 5 intergenic; 2 exonic; 1 upstream; 6 downstream; 1 5'UTR; 1 3'UTR; 1 NA	RBPMS; **GTF2E2**; GSR; TEX15; RP5-1009N12.1; RP11-293D9.2	AFR-LDL-GC	D	RBPMS-AS1 expressed in SAT (5.71E-28)
13	rs144190766	10	108,927,076	10:108927076:C:G	2.52E-08	0.030	1	188	14 intronic; 173 intergenic; 1 NA	SORCS1; RNA5SP326	AFR-HDL-SC	D+R	
14	rs116562538	10	129,845,329	10:129845329:A:C	6.79E-09	0.028	1	32	41 intronic; 1 NA	**PTPRE**	TA-LDL-SC	D+R	
15	rs35287906	11	4,041,010	11:4041010:C:T	7.71E-11	0.014	2	13	11 intronic; 2 intergenic	**STIM1**; OR55B1P; RP11-23F23.3	TA-LDL-GC	D+R	RRM1 expressed in SAT (1.36E-44).
16	rs11230661	11	55,451,313	11:55451313:A:G	3.15E-10	0.185	15	925	3 intronic; 746 intergenic; 22 exonic; 54 upstream; 48 downstream; 1 upstream:downstream; 16 ncRNA intronic; 19 ncRNA exonic; 16 NA	TRIM48; RP11-674C21.9; TRIM51HP; TRIM51; FAM8A2P; RPL5P29; **OR4P1P;** *87 Olfactory Receptor Genes*	EUR-HDL-GC EUR-HDL-SC	D+R	
17	rs148063115	12	14,221,135	12:14221135:T:TTTTCTCCCTCTG	4.86E-08	0.026	1	37	4 intronic; 33 intergenic	GRIN2B; RP11-72J9.1; RP11-298E10.1; RN7SL676P; GUCY2C	AFR-TG-GC	D	
18	rs190746034	13	101,418,063	13:101418063:G:T	1.26E-08	0.022	1	35	1 intronic; 2 intergenic; 31 ncRNA intronic; 1 ncRNA exonic	RP11-151A6.4; TMTC4; **NALCN-AS1**; ARF4P3	AFR-HDL-GC	D	

Notes: 1. rsID based on dbSNP build 146; 2. GWAS p-value of the significant lead SNP for this locus; 3. Minor allele frequency; 4. Number of SNVs significant in GWAS locus; 5. Number of SNVs in LD (r2>0.2) with lead variant from FUMA; 6. Functional consequence of the SNP on the gene obtained from ANNOVAR (NA: not available); 7. Nearest gene of all SNVs in LD (r2>0.2) with lead variant; if SNV not in FUMA, gene or nearest flanking coding genes noted. Bolded genes relfect intragenic lead SNV. 8. The reference panel used in FUMA to obtain functional annotations-trait-exposure; 9. Analysis stage.

New Locus	Locus	MarkerName	rsID	Alt	CHR	POS	P	Ancestry	MAF	N		Comments
1	1	1:180258167:ID	rs139845754	N/A	1	180,258,167	3.40E-08	AA	0.052	10913		
2	2	2:103239356	rs75218113		2	103,239,356	1.72E-08	AA	0.031	6361		
3	3	3:9096107:ID	rs79367750	N/A	3	9,096,107	3.99E-08	AS	0.096	8855		
4	4	3:70605665:ID	rs147731578	N/A	3	70,605,665	1.39E-09	EA	0.018	15584		Missing in FUMA - InDel in 1-SNP Locus
5	6A	4:179620483	rs74620279		4	179,620,483	4.67E-08	AA	0.025	5669		No R 2°/2 filters
6	6B	4:181181245	rs77249395		4	181,181,245	2.41E-08	TA	0.036	123960		
7	6C	4:182664815	rs11132093		4	182,664,815	5.04E-09	AS	0.061	2809		
8	7	5:37160808	rs190502162		5	37,160,808	3.35E-08	EA	0.013	48467		Missing in FUMA - triallelic 1-SNP locus
9	8	5:82514005	rs192718305		5	82,514,005	2.05E-08	AA	0.025	5664		
10	10	6:76693587:ID	rs147892694	N/A	6	76,693,587	8.61E-09	AA	0.034	8599		
11	11	7:97920623	rs7015		7	97,920,623	4.56E-08	TA	0.203	157929		
12	12	8:30437023	rs77655002		8	30,437,023	5.52E-09	AA	0.042	7166		
13	13	10:108927076	rs144190766		10	108,927,076	2.52E-08	AA	0.03	9677		
14	14	10:129845329	rs116562538		10	129,845,329	6.79E-09	TA	0.028	20939		
15	15	11:4041010	rs35287906		11	4,041,010	7.71E-11	TA	0.014	81020		
16	16	11:55451313	rs11230661		11	55,451,313	3.15E-10	EA	0.185	151320		
17	18	12:14221135:ID	rs148063115	N/A	12	14,221,135	4.86E-08	AA	0.026	6168		
18	19	13:101418063	rs190746034		13	101,418,063	1.26E-08	AA	0.022	6361		

Table S16. Previously reported associations (from GWAS Catalog) of SNPs leading the genome-wide significant loci identified in our BP GWAS taking into account effects of psychosocial factors

Notes: 1. SNP leading the significant loci in our GWIS analyses; 2. rsID of reported SNP in GWAS catalog; 3. SNP functional class; 4. Reported risk/effect allele frequency associated with strongest SNP in controls (if not available among all controls, among the control group with the largest sample size). If the associated locus is a haplotype the haplotype frequency will be extracted; 5. Reported odds ratio or beta-coefficient associated with strongest SNP risk allele.

IndSigSNP[1]	Chr	Pos (hgX)	SNP[2]	PMID	FirstAuthor	Trait	InitialN	ReplicationN	ReportedGene	MappedGene	Context[3]	RiskAF[4]	P	OrBeta[5]
rs75218113	2	103215410	rs79387448	27723756	Wang J	Gut microbiota (bacterial taxa)	1,812 European ancestry individuals	NA	SLC9A2	SLC9A4 - SLC9A2	intergenic_variant	NR	8.00E-11	0.31
rs7015	7	97901622	rs58477839	30595370	Kichaev G	Heel bone mineral density	approximately 446,000 European ancestry individuals	NA		BRI3	intron_variant	NR	1.00E-09	
rs7015	7	97993362	rs3779195	22829776	Coviello AD	Sex hormone-binding globulin levels	21,791 European ancestry individuals	8,175 European ancestry individuals	BAIAP2L1	BAIAP2L1	intron_variant	0.17	3.00E-08	0.028
rs11230661	11	54804993	rs28872512	29617998	Gao XR	Intraocular pressure	115,486 European ancestry individuals	NA	TRIM48	OR4A6P - TRIM48	intergenic_variant	0.8666	7.00E-13	0.145
rs11230661	11	54867854	rs74966104	30595370	Kichaev G	Cardiovascular disease	approximately 459,000 European ancestry individuals	NA		OR4A6P - TRIM48	intergenic_variant	NR	2.00E-08	
rs11230661	11	55085501	rs7936821	29785010	Khawaja AP	Intraocular pressure	139,555 European ancestry individuals	NA	TRIM51HP	AP005597.1 - OR4A11P	intergenic_variant	0.865	3.00E-14	0.143
rs11230661	11	55129640	rs11229252	30578418	Giri A	Systolic blood pressure	365,998 European ancestry individuals, 63,490 African ancestry individuals, 22,802 Hispanic individuals, 4,792 Asian ancestry individuals, 2,695 Native American ancestry individuals	299,024 European ancestry individuals, 17,277 individuals	OR4A15	OR4A16 - OR4A15	intergenic_variant	0.0986	7.00E-09	0.3524
rs11230661	11	55314492	rs76154806	29617998	Gao XR	Intraocular pressure	115,486 European ancestry individuals	NA	OR4A15, OR4C15	OR4C14P - OR4C15	intergenic_variant	0.8709	1.00E-11	0.14
rs11230661	11	55355182	rs74237369	27841878	Hoffmann TJ	Diastolic blood pressure	295,529 European ancestry individuals, 8,231 Latino individuals, 3,058 African American individuals, 2,029 African British individuals, 7,701 East Asian ancestry individuals, 2,735 South Asian ancestry individuals, 1,979 mixed and unknown ancestry individuals	NA	OR4C16, OR4C11	OR4C16 - OR4C11	intergenic_variant		4.00E-06	0.162
rs11230661	11	55355182	rs74237369	27841878	Hoffmann TJ	Systolic blood pressure	295,529 European ancestry individuals, 8,231 Latino individuals, 3,058 African American individuals, 2,029 African British individuals, 7,701 East Asian ancestry individuals, 2,735 South Asian ancestry individuals, 1,979 mixed and unknown ancestry individuals	NA	OR4C16, OR4C11	OR4C16 - OR4C11	intergenic_variant		5.00E-08	0.315
rs11230661	11	55355182	rs74237369	27841878	Hoffmann TJ	Systolic blood pressure	80,792 European ancestry individuals, 8,231 Latino individuals, 3,058 African American individuals, 7,243 East Asian ancestry individuals, 461 South Asian ancestry individuals	NA	OR4C16, OR4C11	OR4C16 - OR4C11	intergenic_variant		1.00E-06	0.445
rs11230661	11	55355182	rs74237369	27841878	Hoffmann TJ	Systolic blood pressure	80,792 European ancestry individuals, 8,231 Latino individuals, 3,058 African American individuals, 7,243 East Asian ancestry individuals, 461 South Asian ancestry individuals	NA	OR4C16, OR4C11	OR4C16 - OR4C11	intergenic_variant		4.00E-07	0.423
rs11230661	11	55415618	rs11230523	29617998	Gao XR	Intraocular pressure	115,486 European ancestry individuals	NA	OR4P4, OR4S2	OR4P4 - OR4S2	intergenic_variant	0.8434	3.00E-13	0.15
rs11230661	11	55753915	rs182996269	30578418	Giri A	Systolic blood pressure	365,998 European ancestry individuals, 63,490 African ancestry individuals, 22,802 Hispanic individuals, 4,792 Asian ancestry individuals, 2,695 Native American ancestry individuals	299,024 European ancestry individuals, 17,277 individuals	OR7E5P	OR7E5P - OR5F1	regulatory_region_variant	0.0877	2.00E-09	0.4017
rs11230661	11	55889895	rs61751933	29617998	Gao XR	Intraocular pressure	115,486 European ancestry individuals	NA	OR8H3	OR8H3	missense_variant	0.8757	2.00E-11	0.14
rs11230661	11	56242261	rs7939886	23049088	Meng W	Myopia (pathological)	187 European ancestry cases, 1064 European ancestry controls		intergenic	AP002512.3 - OR5M2P	intergenic_variant	0.906	2.00E-07	