Title: The power of genetically diverse individuals in genome-wide association studies of blood lipid levels

Authors:
Sarah E Graham[1], Shoa L Clarke[2,3‡], Kuan-Han H Wu[4‡], Stavroula Kanoni[5‡], Greg JM Zajac[6‡], Shweta Ramdas[7‡], Ida Surakka[1], Ioanna Ntalla[8], Sailaja Vedantam[9,10], Thomas W Winkler[11], Adam E Locke[12], Eirini Marouli[5], Mi Yeong Hwang[13], Sohee Han[13], Akira Narita[14], Ananyo Choudhury[15], Amy R Bentley[16], Kenneth Ekoru[16], Anurag Verma[17], Bhavi Trivedi[18], Hilary C Martin[19], Karen A Hunt[18], Qin Hui[20,21], Derek Klarin[22,23,24], Xiang Zhu[25,26,27,28], Gudmar Thorleifsson[29], Anna Helgadottir[29], Daniel F Gudbjartsson[29,30], Hilma Holm[29], Isleifur Olafsson[31], Masato Akiyama[32,33], Saori Sakaue[34,32,35], Chikashi Terao[36], Masahiro Kanai[37,38,39], Wei Zhou[40,41,42], Ben M Brumpton[43,44,45], Humaira Rasheed[43,44], Sanni E Ruotsalainen[46], Aki S Havulinna[46,47], Yogasudha Veturi[48], QiPing Feng[49], Elisabeth A Rosenthal[50], Todd Lingren[51], Jennifer Allen Pacheco[52], Sarah A Pendergrass[53], Jeffrey Haessler[54], Franco Giulianini[55], Yuki Bradford[48], Jason E Miller[48], Archie Campbell[56,57], Kuang Lin[58], Iona Y Millwood[58,59], George Hindy[60], Asif Rasheed[61], Jessica D Faul[62], Wei Zhao[63], David R Weir[62], Constance Turman[64], Hongyan Huang[64], Mariaelisa Graff[65], Anubha Mahajan[66#], Michael R Brown[67], Weihua Zhang[68,69,70], Ketian Yu[71], Ellen M Schmidt[71], Anita Pandit[71], Stefan Gustafsson[72], Xianyong Yin[71], Jian'an Luan[73], Jing-Hua Zhao[73], Fumihiko Matsuda[74], Hye-Mi Jang[13], Kyungheon Yoon[13], Carolina Medina-Gomez[75,76], Achilleas Pitsillides[77], Jouke Jan Hottenga[78,79], Gonneke Willemsen[78,80], Andrew R Wood[81], Yingji Ji[81], Zishan Gao[82,83,84], Simon Haworth[85,86], Ruth E Mitchell[85,87], Jin Fang Chai[88], Mette Aadahl[89], Jie Yao[90], Ani Manichaikul[91], Helen R Warren[92,93], Julia Ramirez[92], Jette Bork-Jensen[94], Line L Kårhus[95], Anuj Goel[96,97], Maria Sabater-Lleal[98,99], Raymond Noordam[100], Carlo Sidore[101], Edoardo Fiorillo[102], Aaron F McDaid[103,104], Pedro Marques-Vidal[105], Matthias Wielscher[106], Stella Trompet[107,108], Naveed Sattar[109], Line T Møllehave[89], Betina H Thuesen[89], Matthias Munz[110], Lingyao Zeng[111,112], Jianfeng Huang[113], Bin Yang[113], Alaitz Poveda[114], Azra Kurbasic[114], Claudia Lamina[115], Lukas Forer[115], Markus Scholz[116,117], Tessel E. Galesloot[118], Jonathan P. Bradfield[119], E Warwick Daw[120], Joseph M Zmuda[121], Jonathan S Mitchell[122], Christian Fuchsberger[122], Henry Christensen[123], Jennifer A Brody[124], Mary F Feitosa[120], Mary K Wojczynski[120], Michael Preuss[125], Massimo Mangino[126,127], Paraskevi Christofidou[126], Niek Verweij[128], Jan W Benjamins[128], Jorgen Engmann[129,130], Rachel L Kember[131], Roderick C Slieker[132,133], Ken Sin Lo[134], Nuno R Zilhao[135], Phuong Le[136], Marcus E Kleber[137,138], Graciela E Delgado[137], Shaofeng Huo[139], Daisuke D Ikeda[140], Hiroyuki Iha[140], Jian Yang[141,142], Jun Liu[143], Hampton L Leonard[144,145], Jonathan Marten[146], Börge Schmidt[147], Marina Arendt[147,148], Laura J Smyth[149], Marisa Cañadas-Garre[149], Chaolong Wang[150,151], Masahiro Nakatochi[152], Andrew Wong[153], Nina Hutri-Kähönen[154,155], Xueling Sim[88], Rui Xia[156], Alicia Huerta-Chagoya[157], Juan Carlos Fernandez-Lopez[158], Valeriya Lyssenko[159,160], Meraj Ahmed[161], Anne U Jackson[6], Marguerite R Irvin[162], Christopher Oldmeadow[163], Han-Na Kim[164], Seungho Ryu[165,166], Paul RHJ Timmers[167,146], Liubov Arbeeva[168], Rajkumar Dorajoo[169], Leslie A Lange[170], Xiaoran Chai[171,172], Gauri Prasad[173,174], Laura Lorés-Motta[175], Marc Pauper[175], Jirong Long[176], Xiaohui Li[90], Elizabeth Theusch[177], Fumihiko Takeuchi[178], Cassandra N Spracklen[179,180], Anu Loukola[46], Sailalitha Bollepalli[46], Sophie C Warner[181,182], Ya Xing Wang[183], Wen B. Wei[184], Teresa Nutile[185], Daniela Ruggiero[185,186], Yun Ju Sung[187], Yi-Jen Hung[188], Shufeng Chen[113], Fangchao Liu[113], Jingyun Yang[189,190], Katherine A Kentistou[167], Mathias Gorski[11,191], Marco Brumat[192], Karina Meidtner[193,194], Lawrence F Bielak[195], Jennifer A Smith[195,196], Prashantha Hebbar[197], Aliki-Eleni Farmaki[198,199], Edith Hofer[200,201], Maoxuan Lin[202], Chao Xue[203], Jifeng Zhang[1], Maria Pina Concas[204], Simona Vaccargiu[205], Peter J van der Most[206], Niina

Pitkänen[207,208], Brian E Cade[209,210], Jiwon Lee[209], Sander W. van der Laan[211], Kumaraswamy Naidu Chitrala[212], Stefan Weiss[213], Martina E Zimmermann[11], Jong Young Lee[214], Hyeok Sun Choi[215], Maria Nethander[216,217], Sandra Freitag-Wolf[218], Lorraine Southam[219,220], Nigel W Rayner[221,222,223,219], Carol A Wang[224], Shih-Yi Lin[225,226,227], Jun-Sing Wang[228,229], Christian Couture[230], Leo-Pekka Lyytikäinen[231,232], Kjell Nikus[233,234], Gabriel Cuellar-Partida[235], Henrik Vestergaard[236], Bertha Hildalgo[237], Olga Giannakopoulou[5], Qiuyin Cai[176], Morgan O Obura[132], Jessica van Setten[238], Xiaoyin Li[239], Karen Schwander[240], Natalie Terzikhan[241], Jae Hun Shin[215], Rebecca D Jackson[242], Alexander P Reiner[243], Lisa Warsinger Martin[244], Zhengming Chen[58,59], Liming Li[245], Heather M Highland[65], Kristin L Young[65], Takahisa Kawaguchi[74], Joachim Thiery[246,117], Joshua C Bis[124], Girish N. Nadkarni[125], Lenore J Launer[247], Huaixing Li[139], Mike A Nalls[144,145], Olli T Raitakari[248,249,250], Sahoko Ichihara[251], Sarah H Wild[252], Christopher P Nelson[181,182], Harry Campbell[167], Susanne Jäger[193,194], Toru Nabika[253], Fahd Al-Mulla[197], Harri Niinikoski[254,255], Peter S Braund[181,182], Ivana Kolcic[256], Peter Kovacs[257], Tota Giardoglou[258], Tomohiro Katsuya[259,260], Konain Fatima Bhatti[5], Dominique de Kleijn[261], Gert J. de Borst[261], Eung Kweon Kim[262], Hieab H.H. Adams[241,263], M. Arfan Ikram[241], Xiaofeng Zhu[239], Folkert W Asselbergs[238], Adriaan O Kraaijeveld[238], Joline WJ Beulens[132,264], Xiao-Ou Shu[176], Loukianos S Rallidis[265], Oluf Pedersen[94], Torben Hansen[94], Paul Mitchell[266], Alex W Hewitt[267,268], Mika Kähönen[269,270], Louis Pérusse[230,271], Claude Bouchard[272], Anke Tönjes[273], Yii-Der Ida Chen[90], Craig E Pennell[224], Trevor A Mori[274], Wolfgang Lieb[275], Andre Franke[276], Claes Ohlsson[216,277], Dan Mellström[216,278], Yoon Shin Cho[215], Hyejin Lee[279], Jian-Min Yuan[280,281], Woon-Puay Koh[282,283], Sang Youl Rhee[284], Jeong-Taek Woo[284], Iris M Heid[11], Klaus J Stark[11], Henry Völzke[285], Georg Homuth[213], Michele K Evans[286], Alan B Zonderman[286], Ozren Polasek[256], Gerard Pasterkamp[211], Imo E Hoefer[211], Susan Redline[209,210], Katja Pahkala[207,208,287], Albertine J Oldehinkel[288], Harold Snieder[206], Ginevra Biino[289], Reinhold Schmidt[200], Helena Schmidt[290], Y Eugene Chen[1], Stefania Bandinelli[291], George Dedoussis[198], Thangavel Alphonse Thanaraj[292], Sharon LR Kardia[195], Norihiro Kato[178], Matthias B Schulze[193,194,293], Giorgia Girotto[192,294], Bettina Jung[295], Carsten A Böger[295,296,297], Peter K Joshi[167], David A Bennett[189,190], Philip L De Jager[298,299], Xiangfeng Lu[113], Vasiliki Mamakou[300,301], Morris Brown[302,93], Mark J Caulfield[92,93], Patricia B Munroe[92,93], Xiuqing Guo[90], Marina Ciullo[185,186], Jost B. Jonas[303,183,304], Nilesh J Samani[181,182], Daniel I. Chasman[55,305], Jaakko Kaprio[46], Päivi Pajukanta[306], Teresa Tusié-Luna[307,308], Carlos A Aguilar-Salinas[309], Linda S Adair[310,311], Sonny Augustin Bechayda[312,313], H. Janaka de Silva[314], Ananda R Wickremasinghe[315], Ronald Krauss[316], Jer-Yuarn Wu[317], Wei Zheng[176], Anneke I den Hollander[175], Dwaipayan Bharadwaj[174,318], Adolfo Correa[319], James G Wilson[320], Lars Lind[321], Chew-Kiat Heng[322], Amanda E Nelson[168,323], Yvonne M Golightly[168,324,325,326], James F Wilson[167,146], Brenda Penninx[327,80], Hyung-Lae Kim[328], John Attia[329,163], Rodney J Scott[329,163], D C Rao[330], Donna K Arnett[331], Mark Walker[332], Heikki A Koistinen[333,334,335], Giriraj R Chandak[161,336], Chittaranjan S Yajnik[337], Josep M Mercader[338,339,340], Teresa Tusie-Luna[341], Carlos Aguilar-Salinas[342], Clicerio Gonzalez Villalpando[343], Lorena Orozco[344], Myriam Fornage[156,345], E Shyong Tai[346,88], Rob M van Dam[88,346], Terho Lehtimäki[231,232], Nish Chaturvedi[153], Mitsuhiro Yokota[347], Jianjun Liu[151], Dermot F Reilly[348], Amy Jayne McKnight[149], Frank Kee[149], Karl-Heinz Jöckel[147], Mark I McCarthy[66,349#], Colin NA Palmer[350], Veronique Vitart[146], Caroline Hayward[146], Eleanor Simonsick[351], Cornelia M van Duijn[143], Fan Lu[352], Jia Qu[352], Haretsugu Hishigaki[140], Xu Lin[353], Winfried März[354,355,137], Esteban J Parra[136], Miguel Cruz[356], Vilmundur Gudnason[135,357], Jean-Claude Tardif[134,358], Guillaume Lettre[134,358], Leen M t Hart[133,359,132], Petra JM Elders[360], Daniel J Rader[361], Scott M Damrauer[362,363], Meena Kumari[364], Mika Kivimaki[130], Pim van der Harst[128], Tim D Spector[126], Ruth J.F. Loos[125,365], Michael A Province[120], Bruce M Psaty[366,367], Ivan Brandslund[123,368], Peter P

Pramstaller[122], Kaare Christensen[369], Samuli Ripatti[46,370,371], Elisabeth Widén[46], Hakon Hakonarson[372,373], Struan F.A. Grant[373,374,375], Lambertus ALM Kiemeney[118], Jacqueline de Graaf[118], Markus Loeffler[116,117], Florian Kronenberg[115], Dongfeng Gu[113,376], Jeanette Erdmann[377], Heribert Schunkert[111,112], Paul W Franks[114], Allan Linneberg[89,378], J. Wouter Jukema[107,379], Amit V Khera[380,381,382,383], Minna Männikkö[384], Marjo-Riitta Jarvelin[106,385,386], Zoltan Kutalik[103,104], Francesco Cucca[387,388], Dennis O Mook-Kanamori[389,390], Ko Willems van Dijk[391,392,393], Hugh Watkins[96,97], David P Strachan[394], Niels Grarup[94], Peter Sever[395], Neil Poulter[396], Jerome I Rotter[90], Thomas M Dantoft[89], Fredrik Karpe[397,398], Matt J Neville[397,398], Nicholas J Timpson[85,87], Ching-Yu Cheng[171,399], Tien-Yin Wong[171,399], Chiea Chuen Khor[151], Charumathi Sabanayagam[171,399], Annette Peters[84,400,401], Christian Gieger[83,84,401], Andrew T Hattersley[402], Nancy L Pedersen[403], Patrik KE Magnusson[403], Dorret I Boomsma[404,79,405], Eco JC de Geus[78,80], L Adrienne Cupples[77,406], Joyce B.J. van Meurs[75,76], Mohsen Ghanbari[76,407], Penny Gordon-Larsen[310,311], Wei Huang[408], Young Jin Kim[13], Yasuharu Tabara[74], Nicholas J Wareham[73], Claudia Langenberg[73], Eleftheria Zeggini[219,220,409], Johanna Kuusisto[410], Markku Laakso[410], Erik Ingelsson[411,412,413,72], Goncalo Abecasis[71,414], John C Chambers[415,68,69,416], Jaspal S Kooner[69,70,417,418], Paul S de Vries[67], Alanna C Morrison[67], Kari E. North[65], Martha Davigulus[419], Peter Kraft[64,420], Nicholas G Martin[421], John B Whitfield[421], Shahid Abbas[422], Danish Saleheen[61,423,424], Robin G Walters[425,426,427], Michael V Holmes[58,59,428], Corri Black[429], Blair H Smith[430], Anne E Justice[431], Aris Baras[432], Julie E Buring[55,433], Paul M Ridker[55,433], Daniel I Chasman[55,433], Charles Kooperberg[54], Wei-Qi Wei[434], Gail P Jarvik[435], Bahram Namjou[436], M. Geoffrey Hayes[437,438,439], Marylyn D Ritchie[48], Pekka Jousilahti[47], Veikko Salomaa[47], Kristian Hveem[43,440,441], Bjørn Olav Åsvold[43,440,442], Michiaki Kubo[443], Yoichiro Kamatani[32,444], Yukinori Okada[34,32,445,446], Yoshinori Murakami[447], Unnur Thorsteinsdottir[29,448], Kari Stefansson[29,448], Yuk-Lam Ho[449], Julie A Lynch[450,451], Daniel Rader[452], Phil S Tsao[2,3,453], Kyong-Mi Chang[454,452], Kelly Cho[449,455], Christopher J O'Donnell[449,455], John M Gaziano[449,455], Peter Wilson[456,457], Charles N Rotimi[16], Scott Hazelhurst[15,458], Michèle Ramsay[15,459], Richard C Trembath[460], David A van Heel[18], Gen Tamiya[14], Masayuki Yamamoto[14], Bong-Jo Kim[13], Karen L Mohlke[179], Joel N Hirschhorn[9,10,461], Sekar Kathiresan[462,381,383], Michael Boehnke[6], Pradeep Natarajan[463,464,465,466], Gina M Peloso[467], Christopher D Brown[7†], Andrew P Morris[468†], Themistocles L Assimes[2,3,453†*], Panos Deloukas[5,469†], Yan V Sun[20,21†], Cristen Willer[1,470,471†*]

[1]Department of Internal Medicine, Division of Cardiology, University of Michigan, Ann Arbor, MI 48109, USA, [2]VA Palo Alto Health Care system, Palo Alto, California, USA, [3]Department of Medicine, Division of Cardiovascular Medicine, Stanford University School of Medicine, Stanford, California, USA, [4]Department of Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, MI, USA, [5]William Harvey Research Institute, Barts and the London School of Medicine and Dentistry, Queen Mary University of London, Charterhouse square, EC1M 6BQ, UK, [6]Department of Biostatistics and Center for Statistics Genetics, University of Michigan, Ann Arbor, MI 48109, [7]Department of Genetics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, [8]Clinical Pharmacology, William Harvey Research Institute, Barts and The London School of Medicine and Dentistry, Queen Mary University of London, London, EC1M 6BQ UK, [9]Endocrinology, Boston Children's Hospital, Boston 02115 MA,USA, [10]Medical and Population Genetics, Broad Institute, 75 Ames street, Cambridge, MA 02142,USA, [11]Department of Genetic Epidemiology, University of Regensburg, Regensburg, Germany, [12]McDonnell Genome Institute and Department of Medicine, Washington University, St. Louis, MO, 63108, [13]Division of Genome Research, Center for

Genome Science, National Institute of Health, Chungcheongbuk-do, South Korea, [14]Tohoku Medical Megabank Organization, Tohoku University, Sendai 980-8573, Japan, [15]Sydney Brenner Institute for Molecular Bioscience, Faculty of Health Sciences, University of the Witwatersrand, Johannesburg, South Africa., [16]Center for Research on Genomics and Global Health, National Human Genome Research Institute, National Institutes of Health, 12 South Drive, Room 4047, Bethesda, MD, 20892, USA, [17]Department of Genetics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104, USA., [18]Blizard Institute, Barts and the London School of Medicine and Dentistry, Queen Mary University of London, London, UK, [19]Wellcome Sanger Institute, Hinxton, UK, [20]Department of Epidemiology, Emory University Rollins School of Public Health, Atlanta, Georgia, USA, [21]Atlanta VA Health Care System, Decatur, Georgia, USA, [22]Malcolm Randall VA Medical Center, Gainesville, FL, [23]Division of Vascular Surgery and Endovascular Therapy, University of Florida College of Medicine, Gainesville, FL, [24]Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, Massachusetts, USA, [25]Department of Statistics, The Pennsylvania State University, University Park, PA, USA, [26]Huck Institutes of the Life Sciences, The Pennsylvania State University, University Park, PA, USA, [27]VA Palo Alto Health Care System, Palo Alto, CA, USA, [28]Department of Statistics, Stanford University, Stanford, CA, USA, [29]deCODE genetics/Amgen, Inc. Sturlugata 8, Reykjavik, 102, Iceland, [30]School of Engineering and Natural Sciences, University of Iceland, Sæmundargötu 2, Reykjavik, 102, Iceland, [31]Department of Clinical Biochemistry, Landspitali - National University Hospital of Iceland, Hringbraut, Reykjavik, 101, Iceland, [32]Laboratory for Statistical Analysis, RIKEN Center for Integrative Medical Sciences, [33]Department of Ophthalmology, Graduate School of Medical Sciences, Kyushu University, Fukuoka, Japan, [34]Department of Statistical Genetics, Osaka University Graduate School of Medicine, Osaka, Japan, [35]Department of Allergy and Rheumatology, Graduate School of Medicine, The University of Tokyo, Tokyo, Japan, [36]Laboratory for Statistical and Translational Genetics, RIKEN Center for Integrative Medical Sciences, Yokohama, Japan., [37]Laboratory for Statistical Analysis, RIKEN Center for Integrative Medical Sciences, Yokohama, Japan., [38]Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, MA, USA., [39]Department of Biomedical Informatics, Harvard Medical School, Boston, MA, USA., [40]Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston, Massachusetts, USA, [41]Department of Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, Michigan, USA, [42]Program in Medical and Population Genetics, Broad Institute of Harvard and MIT, Cambridge, Massachusetts, USA, [43]K.G. Jebsen Center for Genetic Epidemiology, Department of Public Health and Nursing, NTNU, Norwegian University of Science and Technology, Trondheim, Norway, [44]MRC Integrative Epidemiology Unit, University of Bristol, UK, [45]Department of Thoracic Medicine, St. Olavs Hospital, Trondheim University Hospital, Trondheim, Norway, [46]Institute for Molecular Medicine Finland (FIMM), HiLIFE, University of Helsinki, Tukholmankatu 8, 00014 Helsinki, Finland, [47]Finnish institute for Health and Welfare, Helsinki, Finland, [48]Department of Genetics, Institute for Biomedical Informatics, University of Pennsylvania, Perelman School of Medicine, Philadelphia, PA 19104, USA, [49]Division of Clinical Pharmacology, Department of Medicine, Vanderbilt University Medical Center, Nashville, TN, [50]Department of Medicine (Medical Genetics), University of Washington, [51]Division of Biomedical Informatics, Cincinnati Children's Hospital Medical Center, [52]Center for Genetic Medicine, Northwestern University, [53]Genentech, 1 DNA Way, South San Francisco, 94084, USA, [54]Fred Hutchinson Cancer Research Center, Division of Public Health Sciences, Seattle WA 9810, USA, [55]Division of Preventive Medicine, Brigham and Women's Hospital, Boston, MA 02215, [56]Centre for Genomic and Experimental

Medicine , Institute of Genetics & Molecular Medicine, University of Edinburgh, Western General Hospital, Edinburgh EH4 2XU, United Kingdom, [57]Usher Institute for Population Health Sciences and Informatics, The University of Edinburgh, Nine, Edinburgh Bioquarter, 9 Little France Road, Edinburgh, EH16 4UX., [58]Clinical Trial Service Unit and Epidemiological Studies Unit, Nuffield Department of Population Health, University of Oxford, Oxford OX3 7LF, [59]Medical Research Council Population Health Research Unit, Nuffield Department of Population Health, University of Oxford, Oxford OX3 7LF, [60]Department of Population Medicine, Qatar University College of Medicine, QU Health, Doha, Qatar, [61]Center for Non-Communicable Diseases, Karachi, Sindh, Pakistan, [62]Survey Research Center, Institute for Social Research, University of Michigan, Ann Arbor, MI, 48104, [63]Department of Epidemiology, School of Public Health, University of Michigan, Ann Arbor, MI, 48109, [64]Program in Genetic Epidemiology and Statistical Genetics, Department of Epidemiology, Harvard T.H. Chan School of Public Health, 677 Huntington Avenue, Boston, MA, 02115, USA, [65]Department of Epidemiology, University of North Carolina, Chapel Hill, NC, [66]Wellcome Centre for Human Genetics, University of Oxford, UK, [67]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, Texas, 77030, USA, [68]Department of Epidemiology and Biostatistics, Imperial College London, London W2 1PG, UK, [69]Department of Cardiology, Ealing Hospital, London North West University Healthcare NHS Trust, Middlesex UB1 3HW, UK, [70]Imperial College Healthcare NHS Trust, London W12 0HS, UK, [71]Department of Biostatistics, Center for Statistical Genetics, University of Michigan, Ann Arbor, [72]Department of Medical Sciences, Molecular Epidemiology and Science for Life Laboratory, Uppsala University, Uppsala, Sweden., [73]MRC Epidemiology Unit, University of Cambridge School of Clinical Medicine, Cambridge, CB2 0QQ, UK, [74]Center for Genomic Medicine, Kyoto University Graduate School of Medicine, [75]Department of Internal Medicine, Erasmus MC, University Medical Center Rotterdam, [76]Department of Epidemiology, Erasmus MC, University Medical Center Rotterdam, [77]Department of Biostatistics, Boston University School of Public Health, 801 Massachusetts Ave, Boston, MA 02118, USA, [78]Department of Biological Psychology, Behavioral and Movement Sciences, Vrije Universiteit Amsterdam, [79]Amsterdam Public Health, VU medical center Amsterdam, [80]Amsterdam Public Health research institute, VU medical center Amsterdam, [81]Genetics of Complex Traits, University of Exeter Medical School, University of Exeter, Exeter, EX2 5DW, UK, [82]Department of Clinical Acupuncture and Moxibustion, Nanjing University of Chinese Medicine, Nanjing, Jiangsu 210029, China, [83]Research Unit of Molecular Epidemiology, Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg, Germany, [84]Institute of Epidemiology, Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg, Germany, [85]MRC Integrative Epidemiology Unit at the University of Bristol, Oakfield Road, Bristol, BS8 2BN, United Kingdom, [86]Bristol Dental School, University of Bristol, Lower Maudlin Street, Bristol BS1 2LY, United Kingdom, [87]Population Health Sciences, Bristol Medical School, University of Bristol, Oakfield Grove, Bristol, BS8 2BN, United Kingdom, [88]Saw Swee Hock School of Public Health, National Univeristy of Singapore and National University Health System, 117549, Singapore, [89]Center for Clinical Research and Prevention, Bispebjerg and Frederiksberg Hospital, Copenhagen, Denmark, [90]The Institute for Translational Genomics and Population Sciences, Department of Pediatrics, Lundquist Institute for Biomedical Innovations (Formerly LABioMed) at Harbor-UCLA Medical Center, Torrance, CA 90502, USA, [91]Center for Public Health Genomics, University of Virginia, Charlottesville, VA 22903 USA, [92]William Harvey Research Institute, Barts and The London School of Medicine and Dentistry, Queen Mary University of London, John Vane Science Centre, Charterhouse Square,

London, EC1M 6BQ, UK, [93]NIHR Barts Cardiovascular Biomedical Research Centre, Barts and The London School of Medicine and Dentistry, Queen Mary University of London, London, EC1M 6BQ, UK, [94]Novo Nordisk Foundation Center for Basic Metabolic Research, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark, [95]Center for Clinical Research and Prevention, Bispebjerg and Frederiksberg Hospital, Copenhagen, Denmark, Nordre Fasanvej 57, DK-2000 Frederiksberg, Denmark, [96]Division of Cardiovascular Medicine, Radcliffe Department of Medicine, John Radcliffe Hospital, University of Oxford, Oxford. UK. OX3 9DU, [97]Wellcome Centre for Human Genetics, University of Oxford, Oxford. UK. OX3 7BN, [98]Group of Genomics of Complex Diseases. Research Institute of Hospital de la Santa Creu i Sant Pau (IIB Sant Pau), Barcelona, Spain, [99]Cardiovascular Medicine Unit, Department of Medicine, Karolinska Institutet, Center for Molecular Medicine, Karolinska University Hospital, Stockholm, Sweden, [100]Department of Internal Medicine, Section Gerontology and Geriatrics, Leiden University Medical Center, Leiden, the Netherlands, [101]Institute for Genetic and Biomedical Research, Italian National Council of Research (IRGB CNR), Cagliari Italy, [102]Institute for Genetic and Biomedical Research, Italian National Council of Research (IRGB CNR), Lanusei, Italy, [103]University Center for Primary Care and Public Health, University of Lausanne, Rte de la Corniche 10, Lausanne, 1010, Switzerland, [104]Swiss Institute of Bioinformatics, Lausanne, 1015, Switzerland, [105]Department of Medicine, Internal Medicine, Lausanne University Hospital and University of Lausanne, Rue du Bugnon 46, Lausanne, 1011, Switzerland, [106]Department of Epidemiology and Biostatistics, MRC-PHE Centre for Environment and Health, School of Public Health, Imperial College London, London, UK, [107]Dept of Cardiology, Leiden University Medical Center, Leiden, the Netherlands, [108]Dept of Internal Medicine, Section of Gerontology and Geriatrics, Leiden university Medical Center, Leiden, the Netherlands, [109]BHF Glasgow Cardiovascular Research Centre, Faculty of Medicine, Glasgow, United Kingdom, [110]Institute for Cardiogenetics, University of Lübeck, DZHK (German Research Centre for Cardiovascular Research), partner site Hamburg/Lübeck/Kiel, University Heart Center Lübeck, Lübeck and Charité – University Medicine Berlin, corporate member of Freie Universität Berlin, Humboldt-Universität zu Berlin, and Berlin Institute of Health, Institute for Dental and Craniofacial Sciences, Department of Periodontology and Synoptic Dentistry, Berlin, Germany, [111]Deutsches Herzzentrum München, Klinik für Herz- und Kreislauferkrankungen, Technische Universität München,Munich, Germany., [112]Deutsches Zentrum für Herz-Kreislauf-Forschung (DZHK) e.V., partner site Munich Heart Alliance,Munich, Germany., [113]Key Laboratory of Cardiovascular Epidemiology & Department of Epidemiology, State Key Laboratory of Cardiovascular Disease, Fuwai Hospital, National Center for Cardiovascular Diseases, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing 100037, China, [114]Lund University Diabetes Centre, Malmö, Sweden, [115]Institute of Genetic Epidemiology, Department of Genetics and Pharmacology, Medical University of Innsbruck, Innsbruck, Austria and German Chronic Kidney Disease study, [116]Institute for Medical Informatics, Statistics and Epidemiology, University of Leipzig, Haertelstrasse 16-18, 04107 Leipzig, Germany, [117]LIFE Research Centre for Civilization Diseases, University of Leipzig, Philipp-Rosenthal-Straße 27, 04103 Leipzig, Germany, [118]Radboud university medical center, Radboud Institute for Health Sciences, Nijmegen, The Netherlands, [119]Quantinuum Research LLC, Wayne, PA, 19087 USA, [120]Division of Statistical Genomics, Department of Genetics; Washington University School of Medicine; St. Louis, MO, USA, [121]Department of Epidemiology; University of Pittsburgh; Pittsburgh, PA, USA, [122]Institute for Biomedicine, Eurac Research, Affiliated Institute of the University of Lübeck, Via Galvani 31, 39100, Bolzano, Italy, [123]Department of Clinical Biochemistry, Lillebaelt Hospital, Vejle,

Denmark, [124]Cardiovascular Health Research Unit, Department of Medicine, University of Washington, Seattle, 98101, USA, [125]The Charles Bronfman Institute for Personalized Medicine, Icahn School of Medicine at Mount Sinai, New York, NY, 10029, USA, [126]Department of Twin Research and Genetic Epidemiology, King's College London, London SE1 7EH, UK, [127]NIHR Biomedical Research Centre at Guy's and St Thomas' Foundation Trust, London SE1 9RT, UK, [128]University of Groningen, University Medical Center Groningen, Department of Cardiology, 9700RB Groningen, The Netherlands, [129]Institute of Cardiovascular Sciences, University College London, Gower Street, WC1E 6BT London, UK, [130]Department of Epidemiology and Public Health, University College London, 1-19 Torrington Place, WC1E 6BT London, United Kingdom, [131]Department of Psychiatry, University of Pennsylvania, Philadelphia, PA, 19104, USA, [132]Amsterdam UMC, Department of Epidemiology and Biostatistics, Amsterdam Public Health Research Institute, Amsterdam, 1081HV, the Netherlands., [133]Leiden University Medical Center, Department of Cell and Chemical Biology, Leiden, 2333ZA, The Netherlands, [134]Montreal Heart Institute, 5000 Belanger Street, Montreal, Quebec, H1T 1C8, Canada, [135]Icelandic Heart Association, 201 Kopavogur, Iceland, [136]Department of Anthropology, University of Toronto at Mississauga, Mississauga, ON L5L 1C6, Canada, [137]Vth Department of Medicine, Medical Faculty Mannheim, Heidelberg University, 68167 Mannheim, Germany, [138]SYNLAB MVZ Humangenetik Mannheim GmbH, 68163 Mannheim, Germany, [139]CAS Key Laboratory of Nutrition, Metabolism and Food Safety, Shanghai Institute of Nutrition and Health, University of Chinese Academy of Sciences, Chinese Academy of Sciences, [140]Biomedical Technology Research Center, Tokushima Research Institute, Otsuka Pharmaceutical Co., Ltd., Tokushima, Japan, [141]Institute for Molecular Bioscience, The University of Queensland, Brisbane, Queensland 4072, Australia, [142]Institute for Advanced Research, Wenzhou Medical University, Wenzhou, Zhejiang 325027, China, [143]Nuffield Department of Population Health, University of Oxford, Oxford, United Kingdom, [144]Laboratory of Neurogenetics, National Institute on Aging, NIH, Bethesda MD, USA, [145]Data Tecnica International, Glen Echo MD, USA, [146]MRC Human Genetics Unit, Institute of Genetics and Molecular Medicine, University of Edinburgh, Western General Hospital, Crewe Road, Edinburgh, EH4 2XU, Scotland, [147]Institute for Medical Informatics, Biometrie and Epidemiology, University of Duisburg-Essen, Essen, Germany, [148]Department of Computer Science, University of Applied Sciences and Arts Dortmund, Emil-Figge-Str. 42, 44227 Dortmund, Germany, [149]Centre for Public Health, Queen's University of Belfast, Northern Ireland, [150]Department of Epidemiology and Biostatistics, School of Public Health, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China, [151]Genome Institute of Singapore, Agency for Science, Technology and Research, 138672, Singapore, [152]Public Health Informatics Unit, Department of Integrated Health Sciences, Nagoya University Graduate School of Medicine, Nagoya, 461-8673, Japan, [153]MRC Unit for Lifelong Health and Ageing at UCL, 1-19 Torrington Place, London, WC1E 7HB, United Kingdom, [154]Department of Pediatrics, Tampere University Hospital, Tampere 33521, Finland, [155]Department of Pediatrics, Faculty of Medicine and Health Technology, Tampere University, Tampere 33014, Finland, [156]Brown Foundation Institute of Molecular Medicine, McGovern Medical School, University of Texas Health Science Center at Houston, Houston TX 77030, USA, [157]CONACYT, Instituto Nacional de Ciencias Médicas y Nutrición Salvador Zubirán, Ciudad de Mexico, Mexico, [158]Departamento de Genómica Computacional, Instituto Nacional de Medicina Genómica, Ciudad de Mexico, Mexico, [159]Center for diabetes research, University of Bergen, Bergen, Norway, [160]Lund University Diabetes Center, Lunds University, Malmö, Sweden, [161]Genomic Research on Complex diseases (GRC Group), CSIR-Centre for Cellular and Molecular Biology, Hyderabad, Telangana, India, [162]University of Alabama at Birmingham,

Epidemiology, School of Public Health, [163]Hunter Medical Research Institute, Newcastle, Australia, [164]Medical Research Institute, Kangbuk Samsung Hospital, Sungkyunkwan University School of Medicine, Seoul 03181, Korea, [165]Center for Cohort Studies, Total Healthcare Center, Kangbuk Samsung Hospital, Sungkyunkwan University School of Medicine, Seoul, 04514, Korea, [166]Department of Occupational and Environmental Medicine, Kangbuk Samsung Hospital, Sungkyunkwan University School of Medicine, Seoul, 03181, Korea, [167]Centre for Global Health Research, Usher Institute, University of Edinburgh, Teviot Place, Edinburgh, EH8 9AG, Scotland, [168]Thurston Arthritis Research Center, University of North Carolina, Chapel Hill, North Carolina, USA, [169]Genome Institute of Singapore, Agency for Science, Technology and Research, Singapore, [170]Division of Biomedical Informatics and Personalized Medicine, Department of Medicine, Anschutz Medical Campus, University of Colorado, Denver, Aurora, CO 80045, USA, [171]Ocular Epidemiology, Singapore Eye Research Institute, Singapore National Eye Centre, 168751, Singapore, [172]Department of Ophthalmology, National University of Singapore and National University Health System, 119228, Singapore, [173]Genomics and Molecular Medicine Unit, CSIR-Institute of Genomics and Integrative Biology, New Delhi - 110020, India., [174]Academy of Scientific and Innovative Research, CSIR-Institute of Genomics and Integrative Biology Campus, New Delhi 110020, India., [175]Departments of Ophthalmology and Human Genetics, Radboud Univeristy Nijmegen Medical Center, Philips van Leydenlaan 15, Nijmegen, 6525 EX, the Netherlands, [176]Vanderbilt Epidemiology Center, Division of Epidemiology, Vanderbilt University Medical Center, [177]Department of Pediatrics, University of California San Francisco, Oakland, CA 94609 USA, [178]National Center for Global Health and Medicine, Tokyo, 1628655, Japan, [179]Department of Genetics, University of North Carolina, Chapel Hill, NC 27599 USA, [180]Department of Biostatistics and Epidemiology, University of Massachusetts-Amherst, Amherst, MA 01003 USA, [181]Department of Cardiovascular Sciences, University of Leicester, Leicester, UK, [182]NIHR Leicester Biomedical Research Centre, Glenfield Hospital, Leicester, UK, [183]Beijing Institute of Ophthalmology, Beijing Key Laboratory of Ophthalmology and Visual Sciences, Beijing Tongren Eye Center, Beijing Tongren Hospital, Capital Medical University, 17 Hougou Lane, Chong Wen Men, Beijing, 100005 , China, [184]Beijing Tongren Eye Center, Beijing Tongren Hospital, Capital Medical University, 1 Dong Jiao Min Xiang, Dong Cheng District, Beijing, 100730, China, [185]Institute of Genetics and Biophysics “Adriano Buzzati-Traverso" - CNR, Naples, Italy, [186]IRCCS Neuromed, Pozzilli, Isernia, Italy, [187]Division of Biostatistics, Washington University, St. Louis, MO 63110, [188]Division of Endocrinology and Metabolism, Tri-Service General Hospital Songshan Branch, Taipei, Taiwan, [189]Rush Alzheimer’s Disease Center, Rush University Medical Center, [190]Department of Neurological Sciences, Rush University Medical Center, [191]Department of Nephrology, University Hospital Regensburg, Regensburg, Germany, [192]Department of Medicine, Surgery and Health Sciences, University of Trieste, Strada di Fiume 447, 34149, Trieste, Italy, [193]Department of Molecular Epidemiology, German Institute of Human Nutrition Potsdam-Rehbruecke, Nuthetal, Germany, [194]German Center for Diabetes Research (DZD), München-Neuherberg, Germany, [195]Department of Epidemiology, School of Public Health, University of Michigan, Ann Arbor, MI 48109, [196]Survey Research Center, Institute for Social Research, University of Michigan, Ann Arbor, MI, 48104, [197]Department of Genetics and Bioinformatics, Dasman Diabetes Institute, Kuwait, [198]Department of Nutrition and Dietetics, School of Health Science and Education, Harokopio University of Athens, Athens, Greece, [199]Department of Population Science and Experimental Medicine, University College London, London, UK, [200]Clinical Division of Neurogeriatrics, Department of Neurology, Medical University of Graz, Graz, Austria, [201]Institute for Medical Informatics, Statistics and Documentation, Medical

University of Graz, Graz, Austria, [202]Department of Bioinformatics and Genomics, University of North Carolina at Charlotte, NC 28223 USA, [203]Department of Internal Medicine, University of Michigan Medical Center, [204]Institute for Maternal and Child Health, IRCCS Burlo Garofolo, Trieste, Italy, [205]Institute of Genetic and Biomedical Research, National Research Council of Italy, UOS of Sassari, Sassari, Italy, [206]University of Groningen, University Medical Center Groningen, Department of Epidemiology, Groningen, 9700 RB, the Netherlands, [207]Research Centre of Applied and Preventive Cardiovascular Medicine, University of Turku, Turku, Finland, [208]Centre for Population Health Research, University of Turku and Turku University Hospital, Turku, Finland, [209]Sleep Medicine and Circadian Disorders, Brigham and Women's Hospital, Boston, Massachusetts 02115, USA, [210]Division of Sleep Medicine, Harvard Medical School, Boston, Massachusetts 02115, USA, [211]Central Diagnostics Laboratory, Division Laboratories, Pharmacy, and Biomedical genetics, University Medical Center Utrecht, Utrecht University, Utrecht, the Netherlands, [212]Laboratory of Epidemiology and Population Science National Institute on Aging Intramural Research Program, NIH 251 Bayview Blvd, NIH Biomedical Research Center, NIA, Baltimore, MD 21224, [213]Interfaculty Institute for Genetics and Functional Genomics, Department of Functional Genomics, University of Greifswald and University Medicine Greifswald, Greifswald, Germany, [214]Oneomics. co. ltd. 2F, Soonchunhyang Mirai Medical Center 173, Buheuyng-ro, Bucheon-si Gyeonggi-do, 14585, Korea, [215]Department of Biomedical Science, Hallym University, Chuncheon, Gangwon-do 24252, Korea, [216]Centre for Bone and Arthritis Research, Department of Internal Medicine and Clinical Nutrition, Institute of Medicine, Sahlgrenska Academy, University of Gothenburg, Gothenburg, Sweden., [217]Bioinformatics Core Facility, Sahlgrenska Academy, University of Gothenburg, Gothenburg, Sweden, [218]Institute of Medical Informatics and Statistics, Kiel University, Kiel, Germany., [219]Institute of Translational Genomics, Helmholtz Zentrum München – German Research Center for Environmental Health, Neuherberg, Germany, [220]Wellcome Trust Sanger Institute, Hinxton, CB10 1SA, UK, [221]Wellcome Centre for Human Genetics, University of Oxford, Oxford, UK, [222]Oxford Centre for Diabetes Endocrinology and Metabolism, Oxford, UK, [223]Wellcome Sanger Institute, Hinxton, Cambridge, HH CB10 1 UK, [224]School of Medicine and Public Health, Faculty of Medicine and Health, University of Newcastle, Newcastle, New South Wales, 2308, Australia, [225]Center for Geriatrics and Gerontology, Division of Endocrinology and Metabolism, Department of Internal Medicine, Taichung Veterans General Hospital, Taichung, Taiwan, [226]School of Medicine, National Yang-Ming University, Taipei, Taiwan, [227]School of Medicine, National Defense Medical Center, Taipei, Taiwan, [228]Division of Endocrinology and Metabolism, Department of Internal Medicine, Taichung Veterans General Hospital, Taichung, Taiwan, [229]Department of Medicine, School of Medicine, National Yang-Ming University, Taipei, Taiwan, [230]Dept of Kinesiology, Université Laval, Québec, Canada, [231]Department of Clinical Chemistry, Fimlab Laboratories, Tampere 33520, Finland, [232]Department of Clinical Chemistry, Finnish Cardiovascular Research Center - Tampere, Faculty of Medicine and Health Technology, Tampere University, Tampere 33014, Finland, [233]Department of Cardiology, Heart Center, Tampere University Hospital, Tampere 33521, Finland, [234]Department of Cardiology, Finnish Cardiovascular Research Center - Tampere, Faculty of Medicine and Health Technology, Tampere University, Tampere 33014, Finland, [235]University of Queensland Diamantina Institute, Translational Research Institute, Kent St, Woolloongabba, Brisbane, QLD, 4102, Australia., [236]Novo Nordisk Foundation Center for Basic Metabolic Research, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark, Department of Medicine, Bornholms Hospital, Rønne, Denmark, [237]School of Public Health, University of Alabama at Birmingham, [238]Cardiology, Division Heart & Lungs, University Medical Center Utrecht, Utrecht

University, Utrecht, the Netherlands, [239]Department of Population and Quantitative Health Sciences, Case Western Reserve University, Cleveland, OH, 44106, USA, [240]Division of Statistical Genomics, Department of Genetics, Washington University School of Medicine, St. Louis, MO, USA, [241]Department of Epidemiology - Erasmus MC - University Medical Center Rotterdam, Rotterdam, the Netherlands., [242]Ohio State University, Division of Endocrinology, Columbus OH 43210, USA, [243]University of Washington, Department of Epidemiology, Seattle WA 98195, USA, [244]George Washington University, School of Medicine and Health Sciences, Washington DC 20037, USA, [245]Department of Epidemiology, School of Public Health, Peking University Health Science Center, Beijing, China, [246]Institute for Laboratory Medicine, University Hospital Leipzig, Paul-List-Strasse 13/15, 04103 Leipzig, Germany, [247]Laboratory of Epidemiology and Population Sciences, National Institute on Aging, NIH, Baltimore,MD, 20892-9205, USA, [248]Centre for Population Health Research, University of Turku and Turku University Hospital, Finland, [249]Research Centre of Applied and Preventive Cardiovascular Medicine, University of Turku, Finland, [250]Department of Clinical Physiology and Nuclear Medicine, Turku University Hospital, Turku, Finland, [251]Department of Environmental and Preventive Medicine, Jichi Medical University School of Medicine, Shimotsuke, 329-0498, Japan, [252]Centre for Population Health Sciences, Usher Institute, University of Edinburgh, Teviot Place, Edinburgh, EH8 9AG, Scotland, [253]Department of Functional Pathology, Shimane University School of Medicine, Izumo, 6938501, Japan, [254]Department of Pediatrics and Adolescent Medicine, Turku University Hospital and University of Turku, Turku, Finland, [255]Department of Physiology, University of Turku, Turku, Finland, [256]Faculty of Medicine, University of Split, Šoltanska 2, HR-21000, Split, Croatia, [257]Medical Department III – Endocrinology, Nephrology, Rheumatology, University of Leipzig Medical Center, Liebigstr. 21, 04103 Leipzig, Germany, [258]Department of Nutrition-Dietetics, Harokopio University, Eleftheriou Venizelou, Athens, 17676, Greece, [259]Department of Clinical Gene Therapy, Osaka University Graduate School of Medicine, Suita, 5650871, Japan, [260]Department of Geriatric and General Medicine, Osaka University Graduate School of Medicine, Suita, 5650871, Japan, [261]Department of Vascular Surgery, Division of Surgical Specialties, University Medical Center Utrecht, Utrecht University, Utrecht, the Netherlands, [262]Corneal Dystrophy Research Institute, Department of Ophthalmology, Yonsei University College of Medicine, Seoul 03722, Korea, [263]Dept of Radiology and Nuclear Medicine, Erasmus MC - University Medical Center Rotterdam, Rotterdam, the Netherlands., [264]Julius Centre for Health Sciences and Primary Care, University Medical Centre Utrecht, 3584CG, the Netherlands, [265]Second Department of Cardiology, Medical School, National and Kapodistrian University of Athens, Attikon University Hospital, Athens, Greece, [266]Center for Vision Research, Department of Ophthalmology and The Westmead Institute, University of Sydney, Hawkesbury Rd, Sydney, New South Wales, 2145, Australia., [267]Menzies Institute for Medical Research, School of Medicine, University of Tasmania, Liverpool St, Hobart, Tasmania, 7000, Australia., [268]Centre for Eye Research Australia, University of Melbourne, Melbourne, Victoria, 3002, Australia, [269]Department of Clinical Physiology, Tampere University Hospital, Tampere 33521, Finland, [270]Department of Clinical Physiology, Finnish Cardiovascular Research Center - Tampere, Faculty of Medicine and Health Technology, Tampere University, Tampere 33014, Finland, [271]Institute of Nutrition and Functional Foods (INAF), Université Laval, Québec, Canada, [272]Pennington Biomedical Research Center, Baton Rouge, LA 70808, USA, [273]Medical Department III – Endocrinology, Nephrology, Rheumatology, University of Leipzig Medical Center, Liebigstr. 18, 04103 Leipzig, Germany, [274]Medical School, Faculty of Health and Medical Sciences, University of Western Australia, Perth, Western Australia, 6000, Australia, [275]Institute of Epidemiology, Kiel University, Kiel, Germany, [276]Institute of Clinical

Molecular Biology, Kiel University, Kiel, Germany, [277]Sahlgrenska University Hospital, Department of Drug Treatment, Gothenburg, Sweden, [278]Geriatric Medicine, Institute of Medicine, Sahlgrenska Academy, University of Gothenburg, Gothenburg, Sweden, [279]Department of Internal Medicine, EwhaWomans University School of Medicine, Seoul, Korea, [280]Division of Cancer Control and Population Sciences, UPMC Hillman Cancer Center, University of Pittsburgh, Pittsburgh, PA 15232, USA, [281]Department of Epidemiology, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, PA 15232, USA, [282]Health Services and Systems Research, Duke-NUS Medical School, 169857, Singapore, [283]Saw Swee Hock School of Public Health, National University of Singapore and National University Health System, [284]Department of Endocrinology and Metabolism, Kyung Hee University School of Medicine, Seoul 02447, Korea, [285]Institute for Community Medicine, University Medicine Greifswald, Germany, [286]Laboratory of Epidemiology and Population Science National Institute on Aging Intramural Research Program, NIH 251 Bayview Blvd, NIH Biomedical Research Center, Baltimore, MD 21224, [287]Paavo Nurmi Centre, Sports and Exercise Medicine Unit, Department of Physical Activity and Health, University of Turku, Turku, Finland, [288]University of Groningen, University Medical Center Groningen, Interdisciplinary Center Psychopathology and Emotion Regulation (ICPE), Groningen, 9700 RB, the Netherlands, [289]Institute of Molecular Genetics, National Research Council of Italy, Pavia, Italy, [290]Gottfried Schatz Research Center for Cell Signaling, Metabolism and Aging, Medical University of Graz, Graz, Austria, [291]Local Health Unit Toscana Centro, Firenze, Italy., [292]Department of Genetics and Bioinformatics, Dasman Diabetes Institute, Kuwait, [293]Institute of Nutritional Science, University of Potsdam, Nuthetal, Germany, [294]Institute for Maternal and Child Health IRCCS "Burlo Garofolo", Via dell'Istria 65/1, 34137, Trieste, Italy, [295]Dept of Nephrology, University Hospital Regensburg, Regensburg, Germany, [296]Dept of Nephrology, Diabetology, Rheumatology; Traunstein Hospital, Traunstein, Germany, [297]KfH Kidney Center Traunstein, Traunstein, Germany, [298]Center for Translational and Systems Neuroimmunology, Department of Neurology, Columbia University Medical Center, New York, NY, USA, [299]Program in Medical and Population Genetics, Broad Institute, Cambridge, MA, USA, [300]Medical School, National and Kapodistrian University Athens, 75 M. Assias Street, 115 27 Athens, Greece, [301]Dromokaiteio Psychiatric Hospital, 124 61 Athens, Greece, [302]Clinical Pharmacology, William Harvey Research Institute, Queen Mary University of London, London, EC1M 6BQ,UK, [303]Department of Ophthalmology, Medical Faculty Mannheim, Heidelberg University, Kutzerufer 1, Mannheim, 68167, Germany, [304]Institute of Clinical and Scientific Ophthalmology and Acupuncture Jonas & Panda, Heidelberg, Germany, [305]Harvard Medical School, Boston MA 02115, [306]Department of Human Genetics, David Geffen School of Medicine at UCLA, University of California, Los Angeles, CA, USA, [307]Unidad de Biología Molecular y Medicina Genómica, Instituto Nacional de Ciencias Médicas y Nutrición Salvador Zubirán, Mexico 14080, Mexico., [308]Instituto de Investigaciones Biomédicas, UNAM, [309]Departamento de Endocrinología y Metabolismo, Instituto Nacional de Ciencias Médicas y Nutrición Salvador Zubirán, Mexico 14080, Mexico., [310]Department of Nutrition, Gillings School of Global Public Health, University of North Carolina, Chapel Hill, North Carolina, 27599 USA, [311]Carolina Population Center, University of North Carolina, Chapel Hill, North Carolina, 27516 USA, [312]USC–Office of Population Studies Foundation, University of San Carlos, Cebu City, 6000, Philippines, [313]Department of Anthropology, Sociology, and History, University of San Carlos, Cebu City, 6000 Philippines, [314]Department of Medicine, Faculty of Medicine, University of Kelaniya, Ragama, 11010, Sri Lanka, [315]Department of Public Health, Faculty of Medicine, University of Kelaniya, Ragama, 11010, Sri Lanka, [316]Children's Hospital Oakland Research Institute, Oakland, CA 94609 USA, [317]Institute of Biomedical Sciences,

Academia Sinica, Taiwan, [318]Systems Genomics Laboratory, School of Biotechnology, Jawaharlal Nehru University, New Delhi - 110067, India, [319]Department of Medicine, University of Mississippi Medical Center, Jackson, MS, 39216, USA, [320]Department of Physiology and Biophysics, University of Mississippi Medical Center, Jackson, MS, 39216, USA, [321]Department of Medical Sciences, Uppsala University, Sweden, [322]Department of Paediatrics, Yong Loo Lin School of Medicine, National University of Singapore; and Khoo Teck Puat - National University Children's Medical Institute, National University Health System, [323]Department of Medicine, University of North Carolina, Chapel Hill, NC, USA, [324]Department of Epidemiology, Gillings School of Global Public Health, University of North Carolina, Chapel Hill, North Carolina, USA, [325]Injury Prevention Research Center, University of North Carolina, Chapel Hill, North Carolina, USA, [326]Division of Physical Therapy, University of North Carolina, Chapel Hill, North Carolina, USA, [327]Department of Psychiatry, Amsterdam UMC, Vrije Universiteit Amsterdam, [328]Department of Biochemistry, College of Medicine, Ewha Womans University, Seoul 07804, Korea, [329]Faculty of Health and Medicine, University of Newcastle, Australia, [330]Washington University School of Medicine, Division of Biostatistics, [331]University of Kentucky, College of Public Health, [332]Institute of Cellular Medicine (Diabetes), The Medical School, Newcastle University, Framlington Place, Newcastle upon Tyne, NE2 4HH, [333]Department of Public Health Solutions, Finnish Institute for Health and Welfare, P.O. Box 30, FI-00271 Helsinki, Finland., [334]University of Helsinki and Department of Medicine, Helsinki University Central Hospital, P.O.Box 340, Haartmaninkatu 4, Helsinki, FI-00029, Finland., [335]Minerva Foundation Institute for Medical Research, Biomedicum 2U, Tukholmankatu 8, Helsinki, FI-00290, Finland., [336]Academy of Scientific and Innovative Research (AcSIR), New Delhi, India, [337]Diabetology Research Centre, KEM Hospital and Research Centre, Pune, Maharashtra, India, [338]Programs in Metabolism and Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, MA, USA, [339]Diabetes Unit and Center for Genomic Medicine, Massachusetts General Hospital, Boston, MA, USA10, [340]Harvard Medical School, Boston, Massachusetts, USA, [341]Unidad de Biología Molecular y Medicina Genómica, Instituto de Investigaciones Bimédicas UNAM/ Instituto Nacional de Ciencias Médicas y Nutrición Salvador Zubirán, Mexico City, [342]Dirección de Nutrición and Unidad de Estudios de Enfermedades Metabólicas, Instituto Nacional de Ciencias Médicas y Nutrición Salvador Zubirán, Mexico City, [343]InstitutoNacional de Salud Publica y Centro de Estudios en Diabetes, Mexico, [344]Instituto Nacional de Medicina Genómica, Mexico, [345]Human Genetics Center, School of Public Health, University of Texas Health Science Center at Houston, Houston TX 77030, USA, [346]Yong Loo Lin School of Medicine, National University of Singapore and National University Health System, 119228, Singapore, [347]Kurume University School of Medicine, Kurume, 830-0011, Japan, [348]Genetics, Merck Sharp & Dohme Corp., Kenilworth, NJ, 07033, US, [349]Oxford Centre for Diabetes, Endocrinology & Metabolism, University of Oxford, UK, [350]Population Health and Genomics, University of Dundee, Ninwells Hospital and Medical School, Dundee, DD1 9SY, UK, [351]Intramural Research Program, National Institute on Aging, 3001 S. Hanover St., Baltimore, MD 21225, [352]The Eye Hospital, School of Ophthalmology & Optometry, Wenzhou Medical University, Wenzhou, Zhejiang 325027, China, [353]Shanghai Institute of Nutrition and Health University of Chinese Academy of Sciences, Chinese Academy of Sciences, [354]Synlab Academy, SYNLAB Holding Deutschland GmbH, Mannheim and Augsburg, Germany, [355]Clinical Institute of Medical and Chemical Laboratory Diagnostics, Medical University of Graz, Austria, [356]Unidad de Investigacion Medica en Bioquimica, Hospital de Especialidades, Centro Medico Nacional Siglo XXI, Instituto Mexicano del Seguro Social, Mexico City, Mexico, [357]Faculty of Medicine, University of Iceland, 101 Reykjavik, Iceland, [358]Department of Medicine,

Faculty of Medecine, Université de Montréal, 2900 Edouard Montpetit Blvd, Montreal, Quebec, H3T 1J4, Canada, [359]Leiden University Medical Center, Department of Biomedical Data Sciences, Section Molecular Epidemiology, Leiden, 2333ZA, The Netherlands, [360]Amsterdam UMC, Department of General Practice and Elderly Care, Amsterdam Public Health Research Institute, Amsterdam, 1081HV, The Netherlands, [361]Department of Genetics, University of Pennsylvania, Philadelphia, PA, 19104, USA, [362]Department of Surgery, University of Pennsylvania, Philadelphia, PA, 19104, USA, [363]Corporal Michael Crescenz VA Medical Center, Philadelphia, Pennsylvania, PA, 19104, USA, [364]Institute of Social and Economic Research, University of Essex, Wivenhoe Park, CO4 3SQ, United Kingdom, [365]The Mindich Child Health and Development Institute, Icahn School of Medicine at Mount Sinai, New York, NY, 10029, USA, [366]Cardiovascular Health Research Unit, Departments of Medicine, Epidemiology and Health Services, University of Washington, Seattle, 98101, WA, USA, [367]Kaiser Permanent Washington Health Research Institute, Seattle, 98101, WA, USA, [368]Institute of Regional Health Research, University of Southern Denmark, Odense, Denmark, [369]Danish Aging Research Center, University of Southern Denmark; Odense C, Denmark, [370]Public Health, Faculty of Medicine, University of Helsinki, Finland, [371]Broad Institute of MIT and Harvard, Cambridge, MA, [372]Center for Applied Genomics, Children's Hospital of Philadelphia, Philadelphia, PA, 19104 USA, [373]Department of Pediatrics, The University of Pennsylvania Perelman School of Medicine, Philadelphia, PA, 19104 USA, [374]Division of Human Genetics, Children’s Hospital of Philadelphia, Philadelphia, PA, 19104 USA, [375]Department of Genetics, University of Pennsylvania, Philadelphia, PA, 19104 USA, [376]School of Medicine, Southern University of Science and Technology, Shenzhen, China, [377]Institute for Cardiogenetics, University of Lübeck, DZHK (German Research Centre for Cardiovascular Research), partner site Hamburg/Lübeck/Kiel, and University Heart Center Lübeck, Lübeck, Germany, [378]Department of Clinical Medicine, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark, [379]Netherlands Heart Institute, Utrecht, the Netherlands, [380]Division of Cardiology, Department of Medicine, Massachusetts General Hospital, Boston, Massachusetts, USA, [381]Program of Medical and Population Genetics, Broad Institute, Cambridge, Massachusetts, USA, [382]Center for Genomic Medicine, Massachusetts General Hospital, Boston, Massachusetts, USA, [383]Department of Medicine, Harvard Medical School, Boston, Massachusetts, USA, [384]Northern Finland Birth Cohorts, Infrastructure for population studies, Faculty of Medicine, University of Oulu, Oulu, Finland, [385]Center for Life Course Health Research, Faculty of Medicine, University of Oulu, Oulu, Fin, [386]Biocenter of Oulu, University of Oulu, Oulu, Finl, [387]Institute for Genetic and Biomedical Research, Italian National Council of Research (IRGB CNR), Cagliari, Italy, [388]University of Sassari, Sassari, Italy, [389]Department of Clinical Epidemiology, Leiden University Medical Center, Leiden, the Netherlands, [390]Department of Public Health and Primary Care, Leiden University Medical Center, Leiden, the Netherlands, [391]Department of Internal Medicine, Division of Endocrinology, Leiden University Medical Center, Leiden, the Netherlands, [392]Einthoven Laboratory for Experimental Vascular Medicine, Leiden University Medical Center, Leiden, the Netherlands, [393]Department of Human Genetics, Leiden University Medical Center, Leiden, the Netherlands, [394]Population Health Research Institute, St George's, University of London, London SW17 0RE, UK, [395]National Heart and Lung Institute, Imperial College London, London, W2 1PG, UK, [396]School of Public Health, Imperial College London, London, W2 1PG, UK, [397]OCDEM, University of Oxford, Churchill Hospital, Oxford OX3 7LE, UK, [398]NIHR Oxford Biomedical Research Centre, Churchill Hospital, Oxford, UK, [399]Ophthalmology & Visual Sciences Academic Clinical Program (Eye ACP), Duke-NUS Medical School, 169857, Singapore, [400]DZHK (German Centre for

Cardiovascular Research), Munich Heart Alliance partner site, Munich, Germany., [401]German Center for Diabetes Research (DZD), Neuherberg, Germany., [402]University of Exeter Medical School, University of Exeter, Exeter, EX2 5DW, UK, [403]Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden, [404]Netherlands Twin Register, Department of Biological Psychology, Behavioral and Movement Sciences, Vrije Universiteit Amsterdam, [405]Amsterdam Reproduction & Development research institute, VU medical center Amsterdam, [406]Framingham Heart Study, National Heart, Lung, and Blood Institute, US National Institutes of Health, Bethesda, MD, USA., [407]Department of Genetics, School of Medicine, Mashhad University of Medical Sciences, Mashhad, Iran., [408]Department of Genetics, Shanghai-MOST Key Laboratory of Health and Disease Genomics, Chinese National Human Genome Center at Shanghai, Shanghai, 201203 China, [409]TUM School of Medicine, Technical University of Munich and Klinikum Rechts der Isar, Munich, Germany, [410]Institute of Clinical Medicine, Internal Medicine, University of Eastern Finland and Kuopio University Hospital, [411]Department of Medicine, Division of Cardiovascular Medicine, Stanford University School of Medicine, Stanford, CA 94305, [412]Stanford Cardiovascular Institute, Stanford University, Stanford, CA 94305, [413]Stanford Diabetes Research Center, Stanford University, Stanford, CA 94305, [414]Regeneron Pharmaceuticals, Tarrytown, NY, USA, [415]Lee Kong Chian School of Medicine, Nanyang Technological University, Singapore 308232, Singapore, [416]Imperial College Healthcare NHS Trust, Imperial College London, London W12 0HS, UK, [417]MRC-PHE Centre for Environment and Health, Imperial College London, London W2 1PG, UK, [418]National Heart and Lung Institute, Imperial College London, London W12 0NN, UK, [419]Institute for Minority Health Research, University of Illinois College of Medicine, Chicago, Illinois, [420]Department of Biostatistics, Harvard T.H. Chan School of Public Health, 677 Huntington Avenue, Boston, MA, 02115, USA, [421]QIMR Berghofer Medical Research Institute, 300 Herston Road, Brisbane, Queensland 4006, Australia, [422]Center for Non-Communicable Diseases, Karachi, Sindh, Pakistan & Faisalabad Institute of Cardiology, Faislabad, Pakistan, [423]Department of Medicine, Columbia University Irving Medical Center, New York, NY, USA, [424]Department of Cardiology, Columbia University Irving Medical Center, New York, NY, USA, [425]Clinical Trial Service Unit and Epidemiological Studies Unit, Nuffield Department of Population Health, University of Oxford, Oxford OX3 7LF, UK, [426]Medical Research Council Population Health Research Unit, Nuffield Department of Population Health, University of Oxford, Oxford OX3 7LF, UK, [427]Big Data Institute, University of Oxford, Oxford OX3 7LF, UK, [428]National Institute for Health Research Oxford Biomedical Research Centre, Oxford University Hospitals, Oxford, UK, [429]Aberdeen Centre for Health Data Science, 1:042 Polwarth Building School of Medicine, Medical Science and Nutrition University of Aberdeen Foresterhill Aberdeen AB25 2ZD, [430]Division of Population Health and Genomics, Ninewells Hospital and Medical School, University of Dundee, Dundee DD1 9SY, United Kingdom, [431]Biomedical and Translational Informatics, Geisinger Health, Danville, PA 17822, [432]Regeneron Pharmaceuticals, Tarrytown, NY, USA., [433]Harvard Medical School, Boston, MA 02115, [434]Department of Biomedical Informatics, Vanderbilt University Medical Center, Nashville, TN, [435]Departments of Medicine (Medical Genetics) and Genome Sciences, University of Washington, [436]Center for Autoimmune Genomics and Etiology, Cincinnati Children's Hospital Medical Center (CCHMC), Cincinnati, OH, USA., [437]Division of Endocrinology, Metabolism, and Molecular Medicine, Department of Medicine, Northwestern University, Feinberg School of Medicine, Chicago, IL 60618, USA, [438]Department of Anthropology, Northwestern University, Evanston, IL 60208, USA, [439]Center for Genetic Medicine, Northwestern University, Feinberg School of Medicine, Chicago, IL 60618, USA, [440]HUNT Research Centre, Department of Public Health and Nursing, NTNU, Norwegian

University of Science and Technology, Levanger, 7600 Norway, [441]Department of Medicine, Levanger Hospital, Nord-Trøndelag Hospital Trust, Levanger, 7600 Norway, [442]Department of Endocrinology, St. Olavs Hospital, Trondheim University Hospital, Trondheim, Norway, [443]RIKEN Center for Integrative Medical Sciences, [444]Laboratory of Complex Trait Genomics, Department of Computational Biology and Medical Sciences, Graduate School of Frontier Sciences, The University of Tokyo, Tokyo, Japan, [445]Laboratory of Statistical Immunology, WPI Immunology Frontier Research Center, Osaka University, Osaka, Japan, [446]Integrated Frontier Research for Medical Science Division, Institute for Open and Transdisciplinary Research Initiatives, Osaka University, Osaka, Japan, [447]Division of Molecular Pathology, Institute of Medical Science, The University of Tokyo, Tokyo, Japan, [448]Faculty of Medicine, University of Iceland, Sæmundargötu 2, Reykjavik, 102, Iceland, [449]VA Boston Healthcare System, Boston, MA, USA, [450]VA Informatics and Computing Infrastructure, VA Salt Lake City Health Care System, Salt Lake City, UT, USA, [451]University of Massachusetts, Boston, MA, USA, [452]Department of Medicine, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA, USA, [453]Cardiovascular Institute, Stanford University School of Medicine, Stanford, California, USA, [454]Corporal Michael J. Crescenz VA Medical Center, Philadelphia, PA, USA, [455]Department of Medicine, Brigham Women's Hospital, Boston, MA, USA, [456]Atlanta VA Medical Center, Atlanta, GA, USA, [457]Division of Cardiology, Emory University School of Medicine, Atlanta, GA, USA, [458]School of Electrical and Information Engineering, University of the Witwatersrand, Johannesburg, South Africa, [459]Division of Human Genetics, National Health Laboratory Service and School of Pathology, Faculty of Health Sciences, University of the Witwatersrand, Johannesburg, South Africa., [460]School of Basic and Medical Biosciences, Faculty of Life Sciences and Medicine, King's College London, London, UK, [461]Departments of Pediatrics and Genetics, Harvard Medical School, Boston, MA, USA, [462]Center for Genomic Medicine, Department of Medicine, Massachusetts General Hospital, Boston, Massachusetts, USA, [463]Cardiology Division, Massachusetts General Hospital, Harvard Medical School, Boston, MA, [464]Department of Medicine, Massachusetts General Hospital, Harvard Medical School, Boston, MA, [465]Program in Medical and Population Genetics, Broad Institute of Harvard and MIT, Cambridge, MA, [466]Cardiovascular Research Center and Center for Genomic Medicine, Massachusetts General Hospital, Boston, MA, [467]Department of Biostatistics, Boston University School of Public Health, Boston, MA, [468]Centre for Genetics and Genomics Versus Arthritis, Centre for Musculoskeletal Research, Division of Musculoskeletal and Dermatological Sciences, The University of Manchester, Manchester, UK, [469]Princess Al-Jawhara Al-Brahim Centre of Excellence in Research of Hereditary Disorders (PACER-HD), King Abdulaziz University, Jeddah, Saudi Arabia, [470]Department of Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, MI 48109, USA, [471]Department of Human Genetics, University of Michigan, Ann Arbor, MI 48019, USA

[‡]These authors contributed equally to this work (as co-second authors)

[†]These authors jointly supervised this work

[#]Present Address: Genentech, 1 DNA Way, South San Francisco, CA 94080

*****Co-corresponding authors:

Cristen Willer, PhD
1241 E Catherine St
Dept of Human Genetics
University of Michigan, USA, 48109
cristen@umich.edu

Themistocles Assimes, MD, PhD
VA Palo Alto HealthCare System, Palo Alto Division
3801 Miranda Avenue
Palo Alto, CA 94304
650-493-5000 | 800-455-0057
tassimes@stanford.edu

Abstract

Genome-wide association studies (GWAS) of blood lipid levels have led to important biologic and clinical insights for cardiovascular disease. To evaluate the improvement in discovery of lipid loci across ancestrally diverse individuals, we performed a GWAS meta-analysis of ~1.65 million individuals, including ~330,000 (20%) of non-European ancestry. We identified 2,727 associated index variants within 941 genetic regions (1 Mb, p-value < $5x10^{-8}$) for LDL-, HDL-, nonHDL-, total-cholesterol, and triglyceride blood levels including 237 novel loci from ancestry-specific meta-analysis and 120 additional novel loci from trans-ancestry meta-analysis. We evaluated the incremental gain in GWAS discovery after including a substantial proportion of non-European ancestry individuals. Meta-analyses of GWAS from non-European individuals identify the greatest proportion of loci relative to the number of included individuals at currently available sample sizes while European and trans-ancestry analyses identify the largest overall number of loci. Furthermore, the identification of likely causal variants compared to ancestry-specific analysis was substantially improved with the inclusion of multiple ancestries. Of the 1,486 independent associations, 56% showed a profound decrease (median 40% reduction) in the number of putative causal variants, twice that expected by the increase in sample size alone. Lastly, we show that a polygenic score developed from trans-ancestry meta-analysis is the best performing score (or nearly) in all ancestries, simplifying future application of PRS in the clinic. Notably, the trans-ancestry polygenic score for LDL-C performed equally well in Americans with African ancestry as it does in those with European ancestry, without the prediction disparity seen with non-ancestry matched scores. These results provide empirical evidence that genetic discovery and risk prediction are poised to markedly benefit from genetic study of larger numbers of previously under-represented ancestry groups and that increased diversity within genetic discovery efforts can eliminate disparities in polygenic score performance across ancestry groups, avoiding exacerbating existing racial disparities in health care.

Main

Levels of blood lipids are heritable risk factors of cardiovascular diseases. Worldwide, differing dietary patterns and medication use have led to variable cholesterol levels between populations[1]. Within the United States, advances in prevention and treatment of coronary artery disease (CAD) have led to decreases in CAD-related mortality over time, particularly through medications that lower low-density lipoprotein cholesterol (LDL-C)[2], yet heart disease remains the leading cause of death globally[3]. Unfortunately, improvement in outcomes of CAD has been uneven across race/ethnic groups with white individuals benefiting the most[4]. While health disparities are a substantial cause of these differences, genome-wide association studies (GWAS) of blood lipid levels provide a complementary opportunity to investigate the genetic basis of such differences. However, the majority of previous GWAS[5-24] have been conducted in populations of European ancestry and may have missed genetic variants contributing to lipid level variation in individuals from other ancestry groups due to differences in allele frequencies, effect sizes, and linkage-disequilibrium (LD) patterns[25]. For example, sequencing of *PCSK9* identified two premature stop variants with large effects on LDL-C that are common in African Americans but rare in European ancestry individuals[26]. Furthermore, differences in LD patterns between ancestry groups, especially African ancestry[27], may aid in the identification of causal variants and facilitate the connection between GWAS signals and the underlying biology for individuals of all ancestries. Recent research has extended GWAS discovery beyond its emphasis on identifying genes and fundamental biology towards using genetic variants for preventative and precision medicine[28]. However, polygenic scores developed from European GWAS generally show worse performance in other ancestries[29]. Therefore, it is imperative to elucidate all associated alleles across diverse ancestries. In the present study, we performed a meta-analysis of 162 GWAS for 5 lipid levels in up to 1.65 million individuals from 5 ancestral-groups from around the world (AdmAFR; Admixed African or African individuals recruited from the US, UK, and Africa, HIS: Hispanic, EUR: European, EAS: East Asian and SAS: South Asian, **Supplementary Table 1**). Using these results, we aimed to quantify the role of ancestral diversity in GWAS discovery and fine-mapping, compare the effect sizes of identified variants between groups, and establish an approach for generating polygenic scores that are most informative in individuals of other ancestries.

Ancestry-specific and trans-ancestry genetic discovery

To evaluate the impact of multi-ancestry genetic discovery efforts relative to those in a single ancestry, we performed GWAS for five blood lipid traits: low-density lipoprotein

cholesterol (LDL-C), high-density lipoprotein cholesterol (HDL-C), triglycerides (TG), total cholesterol (TC), and non-high-density lipoprotein cholesterol (nonHDL-C) in individuals from five ancestry groups (**Table 1**, **Supplementary Tables 1-2**, **Supplementary Figures 1-2**). These analyses included 91 million variants imputed primarily from the Haplotype Reference Consortium or 1000 Genomes Phase 3. Ancestry-specific meta-analyses identified 773 lipid-associated genomic regions (p-value < $5x10^{-8}$, ± 500 kb, **Supplementary Table 3**) containing 1,765 unique index variants across all ancestry groups and lipid traits. Of these regions, 237 were novel based on the most-significant index variant in each region being > 500kb from previously reported variants[5-24,30] for any of the five lipid traits.

Table 1: Meta-analysis sample size by ancestry group

Ancestry Group	Sample Size
European	1,320,016
East Asian	146,492
Admixed African/African	99,432
Hispanic	48,057
South Asian	40,963
Total	1,654,960

A majority (76%) of loci were identified only in the European ancestry-specific analyses, consistent with the large fraction of European ancestry individuals included this study (N~1.3m, 80% of sample) while 21% of loci reached genome-wide significance in more than one ancestry group. Meta-analysis of African-ancestry individuals (N~99k, primarily African American) identified more ancestry-specific loci (15 unique to AdmAFR) than analyses of each of the other non-European ancestry groups (6 loci unique to EAS, 6 to HIS, 1 to SAS), a difference likely attributed to allele frequencies being most different between African and European ancestry populations and to African populations having a greater overall level of genetic diversity[27]. For these variants, the frequency of the African-ancestry minor allele in European ancestry populations ranged from 0.03% to 98.9%, with a median 4.5-fold difference in allele frequencies between these groups (**Supplementary Table 4**, **Figure 1**). The effect sizes of variants identified from each ancestry-specific analysis included common variants of small effect and rare variants with moderate to large effect (**Supplementary Figure 3**).

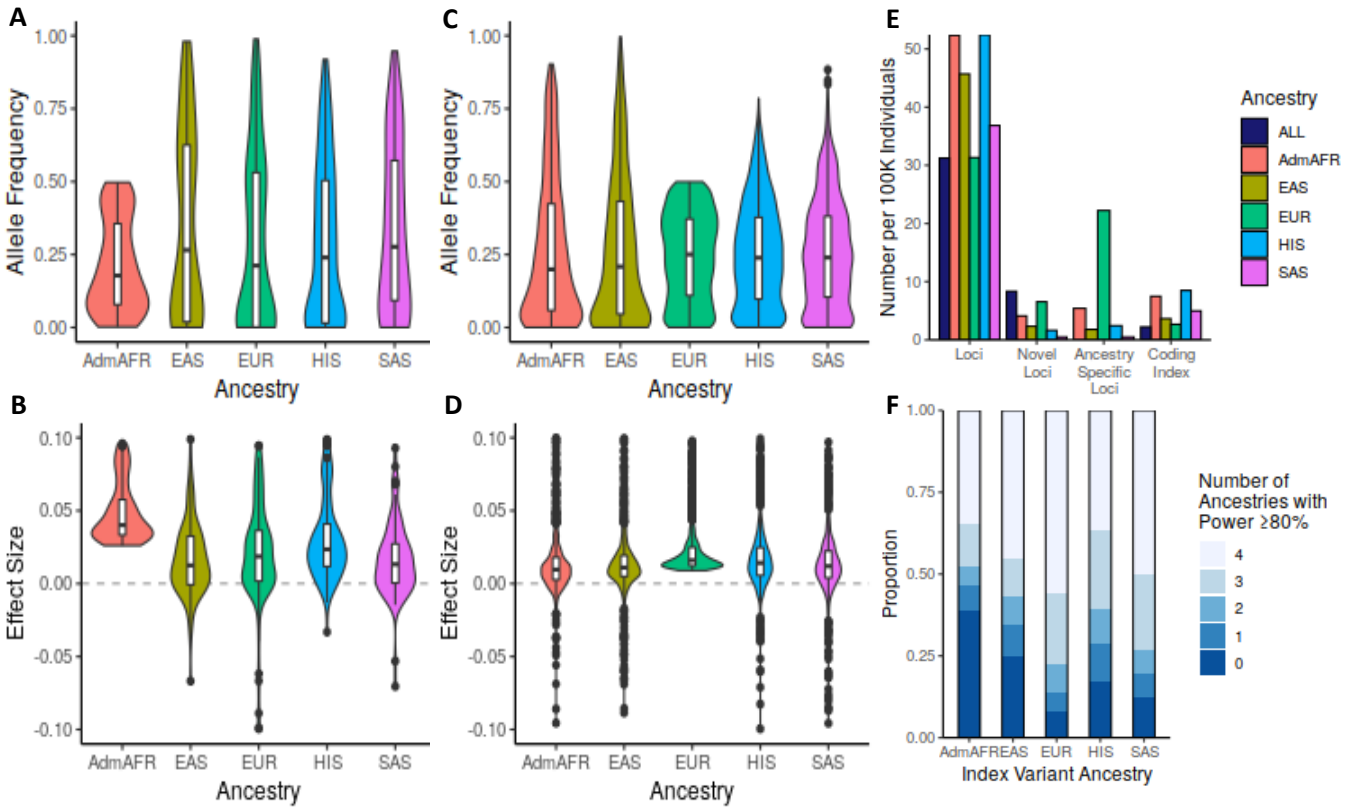

Figure 1: Comparison of identified loci across ancestry groups
A) Allele frequency of African-ancestry index variants. B) Effect size of African-ancestry index variants. C) Allele frequency of European-ancestry index variants. D) Effect size of European-ancestry index variants. The mean effect size of African-ancestry identified index variants is larger than from European-ancestry analysis, reflecting the difference in power to detect an association within each group as a result of the >10-fold difference in sample size. E) Number of loci identified within each ancestry group, normalized to a constant sample size of 100,000 individuals and averaged across lipid traits. At currently available sample sizes, trans-ancestry and European-ancestry analyses identify a lower proportion of loci relative to the number of individuals than analyses of other ancestry groups. However, the larger sample size of European or trans-ancestry analyses leads to a greater relative proportion of novel loci and a higher proportion of loci significant only in European-ancestry analyses. F) Proportion of index variants identified from each ancestry-specific meta-analysis that would be well-powered in other ancestry groups to detect an association. Dark blue regions indicate variants likely to be detected at an equivalent sample size only in the original ancestry group.

Trans-ancestry meta-analysis was then carried out using the meta-regression approach implemented in MR-MEGA[30] to account for heterogeneity in variant effect sizes on lipids between ancestry groups. This identified a total of 923 loci (± 500 kb regions), containing 1,750 unique index variants across all lipid traits, including 168 additional regions (71% novel) not identified by ancestry-specific analysis (**Supplementary Tables 5-6**, **Supplementary Figures 4-5**). The majority (98%) of index variants from ancestry-specific analysis remained significant

(p-value < $5x10^{-8}$) after meta-analysis across all ancestry groups, although 15 AdmAFR, 9 EAS, 3 HIS, and 1 SAS index variants from ancestry-specific analysis did not (trans-ancestry p-value $7.7x10^{-6}$ to $5.9x10^{-8}$, **Supplementary Figure 6**). In total, we identified 355 novel loci between single- and trans-ancestry analyses. Next, we compared the number of loci identified between ancestry groups per 100,000 participants (**Figure 1**). The European and trans-ancestry analyses identified the greatest absolute number of total and novel loci, reflecting their large sample size. However, analyses of non-European ancestry groups identified a greater number of associations relative to the number of included individuals and African-ancestry analyses identified the third highest proportion of novel loci, reflecting the value of including diverse groups even if available sample sizes are smaller. We then estimated the proportion of index variants within each ancestry group that would likely be identified from analyses of an equivalent sample size in other ancestry groups (**Figure 1**). Of the index variants that were significant in the African-ancestry meta-analysis, we estimate that 39% would not reach significance in a meta-analysis of an equivalent number of individuals in any other ancestry group, due to an estimated power of <80%. To further examine trends in the number of loci with increasing study sample size, we compared both the number of identified loci and the mean chi-squared value relative to meta-analysis sample size across ancestry groups (**Supplementary Table 7**, **Supplementary Figure 7**). Both the number of identified loci and the mean chi-squared value were generally linearly related to sample size as ancestry groups were meta-analyzed. Within the European ancestry group, however, the incremental increase in either the number of loci or chi-squared value per increase in sample size was attenuated at the largest sample sizes (**Supplementary Figure 7**). This suggests that once sufficiently well-powered sample sizes are reached within a given ancestry group, increasing sample sizes of other previously under-represented groups will most advance GWAS discovery.

Overall, the trans-ancestry index variants explain, for HDL-C, LDL-C, TG, nonHDL-C, and TC, respectively,11.7%, 12.7%, 9.4%, 13.4%, and 11.7% of the variance across all ancestries. Novel index variants explained 0.8% of the variance in each trait on average. The associated regions from trans-ancestry analysis encompassed 13% of the genome overall across all traits, with 7%, 5%, 6%, 5%, and 6% of the genome associated with HDL-C, LDL-C, TG, nonHDL-C, and TC, respectively. However, it is important to note that genes involved in lipid levels can be effective therapeutic targets even if naturally occurring variation, for example in the non-coding region, has a small effect on the trait (e.g. HMGCR and statins).

Comparison in effect sizes across ancestries

Finding different variants across ancestries could be due to variation in allele frequencies -- with some variants being more common to one ancestry group -- and/or variability in effect size estimates, which could be a reflection of differing patterns of LD with the underlying causal variant or interaction with environmental risk factors. We found that effect sizes of the unique index variants from ancestry-specific analysis were largely similar between ancestry groups based on pairwise comparisons, with a consistent direction of effect observed at 73.5-87.5% of variants across the five lipids (**Supplementary Table 4, Supplementary Figure 8A-C**). This increased to 97.3%-99.8% for index variants that were at least nominally significant (p-value < 0.05) in both ancestry groups, with a correlation in effect sizes of 0.92-0.98 for variants with p-value < $5x10^{-8}$ in both groups (**Supplementary Figure 8D-E**). We tested for genome-wide differences in both the genetic effect correlation and the genetic impact correlation, i.e. the genetic effect after accounting for allele frequency differences between populations, using Popcorn[31]. Only African-ancestry effect sizes showed a significantly decreased correlation (p-value < 0.05) with other ancestry groups, with genetic impact correlations ranging from 0.49-0.80 for TC, 0.52-0.77 for LDL-C, 0.68-0.90 for TG, and 0.54-1 for nonHDL-C (**Supplementary Figures 9** and **10)**. HDL-C effect sizes were highly correlated and not significantly different from 1. Considering the variants identified from trans-ancestry meta-analysis (**Supplementary Table 6**), 7% showed significant heterogeneity due to ancestry (p-value < $2.2x10^{-5}$; Bonferroni correction for 2,286 variants across 5 traits). Of these, 23 variants were present only in European ancestry cohorts and showed significant heterogeneity due to fine-scale differences within populations of European ancestry. In addition, 2% of variants from trans-ancestry meta-analysis showed significant residual heterogeneity not due to ancestry, which could arise due to differences in ascertainment or analysis strategy between cohorts (**Supplementary Table 6**).

Trans-ancestry fine-mapping reduces the number of potential causal variants

We next determined 99% credible sets of variants that encompassed the causal variant with 99% posterior probability (PP), for each of the association signals from trans-ancestry meta-analysis. Using these, we assessed the contribution of non-European ancestries to narrow the set of likely causal variants. The 1,750 unique index variants were grouped into 1,486 independent association signals based on an LD r^2 threshold of 0.7 (using all 1000 Genomes ancestries) to avoid double-counting overlapping association signals. We then considered the smallest credible set across any of the traits for the ± 500 kb region surrounding each of these 1,486 variants, assuming a single causal variant per independent signal (**Supplementary Table**

8). Nearly one-fifth of the association signals (19%) had only one variant in the credible set and 55% (816/1,486) had 10 or fewer. In contrast, 5% (73/1486) had more than 100 variants. As the LD patterns in Africans are expected to be most different from those in other ancestries, we next focused on comparing the credible sets generated from the African, European (the largest ancestry by sample size), and trans-ancestry analysis. Of the 353 regions with p-value < $5x10^{-6}$ in both African and European ancestry-specific analyses, 47% of association signals had the smallest credible set in the trans-ancestry analysis. Considering all regions, a reduction in credible set size in the trans-ancestry analysis was observed at 56% (825/1,486) of association signals compared to the smallest credible set in African or European analyses. At these signals with improved fine-mapping, we observed a median 40% reduction in credible set size. This difference was similar (39% reduction) after recreating the European and trans-ancestry credible sets using only variants present in the African-ancestry analysis, indicating that variants missing from one ancestry were not driving this reduction. These credible sets included a median of 26 variants in European and just 13 after trans-ancestry meta-analysis (**Supplementary Figure 11**). For example, the *FADS1/FADS2* TG locus contained 27 variants in the AdmAFR credible set, 8 in EUR, and a single *FADS2* intronic variant in the trans-ancestry analysis (rs174564, **Supplementary Figure 12**), which has been previously reported as the likely causal variant at this locus[32]. While these findings are at least partially attributed to differences in sample size, we also anticipate that differences in linkage disequilibrium patterns between multiple ancestries improves fine-mapping. After adjusting the Bayes factors to reflect the increased sample size of the trans-ancestry meta-analysis, we estimate that an equivalent number of European individuals would yield a 20% reduction in credible set size, only half of the reduction observed from trans-ancestry analysis. This supports the notion that the efficiency of fine-mapping was strongly driven by the increase in non-European ancestry, narrowing the set of likely causal variants two-fold more than expected by the increase in sample size. For example, across African, European, and trans-ancestry analysis of LDL-C, we identify the *DMTN* intron variant rs900776 as the most significantly associated variant in the region. This variant has PP of 0.86 in the African-ancestry derived credible sets, >0.99 in the trans-ancestry analysis, and just 0.51 in the European specific analysis (**Figure 2**). We estimate that the PP of this variant would reach only 0.59 if the European sample size were increased to match that of the trans-ancestry analysis, due to its strong LD with nearby variants.

Considering the 407 variants with >90% PP of being the causal variant at a locus in the trans-ancestry meta-analysis, 56 (14%) were missense variants, 4 (1%) were stop-gain variants (*CD36, HBB, ANGPTL8, PDE3B*), 7 (2%) were splice-region variants, and 63 (15%) were

significant eQTL variants (p-value < $5x10^{-8}$) within a colocalized eQTL signal (**Supplementary Tables 9-10**). Of these, 23 variants were also associated with CAD (p-value < 0.05, **Table 2, Supplementary Table 11**). While the previous comparisons between ancestries assume a single, shared causal variant, some loci may have additional ancestry-specific causal variants. We further examined the variants with high PP in the African-specific analysis for comparison with those from trans-ancestry analysis. African-specific analysis identified additional independent variants including the *PCSK9* p.Cys679Ter stop-gain variant[33] and the missense variants *ZBTB42* p.Glu232Lys*, APOA5* p.Ser19Trp[33]*, HOOK2* p.His488Gln*,* and *ABCG5* p.Lys262Arg. In some cases, these variants may more directly pinpoint the causal gene in the region. For example, the African-specific credible set for TG included a single 3'UTR indel variant within *ANGPTL3*, the likely causal gene, while the trans-ancestry credible set contained a single variant downstream of *DOCK7* (**Supplementary Figure 13**). These findings highlight the importance of including genetically diverse individuals in population genetic studies to query a more complete spectrum of the genetic variation that exists in the global population.

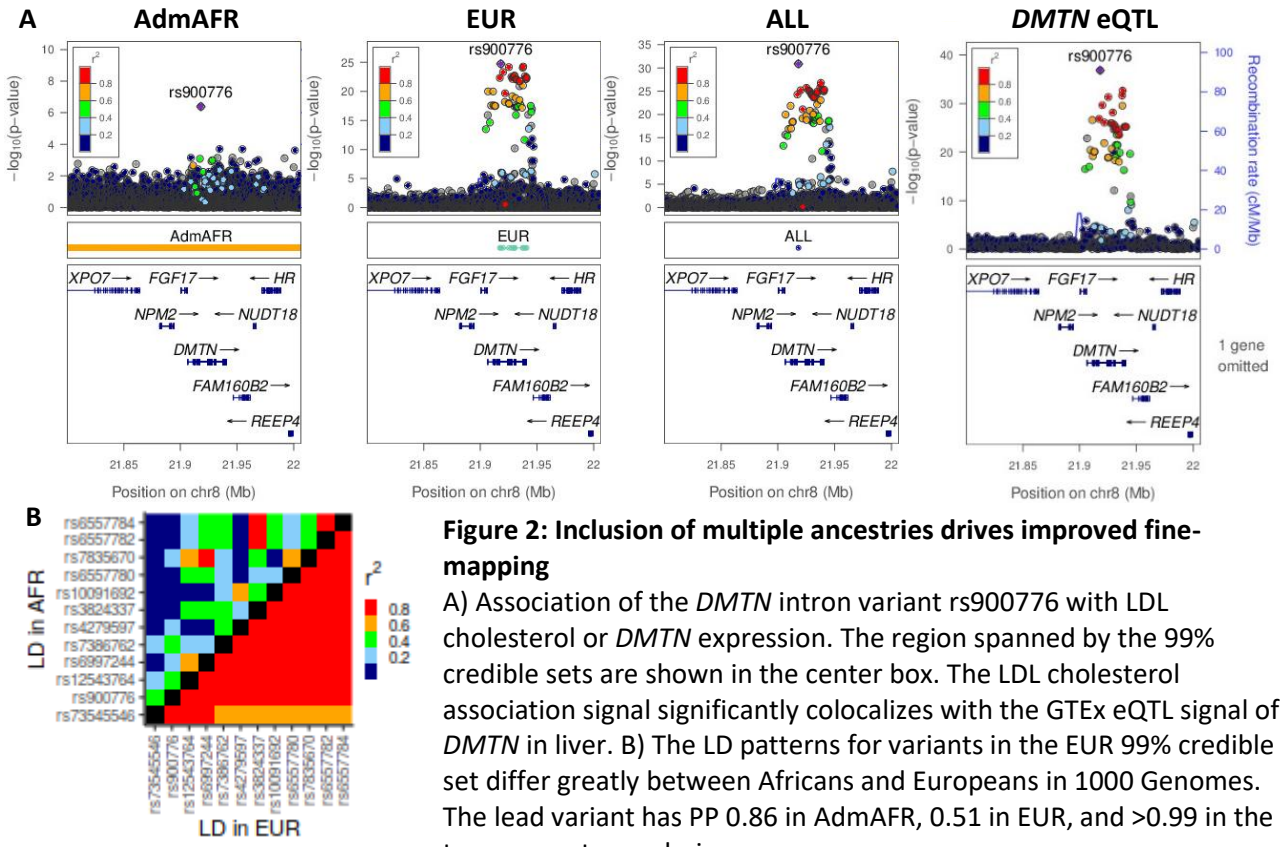

Figure 2: Inclusion of multiple ancestries drives improved fine-mapping

A) Association of the *DMTN* intron variant rs900776 with LDL cholesterol or *DMTN* expression. The region spanned by the 99% credible sets are shown in the center box. The LDL cholesterol association signal significantly colocalizes with the GTEx eQTL signal of *DMTN* in liver. B) The LD patterns for variants in the EUR 99% credible set differ greatly between Africans and Europeans in 1000 Genomes. The lead variant has PP 0.86 in AdmAFR, 0.51 in EUR, and >0.99 in the trans-ancestry analysis.

Table 2: Association of likely causal lipid variants with CAD, NAFLD, and T2D

rsID	Variant	EA	Most significant lipid trait	Lipid effect	CAD effect	CAD p-value	NAFLD effect	NAFLD p-value	T2D effect	T2D p-value
rs7412	APOE p.R176C	T	LDL-C	-0.52	-0.14	$1.3x10^{-27}$	0.02	0.79	0.01	0.55
rs429358	APOE p.C130R	C	LDL-C*	0.21	0.09	$1.9x10^{-23}$	-0.23	$7.2x10^{-5}$	-0.08	$1.8x10^{-18}$
rs3184504	SH2B3 p.W262R	C	TC*	0.03	-0.07	$2.9x10^{-22}$	-0.01	0.83	-0.02	$3.4x10^{-3}$
rs11591147	PCSK9 p.R46L	T	LDL-C	-0.43	-0.25	$8.1x10^{-18}$	-0.15	0.37	0.04	0.16
rs1169288	HNF1A p.I27L; splice variant	C	TC	0.04	0.05	$6.1x10^{-14}$	0.09	0.060	0.05	$7.3x10^{-13}$
rs116843064	ANGPTL4 p.E40K	A	TG	-0.24	-0.16	$3.8x10^{-11}$	-0.04	0.77	-0.10	$2.3x10^{-5}$
rs11601507	TRIM5 p.V112F	A	LDL-C	0.04	0.08	$2.6x10^{-10}$	0.14	0.084	0.01	0.26
rs34931250	ABCA8 splice variant	T	HDL-C	-0.05	0.07	$7.5x10^{-6}$	-0.03	0.72	0.00	0.91
rs4760	PLAUR p.L317P	G	HDL-C	-0.02	0.04	$7.9x10^{-6}$	0.03	0.61	0.01	0.30
rs58542926	TM6SF2 p.E167K	T	TC	-0.12	-0.05	$2.4x10^{-5}$	0.39	$2.7x10^{-6}$	0.09	$2.0x10^{-14}$
rs742493	UNC5CL p.R432G	C	TG	-0.02	-0.04	$2.5x10^{-4}$	-0.10	0.13	-0.03	$4.1x10^{-3}$
rs1265097	PSORS1C1 p.P24T	A	TC*	0.03	0.04	$3.6x10^{-4}$	0.02	0.83	0.03	0.012
rs77960347	LIPG p.N396S	G	HDL-C*	0.25	-0.10	$9.7x10^{-4}$	0.03	0.86	0.02	0.59
rs1047891	CPS1 p.T1406N	A	HDL-C	-0.02	-0.02	$2.5x10^{-3}$	-0.05	0.28	0.00	0.55
rs140201358	PNPLA2 p.N252K; splice variant	G	HDL-C	-0.06	0.10	$4.0x10^{-3}$	0.09	0.60	0.06	0.034
rs72836561	CD300LG p.R82C	T	HDL-C	-0.19	0.06	$4.6x10^{-3}$	0.13	0.29	0.02	0.40
rs1801689	APOH p.C325G	C	LDL-C	0.09	0.06	$9.0x10^{-3}$	0.02	0.89	0.02	0.40
rs235314	PTTG1IP p.A87T	T	HDL-C	-0.02	0.02	0.015	-0.02	0.58	-0.01	0.30
rs1260326	GCKR p.L446P; splice variant	C	TG	-0.11	-0.02	0.021	-0.17	$7.3x10^{-5}$	0.07	$1.3x10^{-24}$
rs77542162	ABCA6 p.C1359R	G	LDL-C	0.18	0.05	0.026	-0.20	0.15	-0.06	0.019
rs41274050	A1CF p.G390S	T	TG	0.09	0.08	0.037	-0.08	0.72	-0.08	0.031
rs114816312	PLA2G12A p.D111N	T	TG	0.13	0.08	0.041	-0.31	0.20	0.06	0.15
rs34311866	TMEM175 p.M393T	C	TG	0.02	0.02	0.044	-0.01	0.82	0.02	$4.5x10^{-3}$

EA: Effect allele. Variant annotation from SnpEff and VEP with numbering from dbSNP. NAFLD: non-alcoholic fatty liver disease (UK Biobank). T2D: type 2 diabetes (DIAGRAM). CAD (CARDIoGRAMplusC4D+HUNT+UK Biobank). Lipid effect sizes are in standard deviation units, NAFLD, T2D, and CAD are in units of log odds. *Denotes variant with PP > 0.9 in a trait other than the most significantly associated.

Polygenic scores are more predictive when derived from multi-ancestry GWAS

We next evaluated the potential of polygenic scores to predict individuals at risk of CAD through elevated LDL-C in blood. We created three datasets with no overlapping individuals to i) perform GWAS to estimate the effect sizes of variants, ii) optimize risk score parameters, and iii) evaluate the utility of the score. Using LDL-C meta-analysis results excluding individuals from the UK Biobank and Michigan Genomics Initiative (MGI) used for score optimization and validation, we generated genome-wide polygenic scores using either ancestry-specific or trans-

ancestry meta-analysis results (**Supplementary Table 12, Supplementary Figure 14**). We first optimized the scores in ancestry-matched UK Biobank (AdmAFR, EAS, EUR, SAS, ALL) or MGI participants (HIS, **Supplementary Figure 15, Supplementary Table 13**). We then evaluated the optimized PRS scores in 8 cohorts of individuals (N=295,577, **Supplementary Table 14**) not included in the discovery GWAS, from 6 ancestral groups: Continental Africa (2,452 East Africa, 4,972 South Africa, 7,309 West Africa), African American (21,730), Hispanic American (7,669), East Asian (146,477), South Asian (15,242), Asian American (4,155), and European American (85,571).

The polygenic score developed from trans-ancestry meta-analysis consistently showed the best or near-best performance across ancestry groups (**Figure 3, Supplementary Table 15**), with improved or equivalent prediction relative to ancestry-matched scores. This was especially evident when limited sample sizes of ancestry-matched cohorts were available in the GWAS, as was the case for HIS and SAS. In contrast, prediction in South and East African individuals was slightly lower (by 0.01-0.02, **Supplementary Table 15**) for the trans-ancestry score relative to the African-ancestry derived score within each of the AADM and AWI-Gen regional subgroups. We further assessed the prediction of polygenic scores relative to the proportion of African admixture in MGI and MVP participants. PRS prediction was similar for participants across all quartiles of African ancestry proportions (**Supplementary Figure 16**). Previous studies have suggested that population stratification may influence the predictive ability of polygenic scores across diverse populations[29]. We tested for correlation between the trans-ancestry polygenic score and principal components (PCs) in 1000 Genomes individuals. Significant correlation was observed between the trans-ancestry polygenic score and principal components 1 and 3 only (p-value < 0.0025; 0.05/20 tested PCs), indicating that the minimum of 4 PCs used in the polygenic score analysis were sufficient to correct for this (**Supplementary Table 16**).

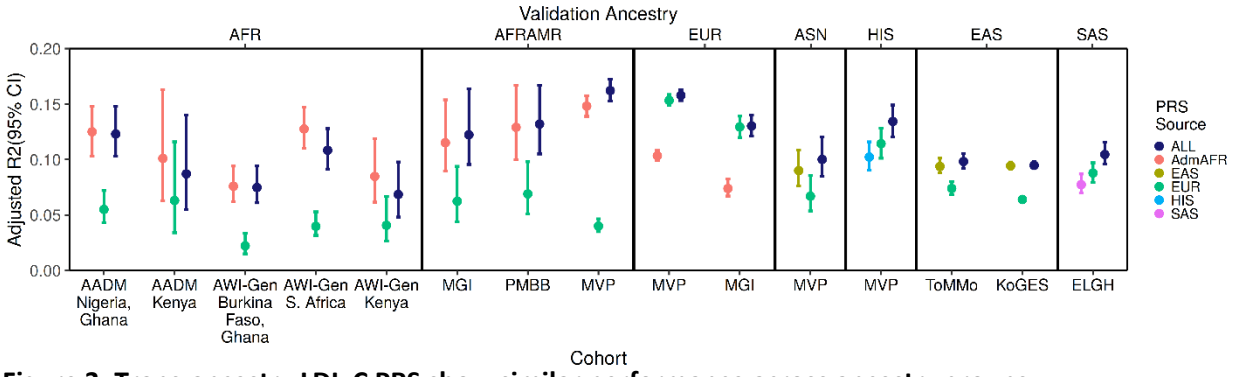

Figure 3: Trans-ancestry LDL-C PRS show similar performance across ancestry groups
Polygenic scores generated from trans-ancestry meta-analysis show equivalent or better performance across most ancestry groups relative to ancestry-specific PRS within each cohort, whereas European-specific scores show less transferability. Adjusted R^2 is calculated with the risk score as a predictor of LDL-C in a linear model with covariates. AFR: African, AFRAMR: African American, ASN: Asian American

We then examined the role of GWAS sample size using the 2010 Global Lipids Genetics Consortium (EUR N=95,454, imputed with HapMap) LDL-C meta-analysis results[5] alone or meta-analyzed with the present AdmAFR or non-European ancestry results. In addition, we generated subsets of the present European ancestry meta-analysis to yield 100K, 200K, or 400K individuals. For European ancestry individuals, we observed a 32% increase in the predictive power of LDL-C polygenic scores (as measured by adjusted R^2) with the nine-fold increase in sample size between the 2010 and present European-specific PRS (adjusted R^2 = 0.12 and 0.15, respectively). However, the prediction of LDL-C in African-American individuals was relatively weak from an entirely European ancestry GWAS, irrespective of the sample size (adjusted R^2 = 0.04, **Supplementary Figure 17, Supplementary Table 15**). We next aimed to investigate the role of variant selection and effect size estimate variability in PRS performance between ancestry groups. We found that European ancestry GWAS-derived scores could be tuned to have improved prediction in African-American individuals by optimizing the score parameters based on testing in Africans rather than Europeans and substituting African ancestry-derived weights for the included variants (**Supplementary Figure 17**). However, the polygenic score derived from trans-ancestry meta-analysis was most predictive in all ancestries examined and performed similarly across most ancestry groups (adjusted R^2 ~ 0.10-0.16), with the trans-ancestry-derived PRS explaining an equal proportion of variance of LDL-C measurement in African American (0.16) and European-ancestry (0.16) individuals. Within the Million Veteran Program, the variance explained by one standard deviation increase in PRS corresponded to a similar increase in LDL-C in 4 separate ancestry groups: 13.2 mg/dL for

African Americans, 8.9 mg/dL for Asians (EAS/SAS), 10.5 mg/dL for Europeans, and 10.6 mg/dL for Hispanics.

Discussion

Genome-wide discovery for blood lipid traits involving 1.65 million individuals from five ancestry groups confirms that the contributions of common genetic variation to blood lipids are largely similar across diverse populations. This study also empirically demonstrates the substantial benefits of including a more diverse set of individuals in GWAS. As sample sizes of GWAS increase, the benefits from genetic discovery efforts will likely not be measured by the number of loci discovered. Rather, the focus will turn to improving our understanding of the biology at established loci, identifying potential therapeutic targets at these loci, and efficiently identifying individuals at high-risk of adverse health outcomes without exacerbating existing health disparities. The present results suggest that diversifying the population under study, rather than simply increasing the sample size, will be the single most efficient approach to achieving these goals for blood lipids and other heritable traits and diseases.

We evaluated the impact of genetic diversity on blood lipid GWAS based on three factors: i) the number of association signals detected, ii) statistical fine-mapping of association signals, and iii) the predictive power of a polygenic score. We found that the number of identified significant loci relative to sample size was similar within each ancestry group, approximately linearly related to sample size, with a small increase in ancestry-specific variants observed in African-ancestry cohorts relative to others. We also demonstrated empirically that including additional ancestries through trans-ancestry fine-mapping reduces the set of potential causal variants and may improve the identification of causal genes underlying the GWAS signals. In turn, this would allow for an expedited focus on potential drug targets and drive biological insight from these results. For example, *UNC5CL* p.Arg432Gly was identified as a likely causal variant (PP > 0.99) associated with decreased TG levels that also exhibited a favorable impact on risk for CAD, type 2 diabetes, and non-alcoholic fatty liver disease. Finally, we found that a polygenic score derived from ~88k African-ancestry and ~830k European-ancestry individuals correlated with observed lipid levels as well in African-ancestry Americans as it did in Americans with solely European ancestry. It is important to note that normalization of the PRS into percentiles should be performed within similar genetic ancestry groups. We hypothesize that the inclusion of African-ancestry individuals to the polygenic score yields improvement in a similar way to that observed in fine-mapping, by focusing the set of strongly-associated variants chosen for inclusion to those more likely to be causal. A key finding from this work is that the trans-

ancestry score was the most informative score across all major population groups examined. This provides useful information for other genetic discovery efforts and investigations of the utility of the PRS in diverse individuals.

However, the LDL-C polygenic score showed greater variability in prediction of LDL-C for cohorts within Africa than it did it African Americans. Mean lipid levels within each cohort also exhibited greater variation between the continental African cohorts compared to all other ancestry groups (AFR range: 71.9-131.5 mg/dL, all others: 113.1-129.2 mg/dL). This suggests that sufficient representation of individuals and genetic variation from the risk prediction population improves performance, even for trans-ancestry derived scores. Therefore, additional studies are needed to better understand both the genetic and environmental factors influencing LDL levels within African populations and elsewhere. This underscores the need to ensure that future GWAS studies are representative of global populations.

GWAS have benefitted from substantially increased sample sizes in recent years, which has improved power to detect rare variation and identified considerably higher numbers of common variants with smaller effects. In light of these results, and those of related studies[34-36], we believe future genetic studies will greatly benefit from inclusion of participants of diverse ancestry. Further gains in the depth and number of sequenced individuals of diverse ancestries[37,38] available for array design and imputation into diverse cohorts may additionally improve discovery of novel variants and loci in diverse cohorts, particularly variants absent from other ancestries. As the cost of genotyping and sequencing continues to drop, we may soon find studies limited not by the cost of genetic technologies but rather the genetic diversity of previously collected study participants. Careful thought must be given moving forward on the optimal approach for recruitment of study participants into biorepository and other genetic research studies. We suggest that the primary focus of future recruitment efforts (i.e. DNA collection) should be the enrollment of non-European ancestry participants to increase previously underrepresented populations in genetic studies. In this respect, the results of this study strongly support recent efforts such as the All of Us program and H3Africa to extensively diversify genetic studies and provide compelling evidence that these efforts will lead to accelerated biological and pharmaceutical translation of findings from genetic discovery studies, and possibly a reduction in existing health disparities through the application of genetic risk prediction algorithms to clinical practice that perform equally well across all major race/ethnic groups.

Data Availability and Code Availability

The GWAS meta-analysis results (including both ancestry-specific and trans-ancestry analyses) and risk score weights are available at: http://csg.sph.umich.edu/willer/public/XXXXXXXX. All custom scripts used for analysis and summary of results are available upon request. The optimized trans-ancestry and single-ancestry PRS weights will be deposited within the PGS Catalog (https://www.pgscatalog.org/).

Acknowledgments

Funding for the Global Lipids Genetics Consortium was provided by the NIH (R01-HL127564). This research has been conducted using the UK Biobank Resource under application number 24460. Computing support and file management for central meta-analysis by Sean Caron is gratefully acknowledged. This research is based on data from the Million Veteran Program, Office of Research and Development, Veterans Health Administration, and was supported by awards #2I01BX003362-03A1 and 1I01BX004821-01A1#. This publication does not represent the views of the Department of Veteran Affairs or the United States Government. Study-specific acknowledgements are provided in the supplemental material.

Methods:

Cohort level analysis

Each cohort contributed summary statistics, imputation quality statistics, and analysis metrics for QC, following a detailed analysis plan (**Supplementary File 1**). Briefly, we requested that each cohort perform imputation to 1000 Genomes Phase 3 (1KGP3), with European cohorts additionally imputing with the Haplotype Reference Consortium (HRC) panel using the Michigan Imputation Server (https://imputationserver.sph.umich.edu/index.html#!) which uses Minimac software[39]. Detailed pre-imputation QC guidelines were provided, these included removing samples with call rate < 95%, samples with heterozygosity > median + 3(interquartile range), PCA outliers within each ancestry group, and variants deviating from Hardy-Weinberg equilibrium (p-value < 10^{-6}) or with variant call rate < 98%. Analysis was to be carried out separately in each ancestry group and stratified by cases and controls where appropriate (i.e., based on ascertainment). Residuals were generated separately in males and females adjusting for age, age^2, principal components of ancestry, and any necessary study-specific covariates. Triglyceride levels were natural-log transformed before generating residuals. Inverse normalization was then done on the residual values. Individuals on cholesterol lowering medication had their pre-medication levels[40] approximated by dividing the LDL-C value by 0.7 and the TC value by 0.8. Association analysis of the residuals for the majority of cohorts was carried out using a linear mixed-model approach in rvtests or with other similar software including BOLT-LMM[41], SAIGE[42], or deCode association software.

Quality Control

Each input file was assessed for quality control using the EasyQC software[43]. We generated QQ plots by frequency bins, assessed trends in standard errors relative to sample size for each cohort, and checked allele frequency of submitted variants relative to their expected value based on the imputation reference panel. In addition, we checked that each cohort reproduced the expected direction of effect at the majority of known loci relative to the cohort sample size. Cohorts identified to have issues with the submitted files were contacted and corrected files were submitted or the cohort was excluded from meta-analysis. In some cases, cohorts had generated residuals of males and females together, without accounting for sex. In these cases, sex-stratified files were used for meta-analysis. During the QC process, within each cohort we removed poorly imputed variants (info score or r^2 < 0.3), variants deviating from Hardy-Weinberg Equilibrium (HWE p-value < $1x10^{-8}$, except for MVP which used

HWE p-value < $1x10^{-20}$), and variants with minor allele count < 3. An imputation info score threshold of 0.3 was selected in order to balance the inclusion of variants across diverse studies while removing poorly imputed variants. Summary statistics were then genomic-control (GC) corrected using the λ_{GC} value calculated from the median p-value of variants with MAF > 0.5%. To capture all potential variants, summary statistics from cohorts that had submitted both HRC and 1KGP3 imputed files were joined, favoring variants imputed from HRC. For variants imputed by both panels, we observed that variants imputed from the HRC panel resulted in a higher imputation info score for 94% of variants when compared to the imputation info score from 1KGP3.

Meta-analysis

Ancestry-specific meta-analysis was performed using RAREMETAL[44]. Trans-ancestry meta-analysis was performed using MR-MEGA[45] with 5 principal components of ancestry. The choice of 5 principal components was made after comparing the λ_{GC} values across minor allele frequency bins from meta-analysis of HDL-C with MR-MEGA using from 2 up to 10 principal components. In addition, fixed-effects meta-analysis was carried out with METAL[46] to calculate effect sizes for use in the creation of polygenic scores. Study-level principal components were plotted for each cohort by ancestry group to verify that the reported ancestry for each cohort was as expected. Following meta-analysis, we identified loci based on a genome-wide significance threshold of $5x10^{-8}$ after GC correction using the λ_{GC} value calculated from the median p-value of variants with MAF > 0.5%. The choice of double-GC correction was made to be most conservative and to minimize potential false-positive findings. Observed λ_{GC} values were within the expected range for similarly sized studies and are included in **Supplementary Tables 2** and **5**. Index variants were identified following an iterative procedure starting with the most significant variant and grouping the surrounding region into a locus based on the larger of either ± 500 kb or ± 0.25 cM. cM positions were interpolated using the genetic map distributed with Eagle v2.3.2 (genetic_map_hg19_withX.txt)[47]. Variants were annotated using WGSA[48] including the summary of each variant from SnpEff[49] and the closest genes for intergenic variants from ANNOVAR[50]. Annotation of variants as known or novel was done based on manual review of previously published variants and with variants reported in the GWAS catalog[30] for any of the studied lipid traits (accessed May 2020, provided as **Supplementary Table 17**). For comparison between ancestries and lipid traits, index variants were grouped into genomic regions starting with the most significantly associated variant and grouping all surrounding index variants within ± 500 kb into a single region. Power to detect an association

in other ancestries was determined using the effect size and sample size of the variant within the original ancestry group and the observed allele frequency from the other ancestry groups with alpha set to $5x10^{-8}$. The proportion of variance explained by each variant was estimated as $2\beta^2(1-f)f$ where β is the effect size from METAL and f is the effect allele frequency. Coverage of the genome by associated genetic regions was calculated using BEDTools[51] for the regions defined by the minimum and maximum position within each locus having p-value < $5x10^{-8}$.

Conditional analysis to flag non-independent index variants

Approximate conditional analysis was performed using rareGWAMA[52] to identify index variants that were shadows of nearby, more significant associations. LD reference populations were taken from UK Biobank specific to AdmAFR, EUR (subset of 40,000), or SAS individuals or from the 1000 Genomes project (1KGP3) for EAS or HIS individuals. Conditional analysis was carried out using the individual cohort level summary statistics as was done for meta-analysis with RAREMETAL. rareGWAMA requires imputation quality scores which were set to 1 for all variants that had previously passed quality control (pre-filtered at imputation info/r^2 > 0.3). The EUR subset of UK Biobank was used as the reference population for the conditional analysis of the trans-ancestry meta-analysis (~80% European). Step-wise conditional analysis was performed sequentially for the index variants within each chromosome ranked by most to least significant. Index variants were then flagged as not independent from other more significant variants if the absolute value of the ratio of the original effect size to the effect size after conditional analysis was greater than the 95th percentile (**Supplementary Figure 18**). This corresponded to a ratio of original to conditional effect size of 1.6 for ancestry-specific conditional analysis and a ratio of 1.7 for the trans-ancestry conditional analysis. The effect sizes from meta-analysis with METAL were used for comparison with the trans-ancestry conditional analysis results. Variants flagged as non-independent were excluded from the summary results in the manuscript and are flagged as non-independent in **Supplementary Tables 3** and **5**.

Genetic correlation

Popcorn was used to assess the degree of correlation in effect sizes between ancestry groups for each of the lipid traits. 1000 Genomes phase 3 was used as the reference LD panel. Only variants with MAF > 0.01 in each ancestry individually were included in the comparison. Both the genetic effect and genetic impact models were tested.

Credible sets

Credible sets of potentially causal variants were generated for each of the loci identified in the trans-ancestry meta-analysis. Regions for construction of credible sets were defined as the ± 500 kb region around each index variant. Bayes factors[53,54] (BF) for each variant in the ancestry-specific meta-analysis were approximated by:

$$ln(BF) \propto 0.5 \frac{\beta^2}{SE^2}$$

where β and SE are the effect sizes and standard errors from the RAREMETAL meta-analysis. To account for the difference in sample sizes between ancestry groups, we additionally approximated the Bayes factors after adjustment for the total trans-ancestry sample size for each trait (N_{TE}) relative to the ancestry-specific sample size for that trait (N_{AS}) using the following equation:

$$ln(BF) \propto 0.5 \frac{\beta^2 N_{TE}}{SE^2 N_{AS}}$$

Credible sets for the trans-ancestry meta-analysis were generated using the Bayes factors as output by MR-MEGA. The credible sets within each region were generated by ranking all variants by Bayes factor and calculating the number of variants required to reach a cumulative probability of 99%. In addition, we calculated credible sets in the same manner using the European and trans-ancestry meta-analysis results but including only the set of variants present in the AdmAFR meta-analysis. To summarize the size of the credible sets across the 5 lipid traits examined, we identified the set of independent index variants from the trans-ancestry meta-analysis after grouping variants based on LD. For each ± 500kb region centered around the most-significantly associated index variant for any trait, we determined the pairwise LD between all index variants in this region using LDpair[55] with all reference populations (1000 Genomes AFR, AMR, EAS, EUR, and SAS) included. We considered variants to be independent if they were outside of this region, had LD $r^2 < 0.7$, or were not available in the LDpair reference populations. Variants within the credible sets were annotated with SnpEff[49] using WGSA[48] and with VEP[56]. Protein numbering was taken from dbSNP[57].

eQTL colocalization was performed using coloc[58] version 3.2.1 with R version 3.4.3 using the default parameters. Results from GTEx V8[59] were compared with the GWAS signals in the region defined by the larger of ±0.25cM or ±500kb surrounding each index variant. The

eQTL and GWAS signals (based on p-values from MR-MEGA) were considered to be colocalized if PP3 + PP4 ≥ 0.8 and if PP4/(PP3+PP4) > 0.9, where PP3 is the probability of two independent causal variants while PP4 is the probability of a single, shared causal variant. Association of variants with CAD was assessed based on a fixed-effects meta-analysis of results from the CARDIoGRAMplusC4D 1000 Genomes-based GWAS[60], GWAS of CAD in the HUNT Biobank, and GWAS of IHD (phecode 411) in UK Biobank. Type 2 diabetes association results were taken from the 2018 DIAGRAM consortium results[61]. GWAS for non-alcoholic fatty liver disease was performed in the white British subset of UK Biobank using sex, birth year, and PC1-4 as covariates using SAIGE.

LDL-C polygenic scores

Weights for the LDL-C polygenic scores were derived from beta estimates generated from each of the ancestry-specific meta-analyses and from the trans-ancestry results using METAL. Additional meta-analyses were carried out using the 2010 Global Lipids Genetics Consortium LDL-C meta-analysis results[5] in combination with the i) AdmAFR or ii) AdmAFR, EAS, HIS, and SAS results from the present meta-analysis for comparison. Furthermore, we performed a meta-analysis of European cohorts randomly selected to reach a total sample size near 100K, 200K, or 400K to understand the role of increasing European sample size and the influence of imputation panel. In addition, we tested possible methods for improving performance of European-derived scores in African-ancestry individuals by separately fitting the EUR polygenic scores in the UK Biobank AdmAFR subset to determine the best set of risk score parameters (either pruning and thresholding or PRS-CS).

We generated polygenic score weights using both: i) significant variants only (at a variety of p-value thresholds) and ii) using genome-wide methods. Meta-analysis results were first filtered to variants present in UK Biobank, MGI, and MVP with imputation info score > 0.3. Pruning and thresholding was performed in PLINK[62] with ancestry-matched subsets of UK Biobank individuals (AdmAFR N=7,324, EUR N=40,000, SAS N=7,193, trans-ancestry: N=10,000 (80% EUR, 15% AdmAFR, 5% SAS)) or 1KGP3 (HIS N=347 , EAS N=504) used for LD reference. We additionally tested 1000 Genomes phase 3 with all populations included as the LD reference panel for the trans-ancestry score (results not shown), which gave very similar results to those of the UK Biobank trans-ancestry reference set originally selected for its larger sample size. P-value thresholds (after GC correction) of $5x10^{-10}$, $5x10^{-9}$, $5x10^{-8}$, $5x10^{-7}$, $5x10^{-6}$, $5x10^{-5}$, $5x10^{-4}$, $5x10^{-3}$, and $5x10^{-2}$ were tested with distance thresholds of 250 and 500 kb and LD r^2 thresholds of 0.1 and 0.2. Polygenic score weights were also generated using PRS-CS[63]

with the LD reference panels for AFR, EAS, and EUR populations from 1000 Genomes provided by the developers. PRS-CS LD reference panels for the other ancestries were generated using 1000 Genomes following the same protocol as provided by the PRS-CS authors[63]. This included removing variants with MAF ≤ 0.01, ambiguous A/T or G/C variants, and restricting to variants included in HapMap3. Pairwise LD matrices within pre-defined LD blocks[64] (using EUR LDetect blocks for HIS and trans-ancestry LD calculations and ASN blocks for SAS) were then calculated using PLINK and converted to HDF5 format.

For each individual in the PRS testing cohorts, polygenic scores were calculated as the sum of the dosages multiplied by the given weight at each variant. UK Biobank individuals not present in datasets used to generate the summary statistics (either AdmAFR, white British, both AdmAFR and white British, EAS, SAS, or all individuals excluding SAS) were used to select the best performing AdmAFR, EUR, AdmAFR+EUR, EAS, SAS, and trans-ancestry polygenic scores, respectively. UK Biobank SAS individuals were included in the trans-ancestry risk score weights but excluded from the UK Biobank trans-ancestry testing set due to an initial focus on comparing predictions among European- and African-ancestry individuals. Sample sizes of the ancestry groups in UK Biobank used to test PRS performance included: AdmAFR N=6,863; EAS N=1,441; EUR N=389,158; SAS N=6,814; ALL=461,918. The best performing HIS polygenic score weights were selected based on performance in Hispanic individuals in the Michigan Genomics Initiative dataset. Model fit was assessed by the adjusted R^2 of a linear model for LDL-C value at initial assessment adjusted for cholesterol medication (divided by 0.7 to estimate pre-medication levels) with sex, batch, age at initial assessment, and PCs1-4 as covariates.

The best performing polygenic score in each ancestry group was then tested in the validation cohorts: the Michigan Genomics Initiative (EUR N=17,190; AFRAMR N=1,341), East London Genes and Health[65] (ELGH; SAS N=15,242), Tohoku Medical Megabank Community Cohort Study (ToMMo; EAS N=28,217), Korean Genome and Epidemiology Study[66] (KoGES; EAS N=118,260), Penn Medicine BioBank (PMBB; AFRAMR=2,138), Africa America Diabetes Mellitus (AADM; 3,566 West AFR; 707 East AFR), Africa Wits-INDEPTH partnership for Genomic Studies (AWI-Gen; 1,744 East AFR; 4,972 South AFR; 3,744 West AFR) and Million Veterans Program participants not included in the discovery meta-analysis (MVP; EUR N=68,381; AFRAMR N=18,251; EAS/SAS N=4,155; HIS N=7,669). Adjusted R^2 values were reported for each cohort and ancestry group, with 95% confidence intervals for the adjusted R^2 values calculated using bootstrapping. Within each cohort, covariates used were: MGI- sex, batch, PC1-4, and birth year; PMBB- birth year, sex, and PC1-4; ELGH- age, sex, and PC1-10;

MVP- sex, PC1-4, birth year, and mean age; ToMMo-sex, age, recruitment method, and PC1-20 (only participants from Miyagi Prefecture were included); KoGES-age, sex, and recruitment area, AADM-age, sex, PC1-3, AWI-Gen East Africa- age, sex, PC1-6, AWI-Gen South Africa-age, sex, PC1-6, and AWI-Gen West Africa- age, sex, and PC1-4. The type of LDL-C value used in the model varied depending on the measurements selected by each cohort. Mean LDL-C values were used for MGI, MVP and PMBB, maximum LDL-C values for ELGH, and baseline measurements for AADM, AWI-Gen, ToMMo and KoGES. A descriptive summary of each replication cohort is included in **Supplementary Table 14**. African admixture for MGI was calculated using all African-ancestry individuals in 1000 Genomes with ADMIXTURE v1.3[67]. African admixture for MVP was calculated using the YRI and LWK African-ancestry individuals in 1000 Genomes.

References

1. Taddei, C. *et al.* Repositioning of the global epicentre of non-optimal cholesterol. *Nature* **582**, 73-77 (2020).
2. Ference, B.A. *et al.* Low-density lipoproteins cause atherosclerotic cardiovascular disease. 1. Evidence from genetic, epidemiologic, and clinical studies. A consensus statement from the European Atherosclerosis Society Consensus Panel. *European heart journal* **38**, 2459-2472 (2017).
3. Roth, G.A. *et al.* Global, regional, and national age-sex-specific mortality for 282 causes of death in 195 countries and territories, 1980–2017: a systematic analysis for the Global Burden of Disease Study 2017. *The Lancet* **392**, 1736-1788 (2018).
4. Van Dyke, M. *et al.* Heart Disease Death Rates Among Blacks and Whites Aged ≥35 Years - United States, 1968-2015. *MMWR Surveill Summ* **67**, 1-11 (2018).
5. Teslovich, T.M. *et al.* Biological, clinical and population relevance of 95 loci for blood lipids. *Nature* **466**, 707-13 (2010).
6. Willer, C.J. *et al.* Discovery and refinement of loci associated with lipid levels. *Nat Genet* **45**, 1274-1283 (2013).
7. Liu, D.J. *et al.* Exome-wide association study of plasma lipids in >300,000 individuals. *Nat Genet* **49**, 1758-1766 (2017).
8. Lu, X. *et al.* Exome chip meta-analysis identifies novel loci and East Asian-specific coding variants that contribute to lipid levels and coronary artery disease. *Nat Genet* **49**, 1722-1730 (2017).
9. Kathiresan, S. *et al.* A genome-wide association study for blood lipid phenotypes in the Framingham Heart Study. *BMC medical genetics* **8 Suppl 1**, S17-S17 (2007).
10. Kathiresan, S. *et al.* Polymorphisms Associated with Cholesterol and Risk of Cardiovascular Events. *New England Journal of Medicine* **358**, 1240-1249 (2008).
11. Peloso, G.M. *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* **94**, 223-32 (2014).
12. Hoffmann, T.J. *et al.* A large electronic-health-record-based genome-wide study of serum lipids. *Nat Genet* **50**, 401-413 (2018).
13. Surakka, I. *et al.* The impact of low-frequency and rare variants on lipid levels. *Nat Genet* **47**, 589-97 (2015).
14. Klarin, D. *et al.* Genetics of blood lipids among ~300,000 multi-ethnic participants of the Million Veteran Program. *Nature Genetics* **50**, 1514-1523 (2018).
15. Holmen, O.L. *et al.* Systematic evaluation of coding variation identifies a candidate causal variant in TM6SF2 influencing total cholesterol and myocardial infarction risk. *Nat Genet* **46**, 345-51 (2014).
16. Asselbergs, F.W. *et al.* Large-scale gene-centric meta-analysis across 32 studies identifies multiple lipid loci. *Am J Hum Genet* **91**, 823-38 (2012).
17. Albrechtsen, A. *et al.* Exome sequencing-driven discovery of coding polymorphisms associated with common metabolic phenotypes. *Diabetologia* **56**, 298-310 (2013).
18. Saxena, R. *et al.* Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels. *Science* **316**, 1331-6 (2007).
19. Iotchkova, V. *et al.* Discovery and refinement of genetic loci associated with cardiometabolic risk using dense imputation maps. *Nat Genet* **48**, 1303-1312 (2016).
20. Tachmazidou, I. *et al.* A rare functional cardioprotective APOC3 variant has risen in frequency in distinct population isolates. *Nature Communications* **4**, 2872 (2013).

21. Tang, C.S. *et al.* Exome-wide association analysis reveals novel coding sequence variants associated with lipid traits in Chinese. *Nature Communications* **6**, 10206 (2015).
22. van Leeuwen, E.M. *et al.* Genome of the Netherlands population-specific imputations identify an ABCA6 variant associated with cholesterol levels. *Nature Communications* **6**, 6065 (2015).
23. Spracklen, C.N. *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* **26**, 1770-1784 (2017).
24. Kanai, M. *et al.* Genetic analysis of quantitative traits in the Japanese population links cell types to complex human diseases. *Nat Genet* **50**, 390-400 (2018).
25. Sirugo, G., Williams, S.M. & Tishkoff, S.A. The Missing Diversity in Human Genetic Studies. *Cell* **177**, 26-31 (2019).
26. Cohen, J. *et al.* Low LDL cholesterol in individuals of African descent resulting from frequent nonsense mutations in PCSK9. *Nature Genetics* **37**, 161-165 (2005).
27. Tishkoff, S.A. *et al.* The genetic structure and history of Africans and African Americans. *Science (New York, N.Y.)* **324**, 1035-1044 (2009).
28. Khera, A.V. *et al.* Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations. *Nature Genetics* **50**, 1219-1224 (2018).
29. Duncan, L. *et al.* Analysis of polygenic risk score usage and performance in diverse human populations. *Nature Communications* **10**, 3328 (2019).
30. Buniello, A. *et al.* The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic Acids Res* **47**, D1005-d1012 (2019).
31. Brown, B.C., Ye, C.J., Price, A.L. & Zaitlen, N. Transethnic Genetic-Correlation Estimates from Summary Statistics. *Am J Hum Genet* **99**, 76-88 (2016).
32. Çalışkan, M. *et al.* Genetic and Epigenetic Fine Mapping of Complex Trait Associated Loci in the Human Liver. *The American Journal of Human Genetics* **105**, 89-107 (2019).
33. Wu, Y. *et al.* Trans-ethnic fine-mapping of lipid loci identifies population-specific signals and allelic heterogeneity that increases the trait variance explained. *PLoS Genet* **9**, e1003379 (2013).
34. Cavazos, T.B. & Witte, J.S. Inclusion of Variants Discovered from Diverse Populations Improves Polygenic Risk Score Transferability. *bioRxiv*, 2020.05.21.108845 (2020).
35. Wojcik, G.L. *et al.* Genetic analyses of diverse populations improves discovery for complex traits. *Nature* **570**, 514-518 (2019).
36. Bentley, A.R. *et al.* Multi-ancestry genome-wide gene–smoking interaction study of 387,272 individuals identifies new loci associated with serum lipids. *Nature Genetics* **51**, 636-648 (2019).
37. Taliun, D. *et al.* Sequencing of 53,831 diverse genomes from the NHLBI TOPMed Program. *bioRxiv*, 563866 (2019).
38. Kowalski, M.H. *et al.* Use of >100,000 NHLBI Trans-Omics for Precision Medicine (TOPMed) Consortium whole genome sequences improves imputation quality and detection of rare variant associations in admixed African and Hispanic/Latino populations. *PLOS Genetics* **15**, e1008500 (2019).
39. Das, S. *et al.* Next-generation genotype imputation service and methods. *Nature Genetics* **48**, 1284-1287 (2016).
40. Efficacy and safety of cholesterol-lowering treatment: prospective meta-analysis of data from 90 056 participants in 14 randomised trials of statins. *The Lancet* **366**, 1267-1278 (2005).
41. Loh, P.-R. *et al.* Efficient Bayesian mixed-model analysis increases association power in large cohorts. *Nature Genetics* **47**, 284-290 (2015).
42. Zhou, W. *et al.* Efficiently controlling for case-control imbalance and sample relatedness in large-scale genetic association studies. *Nat Genet* **50**, 1335-1341 (2018).

43. Winkler, T.W. *et al.* Quality control and conduct of genome-wide association meta-analyses. *Nature protocols* **9**, 1192-1212 (2014).
44. Feng, S., Liu, D., Zhan, X., Wing, M.K. & Abecasis, G.R. RAREMETAL: fast and powerful meta-analysis for rare variants. *Bioinformatics (Oxford, England)* **30**, 2828-2829 (2014).
45. Mägi, R. *et al.* Trans-ethnic meta-regression of genome-wide association studies accounting for ancestry increases power for discovery and improves fine-mapping resolution. *Human Molecular Genetics* **26**, 3639-3650 (2017).
46. Willer, C.J., Li, Y. & Abecasis, G.R. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics (Oxford, England)* **26**, 2190-2191 (2010).
47. Loh, P.-R., Palamara, P.F. & Price, A.L. Fast and accurate long-range phasing in a UK Biobank cohort. *Nature Genetics* **48**, 811-816 (2016).
48. Liu, X. *et al.* WGSA: an annotation pipeline for human genome sequencing studies. *Journal of Medical Genetics* **53**, 111-112 (2016).
49. Cingolani, P. *et al.* A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. *Fly (Austin)* **6**, 80-92 (2012).
50. Wang, K., Li, M. & Hakonarson, H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Research* **38**, e164-e164 (2010).
51. Quinlan, A.R. & Hall, I.M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841-2 (2010).
52. Liu, D.J. *et al.* Meta-analysis of gene-level tests for rare variant association. *Nature Genetics* **46**, 200-204 (2014).
53. Maller, J.B. *et al.* Bayesian refinement of association signals for 14 loci in 3 common diseases. *Nature Genetics* **44**, 1294-1301 (2012).
54. Schwarz, G. Estimating the Dimension of a Model. *Annals of Statistics* **6**, 461-464 (1978).
55. Machiela, M.J. & Chanock, S.J. LDlink: a web-based application for exploring population-specific haplotype structure and linking correlated alleles of possible functional variants. *Bioinformatics* **31**, 3555-7 (2015).
56. McLaren, W. *et al.* The Ensembl Variant Effect Predictor. *Genome Biology* **17**, 122 (2016).
57. Sherry, S.T. *et al.* dbSNP: the NCBI database of genetic variation. *Nucleic acids research* **29**, 308-311 (2001).
58. Giambartolomei, C. *et al.* Bayesian Test for Colocalisation between Pairs of Genetic Association Studies Using Summary Statistics. *PLOS Genetics* **10**, e1004383 (2014).
59. Aguet, F. *et al.* The GTEx Consortium atlas of genetic regulatory effects across human tissues. *bioRxiv*, 787903 (2019).
60. Nikpay, M. *et al.* A comprehensive 1000 Genomes–based genome-wide association meta-analysis of coronary artery disease. *Nature Genetics* **47**, 1121-1130 (2015).
61. Mahajan, A. *et al.* Fine-mapping type 2 diabetes loci to single-variant resolution using high-density imputation and islet-specific epigenome maps. *Nature Genetics* **50**, 1505-1513 (2018).
62. Purcell, S. *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* **81**, 559-75 (2007).
63. Ge, T., Chen, C.-Y., Ni, Y., Feng, Y.-C.A. & Smoller, J.W. Polygenic prediction via Bayesian regression and continuous shrinkage priors. *Nature Communications* **10**, 1776 (2019).
64. Berisa, T. & Pickrell, J.K. Approximately independent linkage disequilibrium blocks in human populations. *Bioinformatics* **32**, 283-5 (2016).
65. Finer, S. *et al.* Cohort Profile: East London Genes & Health (ELGH), a community-based population genomics and health study in British Bangladeshi and British Pakistani people. *International Journal of Epidemiology* **49**, 20-21i (2019).

66. Moon, S. *et al.* The Korea Biobank Array: Design and Identification of Coding Variants Associated with Blood Biochemical Traits. *Sci Rep* **9**, 1382 (2019).
67. Alexander, D.H., Novembre, J. & Lange, K. Fast model-based estimation of ancestry in unrelated individuals. *Genome research* **19**, 1655-1664 (2009).