



et al. (2021). The power of genetic diversity in genome-wide association studies of lipids. *Nature*, 600, 675–679 (2021).
<https://doi.org/10.1038/s41586-021-04064-3>

Peer reviewed version

Link to published version (if available):
[10.1038/s41586-021-04064-3](https://doi.org/10.1038/s41586-021-04064-3)

[Link to publication record in Explore Bristol Research](#)
PDF-document

This is the accepted author manuscript (AAM). The final published version (version of record) is available online via Nature Research at <https://doi.org/10.1038/s41586-021-04064-3>. Please refer to any applicable terms of use of the publisher.

University of Bristol - Explore Bristol Research

General rights

This document is made available in accordance with publisher policies. Please cite only the published version using the reference above. Full terms of use are available:
<http://www.bristol.ac.uk/red/research-policy/pure/user-guides/ebr-terms/>

Title: The power of genetic diversity in genome-wide association studies of lipids

Authors:

Sarah E Graham¹, Shoa L Clarke^{2,3†}, Kuan-Han H Wu^{4†}, Stavroula Kanoni^{5†}, Greg JM Zajac^{6†}, Shweta Ramdas^{7†}, Ida Surakka¹, Ioanna Ntalla⁸, Sailaja Vedantam^{9,10}, Thomas W Winkler¹¹, Adam E Locke¹², Eirini Marouli⁵, Mi Yeong Hwang¹³, Sohee Han¹³, Akira Narita¹⁴, Ananyo Choudhury¹⁵, Amy R Bentley¹⁶, Kenneth Ekoru¹⁶, Anurag Verma¹⁷, Bhavi Trivedi¹⁸, Hilary C Martin¹⁹, Karen A Hunt¹⁸, Qin Hui^{20,21}, Derek Klarin^{22,23,24}, Xiang Zhu^{25,26,27,28}, Gudmar Thorleifsson²⁹, Anna Helgadottir²⁹, Daniel F Gudbjartsson^{29,30}, Hilma Holm²⁹, Isleifur Olafsson³¹, Masato Akiyama^{32,33}, Saori Sakaue^{34,32,35}, Chikashi Terao³⁶, Masahiro Kanai^{37,38,39}, Wei Zhou^{40,41,42}, Ben M Brumpton^{43,44,45}, Humaira Rasheed^{43,44}, Sanni E Ruotsalainen⁴⁶, Aki S Havulinna^{46,47}, Yogasudha Veturi⁴⁸, QiPing Feng⁴⁹, Elisabeth A Rosenthal⁵⁰, Todd Lingren⁵¹, Jennifer Allen Pacheco⁵², Sarah A Pendergrass⁵³, Jeffrey Haessler⁵⁴, Franco Giulianini⁵⁵, Yuki Bradford⁴⁸, Jason E Miller⁴⁸, Archie Campbell^{56,57}, Kuang Lin⁵⁸, Iona Y Millwood^{58,59}, George Hindy⁶⁰, Asif Rasheed⁶¹, Jessica D Faul⁶², Wei Zhao⁶³, David R Weir⁶², Constance Turman⁶⁴, Hongyan Huang⁶⁴, Mariaelisa Graff⁶⁵, Anubha Mahajan^{66#}, Michael R Brown⁶⁷, Weihua Zhang^{68,69,70}, Ketian Yu⁷¹, Ellen M Schmidt⁷¹, Anita Pandit⁷¹, Stefan Gustafsson⁷², Xianyong Yin⁷¹, Jian'an Luan⁷³, Jing-Hua Zhao⁷³, Fumihiko Matsuda⁷⁴, Hye-Mi Jang¹³, Kyunghoon Yoon¹³, Carolina Medina-Gomez^{75,76}, Achilles Pitsillides⁷⁷, Jouke Jan Hottenga^{78,79}, Gonneke Willemssen^{78,80}, Andrew R Wood⁸¹, Yingji Ji⁸¹, Zishan Gao^{82,83,84}, Simon Haworth^{85,86}, Ruth E Mitchell^{85,87}, Jin Fang Chai⁸⁸, Mette Aadahl⁸⁹, Jie Yao⁹⁰, Ani Manichaikul⁹¹, Helen R Warren^{92,93}, Julia Ramirez⁹², Jette Bork-Jensen⁹⁴, Line L Kårhus⁹⁵, Anuj Goel^{96,97}, Maria Sabater-Lleal^{98,99}, Raymond Noordam¹⁰⁰, Carlo Sidore¹⁰¹, Edoardo Fiorillo¹⁰², Aaron F McDaid^{103,104}, Pedro Marques-Vidal¹⁰⁵, Matthias Wielscher¹⁰⁶, Stella Trompet^{107,108}, Naveed Sattar¹⁰⁹, Line T Møllehave⁸⁹, Betina H Thuesen⁸⁹, Matthias Munz¹¹⁰, Lingyao Zeng^{111,112}, Jianfeng Huang¹¹³, Bin Yang¹¹³, Alaitz Poveda¹¹⁴, Azra Kurbasic¹¹⁴, Claudia Lamina¹¹⁵, Lukas Forer¹¹⁵, Markus Scholz^{116,117}, Tessel E. Galesloot¹¹⁸, Jonathan P. Bradfield¹¹⁹, E Warwick Daw¹²⁰, Joseph M Zmuda¹²¹, Jonathan S Mitchell¹²², Christian Fuchsberger¹²², Henry Christensen¹²³, Jennifer A Brody¹²⁴, Mary F Feitosa¹²⁰, Mary K Wojczynski¹²⁰, Michael Preuss¹²⁵, Massimo Mangino^{126,127}, Paraskevi Christofidou¹²⁶, Niek Verweij¹²⁸, Jan W Benjamins¹²⁸, Jorgen Engmann^{129,130}, Rachel L Kember¹³¹, Roderick C Sliker^{132,133}, Ken Sin Lo¹³⁴, Nuno R Zilhao¹³⁵, Phuong Le¹³⁶, Marcus E Kleber^{137,138}, Graciela E Delgado¹³⁷, Shaofeng Huo¹³⁹, Daisuke D Ikeda¹⁴⁰, Hiroyuki Iha¹⁴⁰, Jian Yang^{141,142}, Jun Liu¹⁴³, Hampton L Leonard^{144,145}, Jonathan Marten¹⁴⁶, Børge Schmidt¹⁴⁷, Marina Arendt^{147,148}, Laura J Smyth¹⁴⁹, Marisa Cañadas-Garre¹⁴⁹, Chaolong Wang^{150,151}, Masahiro Nakatochi¹⁵², Andrew Wong¹⁵³, Nina Hutri-Kähönen^{154,155}, Xueling Sim⁸⁸, Rui Xia¹⁵⁶, Alicia Huerta-Chagoya¹⁵⁷, Juan Carlos Fernandez-Lopez¹⁵⁸, Valeriya Lyssenko^{159,160}, Meraj Ahmed¹⁶¹, Anne U Jackson⁶, Marguerite R Irvin¹⁶², Christopher Oldmeadow¹⁶³, Han-Na Kim¹⁶⁴, Seunggho Ryu^{165,166}, Paul RHJ Timmers^{167,146}, Liubov Arbeeve¹⁶⁸, Rajkumar Dorajoo¹⁶⁹, Leslie A Lange¹⁷⁰, Xiaoran Chai^{171,172}, Gauri Prasad^{173,174}, Laura Lorés-Motta¹⁷⁵, Marc Pauper¹⁷⁵, Jirong Long¹⁷⁶, Xiaohui Li⁹⁰, Elizabeth Theusch¹⁷⁷, Fumihiko Takeuchi¹⁷⁸, Cassandra N Spracklen^{179,180}, Anu Loukola⁴⁶, Sailalitha Bollepalli⁴⁶, Sophie C Warner^{181,182}, Ya Xing Wang¹⁸³, Wen B. Wei¹⁸⁴, Teresa Nutile¹⁸⁵, Daniela Ruggiero^{185,186}, Yun Ju Sung¹⁸⁷, Yi-Jen Hung¹⁸⁸, Shufeng Chen¹¹³, Fangchao Liu¹¹³, Jingyun Yang^{189,190}, Katherine A Kentistou¹⁶⁷, Mathias Gorski^{11,191}, Marco Brumat¹⁹², Karina Meidtner^{193,194}, Lawrence F Bielak¹⁹⁵, Jennifer A Smith^{195,196}, Prashantha Hebbar¹⁹⁷, Aliko-Eleni Farmaki^{198,199}, Edith Hofer^{200,201}, Maoxuan Lin²⁰², Chao Xue²⁰³, Jifeng Zhang¹, Maria Pina Concac²⁰⁴, Simona Vaccargiu²⁰⁵, Peter J van der Most²⁰⁶, Niina Pitkänen^{207,208}, Brian E Cade^{209,210}, Jiwon Lee²⁰⁹, Sander W. van der Laan²¹¹, Kumaraswamy

Naidu Chitrana²¹², Stefan Weiss²¹³, Martina E Zimmermann¹¹, Jong Young Lee²¹⁴, Hyeok Sun Choi²¹⁵, Maria Nethander^{216,217}, Sandra Freitag-Wolf²¹⁸, Lorraine Southam^{219,220}, Nigel W Rayner^{221,222,223,219}, Carol A Wang²²⁴, Shih-Yi Lin^{225,226,227}, Jun-Sing Wang^{228,229}, Christian Couture²³⁰, Leo-Pekka Lyytikäinen^{231,232}, Kjell Nikus^{233,234}, Gabriel Cuellar-Partida²³⁵, Henrik Vestergaard²³⁶, Bertha Hildalgo²³⁷, Olga Giannakopoulou⁵, Qiuyin Cai¹⁷⁶, Morgan O Obura¹³², Jessica van Setten²³⁸, Xiaoyin Li²³⁹, Karen Schwander²⁴⁰, Natalie Terzikhan²⁴¹, Jae Hun Shin²¹⁵, Rebecca D Jackson²⁴², Alexander P Reiner²⁴³, Lisa Warsinger Martin²⁴⁴, Zhengming Chen^{58,59}, Liming Li²⁴⁵, Heather M Highland⁶⁵, Kristin L Young⁶⁵, Takahisa Kawaguchi⁷⁴, Joachim Thiery^{246,117}, Joshua C Bis¹²⁴, Girish N. Nadkarni¹²⁵, Lenore J Launer²⁴⁷, Huaixing Li¹³⁹, Mike A Nalls^{144,145}, Olli T Raitakari^{248,249,250}, Sahoko Ichihara²⁵¹, Sarah H Wild²⁵², Christopher P Nelson^{181,182}, Harry Campbell¹⁶⁷, Susanne Jäger^{193,194}, Toru Nabika²⁵³, Fahd Al-Mulla¹⁹⁷, Harri Niinikoski^{254,255}, Peter S Braund^{181,182}, Ivana Kolcic²⁵⁶, Peter Kovacs²⁵⁷, Tota Giardoglou²⁵⁸, Tomohiro Katsuya^{259,260}, Konain Fatima Bhatti⁵, Dominique de Kleijn²⁶¹, Gert J. de Borst²⁶¹, Eung Kweon Kim²⁶², Hieab H.H. Adams^{241,263}, M. Arfan Ikram²⁴¹, Xiaofeng Zhu²³⁹, Folkert W Asselbergs²³⁸, Adriaan O Kraaijeveld²³⁸, Joline WJ Beulens^{132,264}, Xiao-Ou Shu¹⁷⁶, Loukianos S Rallidis²⁶⁵, Oluf Pedersen⁹⁴, Torben Hansen⁹⁴, Paul Mitchell²⁶⁶, Alex W Hewitt^{267,268}, Mika Kähönen^{269,270}, Louis Pérusse^{230,271}, Claude Bouchard²⁷², Anke Tönjes²⁷³, Yii-Der Ida Chen⁹⁰, Craig E Pennell²²⁴, Trevor A Mori²⁷⁴, Wolfgang Lieb²⁷⁵, Andre Franke²⁷⁶, Claes Ohlsson^{216,277}, Dan Mellström^{216,278}, Yoon Shin Cho²¹⁵, Hyejin Lee²⁷⁹, Jian-Min Yuan^{280,281}, Woon-Puay Koh^{282,283}, Sang Youl Rhee²⁸⁴, Jeong-Taek Woo²⁸⁴, Iris M Heid¹¹, Klaus J Stark¹¹, Henry Völzke²⁸⁵, Georg Homuth²¹³, Michele K Evans²⁸⁶, Alan B Zonderman²⁸⁶, Ozren Polasek²⁵⁶, Gerard Pasterkamp²¹¹, Imo E Hofer²¹¹, Susan Redline^{209,210}, Katja Pahkala^{207,208,287}, Albertine J Oldehinkel²⁸⁸, Harold Snieder²⁰⁶, Ginevra Biino²⁸⁹, Reinhold Schmidt²⁰⁰, Helena Schmidt²⁹⁰, Y Eugene Chen¹, Stefania Bandinelli²⁹¹, George Dedoussis¹⁹⁸, Thangavel Alphonse Thanaraj²⁹², Sharon LR Kardia¹⁹⁵, Norihiro Kato¹⁷⁸, Matthias B Schulze^{193,194,293}, Giorgia Grotto^{192,294}, Bettina Jung²⁹⁵, Carsten A Böger^{295,296,297}, Peter K Joshi¹⁶⁷, David A Bennett^{189,190}, Philip L De Jager^{298,299}, Xiangfeng Lu¹¹³, Vasiliki Mamakou^{300,301}, Morris Brown^{302,93}, Mark J Caulfield^{192,93}, Patricia B Munroe^{92,93}, Xiuqing Guo⁹⁰, Marina Ciullo^{185,186}, Jost B. Jonas^{303,183,304}, Nilesh J Samani^{181,182}, Daniel I. Chasman^{55,305}, Jaakko Kaprio⁴⁶, Päivi Pajukanta³⁰⁶, Teresa Tusié-Luna^{307,308}, Carlos A Aguilar-Salinas³⁰⁹, Linda S Adair^{310,311}, Sonny Augustin Bechayda^{312,313}, H. Janaka de Silva³¹⁴, Ananda R Wickremasinghe³¹⁵, Ronald Krauss³¹⁶, Jer-Yuarn Wu³¹⁷, Wei Zheng¹⁷⁶, Anneke I den Hollander¹⁷⁵, Dwaipayan Bharadwaj^{174,318}, Adolfo Correa³¹⁹, James G Wilson³²⁰, Lars Lind³²¹, Chew-Kiat Heng³²², Amanda E Nelson^{168,323}, Yvonne M Golightly^{168,324,325,326}, James F Wilson^{167,146}, Brenda Penninx^{327,80}, Hyung-Lae Kim³²⁸, John Attia^{329,163}, Rodney J Scott^{329,163}, D C Rao³³⁰, Donna K Arnett³³¹, Mark Walker³³², Heikki A Koistinen^{333,334,335}, Giriraj R Chandak^{161,336}, Chittaranjan S Yajnik³³⁷, Josep M Mercader^{338,339,340}, Teresa Tusié-Luna³⁴¹, Carlos Aguilar-Salinas³⁴², Clicerio Gonzalez Villalpando³⁴³, Lorena Orozco³⁴⁴, Myriam Fornage^{156,345}, E Shyong Tai^{346,88}, Rob M van Dam^{88,346}, Terho Lehtimäki^{231,232}, Nish Chaturvedi¹⁵³, Mitsuhiro Yokota³⁴⁷, Jianjun Liu¹⁵¹, Dermot F Reilly³⁴⁸, Amy Jayne McKnight¹⁴⁹, Frank Kee¹⁴⁹, Karl-Heinz Jöckel¹⁴⁷, Mark I McCarthy^{66,349#}, Colin NA Palmer³⁵⁰, Veronique Vitart¹⁴⁶, Caroline Hayward¹⁴⁶, Eleanor Simonsick³⁵¹, Cornelia M van Duijn¹⁴³, Fan Lu³⁵², Jia Qu³⁵², Haretsugu Hishigaki¹⁴⁰, Xu Lin³⁵³, Winfried März^{354,355,137}, Esteban J Parra¹³⁶, Miguel Cruz³⁵⁶, Vilmundur Gudnason^{135,357}, Jean-Claude Tardif^{134,358}, Guillaume Lettre^{134,358}, Leen M 't Hart^{133,359,132}, Petra JM Elders³⁶⁰, Daniel J Rader³⁶¹, Scott M Damrauer^{362,363}, Meena Kumari³⁶⁴, Mika Kivimäki¹³⁰, Pim van der Harst¹²⁸, Tim D Spector¹²⁶, Ruth J.F. Loos^{125,365}, Michael A Province¹²⁰, Bruce M Psaty^{366,367}, Ivan Brandslund^{123,368}, Peter P Pramstaller¹²², Kaare Christensen³⁶⁹, Samuli Ripatti^{46,370,371}, Elisabeth Widén⁴⁶, Hakon

Hakonarson^{372,373}, Struan F.A. Grant^{373,374,375}, Lambertus ALM Kiemeneij¹¹⁸, Jacqueline de Graaf¹¹⁸, Markus Loeffler^{116,117}, Florian Kronenberg¹¹⁵, Dongfeng Gu^{113,376}, Jeanette Erdmann³⁷⁷, Heribert Schunkert^{111,112}, Paul W Franks¹¹⁴, Allan Linneberg^{89,378}, J. Wouter Jukema^{107,379}, Amit V Khera^{380,381,382,383}, Minna Männikkö³⁸⁴, 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 Strachan³⁹⁴, Niels Grarup⁹⁴, Peter Sever³⁹⁵, Neil Poulter³⁹⁶, Jerome I Rotter⁹⁰, Thomas M Dantoff⁸⁹, 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 Khor¹⁵¹, Charumathi Sabanayagam^{171,399}, Annette Peters^{84,400,401}, Christian Gieger^{83,84,401}, Andrew T Hattersley⁴⁰², Nancy L Pedersen⁴⁰³, Patrik KE Magnusson⁴⁰³, 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 Huang⁴⁰⁸, Young Jin Kim¹³, Yasuharu Tabara⁷⁴, Nicholas J Wareham⁷³, Claudia Langenberg⁷³, Eleftheria Zeggini^{219,220,409}, Johanna Kuusisto⁴¹⁰, Markku Laakso⁴¹⁰, 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 Vries⁶⁷, Alanna C Morrison⁶⁷, Kari E. North⁶⁵, Martha Davigulus⁴¹⁹, Peter Kraft^{64,420}, Nicholas G Martin⁴²¹, John B Whitfield⁴²¹, Shahid Abbas⁴²², Danish Saleheen^{61,423,424}, Robin G Walters^{425,426,427}, Michael V Holmes^{58,59,428}, Corri Black⁴²⁹, Blair H Smith⁴³⁰, Anne E Justice⁴³¹, Aris Baras⁴³², Julie E Buring^{55,433}, Paul M Ridker^{55,433}, Daniel I Chasman^{55,433}, Charles Kooperberg⁵⁴, Wei-Qi Wei⁴³⁴, Gail P Jarvik⁴³⁵, Bahram Namjou⁴³⁶, M. Geoffrey Hayes^{437,438,439}, Marylyn D Ritchie⁴⁸, Pekka Jousilahti⁴⁷, Veikko Salomaa⁴⁷, Kristian Hveem^{43,440,441}, Bjørn Olav Åsvold^{43,440,442}, Michiaki Kubo⁴⁴³, Yoichiro Kamatani^{32,444}, Yukinori Okada^{34,32,445,446}, Yoshinori Murakami⁴⁴⁷, Unnur Thorsteinsdottir^{29,448}, Kari Stefansson^{29,448}, Yuk-Lam Ho⁴⁴⁹, Julie A Lynch^{450,451}, Daniel Rader⁴⁵², 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 Rotimi¹⁶, Scott Hazelhurst^{15,458}, Michèle Ramsay^{15,459}, Richard C Trembath⁴⁶⁰, David A van Heel¹⁸, Gen Tamiya¹⁴, Masayuki Yamamoto¹⁴, Bong-Jo Kim¹³, Karen L Mohlke¹⁷⁹, Joel N Hirschhorn^{9,10,461}, Sekar Kathiresan^{462,381,383}, Michael Boehnke⁶, Pradeep Natarajan^{463,464,465,466}, Gina M Peloso⁴⁶⁷, 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†*}

¹Department of Internal Medicine, Division of Cardiology, University of Michigan, Ann Arbor, MI 48109, USA, ²VA Palo Alto Health Care system, Palo Alto, California, USA, ³Department of Medicine, Division of Cardiovascular Medicine, Stanford University School of Medicine, Stanford, California, USA, ⁴Department of Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, MI, USA, ⁵William Harvey Research Institute, Barts and the London School of Medicine and Dentistry, Queen Mary University of London, Charterhouse square, EC1M 6BQ, UK, ⁶Department of Biostatistics and Center for Statistics Genetics, University of Michigan, Ann Arbor, MI 48109, ⁷Department of Genetics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, ⁸Clinical Pharmacology, William Harvey Research Institute, Barts and The London School of Medicine and Dentistry, Queen Mary University of London, London, EC1M 6BQ UK, ⁹Endocrinology, Boston Children's Hospital, Boston 02115 MA, USA, ¹⁰Medical and Population Genetics, Broad Institute, 75 Ames street, Cambridge, MA 02142, USA, ¹¹Department of Genetic Epidemiology, University of Regensburg, Regensburg, Germany, ¹²McDonnell Genome Institute and Department of Medicine, Washington University, St. Louis, MO, 63108, ¹³Division of Genome Research, Center for Genome Science, National Institute of Health, Chungcheongbuk-do, South Korea, ¹⁴Tohoku

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

Hospital, Edinburgh EH4 2XU, United Kingdom, ⁵⁷Usher Institute for Population Health Sciences and Informatics, The University of Edinburgh, Nine, Edinburgh Bioquarter, 9 Little France Road, Edinburgh, EH16 4UX., ⁵⁸Clinical Trial Service Unit and Epidemiological Studies Unit, Nuffield Department of Population Health, University of Oxford, Oxford OX3 7LF, ⁵⁹Medical Research Council Population Health Research Unit, Nuffield Department of Population Health, University of Oxford, Oxford OX3 7LF, ⁶⁰Department of Population Medicine, Qatar University College of Medicine, QU Health, Doha, Qatar, ⁶¹Center for Non-Communicable Diseases, Karachi, Sindh, Pakistan, ⁶²Survey Research Center, Institute for Social Research, University of Michigan, Ann Arbor, MI, 48104, ⁶³Department of Epidemiology, School of Public Health, University of Michigan, Ann Arbor, MI, 48109, ⁶⁴Program in Genetic Epidemiology and Statistical Genetics, Department of Epidemiology, Harvard T.H. Chan School of Public Health, 677 Huntington Avenue, Boston, MA, 02115, USA, ⁶⁵Department of Epidemiology, University of North Carolina, Chapel Hill, NC, ⁶⁶Wellcome Centre for Human Genetics, University of Oxford, UK, ⁶⁷Human 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, ⁶⁸Department of Epidemiology and Biostatistics, Imperial College London, London W2 1PG, UK, ⁶⁹Department of Cardiology, Ealing Hospital, London North West University Healthcare NHS Trust, Middlesex UB1 3HW, UK, ⁷⁰Imperial College Healthcare NHS Trust, London W12 0HS, UK, ⁷¹Department of Biostatistics, Center for Statistical Genetics, University of Michigan, Ann Arbor, ⁷²Department of Medical Sciences, Molecular Epidemiology and Science for Life Laboratory, Uppsala University, Uppsala, Sweden., ⁷³MRC Epidemiology Unit, University of Cambridge School of Clinical Medicine, Cambridge, CB2 0QQ, UK, ⁷⁴Center for Genomic Medicine, Kyoto University Graduate School of Medicine, ⁷⁵Department of Internal Medicine, Erasmus MC, University Medical Center Rotterdam, ⁷⁶Department of Epidemiology, Erasmus MC, University Medical Center Rotterdam, ⁷⁷Department of Biostatistics, Boston University School of Public Health, 801 Massachusetts Ave, Boston, MA 02118, USA, ⁷⁸Department of Biological Psychology, Behavioral and Movement Sciences, Vrije Universiteit Amsterdam, ⁷⁹Amsterdam Public Health, VU medical center Amsterdam, ⁸⁰Amsterdam Public Health research institute, VU medical center Amsterdam, ⁸¹Genetics of Complex Traits, University of Exeter Medical School, University of Exeter, Exeter, EX2 5DW, UK, ⁸²Department of Clinical Acupuncture and Moxibustion, Nanjing University of Chinese Medicine, Nanjing, Jiangsu 210029, China, ⁸³Research Unit of Molecular Epidemiology, Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg, Germany, ⁸⁴Institute of Epidemiology, Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg, Germany, ⁸⁵MRC Integrative Epidemiology Unit at the University of Bristol, Oakfield Road, Bristol, BS8 2BN, United Kingdom, ⁸⁶Bristol Dental School, University of Bristol, Lower Maudlin Street, Bristol BS1 2LY, United Kingdom, ⁸⁷Population Health Sciences, Bristol Medical School, University of Bristol, Oakfield Grove, Bristol, BS8 2BN, United Kingdom, ⁸⁸Saw Swee Hock School of Public Health, National University of Singapore and National University Health System, 117549, Singapore, ⁸⁹Center for Clinical Research and Prevention, Bispebjerg and Frederiksberg Hospital, Copenhagen, Denmark, ⁹⁰The 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, ⁹¹Center for Public Health Genomics, University of Virginia, Charlottesville, VA 22903 USA, ⁹²William 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, ⁹³NIHR Barts Cardiovascular Biomedical Research Centre, Barts and

The London School of Medicine and Dentistry, Queen Mary University of London, London, EC1M 6BQ, UK, ⁹⁴Novo Nordisk Foundation Center for Basic Metabolic Research, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark, ⁹⁵Center for Clinical Research and Prevention, Bispebjerg and Frederiksberg Hospital, Copenhagen, Denmark, Nordre Fasanvej 57, DK-2000 Frederiksberg, Denmark, ⁹⁶Division of Cardiovascular Medicine, Radcliffe Department of Medicine, John Radcliffe Hospital, University of Oxford, Oxford. UK. OX3 9DU, ⁹⁷Wellcome Centre for Human Genetics, University of Oxford, Oxford. UK. OX3 7BN, ⁹⁸Group of Genomics of Complex Diseases. Research Institute of Hospital de la Santa Creu i Sant Pau (IIB Sant Pau), Barcelona, Spain, ⁹⁹Cardiovascular Medicine Unit, Department of Medicine, Karolinska Institutet, Center for Molecular Medicine, Karolinska University Hospital, Stockholm, Sweden, ¹⁰⁰Department of Internal Medicine, Section Gerontology and Geriatrics, Leiden University Medical Center, Leiden, the Netherlands, ¹⁰¹Institute for Genetic and Biomedical Research, Italian National Council of Research (IRGB CNR), Cagliari Italy, ¹⁰²Institute for Genetic and Biomedical Research, Italian National Council of Research (IRGB CNR), Lanusei, Italy, ¹⁰³University Center for Primary Care and Public Health, University of Lausanne, Rte de la Corniche 10, Lausanne, 1010, Switzerland, ¹⁰⁴Swiss Institute of Bioinformatics, Lausanne, 1015, Switzerland, ¹⁰⁵Department of Medicine, Internal Medicine, Lausanne University Hospital and University of Lausanne, Rue du Bugnon 46, Lausanne, 1011, Switzerland, ¹⁰⁶Department of Epidemiology and Biostatistics, MRC-PHE Centre for Environment and Health, School of Public Health, Imperial College London, London, UK, ¹⁰⁷Dept of Cardiology, Leiden University Medical Center, Leiden, the Netherlands, ¹⁰⁸Dept of Internal Medicine, Section of Gerontology and Geriatrics, Leiden university Medical Center, Leiden, the Netherlands, ¹⁰⁹BHF Glasgow Cardiovascular Research Centre, Faculty of Medicine, Glasgow, United Kingdom, ¹¹⁰Institute 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, ¹¹¹Deutsches Herzzentrum München, Klinik für Herz- und Kreislauferkrankungen, Technische Universität München, Munich, Germany., ¹¹²Deutsches Zentrum für Herz-Kreislauf-Forschung (DZHK) e.V., partner site Munich Heart Alliance, Munich, Germany., ¹¹³Key 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, ¹¹⁴Lund University Diabetes Centre, Malmö, Sweden, ¹¹⁵Institute of Genetic Epidemiology, Department of Genetics and Pharmacology, Medical University of Innsbruck, Innsbruck, Austria and German Chronic Kidney Disease study, ¹¹⁶Institute for Medical Informatics, Statistics and Epidemiology, University of Leipzig, Haertelstrasse 16-18, 04107 Leipzig, Germany, ¹¹⁷LIFE Research Centre for Civilization Diseases, University of Leipzig, Philipp-Rosenthal-Straße 27, 04103 Leipzig, Germany, ¹¹⁸Radboud university medical center, Radboud Institute for Health Sciences, Nijmegen, The Netherlands, ¹¹⁹Quantinuum Research LLC, Wayne, PA, 19087 USA, ¹²⁰Division of Statistical Genomics, Department of Genetics; Washington University School of Medicine; St. Louis, MO, USA, ¹²¹Department of Epidemiology; University of Pittsburgh; Pittsburgh, PA, USA, ¹²²Institute for Biomedicine, Eurac Research, Affiliated Institute of the University of Lübeck, Via Galvani 31, 39100, Bolzano, Italy, ¹²³Department of Clinical Biochemistry, Lillebaelt Hospital, Vejle, Denmark, ¹²⁴Cardiovascular Health Research Unit, Department of Medicine, University of

Washington, Seattle, 98101, USA, ¹²⁵The Charles Bronfman Institute for Personalized Medicine, Icahn School of Medicine at Mount Sinai, New York, NY, 10029, USA, ¹²⁶Department of Twin Research and Genetic Epidemiology, King's College London, London SE1 7EH, UK, ¹²⁷NIHR Biomedical Research Centre at Guy's and St Thomas' Foundation Trust, London SE1 9RT, UK, ¹²⁸University of Groningen, University Medical Center Groningen, Department of Cardiology, 9700RB Groningen, The Netherlands, ¹²⁹Institute of Cardiovascular Sciences, University College London, Gower Street, WC1E 6BT London, UK, ¹³⁰Department of Epidemiology and Public Health, University College London, 1-19 Torrington Place, WC1E 6BT London, United Kingdom, ¹³¹Department of Psychiatry, University of Pennsylvania, Philadelphia, PA, 19104, USA, ¹³²Amsterdam UMC, Department of Epidemiology and Biostatistics, Amsterdam Public Health Research Institute, Amsterdam, 1081HV, the Netherlands., ¹³³Leiden University Medical Center, Department of Cell and Chemical Biology, Leiden, 2333ZA, The Netherlands, ¹³⁴Montreal Heart Institute, 5000 Belanger Street, Montreal, Quebec, H1T 1C8, Canada, ¹³⁵Icelandic Heart Association, 201 Kopavogur, Iceland, ¹³⁶Department of Anthropology, University of Toronto at Mississauga, Mississauga, ON L5L 1C6, Canada, ¹³⁷Vth Department of Medicine, Medical Faculty Mannheim, Heidelberg University, 68167 Mannheim, Germany, ¹³⁸SYNLAB MVZ Humangenetik Mannheim GmbH, 68163 Mannheim, Germany, ¹³⁹CAS Key Laboratory of Nutrition, Metabolism and Food Safety, Shanghai Institute of Nutrition and Health, University of Chinese Academy of Sciences, Chinese Academy of Sciences, ¹⁴⁰Biomedical Technology Research Center, Tokushima Research Institute, Otsuka Pharmaceutical Co., Ltd., Tokushima, Japan, ¹⁴¹Institute for Molecular Bioscience, The University of Queensland, Brisbane, Queensland 4072, Australia, ¹⁴²Institute for Advanced Research, Wenzhou Medical University, Wenzhou, Zhejiang 325027, China, ¹⁴³Nuffield Department of Population Health, University of Oxford, Oxford, United Kingdom, ¹⁴⁴Laboratory of Neurogenetics, National Institute on Aging, NIH, Bethesda MD, USA, ¹⁴⁵Data Tecnica International, Glen Echo MD, USA, ¹⁴⁶MRC Human Genetics Unit, Institute of Genetics and Molecular Medicine, University of Edinburgh, Western General Hospital, Crewe Road, Edinburgh, EH4 2XU, Scotland, ¹⁴⁷Institute for Medical Informatics, Biometrie and Epidemiology, University of Duisburg-Essen, Essen, Germany, ¹⁴⁸Department of Computer Science, University of Applied Sciences and Arts Dortmund, Emil-Figge-Str. 42, 44227 Dortmund, Germany, ¹⁴⁹Centre for Public Health, Queen's University of Belfast, Northern Ireland, ¹⁵⁰Department of Epidemiology and Biostatistics, School of Public Health, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China, ¹⁵¹Genome Institute of Singapore, Agency for Science, Technology and Research, 138672, Singapore, ¹⁵²Public Health Informatics Unit, Department of Integrated Health Sciences, Nagoya University Graduate School of Medicine, Nagoya, 461-8673, Japan, ¹⁵³MRC Unit for Lifelong Health and Ageing at UCL, 1-19 Torrington Place, London, WC1E 7HB, United Kingdom, ¹⁵⁴Department of Pediatrics, Tampere University Hospital, Tampere 33521, Finland, ¹⁵⁵Department of Pediatrics, Faculty of Medicine and Health Technology, Tampere University, Tampere 33014, Finland, ¹⁵⁶Brown Foundation Institute of Molecular Medicine, McGovern Medical School, University of Texas Health Science Center at Houston, Houston TX 77030, USA, ¹⁵⁷CONACYT, Instituto Nacional de Ciencias Médicas y Nutrición Salvador Zubirán, Ciudad de Mexico, Mexico, ¹⁵⁸Departamento de Genómica Computacional, Instituto Nacional de Medicina Genómica, Ciudad de Mexico, Mexico, ¹⁵⁹Center for diabetes research, University of Bergen, Bergen, Norway, ¹⁶⁰Lund University Diabetes Center, Lunds University, Malmö, Sweden, ¹⁶¹Genomic Research on Complex diseases (GRC Group), CSIR-Centre for Cellular and Molecular Biology, Hyderabad, Telangana, India, ¹⁶²University of Alabama at Birmingham, Epidemiology, School of Public Health, ¹⁶³Hunter Medical Research Institute, Newcastle,

Australia, ¹⁶⁴Medical Research Institute, Kangbuk Samsung Hospital, Sungkyunkwan University School of Medicine, Seoul 03181, Korea, ¹⁶⁵Center for Cohort Studies, Total Healthcare Center, Kangbuk Samsung Hospital, Sungkyunkwan University School of Medicine, Seoul, 04514, Korea, ¹⁶⁶Department of Occupational and Environmental Medicine, Kangbuk Samsung Hospital, Sungkyunkwan University School of Medicine, Seoul, 03181, Korea, ¹⁶⁷Centre for Global Health Research, Usher Institute, University of Edinburgh, Teviot Place, Edinburgh, EH8 9AG, Scotland, ¹⁶⁸Thurston Arthritis Research Center, University of North Carolina, Chapel Hill, North Carolina, USA, ¹⁶⁹Genome Institute of Singapore, Agency for Science, Technology and Research, Singapore, ¹⁷⁰Division of Biomedical Informatics and Personalized Medicine, Department of Medicine, Anschutz Medical Campus, University of Colorado, Denver, Aurora, CO 80045, USA, ¹⁷¹Ocular Epidemiology, Singapore Eye Research Institute, Singapore National Eye Centre, 168751, Singapore, ¹⁷²Department of Ophthalmology, National University of Singapore and National University Health System, 119228, Singapore, ¹⁷³Genomics and Molecular Medicine Unit, CSIR-Institute of Genomics and Integrative Biology, New Delhi - 110020, India., ¹⁷⁴Academy of Scientific and Innovative Research, CSIR-Institute of Genomics and Integrative Biology Campus, New Delhi 110020, India., ¹⁷⁵Departments of Ophthalmology and Human Genetics, Radboud Univeristy Nijmegen Medical Center, Philips van Leydenlaan 15, Nijmegen, 6525 EX, the Netherlands, ¹⁷⁶Vanderbilt Epidemiology Center, Division of Epidemiology, Vanderbilt University Medical Center, ¹⁷⁷Department of Pediatrics, University of California San Francisco, Oakland, CA 94609 USA, ¹⁷⁸National Center for Global Health and Medicine, Tokyo, 1628655, Japan, ¹⁷⁹Department of Genetics, University of North Carolina, Chapel Hill, NC 27599 USA, ¹⁸⁰Department of Biostatistics and Epidemiology, University of Massachusetts-Amherst, Amherst, MA 01003 USA, ¹⁸¹Department of Cardiovascular Sciences, University of Leicester, Leicester, UK, ¹⁸²NIHR Leicester Biomedical Research Centre, Glenfield Hospital, Leicester, UK, ¹⁸³Beijing 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, ¹⁸⁴Beijing Tongren Eye Center, Beijing Tongren Hospital, Capital Medical University, 1 Dong Jiao Min Xiang, Dong Cheng District, Beijing, 100730, China, ¹⁸⁵Institute of Genetics and Biophysics "Adriano Buzzati-Traverso" - CNR, Naples, Italy, ¹⁸⁶IRCCS Neuromed, Pozzilli, Isernia, Italy, ¹⁸⁷Division of Biostatistics, Washington University, St. Louis, MO 63110, ¹⁸⁸Division of Endocrinology and Metabolism, Tri-Service General Hospital Songshan Branch, Taipei, Taiwan, ¹⁸⁹Rush Alzheimer's Disease Center, Rush University Medical Center, ¹⁹⁰Department of Neurological Sciences, Rush University Medical Center, ¹⁹¹Department of Nephrology, University Hospital Regensburg, Regensburg, Germany, ¹⁹²Department of Medicine, Surgery and Health Sciences, University of Trieste, Strada di Fiume 447, 34149, Trieste, Italy, ¹⁹³Department of Molecular Epidemiology, German Institute of Human Nutrition Potsdam-Rehbruecke, Nuthetal, Germany, ¹⁹⁴German Center for Diabetes Research (DZD), München-Neuherberg, Germany, ¹⁹⁵Department of Epidemiology, School of Public Health, University of Michigan, Ann Arbor, MI 48109, ¹⁹⁶Survey Research Center, Institute for Social Research, University of Michigan, Ann Arbor, MI, 48104, ¹⁹⁷Department of Genetics and Bioinformatics, Dasman Diabetes Institute, Kuwait, ¹⁹⁸Department of Nutrition and Dietetics, School of Health Science and Education, Harokopio University of Athens, Athens, Greece, ¹⁹⁹Department of Population Science and Experimental Medicine, University College London, London, UK, ²⁰⁰Clinical Division of Neurogeriatrics, Department of Neurology, Medical University of Graz, Graz, Austria, ²⁰¹Institute for Medical Informatics, Statistics and Documentation, Medical University of Graz, Graz, Austria, ²⁰²Department of Bioinformatics and Genomics, University of

North Carolina at Charlotte, NC 28223 USA, ²⁰³Department of Internal Medicine, University of Michigan Medical Center, ²⁰⁴Institute for Maternal and Child Health, IRCCS Burlo Garofolo, Trieste, Italy, ²⁰⁵Institute of Genetic and Biomedical Research, National Research Council of Italy, UOS of Sassari, Sassari, Italy, ²⁰⁶University of Groningen, University Medical Center Groningen, Department of Epidemiology, Groningen, 9700 RB, the Netherlands, ²⁰⁷Research Centre of Applied and Preventive Cardiovascular Medicine, University of Turku, Turku, Finland, ²⁰⁸Centre for Population Health Research, University of Turku and Turku University Hospital, Turku, Finland, ²⁰⁹Sleep Medicine and Circadian Disorders, Brigham and Women's Hospital, Boston, Massachusetts 02115, USA, ²¹⁰Division of Sleep Medicine, Harvard Medical School, Boston, Massachusetts 02115, USA, ²¹¹Central Diagnostics Laboratory, Division Laboratories, Pharmacy, and Biomedical genetics, University Medical Center Utrecht, Utrecht University, Utrecht, the Netherlands, ²¹²Laboratory of Epidemiology and Population Science National Institute on Aging Intramural Research Program, NIH 251 Bayview Blvd, NIH Biomedical Research Center, NIA, Baltimore, MD 21224, ²¹³Interfaculty Institute for Genetics and Functional Genomics, Department of Functional Genomics, University of Greifswald and University Medicine Greifswald, Greifswald, Germany, ²¹⁴Oneomics. co. Ltd. 2F, Soonchunhyang Mirai Medical Center 173, Buheuyng-ro, Bucheon-si Gyeonggi-do, 14585, Korea, ²¹⁵Department of Biomedical Science, Hallym University, Chuncheon, Gangwon-do 24252, Korea, ²¹⁶Centre for Bone and Arthritis Research, Department of Internal Medicine and Clinical Nutrition, Institute of Medicine, Sahlgrenska Academy, University of Gothenburg, Gothenburg, Sweden., ²¹⁷Bioinformatics Core Facility, Sahlgrenska Academy, University of Gothenburg, Gothenburg, Sweden, ²¹⁸Institute of Medical Informatics and Statistics, Kiel University, Kiel, Germany., ²¹⁹Institute of Translational Genomics, Helmholtz Zentrum München – German Research Center for Environmental Health, Neuherberg, Germany, ²²⁰Wellcome Trust Sanger Institute, Hinxton, CB10 1SA, UK, ²²¹Wellcome Centre for Human Genetics, University of Oxford, Oxford, UK, ²²²Oxford Centre for Diabetes Endocrinology and Metabolism, Oxford, UK, ²²³Wellcome Sanger Institute, Hinxton, Cambridge, HH CB10 1 UK, ²²⁴School of Medicine and Public Health, Faculty of Medicine and Health, University of Newcastle, Newcastle, New South Wales, 2308, Australia, ²²⁵Center for Geriatrics and Gerontology, Division of Endocrinology and Metabolism, Department of Internal Medicine, Taichung Veterans General Hospital, Taichung, Taiwan, ²²⁶School of Medicine, National Yang-Ming University, Taipei, Taiwan, ²²⁷School of Medicine, National Defense Medical Center, Taipei, Taiwan, ²²⁸Division of Endocrinology and Metabolism, Department of Internal Medicine, Taichung Veterans General Hospital, Taichung, Taiwan, ²²⁹Department of Medicine, School of Medicine, National Yang-Ming University, Taipei, Taiwan, ²³⁰Dept of Kinesiology, Université Laval, Québec, Canada, ²³¹Department of Clinical Chemistry, Fimlab Laboratories, Tampere 33520, Finland, ²³²Department of Clinical Chemistry, Finnish Cardiovascular Research Center - Tampere, Faculty of Medicine and Health Technology, Tampere University, Tampere 33014, Finland, ²³³Department of Cardiology, Heart Center, Tampere University Hospital, Tampere 33521, Finland, ²³⁴Department of Cardiology, Finnish Cardiovascular Research Center - Tampere, Faculty of Medicine and Health Technology, Tampere University, Tampere 33014, Finland, ²³⁵University of Queensland Diamantina Institute, Translational Research Institute, Kent St, Woolloongabba, Brisbane, QLD, 4102, Australia., ²³⁶Novo 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, ²³⁷School of Public Health, University of Alabama at Birmingham, ²³⁸Cardiology, Division Heart & Lungs, University Medical Center Utrecht, Utrecht University, Utrecht, the Netherlands, ²³⁹Department of Population and Quantitative Health

Sciences, Case Western Reserve University, Cleveland, OH, 44106, USA, ²⁴⁰Division of Statistical Genomics, Department of Genetics, Washington University School of Medicine, St. Louis, MO, USA, ²⁴¹Department of Epidemiology - Erasmus MC - University Medical Center Rotterdam, Rotterdam, the Netherlands., ²⁴²Ohio State University, Division of Endocrinology, Columbus OH 43210, USA, ²⁴³University of Washington, Department of Epidemiology, Seattle WA 98195, USA, ²⁴⁴George Washington University, School of Medicine and Health Sciences, Washington DC 20037, USA, ²⁴⁵Department of Epidemiology, School of Public Health, Peking University Health Science Center, Beijing, China, ²⁴⁶Institute for Laboratory Medicine, University Hospital Leipzig, Paul-List-Strasse 13/15, 04103 Leipzig, Germany, ²⁴⁷Laboratory of Epidemiology and Population Sciences, National Institute on Aging, NIH, Baltimore, MD, 20892-9205, USA, ²⁴⁸Centre for Population Health Research, University of Turku and Turku University Hospital, Finland, ²⁴⁹Research Centre of Applied and Preventive Cardiovascular Medicine, University of Turku, Finland, ²⁵⁰Department of Clinical Physiology and Nuclear Medicine, Turku University Hospital, Turku, Finland, ²⁵¹Department of Environmental and Preventive Medicine, Jichi Medical University School of Medicine, Shimotsuke, 329-0498, Japan, ²⁵²Centre for Population Health Sciences, Usher Institute, University of Edinburgh, Teviot Place, Edinburgh, EH8 9AG, Scotland, ²⁵³Department of Functional Pathology, Shimane University School of Medicine, Izumo, 6938501, Japan, ²⁵⁴Department of Pediatrics and Adolescent Medicine, Turku University Hospital and University of Turku, Turku, Finland, ²⁵⁵Department of Physiology, University of Turku, Turku, Finland, ²⁵⁶Faculty of Medicine, University of Split, Šoltanska 2, HR-21000, Split, Croatia, ²⁵⁷Medical Department III – Endocrinology, Nephrology, Rheumatology, University of Leipzig Medical Center, Liebigstr. 21, 04103 Leipzig, Germany, ²⁵⁸Department of Nutrition-Dietetics, Harokopio University, Eleftheriou Venizelou, Athens, 17676, Greece, ²⁵⁹Department of Clinical Gene Therapy, Osaka University Graduate School of Medicine, Suita, 5650871, Japan, ²⁶⁰Department of Geriatric and General Medicine, Osaka University Graduate School of Medicine, Suita, 5650871, Japan, ²⁶¹Department of Vascular Surgery, Division of Surgical Specialties, University Medical Center Utrecht, Utrecht University, Utrecht, the Netherlands, ²⁶²Corneal Dystrophy Research Institute, Department of Ophthalmology, Yonsei University College of Medicine, Seoul 03722, Korea, ²⁶³Dept of Radiology and Nuclear Medicine, Erasmus MC - University Medical Center Rotterdam, Rotterdam, the Netherlands., ²⁶⁴Julius Centre for Health Sciences and Primary Care, University Medical Centre Utrecht, 3584CG, the Netherlands, ²⁶⁵Second Department of Cardiology, Medical School, National and Kapodistrian University of Athens, Attikon University Hospital, Athens, Greece, ²⁶⁶Center for Vision Research, Department of Ophthalmology and The Westmead Institute, University of Sydney, Hawkesbury Rd, Sydney, New South Wales, 2145, Australia., ²⁶⁷Menzies Institute for Medical Research, School of Medicine, University of Tasmania, Liverpool St, Hobart, Tasmania, 7000, Australia., ²⁶⁸Centre for Eye Research Australia, University of Melbourne, Melbourne, Victoria, 3002, Australia, ²⁶⁹Department of Clinical Physiology, Tampere University Hospital, Tampere 33521, Finland, ²⁷⁰Department of Clinical Physiology, Finnish Cardiovascular Research Center - Tampere, Faculty of Medicine and Health Technology, Tampere University, Tampere 33014, Finland, ²⁷¹Institute of Nutrition and Functional Foods (INAF), Université Laval, Québec, Canada, ²⁷²Pennington Biomedical Research Center, Baton Rouge, LA 70808, USA, ²⁷³Medical Department III – Endocrinology, Nephrology, Rheumatology, University of Leipzig Medical Center, Liebigstr. 18, 04103 Leipzig, Germany, ²⁷⁴Medical School, Faculty of Health and Medical Sciences, University of Western Australia, Perth, Western Australia, 6000, Australia, ²⁷⁵Institute of Epidemiology, Kiel University, Kiel, Germany, ²⁷⁶Institute of Clinical Molecular Biology, Kiel University, Kiel, Germany, ²⁷⁷Sahlgrenska University Hospital,

Department of Drug Treatment, Gothenburg, Sweden, ²⁷⁸Geriatric Medicine, Institute of Medicine, Sahlgrenska Academy, University of Gothenburg, Gothenburg, Sweden, ²⁷⁹Department of Internal Medicine, EwhaWomans University School of Medicine, Seoul, Korea, ²⁸⁰Division of Cancer Control and Population Sciences, UPMC Hillman Cancer Center, University of Pittsburgh, Pittsburgh, PA 15232, USA, ²⁸¹Department of Epidemiology, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, PA 15232, USA, ²⁸²Health Services and Systems Research, Duke-NUS Medical School, 169857, Singapore, ²⁸³Saw Swee Hock School of Public Health, National University of Singapore and National University Health System, ²⁸⁴Department of Endocrinology and Metabolism, Kyung Hee University School of Medicine, Seoul 02447, Korea, ²⁸⁵Institute for Community Medicine, University Medicine Greifswald, Germany, ²⁸⁶Laboratory of Epidemiology and Population Science National Institute on Aging Intramural Research Program, NIH 251 Bayview Blvd, NIH Biomedical Research Center, Baltimore, MD 21224, ²⁸⁷Paavo Nurmi Centre, Sports and Exercise Medicine Unit, Department of Physical Activity and Health, University of Turku, Turku, Finland, ²⁸⁸University of Groningen, University Medical Center Groningen, Interdisciplinary Center Psychopathology and Emotion Regulation (ICPE), Groningen, 9700 RB, the Netherlands, ²⁸⁹Institute of Molecular Genetics, National Research Council of Italy, Pavia, Italy, ²⁹⁰Gottfried Schatz Research Center for Cell Signaling, Metabolism and Aging, Medical University of Graz, Graz, Austria, ²⁹¹Local Health Unit Toscana Centro, Firenze, Italy., ²⁹²Department of Genetics and Bioinformatics, Dasman Diabetes Institute, Kuwait, ²⁹³Institute of Nutritional Science, University of Potsdam, Nuthetal, Germany, ²⁹⁴Institute for Maternal and Child Health IRCCS "Burlo Garofolo", Via dell'Istria 65/1, 34137, Trieste, Italy, ²⁹⁵Dept of Nephrology, University Hospital Regensburg, Regensburg, Germany, ²⁹⁶Dept of Nephrology, Diabetology, Rheumatology; Traunstein Hospital, Traunstein, Germany, ²⁹⁷KfH Kidney Center Traunstein, Traunstein, Germany, ²⁹⁸Center for Translational and Systems Neuroimmunology, Department of Neurology, Columbia University Medical Center, New York, NY, USA, ²⁹⁹Program in Medical and Population Genetics, Broad Institute, Cambridge, MA, USA, ³⁰⁰Medical School, National and Kapodistrian University Athens, 75 M. Assias Street, 115 27 Athens, Greece, ³⁰¹Dromokaiteio Psychiatric Hospital, 124 61 Athens, Greece, ³⁰²Clinical Pharmacology, William Harvey Research Institute, Queen Mary University of London, London, EC1M 6BQ, UK, ³⁰³Department of Ophthalmology, Medical Faculty Mannheim, Heidelberg University, Kutzerufer 1, Mannheim, 68167, Germany, ³⁰⁴Institute of Clinical and Scientific Ophthalmology and Acupuncture Jonas & Panda, Heidelberg, Germany, ³⁰⁵Harvard Medical School, Boston MA 02115, ³⁰⁶Department of Human Genetics, David Geffen School of Medicine at UCLA, University of California, Los Angeles, CA, USA, ³⁰⁷Unidad de Biología Molecular y Medicina Genómica, Instituto Nacional de Ciencias Médicas y Nutrición Salvador Zubirán, Mexico 14080, Mexico., ³⁰⁸Instituto de Investigaciones Biomédicas, UNAM, ³⁰⁹Departamento de Endocrinología y Metabolismo, Instituto Nacional de Ciencias Médicas y Nutrición Salvador Zubirán, Mexico 14080, Mexico., ³¹⁰Department of Nutrition, Gillings School of Global Public Health, University of North Carolina, Chapel Hill, North Carolina, 27599 USA, ³¹¹Carolina Population Center, University of North Carolina, Chapel Hill, North Carolina, 27516 USA, ³¹²USC–Office of Population Studies Foundation, University of San Carlos, Cebu City, 6000, Philippines, ³¹³Department of Anthropology, Sociology, and History, University of San Carlos, Cebu City, 6000 Philippines, ³¹⁴Department of Medicine, Faculty of Medicine, University of Kelaniya, Ragama, 11010, Sri Lanka, ³¹⁵Department of Public Health, Faculty of Medicine, University of Kelaniya, Ragama, 11010, Sri Lanka, ³¹⁶Children's Hospital Oakland Research Institute, Oakland, CA 94609 USA, ³¹⁷Institute of Biomedical Sciences, Academia Sinica, Taiwan, ³¹⁸Systems Genomics Laboratory, School of Biotechnology,

Jawaharlal Nehru University, New Delhi - 110067, India, ³¹⁹Department of Medicine, University of Mississippi Medical Center, Jackson, MS, 39216, USA, ³²⁰Department of Physiology and Biophysics, University of Mississippi Medical Center, Jackson, MS, 39216, USA, ³²¹Department of Medical Sciences, Uppsala University, Sweden, ³²²Department 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, ³²³Department of Medicine, University of North Carolina, Chapel Hill, NC, USA, ³²⁴Department of Epidemiology, Gillings School of Global Public Health, University of North Carolina, Chapel Hill, North Carolina, USA, ³²⁵Injury Prevention Research Center, University of North Carolina, Chapel Hill, North Carolina, USA, ³²⁶Division of Physical Therapy, University of North Carolina, Chapel Hill, North Carolina, USA, ³²⁷Department of Psychiatry, Amsterdam UMC, Vrije Universiteit Amsterdam, ³²⁸Department of Biochemistry, College of Medicine, Ewha Womans University, Seoul 07804, Korea, ³²⁹Faculty of Health and Medicine, University of Newcastle, Australia, ³³⁰Washington University School of Medicine, Division of Biostatistics, ³³¹University of Kentucky, College of Public Health, ³³²Institute of Cellular Medicine (Diabetes), The Medical School, Newcastle University, Framlington Place, Newcastle upon Tyne, NE2 4HH, ³³³Department of Public Health Solutions, Finnish Institute for Health and Welfare, P.O. Box 30, FI-00271 Helsinki, Finland., ³³⁴University of Helsinki and Department of Medicine, Helsinki University Central Hospital, P.O.Box 340, Haartmaninkatu 4, Helsinki, FI-00029, Finland., ³³⁵Minerva Foundation Institute for Medical Research, Biomedicum 2U, Tukholmantatu 8, Helsinki, FI-00290, Finland., ³³⁶Academy of Scientific and Innovative Research (AcSIR), New Delhi, India, ³³⁷Diabetology Research Centre, KEM Hospital and Research Centre, Pune, Maharashtra, India, ³³⁸Programs in Metabolism and Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, MA, USA, ³³⁹Diabetes Unit and Center for Genomic Medicine, Massachusetts General Hospital, Boston, MA, USA, ³⁴⁰Harvard Medical School, Boston, Massachusetts, USA, ³⁴¹Unidad de Biología Molecular y Medicina Genómica, Instituto de Investigaciones Bimédicas UNAM/ Instituto Nacional de Ciencias Médicas y Nutrición Salvador Zubirán, Mexico City, ³⁴²Dirección de Nutrición and Unidad de Estudios de Enfermedades Metabólicas, Instituto Nacional de Ciencias Médicas y Nutrición Salvador Zubirán, Mexico City, ³⁴³Instituto Nacional de Salud Pública y Centro de Estudios en Diabetes, Mexico, ³⁴⁴Instituto Nacional de Medicina Genómica, Mexico, ³⁴⁵Human Genetics Center, School of Public Health, University of Texas Health Science Center at Houston, Houston TX 77030, USA, ³⁴⁶Yong Loo Lin School of Medicine, National University of Singapore and National University Health System, 119228, Singapore, ³⁴⁷Kurume University School of Medicine, Kurume, 830-0011, Japan, ³⁴⁸Genetics, Merck Sharp & Dohme Corp., Kenilworth, NJ, 07033, US, ³⁴⁹Oxford Centre for Diabetes, Endocrinology & Metabolism, University of Oxford, UK, ³⁵⁰Population Health and Genomics, University of Dundee, Ninewells Hospital and Medical School, Dundee, DD1 9SY, UK, ³⁵¹Intramural Research Program, National Institute on Aging, 3001 S. Hanover St., Baltimore, MD 21225, ³⁵²The Eye Hospital, School of Ophthalmology & Optometry, Wenzhou Medical University, Wenzhou, Zhejiang 325027, China, ³⁵³Shanghai Institute of Nutrition and Health University of Chinese Academy of Sciences, Chinese Academy of Sciences, ³⁵⁴Synlab Academy, SYNLAB Holding Deutschland GmbH, Mannheim and Augsburg, Germany, ³⁵⁵Clinical Institute of Medical and Chemical Laboratory Diagnostics, Medical University of Graz, Austria, ³⁵⁶Unidad de Investigación Médica en Bioquímica, Hospital de Especialidades, Centro Médico Nacional Siglo XXI, Instituto Mexicano del Seguro Social, Mexico City, Mexico, ³⁵⁷Faculty of Medicine, University of Iceland, 101 Reykjavik, Iceland, ³⁵⁸Department of Medicine, Faculty of Medicine, Université de Montréal, 2900 Edouard Montpetit Blvd, Montreal, Quebec,

H3T 1J4, Canada, ³⁵⁹Leiden University Medical Center, Department of Biomedical Data Sciences, Section Molecular Epidemiology, Leiden, 2333ZA, The Netherlands, ³⁶⁰Amsterdam UMC, Department of General Practice and Elderly Care, Amsterdam Public Health Research Institute, Amsterdam, 1081HV, The Netherlands, ³⁶¹Department of Genetics, University of Pennsylvania, Philadelphia, PA, 19104, USA, ³⁶²Department of Surgery, University of Pennsylvania, Philadelphia, PA, 19104, USA, ³⁶³Corporal Michael Crescenz VA Medical Center, Philadelphia, Pennsylvania, PA, 19104, USA, ³⁶⁴Institute of Social and Economic Research, University of Essex, Wivenhoe Park, CO4 3SQ, United Kingdom, ³⁶⁵The Mindich Child Health and Development Institute, Icahn School of Medicine at Mount Sinai, New York, NY, 10029, USA, ³⁶⁶Cardiovascular Health Research Unit, Departments of Medicine, Epidemiology and Health Services, University of Washington, Seattle, 98101, WA, USA, ³⁶⁷Kaiser Permanent Washington Health Research Institute, Seattle, 98101, WA, USA, ³⁶⁸Institute of Regional Health Research, University of Southern Denmark, Odense, Denmark, ³⁶⁹Danish Aging Research Center, University of Southern Denmark; Odense C, Denmark, ³⁷⁰Public Health, Faculty of Medicine, University of Helsinki, Finland, ³⁷¹Broad Institute of MIT and Harvard, Cambridge, MA, ³⁷²Center for Applied Genomics, Children's Hospital of Philadelphia, Philadelphia, PA, 19104 USA, ³⁷³Department of Pediatrics, The University of Pennsylvania Perelman School of Medicine, Philadelphia, PA, 19104 USA, ³⁷⁴Division of Human Genetics, Children's Hospital of Philadelphia, Philadelphia, PA, 19104 USA, ³⁷⁵Department of Genetics, University of Pennsylvania, Philadelphia, PA, 19104 USA, ³⁷⁶School of Medicine, Southern University of Science and Technology, Shenzhen, China, ³⁷⁷Institute for Cardiogenetics, University of Lübeck, DZHK (German Research Centre for Cardiovascular Research), partner site Hamburg/Lübeck/Kiel, and University Heart Center Lübeck, Lübeck, Germany, ³⁷⁸Department of Clinical Medicine, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark, ³⁷⁹Netherlands Heart Institute, Utrecht, the Netherlands, ³⁸⁰Division of Cardiology, Department of Medicine, Massachusetts General Hospital, Boston, Massachusetts, USA, ³⁸¹Program of Medical and Population Genetics, Broad Institute, Cambridge, Massachusetts, USA, ³⁸²Center for Genomic Medicine, Massachusetts General Hospital, Boston, Massachusetts, USA, ³⁸³Department of Medicine, Harvard Medical School, Boston, Massachusetts, USA, ³⁸⁴Northern Finland Birth Cohorts, Infrastructure for population studies, Faculty of Medicine, University of Oulu, Oulu, Finland, ³⁸⁵Center for Life Course Health Research, Faculty of Medicine, University of Oulu, Oulu, Fin, ³⁸⁶Biocenter of Oulu, University of Oulu, Oulu, Finl, ³⁸⁷Institute for Genetic and Biomedical Research, Italian National Council of Research (IRGB CNR), Cagliari, Italy, ³⁸⁸University of Sassari, Sassari, Italy, ³⁸⁹Department of Clinical Epidemiology, Leiden University Medical Center, Leiden, the Netherlands, ³⁹⁰Department of Public Health and Primary Care, Leiden University Medical Center, Leiden, the Netherlands, ³⁹¹Department of Internal Medicine, Division of Endocrinology, Leiden University Medical Center, Leiden, the Netherlands, ³⁹²Eindhoven Laboratory for Experimental Vascular Medicine, Leiden University Medical Center, Leiden, the Netherlands, ³⁹³Department of Human Genetics, Leiden University Medical Center, Leiden, the Netherlands, ³⁹⁴Population Health Research Institute, St George's, University of London, London SW17 0RE, UK, ³⁹⁵National Heart and Lung Institute, Imperial College London, London, W2 1PG, UK, ³⁹⁶School of Public Health, Imperial College London, London, W2 1PG, UK, ³⁹⁷OCDEM, University of Oxford, Churchill Hospital, Oxford OX3 7LE, UK, ³⁹⁸NIHR Oxford Biomedical Research Centre, Churchill Hospital, Oxford, UK, ³⁹⁹Ophthalmology & Visual Sciences Academic Clinical Program (Eye ACP), Duke-NUS Medical School, 169857, Singapore, ⁴⁰⁰DZHK (German Centre for Cardiovascular Research), Munich Heart Alliance partner site, Munich, Germany., ⁴⁰¹German

Center for Diabetes Research (DZD), Neuherberg, Germany., ⁴⁰²University of Exeter Medical School, University of Exeter, Exeter, EX2 5DW, UK, ⁴⁰³Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden, ⁴⁰⁴Netherlands Twin Register, Department of Biological Psychology, Behavioral and Movement Sciences, Vrije Universiteit Amsterdam, ⁴⁰⁵Amsterdam Reproduction & Development research institute, VU medical center Amsterdam, ⁴⁰⁶Framingham Heart Study, National Heart, Lung, and Blood Institute, US National Institutes of Health, Bethesda, MD, USA., ⁴⁰⁷Department of Genetics, School of Medicine, Mashhad University of Medical Sciences, Mashhad, Iran., ⁴⁰⁸Department of Genetics, Shanghai-MOST Key Laboratory of Health and Disease Genomics, Chinese National Human Genome Center at Shanghai, Shanghai, 201203 China, ⁴⁰⁹TUM School of Medicine, Technical University of Munich and Klinikum Rechts der Isar, Munich, Germany, ⁴¹⁰Institute of Clinical Medicine, Internal Medicine, University of Eastern Finland and Kuopio University Hospital, ⁴¹¹Department of Medicine, Division of Cardiovascular Medicine, Stanford University School of Medicine, Stanford, CA 94305, ⁴¹²Stanford Cardiovascular Institute, Stanford University, Stanford, CA 94305, ⁴¹³Stanford Diabetes Research Center, Stanford University, Stanford, CA 94305, ⁴¹⁴Regeneron Pharmaceuticals, Tarrytown, NY, USA, ⁴¹⁵Lee Kong Chian School of Medicine, Nanyang Technological University, Singapore 308232, Singapore, ⁴¹⁶Imperial College Healthcare NHS Trust, Imperial College London, London W12 0HS, UK, ⁴¹⁷MRC-PHE Centre for Environment and Health, Imperial College London, London W2 1PG, UK, ⁴¹⁸National Heart and Lung Institute, Imperial College London, London W12 0NN, UK, ⁴¹⁹Institute for Minority Health Research, University of Illinois College of Medicine, Chicago, Illinois, ⁴²⁰Department of Biostatistics, Harvard T.H. Chan School of Public Health, 677 Huntington Avenue, Boston, MA, 02115, USA, ⁴²¹QIMR Berghofer Medical Research Institute, 300 Herston Road, Brisbane, Queensland 4006, Australia, ⁴²²Center for Non-Communicable Diseases, Karachi, Sindh, Pakistan & Faisalabad Institute of Cardiology, Faisalabad, Pakistan, ⁴²³Department of Medicine, Columbia University Irving Medical Center, New York, NY, USA, ⁴²⁴Department of Cardiology, Columbia University Irving Medical Center, New York, NY, USA, ⁴²⁵Clinical Trial Service Unit and Epidemiological Studies Unit, Nuffield Department of Population Health, University of Oxford, Oxford OX3 7LF, UK, ⁴²⁶Medical Research Council Population Health Research Unit, Nuffield Department of Population Health, University of Oxford, Oxford OX3 7LF, UK, ⁴²⁷Big Data Institute, University of Oxford, Oxford OX3 7LF, UK, ⁴²⁸National Institute for Health Research Oxford Biomedical Research Centre, Oxford University Hospitals, Oxford, UK, ⁴²⁹Aberdeen Centre for Health Data Science, 1:042 Polwarth Building School of Medicine, Medical Science and Nutrition University of Aberdeen Foresterhill Aberdeen AB25 2ZD, ⁴³⁰Division of Population Health and Genomics, Ninewells Hospital and Medical School, University of Dundee, Dundee DD1 9SY, United Kingdom, ⁴³¹Biomedical and Translational Informatics, Geisinger Health, Danville, PA 17822, ⁴³²Regeneron Pharmaceuticals, Tarrytown, NY, USA., ⁴³³Harvard Medical School, Boston, MA 02115, ⁴³⁴Department of Biomedical Informatics, Vanderbilt University Medical Center, Nashville, TN, ⁴³⁵Departments of Medicine (Medical Genetics) and Genome Sciences, University of Washington, ⁴³⁶Center for Autoimmune Genomics and Etiology, Cincinnati Children's Hospital Medical Center (CCHMC), Cincinnati, OH, USA., ⁴³⁷Division of Endocrinology, Metabolism, and Molecular Medicine, Department of Medicine, Northwestern University, Feinberg School of Medicine, Chicago, IL 60618, USA, ⁴³⁸Department of Anthropology, Northwestern University, Evanston, IL 60208, USA, ⁴³⁹Center for Genetic Medicine, Northwestern University, Feinberg School of Medicine, Chicago, IL 60618, USA, ⁴⁴⁰HUNT Research Centre, Department of Public Health and Nursing, NTNU, Norwegian University of Science and Technology, Levanger, 7600 Norway, ⁴⁴¹Department of Medicine,

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

Dept of Human Genetics University of Michigan, Ann Arbor, MI, USA

cristen@umich.edu

Themistocles L. Assimes, MD, PhD

VA Palo Alto Health Care System, Palo Alto, CA, USA &

Dept of Medicine, Stanford University School of Medicine, Stanford, CA, USA

tassimes@stanford.edu

Summary

Elevated blood lipid levels are heritable risk factors of cardiovascular disease with varying prevalence worldwide due to differing dietary patterns and medication use¹. Despite advances in prevention and treatment, particularly through the lowering of low-density lipoprotein cholesterol levels², heart disease remains the leading cause of death worldwide³. Genome-wide association studies (GWAS) of blood lipid levels have led to important biological and clinical insights, as well as new drug targets, for cardiovascular disease. However, most previous GWAS⁴⁻²³ have been conducted in European ancestry populations and may have missed genetic variants contributing to lipid level variation in other ancestry groups due to differences in allele frequencies, effect sizes, and linkage-disequilibrium (LD) patterns²⁴. Here we conduct a multi-ancestry genome-wide genetic discovery meta-analysis of lipid levels in ~1.65 million individuals, including 350,000 of non-European ancestry. We quantify the gain in studying non-European ancestries and provide evidence to support expanding recruitment into new ancestries even with relatively smaller sample sizes. We find that increasing diversity rather than studying additional European-ancestry individuals results in substantial improvements in fine-mapping functional variants and portability of polygenic prediction, with modest gains in the number of discovered loci and ancestry-specific variants. As GWAS expands its emphasis beyond identifying genes and fundamental biology towards using genetic variants for preventive and precision medicine²⁵, we anticipate that increased participant diversity will lead to more accurate and equitable²⁶ application of polygenic scores in clinical practice.

Main Text

The Global Lipids Genetics Consortium aggregated GWAS results from 1,654,960 individuals from 164 primary studies representing five genetic ancestry groups: Admixed African or African (AdmAFR, N=99.4k, 6.0% of sample), East Asian (EAS, N=146.5k, 8.9%), European (EUR, N=1.32m, 79.8%), Hispanic (HIS, N=48.1k, 2.9%), and South Asian (SAS, N=41.0k, 2.5%) (**Table 1, Supplementary Table 1, Supplementary Figure 1**). 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). These analyses included 91 million variants imputed primarily from the Haplotype Reference Consortium or 1000 Genomes Phase 3.

Ancestry-specific genetic discovery

We first quantified the number of genome-wide significant loci identified in at least one of the five ancestry-specific meta-analyses. We found 773 lipid-associated genomic regions containing 1,765 distinct index variants that reached genome-wide significance (p -value $< 5 \times 10^{-8}$, ± 500 kb, **Supplementary Tables 2-3, Supplementary Figures 2-3**) for at least one ancestry group and lipid trait. Of these regions, 237 were novel based on the most-significant index variant in each region being >500 kb from variants previously reported as associated with any of the five lipid traits^{4-23,27}. Of these loci, 76% were identified only in the European ancestry-specific analyses ($N \sim 1.3$ m, 80% of sample). Of the non-European ancestry groups, the African-ancestry GWAS ($N \sim 99$ k, primarily African American) identified more ancestry-specific loci (15 unique to AdmAFR) than any other non-European ancestry group (six loci unique to EAS, six to HIS, one to SAS). The difference is likely attributable to allele frequencies being most different between African and European ancestry populations (**Figure 1A-D**) and to African populations having greater genetic diversity²⁸.

Trans-ancestry genetic discovery

We next performed trans-ancestry meta-analyses using the meta-regression approach implemented in MR-MEGA³⁰ to account for heterogeneity in variant effect sizes on lipids between ancestry groups. A total of 1,750 index variants at 923 loci (± 500 kb regions) reached genome-wide significance for at least one lipid trait. These included 168 regions not identified by ancestry-specific analysis, 120 (71%) of which were novel (**Supplementary Tables 4-5, Supplementary Figures 4-5**). Almost all (98%) of index variants from the ancestry-specific analysis remained significant (p -value $< 5 \times 10^{-8}$) after meta-analysis across all ancestry groups, although fifteen AdmAFR, nine EAS, three HIS, and one SAS index variants from ancestry-specific analysis did not (trans-ancestry p -value 7.7×10^{-6} to 5.9×10^{-8} , **Supplementary Figure 6, Supplementary Note**). In total, we identified 355 novel loci from either single- or trans-ancestry analyses.

Next, we compared the number of loci identified per 100,000 participants in each ancestry group and the combined dataset (**Figure 1E**). African and Hispanic ancestry-specific analyses identified the most loci per genotyped individual, perhaps due to greater recent admixture; European and trans-ancestry analyses identified the fewest, likely reflecting a slight reduction in the benefit from new samples added to very large sample sizes (>1 m). For the genome-wide significant variants discovered in each ancestry, we estimated the proportion of ancestry-enriched variants by enumerating the number of other ancestries with sufficient power to detect association (range 0 to 4). We estimated the power for discovery of each variant by

assuming an equivalent discovery sample size in the other ancestries, fixed effect size, and observed allele frequencies from the other ancestries (**Figure 1F**). To allow for comparison at similar sample sizes across ancestry groups, we selected European index variants identified from a meta-analysis of ~100,000 individuals subsampled from the present study. African-ancestry index-variants were most ancestry-enriched, with only 61% of index variants demonstrating sufficient power in at least one other ancestry group (equal N, power>80% to reach $\alpha=5 \times 10^{-8}$), likely due to population-enriched allele frequencies. In comparison, 88% of South Asian index variants had estimated power >80% in at least one other ancestry.

Finally, we found that both the number of identified variants and the mean observed chi-squared values from genome-wide lipid association tests were approximately linearly related to meta-analysis sample size across ancestries (**Supplementary Table 6, Supplementary Figure 7**). However, in Europeans the incremental increase in either the number of loci or chi-squared value was slightly attenuated at the largest sample sizes. Taken together, these results suggest that once sufficiently well-powered GWAS sample sizes are reached within a given ancestry group, assembling large sample sizes of other under-represented groups will modestly enhance variant discovery relative to increasing the sample size of the dominant ancestry.

Comparison of effect sizes across ancestries

Differences in association signals across ancestries despite similar sample sizes could be due to variation in allele frequencies and/or effect sizes. This could reflect differing patterns of LD with the underlying causal variant or an interaction with an environmental risk factor whose prevalence varies by ancestry and/or geography. We found that effect size estimates of individual variants were largely similar based on pairwise comparison between ancestries ($r^2=0.93$ for variants with $p\text{-value}<5 \times 10^{-8}$) (**Supplementary Table 7, Supplementary Figure 8**). We additionally tested for genome-level differences in effect size correlation between East Asian, European, and South Asian ancestry groups using Popcorn²⁹, which were not significantly different from 1 ($p\text{-value}>0.05$, **Supplementary Figures 9 and 10**). We tested for differences in genetic correlation between Admixed Africans and Europeans in the UK Biobank and Million Veteran Program (MVP) using bivariate GREML^{30,31} as the Popcorn method does not account for long-range LD in admixed populations. Genetic correlation between AdmAFR and EUR for HDL-C ($r=0.84$) was not significantly different from 1 in the UK Biobank (possibly due to relatively small numbers of African ancestry individuals), while correlations for the other traits ranged from 0.52-0.60 in UK Biobank and 0.47-0.69 in MVP (**Supplementary Table 8**).

Of the 2,286 variants that reached genome-wide significance in the trans-ancestry meta-analysis across all five lipid traits, 159 (7%) showed significant heterogeneity of effect size due to ancestry ($p\text{-value} < 2.2 \times 10^{-5}$; Bonferroni correction for 2,286 variants, **Supplementary Table 5**). Of these 159, 31 showed the largest effect in African ancestry analyses, 24 in East Asian, 67 in European, 20 in Hispanic, and 17 in South Asian. Only 49 (2%) of these variants from trans-ancestry meta-analysis showed significant residual heterogeneity not due to ancestry, which may be attributable to differences in ascertainment or analysis strategy between cohorts (**Supplementary Table 5**), suggesting cohort-related factors are a less important driver of heterogeneity than genetic ancestry.

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

We next assessed whether trans-ancestry fine-mapping narrowed the set of likely causal variants at each of the independent trans-ancestry association signals ($LD\ r^2 < 0.7$), assuming one shared causal variant per ± 500 kb region (**Supplementary Table 9**). 19% of the association signals had only one variant in the 99% credible set and 55% (816/1,486) had ≤ 10 . In contrast, 5% (73/1486) had > 100 . Of the 407 variants with $> 90\%$ posterior probability of being the causal variant at a locus in the trans-ancestry meta-analysis, 56 (14%) were missense variants, 7 (2%) were splice-region variants, and 4 (1%) were stop-gain variants (*CD36*, *HBB*, *ANGPTL8*, *PDE3B*). (**Supplementary Tables 10-12**).

The median number of variants in 99% credible sets from Europeans was 13; this was reduced to 8 in the trans-ancestry analysis. Of 1,486 association signals, 825 (56%) had reduced credible set size in the trans-ancestry analysis. At these 825 loci, the number of variants in the trans-ancestry credible sets were reduced by 40% relative to the minimum credible set size in either African (the most genetically diverse group) or European analyses (**Supplementary Figure 11**). We estimate that increasing the sample size of European samples to that of the trans-ancestry analysis would yield a 20% reduction in credible set size, approximately half of the 40% reduction observed in trans-ancestry analysis. This suggests that sample size differences alone do not explain the reduction, rather differences in LD patterns and effect sizes across ancestries likely contribute to the improved fine-mapping (**Supplementary Note**). For example, rs900776, an intronic variant in the *DMTN* region with many high LD variants in Europeans, has a posterior probability of being causal of 0.86 in the African-ancestry derived credible sets, > 0.99 in the trans-ancestry analysis, but only 0.51 in the European specific analysis (**Figure 2**).

Polygenic scores of LDL-C are most predictive when derived from multi-ancestry GWAS

We evaluated the potential of polygenic scores to predict elevated LDL-C, a major causal risk factor of CAD, in diverse ancestry groups. We created three non-overlapping datasets to separately: i) perform ancestry-specific or trans-ancestry GWAS to estimate variant effect sizes, ii) optimize risk score parameters, and iii) evaluate the utility of the resulting scores. For each ancestry-specific or trans-ancestry GWAS we created multiple polygenic score weights -- either genome-wide with PRS-CS³² or using pruning and thresholding to select independent variants. We tested each score in the optimizing dataset, which was matched for ancestry to the GWAS (AdmAFR, EAS, EUR, SAS, ALL from UK biobank or HIS from Michigan Genomics Initiative (MGI), **Supplementary Figures 12-14, Supplementary Tables 13-15**). The top-performing score from each GWAS was selected: PRS-CS for East Asian, European, and European 2010 scores from a previous GLGC GWAS⁴, and an optimized pruning and threshold-based score for all others. We then evaluated the polygenic scores in 8 cohorts of individuals (N=295,577, **Supplementary Table 16**), not included in the discovery GWAS, from 6 ancestral groups: East Asian (146,477), European American (85,571), African American (21,730), African (2,452 East Africa, 4,972 South Africa, 7,309 West Africa), South Asian (15,242), Hispanic American (7,669), and Asian American (4,155).

The polygenic score developed from trans-ancestry meta-analysis consistently showed the best or near-best performance in each group tested, with improved or comparable prediction relative to ancestry-matched scores (adjusted $R^2 \sim 0.10-0.16$, **Figure 3, Supplementary Table 17**). This observation was especially evident for ancestries with smaller GWAS sample sizes, as was the case for HIS and SAS. For African Americans in MGI and MVP, polygenic prediction was similar for individuals with different levels of recent African-ancestry admixture (**Supplementary Figure 14**) and reached the level of prediction observed for European-ancestry individuals from the same dataset. The variance explained by each standard deviation increase in the polygenic score was also similar between ancestry groups in MVP: 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. We repeated the evaluation of trans-ancestry vs single-ancestry polygenic scores with a set GWAS with sample size of ~100k individuals and with fixed methodology; results were consistent with those from the full dataset (**Figure 3b**). Thus, polygenic prediction for LDL-C in all ancestries appears to benefit the most from adding samples of diverse ancestries once relatively large numbers of Europeans have already been included. Additional

studies are needed to determine if this applies to other phenotypes with different genetic architectures and heritabilities.

Discussion

Genome-wide discovery for blood lipid traits based on ~1.65 million individuals from five ancestry groups confirmed that the contributions of common genetic variation to blood lipids are largely similar across diverse populations. First, we found that the number of significant loci relative to sample size was similar within each ancestry group, and approximately linearly related to sample size, with a small increase in ancestry-specific variants observed in African-ancestry cohorts relative to the others. Second, we demonstrated that inclusion of additional ancestries through trans-ancestry fine-mapping reduces the set of candidate causal variants in credible sets and does so more rapidly than in single-ancestry analysis. Trans-ancestry GWAS should therefore facilitate identification of effector genes at GWAS loci and allow for accelerated biological insight and identification of potential drug targets. Third, we found that a polygenic score derived from ~88k African-ancestry and ~830k European-ancestry individuals was correlated with observed lipid levels among individuals with admixed African-ancestry as well as among individuals with European-ancestry. We hypothesize that the inclusion of African-ancestry individuals in the GWAS yields improvement in polygenic prediction performance through the general fine-mapping of loci and the improved prioritization of trans-ancestry causal variants. Fourth, and perhaps most important, the trans-ancestry score was generally most informative across all major population groups examined. This provides useful information for other genetic discovery efforts and investigations of the utility of the polygenic scores in diverse populations.

Generalizability of these findings regarding portability of polygenic scores from the trans-ancestry meta-analysis to other traits may depend on the heritability, degree of polygenicity, level of genetic correlation, allele frequencies of causal variants across ancestry groups, gene-environment interactions, and representation of diverse populations in the GWAS^{33,34}. While many traits show a high degree of shared genetic correlation across ancestries^{31,35,36} others have distinct genetic variants with large effects that are more common in specific ancestry groups³³ which may limit the utility of trans-ancestry polygenic scores for particular phenotypes in some ancestries.

The benefits from genetic discovery efforts as GWAS sample sizes increase will likely not be measured just by the number of loci discovered. Rather, the focus will increasingly turn to improving our understanding of the biology at established loci, identifying potential therapeutic targets, and efficiently identifying individuals at high-risk of adverse health outcomes across population groups without exacerbating existing health disparities. Considering the results presented here, and those of related studies³⁷⁻³⁹, we believe future genetic studies will benefit substantially from meta-analysis across participants of diverse ancestries. Further gains in the depth and number of sequenced individuals of diverse ancestries^{40,41} may additionally improve discovery of novel variants and loci in diverse cohorts, particularly variants absent from arrays and imputation reference panels. Our results suggest that diversifying the populations under study, rather than simply increasing the sample size, is now the single most efficient approach to achieving these goals, at least for blood lipids and likely for tightly related downstream adverse health outcomes such as cardiovascular disease. However, if costs for recruitment of diverse populations are higher than recruitment of individuals from previously studied ancestry groups, and total number of genome-wide significant index variants is the goal, then continued low-cost recruitment of majority ancestry groups is expected to still provide some benefit. Taken together, our results also strongly support ongoing and future large-scale recruitment efforts targeted at the enrollment and DNA collection of non-European ancestry participants. Geneticists and those responsible for cohort development must continue diversifying genetic discovery datasets, while increasing sample size in a cost-effective manner, to ensure genetic studies reduce rather than exacerbate existing health inequities across race, ancestry, geographic region, and nationality.

References Cited in Main Text

- 1 Taddei, C. *et al.* Repositioning of the global epicentre of non-optimal cholesterol. *Nature* **582**, 73-77, doi:10.1038/s41586-020-2338-1 (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. *Eur Heart J* **38**, 2459-2472, doi:10.1093/eurheartj/ehx144 (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, doi:10.1016/S0140-6736(18)32203-7 (2018).
- 4 Teslovich, T. M. *et al.* Biological, clinical and population relevance of 95 loci for blood lipids. *Nature* **466**, 707-713, doi:10.1038/nature09270 (2010).
- 5 Willer, C. J. *et al.* Discovery and refinement of loci associated with lipid levels. *Nature genetics* **45**, 1274-1283, doi:10.1038/ng.2797 (2013).
- 6 Liu, D. J. *et al.* Exome-wide association study of plasma lipids in >300,000 individuals. *Nature genetics* **49**, 1758-1766, doi:10.1038/ng.3977 (2017).
- 7 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. *Nature genetics* **49**, 1722-1730, doi:10.1038/ng.3978 (2017).
- 8 Kathiresan, S. *et al.* A genome-wide association study for blood lipid phenotypes in the Framingham Heart Study. *BMC Med Genet* **8 Suppl 1**, S17-S17, doi:10.1186/1471-2350-8-S1-S17 (2007).
- 9 Kathiresan, S. *et al.* Polymorphisms Associated with Cholesterol and Risk of Cardiovascular Events. *New England Journal of Medicine* **358**, 1240-1249, doi:10.1056/NEJMoa0706728 (2008).
- 10 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. *American journal of human genetics* **94**, 223-232, doi:10.1016/j.ajhg.2014.01.009 (2014).
- 11 Hoffmann, T. J. *et al.* A large electronic-health-record-based genome-wide study of serum lipids. *Nature genetics* **50**, 401-413, doi:10.1038/s41588-018-0064-5 (2018).
- 12 Surakka, I. *et al.* The impact of low-frequency and rare variants on lipid levels. *Nature genetics* **47**, 589-597, doi:10.1038/ng.3300 (2015).
- 13 Klarin, D. *et al.* Genetics of blood lipids among ~300,000 multi-ethnic participants of the Million Veteran Program. *Nature genetics* **50**, 1514-1523, doi:10.1038/s41588-018-0222-9 (2018).
- 14 Holmen, O. L. *et al.* Systematic evaluation of coding variation identifies a candidate causal variant in TM6SF2 influencing total cholesterol and myocardial infarction risk. *Nature genetics* **46**, 345-351, doi:10.1038/ng.2926 (2014).
- 15 Asselbergs, F. W. *et al.* Large-scale gene-centric meta-analysis across 32 studies identifies multiple lipid loci. *American journal of human genetics* **91**, 823-838, doi:10.1016/j.ajhg.2012.08.032 (2012).
- 16 Albrechtsen, A. *et al.* Exome sequencing-driven discovery of coding polymorphisms associated with common metabolic phenotypes. *Diabetologia* **56**, 298-310, doi:10.1007/s00125-012-2756-1 (2013).
- 17 Saxena, R. *et al.* Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels. *Science (New York, N.Y.)* **316**, 1331-1336, doi:10.1126/science.1142358 (2007).

- 18 lotchkova, V. *et al.* Discovery and refinement of genetic loci associated with cardiometabolic risk using dense imputation maps. *Nature genetics* **48**, 1303-1312, doi:10.1038/ng.3668 (2016).
- 19 Tachmazidou, I. *et al.* A rare functional cardioprotective APOC3 variant has risen in frequency in distinct population isolates. *Nature Communications* **4**, 2872, doi:10.1038/ncomms3872 (2013).
- 20 Tang, C. S. *et al.* Exome-wide association analysis reveals novel coding sequence variants associated with lipid traits in Chinese. *Nature Communications* **6**, 10206, doi:10.1038/ncomms10206 (2015).
- 21 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, doi:10.1038/ncomms7065 (2015).
- 22 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, doi:10.1093/hmg/ddx062 (2017).
- 23 Kanai, M. *et al.* Genetic analysis of quantitative traits in the Japanese population links cell types to complex human diseases. *Nature genetics* **50**, 390-400, doi:10.1038/s41588-018-0047-6 (2018).
- 24 Sirugo, G., Williams, S. M. & Tishkoff, S. A. The Missing Diversity in Human Genetic Studies. *Cell* **177**, 26-31, doi:10.1016/j.cell.2019.02.048 (2019).
- 25 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, doi:10.1038/s41588-018-0183-z (2018).
- 26 Duncan, L. *et al.* Analysis of polygenic risk score usage and performance in diverse human populations. *Nature Communications* **10**, 3328, doi:10.1038/s41467-019-11112-0 (2019).
- 27 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, doi:10.1093/nar/gky1120 (2019).
- 28 Tishkoff, S. A. *et al.* The genetic structure and history of Africans and African Americans. *Science (New York, N.Y.)* **324**, 1035-1044, doi:10.1126/science.1172257 (2009).
- 29 Brown, B. C., Ye, C. J., Price, A. L. & Zaitlen, N. Transethnic Genetic-Correlation Estimates from Summary Statistics. *American journal of human genetics* **99**, 76-88, doi:10.1016/j.ajhg.2016.05.001 (2016).
- 30 Lee, S. H., Yang, J., Goddard, M. E., Visscher, P. M. & Wray, N. R. Estimation of pleiotropy between complex diseases using single-nucleotide polymorphism-derived genomic relationships and restricted maximum likelihood. *Bioinformatics* **28**, 2540-2542, doi:10.1093/bioinformatics/bts474 (2012).
- 31 Guo, J. *et al.* Quantifying genetic heterogeneity between continental populations for human height and body mass index. *Scientific reports* **11**, 5240, doi:10.1038/s41598-021-84739-z (2021).
- 32 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, doi:10.1038/s41467-019-09718-5 (2019).
- 33 Majara, L. *et al.* Low generalizability of polygenic scores in African populations due to genetic and environmental diversity. *bioRxiv*, 2021.2001.2012.426453, doi:10.1101/2021.01.12.426453 (2021).
- 34 Lehmann, B. C. L., Mackintosh, M., McVean, G. & Holmes, C. C. High trait variability in optimal polygenic prediction strategy within multiple-ancestry cohorts. *bioRxiv*, 2021.2001.2015.426781, doi:10.1101/2021.01.15.426781 (2021).

- 35 Shi, H. *et al.* Population-specific causal disease effect sizes in functionally important regions impacted by selection. *Nature Communications* **12**, 1098, doi:10.1038/s41467-021-21286-1 (2021).
- 36 Martin, A. R. *et al.* Clinical use of current polygenic risk scores may exacerbate health disparities. *Nature genetics* **51**, 584-591, doi:10.1038/s41588-019-0379-x (2019).
- 37 Cavazos, T. B. & Witte, J. S. Inclusion of Variants Discovered from Diverse Populations Improves Polygenic Risk Score Transferability. *bioRxiv*, 2020.2005.2021.108845, doi:10.1101/2020.05.21.108845 (2020).
- 38 Wojcik, G. L. *et al.* Genetic analyses of diverse populations improves discovery for complex traits. *Nature* **570**, 514-518, doi:10.1038/s41586-019-1310-4 (2019).
- 39 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, doi:10.1038/s41588-019-0378-y (2019).
- 40 Taliun, D. *et al.* Sequencing of 53,831 diverse genomes from the NHLBI TOPMed Program. *bioRxiv*, 563866, doi:10.1101/563866 (2019).
- 41 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, doi:10.1371/journal.pgen.1008500 (2019).

Table 1: Meta-analysis sample size by ancestry group

| Ancestry Group | Sample Size | Number of Cohorts | Mean Sample Size per Cohort (range) | Number of Variants |
|-----------------------------|--------------------|--------------------------|--|---------------------------|
| European | 1,320,016 | 123 | 10,928 (173-389,344) | 75 M |
| East Asian | 146,492 | 30 | 7,448 (150-131,050) | 27 M |
| Admixed African/ African | 99,432 | 19 | 5,330 (473-62,022) | 48 M |
| Hispanic | 48,057 | 8 | 6,032 (1,496-22,302) | 37 M |
| South Asian | 40,963 | 6 | 6,413 (1,796-16,110) | 37 M |
| Total | 1,654,960 | 164 | | 91 M |

The present meta-analysis represents a 6-fold overall increase in sample size relative to the most recent 2018 Million Veteran Program blood lipid meta-analysis¹, with a 2-fold increase in sample size of Admixed African and Hispanic individuals.

Figure Legends

Figure 1: Comparison of identified loci across ancestry groups

A) Allele frequency distribution and B) effect sizes of African-ancestry index variants in non-African populations. C) Allele frequency distribution and D) effect sizes of European-ancestry index variants in non-European populations. 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 to detect an association of the same effect size but with ancestry-specific frequencies in the other ancestry groups. Dark blue regions indicate variants likely to be detected at an equivalent sample size only in the original ancestry group (i.e. ancestry-specific). Additional comparisons of allele frequencies and effect sizes across ancestries are provided in Supplementary Figure 3.

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

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

Figure 3: Trans-ancestry LDL-C PRS show similar performance across ancestry groups

A) 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 B) Trans-ancestry scores derived from equal proportions of each ancestry group predict LDL-C better for African Americans in MGI than predominantly European scores at constant sample size.

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/glgc-lipids2021>. All custom scripts used for analysis and summary of results are available upon request. The optimized trans-ancestry and single-ancestry polygenic score 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 GWAS summary statistics for HDL-C, LDL-C, nonHDL-C, TC and TG, imputation quality statistics, and analysis metrics for quality control (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 ancestry 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 software². Detailed pre-imputation QC guidelines were provided; these included removing samples with call rate < 95%, samples with heterozygosity > median + 3(interquartile range), ancestry outliers from principal component analysis within each ancestry group, and variants deviating from Hardy-Weinberg equilibrium (p-value < 10⁻⁶) or with variant call rate < 98%. Analyses were carried out separately by ancestry group and were additionally stratified by cases and controls where appropriate (i.e. for a disease-focused cohort such as CAD). Residuals were generated separately in males and females adjusting for age, age², 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 levels³ 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-LMM⁴, SAIGE⁵, or deCode association software.

Quality Control

Each input file was assessed for quality control using the EasyQC software⁶. We generated QQ plots by minor allele frequency (MAF) bins, assessed trends in standard errors relative to sample size for each cohort, and checked MAF 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 most 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. Results from either sex-stratified analysis or sex-combined analysis with sex as a covariate were used. During the QC process, within each cohort we removed poorly imputed variants (info score or r² < 0.3), variants deviating from Hardy-Weinberg Equilibrium (HWE p-value < 10⁻⁸, except for

MVP which used HWE p -value $< 10^{-20}$), and variants with minor allele count < 3 . An imputation info score threshold of 0.3 was selected 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 as many variants as possible, summary statistics from cohorts that had submitted both HRC and 1KGP3 imputed files were joined, selecting variants imputed from HRC where both imputed versions of a variant existed. 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 RAREMETAL⁷. Trans-ancestry meta-analysis was performed using MR-MEGA⁸ 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 METAL⁹ 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 5×10^{-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 4**. 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)¹⁰. Variants were annotated using WGSAs¹¹ including the summary of each variant from SnpEff¹² and the closest genes for intergenic variants from ANNOVAR¹³. Annotation of variants as known or novel was done based on manual review of previously published variants and with variants reported in the GWAS catalog¹⁴ for any of the studied lipid traits (accessed May 2020, provided as **Supplementary Table 18**). 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 association within each ancestry was determined using the effect size and sample size of the variant within the original discovery ancestry group and the observed allele frequency from the other ancestry groups with alpha set to 5×10^{-8} . We excluded variants that were only successfully imputed in a single ancestry group to account for imputation panel differences between groups (ie. Haplotype Reference Consortium for Europeans and 1000 Genomes for other ancestries). Variants that were successfully imputed in 2 or more ancestries were assumed to have zero power in any other ancestry where the variant was not successfully imputed. 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. The proportion of variance explained within each ancestry was estimated using the trans-ancestry effect size from METAL with the ancestry-specific allele frequency. Coverage of the genome by associated genetic regions was calculated using BEDTools¹⁵ for the regions defined by the minimum and maximum position within each locus having p -value $< 5 \times 10^{-8}$.

Conditional analysis to flag non-independent index variants

Approximate conditional analysis was performed using rareGWAMA¹⁶ 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 $\text{info}/r^2 > 0.3$). The EUR subset of UK Biobank was used as the reference population for the conditional analysis of the trans-ancestry meta-analysis (~80% European). Stepwise 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 of all values (**Supplementary Figure 16**). This threshold was selected to remove variants whose effects were driven by nearby, more strongly associated variants in LD. 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

Popcorn¹⁷ was used to assess the degree of correlation in effect sizes between ancestry groups for each of the lipid traits with 1000 Genomes phase 3 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. Bivariate GREML from GCTA was used to calculate the genetic correlation between unrelated Admixed Africans and a subset of white British individuals in the UK Biobank following the method of Guo et al^{18,19}. HapMap3 variants with MAF > 0.01 in each ancestry were used to construct the genetic relationship matrix (GRM) with the allele frequencies standardized in each population. Individuals with genetic relatedness > 0.05 were removed. A total of up to 5,575 Adm Afr and 38,668 white British individuals from UK Biobank were included in the analysis of each trait after removal of related individuals. The measured lipid traits were corrected for medication use and were inverse-normalized after correction for age, sex, and batch. Principal components 1-20 constructed from the GRM were included as covariates in the calculation of genetic correlation. Analysis within the Million Veteran Program included 24,502 European and 21,950 African American unrelated individuals. Maximum measured values were used for LDL-C, TC, and triglycerides and minimum values for HDL-C. Lipid traits were inverse-normalized after correction for age and sex with principal components 1-20 included as covariates in the calculation of genetic correlation.

Credible sets

Credible sets of potentially causal variants were generated for each of the loci identified in the trans-ancestry meta-analysis. We determined 99% credible sets of variants that encompassed the causal variant with 99% posterior probability. Regions for construction of credible sets were defined as the \pm 500 kb region around each index variant. Bayes factors^{20,21} (BF) for each variant in the ancestry-specific meta-analysis were approximated by:

$$BF \approx \exp \left[0.5 \left(\frac{\beta^2}{SE^2} - \log(N_{AS}) \right) \right]$$

where β and SE are the effect sizes and standard errors from the RAREMETAL meta-analysis, and N_{AS} is the ancestry-specific sample size. A full derivation is included in the **Supplementary**

Methods. 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 using the following equation:

$$BF \approx \exp \left[0.5 \left(\frac{\beta^2 N_{TE}}{SE^2 N_{AS}} - \log(N_{TE}) \right) \right]$$

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 $\pm 500\text{kb}$ 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 LDpair²² 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 SnpEff¹² using WGS¹¹ and with VEP²³. Protein numbering was taken from dbSNP²⁴. eQTL colocalization was performed using coloc²⁵ version 3.2.1 with R version 3.4.3 using the default parameters. Results from GTEx V8²⁶ were compared with the GWAS signals in the region defined by the larger of $\pm 0.25\text{cM}$ or $\pm 500\text{kb}$ surrounding each index variant. The eQTL and GWAS signals (based on p-values from MR-MEGA) were considered to be colocalized if $PP3 + PP4 \geq 0.8$ 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.

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 results²⁷ 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 (various pruning and thresholding parameters or PRS-CS, **Supplementary Note**).

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 PLINK²⁸ 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 5×10^{-10} , 5×10^{-9} , 5×10^{-8} , 5×10^{-7} , 5×10^{-6} , 5×10^{-5} , 5×10^{-4} , 5×10^{-3} , and 5×10^{-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-CS²⁹ 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 authors²⁹. This included removing variants with $MAF \leq 0.01$, ambiguous A/T or G/C variants, and restricting to variants included in HapMap3. Pairwise LD matrices within pre-defined LD blocks³⁰ (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 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 Health³¹ (ELGH; SAS N=15,242), Tohoku Medical Megabank Community Cohort Study (ToMMo; EAS N=28,217), Korean Genome and Epidemiology Study³² (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 Veteran 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 validation cohort is included in **Supplementary Table 16**. African admixture for MGI was calculated using all African-ancestry individuals in 1000 Genomes with ADMIXTURE v1.3³³. African admixture for MVP was calculated using the YRI and LWK African-ancestry individuals in 1000 Genomes.

References Cited in the Methods section

- 1 Klarin, D. *et al.* Genetics of blood lipids among ~300,000 multi-ethnic participants of the Million Veteran Program. *Nature genetics* **50**, 1514-1523, doi:10.1038/s41588-018-0222-9 (2018).
- 2 Das, S. *et al.* Next-generation genotype imputation service and methods. *Nature genetics* **48**, 1284-1287, doi:10.1038/ng.3656 (2016).
- 3 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, doi:[https://doi.org/10.1016/S0140-6736\(05\)67394-1](https://doi.org/10.1016/S0140-6736(05)67394-1) (2005).
- 4 Loh, P.-R. *et al.* Efficient Bayesian mixed-model analysis increases association power in large cohorts. *Nature genetics* **47**, 284-290, doi:10.1038/ng.3190 (2015).
- 5 Zhou, W. *et al.* Efficiently controlling for case-control imbalance and sample relatedness in large-scale genetic association studies. *Nature genetics* **50**, 1335-1341, doi:10.1038/s41588-018-0184-y (2018).
- 6 Winkler, T. W. *et al.* Quality control and conduct of genome-wide association meta-analyses. *Nat Protoc* **9**, 1192-1212, doi:10.1038/nprot.2014.071 (2014).
- 7 Feng, S., Liu, D., Zhan, X., Wing, M. K. & Abecasis, G. R. RAREMETAL: fast and powerful meta-analysis for rare variants. *Bioinformatics* **30**, 2828-2829, doi:10.1093/bioinformatics/btu367 (2014).
- 8 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, doi:10.1093/hmg/ddx280 (2017).
- 9 Willer, C. J., Li, Y. & Abecasis, G. R. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **26**, 2190-2191, doi:10.1093/bioinformatics/btq340 (2010).
- 10 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, doi:10.1038/ng.3571 (2016).
- 11 Liu, X. *et al.* WGSAs: an annotation pipeline for human genome sequencing studies. *Journal of Medical Genetics* **53**, 111-112, doi:10.1136/jmedgenet-2015-103423 (2016).
- 12 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* **6**, 80-92, doi:10.4161/fly.19695 (2012).
- 13 Wang, K., Li, M. & Hakonarson, H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Research* **38**, e164-e164, doi:10.1093/nar/gkq603 (2010).
- 14 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, doi:10.1093/nar/gky1120 (2019).
- 15 Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841-842, doi:10.1093/bioinformatics/btq033 (2010).
- 16 Liu, D. J. *et al.* Meta-analysis of gene-level tests for rare variant association. *Nature genetics* **46**, 200-204, doi:10.1038/ng.2852 (2014).
- 17 Brown, B. C., Ye, C. J., Price, A. L. & Zaitlen, N. Transethnic Genetic-Correlation Estimates from Summary Statistics. *American journal of human genetics* **99**, 76-88, doi:10.1016/j.ajhg.2016.05.001 (2016).
- 18 Guo, J. *et al.* Quantifying genetic heterogeneity between continental populations for human height and body mass index. *Scientific reports* **11**, 5240, doi:10.1038/s41598-021-84739-z (2021).

- 19 Lee, S. H., Yang, J., Goddard, M. E., Visscher, P. M. & Wray, N. R. Estimation of pleiotropy between complex diseases using single-nucleotide polymorphism-derived genomic relationships and restricted maximum likelihood. *Bioinformatics* **28**, 2540-2542, doi:10.1093/bioinformatics/bts474 (2012).
- 20 Maller, J. B. *et al.* Bayesian refinement of association signals for 14 loci in 3 common diseases. *Nature genetics* **44**, 1294-1301, doi:10.1038/ng.2435 (2012).
- 21 Kass, R. E. & Raftery, A. E. Bayes Factors. *Journal of the American Statistical Association* **90**, 773-795, doi:10.1080/01621459.1995.10476572 (1995).
- 22 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**, 3555-3557, doi:10.1093/bioinformatics/btv402 (2015).
- 23 McLaren, W. *et al.* The Ensembl Variant Effect Predictor. *Genome Biology* **17**, 122, doi:10.1186/s13059-016-0974-4 (2016).
- 24 Sherry, S. T. *et al.* dbSNP: the NCBI database of genetic variation. *Nucleic acids research* **29**, 308-311, doi:10.1093/nar/29.1.308 (2001).
- 25 Giambartolomei, C. *et al.* Bayesian Test for Colocalisation between Pairs of Genetic Association Studies Using Summary Statistics. *PLoS genetics* **10**, e1004383, doi:10.1371/journal.pgen.1004383 (2014).
- 26 Aguet, F. *et al.* The GTEx Consortium atlas of genetic regulatory effects across human tissues. *bioRxiv*, 787903, doi:10.1101/787903 (2019).
- 27 Teslovich, T. M. *et al.* Biological, clinical and population relevance of 95 loci for blood lipids. *Nature* **466**, 707-713, doi:10.1038/nature09270 (2010).
- 28 Purcell, S. *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *American journal of human genetics* **81**, 559-575, doi:10.1086/519795 (2007).
- 29 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, doi:10.1038/s41467-019-09718-5 (2019).
- 30 Berisa, T. & Pickrell, J. K. Approximately independent linkage disequilibrium blocks in human populations. *Bioinformatics* **32**, 283-285, doi:10.1093/bioinformatics/btv546 (2016).
- 31 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, doi:10.1093/ije/dyz174 (2019).
- 32 Moon, S. *et al.* The Korea Biobank Array: Design and Identification of Coding Variants Associated with Blood Biochemical Traits. *Scientific reports* **9**, 1382, doi:10.1038/s41598-018-37832-9 (2019).
- 33 Alexander, D. H., Novembre, J. & Lange, K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res* **19**, 1655-1664, doi:10.1101/gr.094052.109 (2009).