The Trans-Ancestral Genomic Architecture of Glycaemic Traits

Ji Chen[1,2#], Cassandra N. Spracklen[3,4#], Gaëlle Marenne[1,5#], Arushi Varshney[6#], Laura J Corbin[7,8#], Jian'an Luan[9], Sara Willems[9], Ying Wu[3], Xiaoshuai Zhang[9,10], Momoko Horikoshi[11,12,13], Thibaud S Boutin[14], Reedik Mägi[15], Johannes Waage[16,17], Achilleas Pitsilides[18], Ruifang Li-Gao[19], Kei Hang[20,21,22], Jie Yao[23], Mila D Anasanti[24], Audrey Y Chu[25], Annique Claringbould[26], Jani Heikkinen[24], Jaeyoung Hong[18], Jouke-Jan Hottenga[27,28], Shaofeng Huo[29], Marika A. Kaakinen[30,24], Tin Louie[31], Winfried März[32,33,34], Hortensia Moreno-Macias[35], Anne Ndungu[12], Sarah C. Nelson[31], Ilja M. Nolte[36], Kari North[37], Chelsea K. Raulerson[3], Debashree Ray[38], Rebecca Rohde[37], Denis Rybin[18], Claudia Schurmann[39,40], Xueling Sim[41,42,43], Loz Southam[1], Isobel Stewart[9], Carol A. Wang[44], Yujie Wang[37], Peitao Wu[18], Weihua Zhang[45,46], Tarunveer S. Ahluwalia[16,17,47], Emil VR Appel[48], Lawrence F. Bielak[49], Jennifer A. Brody[50], Noel P Burtt[51], Claudia P Cabrera[52,53], Brian E Cade[54,55], Jin Fang Chai[56], Xiaoran Chai[57,58], Li-Ching Chang[59], Chien-Hsiun Chen[59], Brian H Chen[60], Kumaraswamy N Chitrala[61], Yen-Feng Chiu[62], Hugoline G. de Haan[19], Graciela E Delgado[34], Ayse Demirkan[63], Qing Duan[3,64], Jorgen Engmann[65], Segun A Fatumo[66,67,68], Javier Gayán[69], Franco Giulianini[70], Jung Ho Gong[20], Stefan Gustafsson[71], Yang Hai[72], Fernando P Hartwig[73,74], Jing He[75], Yoriko Heianza[76], Tao Huang[77], Alicia Huerta-Chagoya[78,79], Mi Yeong Hwang[80], Richard A. Jensen[50], Takahisa Kawaguchi[81], Katherine A Kentistou[82,83], Young Jin Kim[80], Marcus E Kleber[34], Ishminder K Kooner[84], Shuiqing Lai[85], Leslie A Lange[86], Carl D Langefeld[87], Marie Lauzon[23], Man Li[88], Symen Ligthart[63], Jun Liu[63,89], Marie Loh[90,45], Jirong Long[75], Valeriya Lyssenko[91,92], Massimo Mangino[93,94], Carola Marzi[95,96], May E Montasser[97], Abhishek Nag[12], Masahiro Nakatochi[98], Damia Noce[99], Raymond Noordam[100], Giorgio Pistis[101], Michael Preuss[39,102], Laura Raffield[3], Laura J. Rasmussen-Torvik[103], Stephen S Rich[104,105], Neil R Robertson[12], Rico Rueedi[106,107], Kathleen Ryan[97], Serena Sanna[101,26], Richa Saxena[108,109,110], Katharina E Schraut[82,83], Bengt Sennblad[111], Kazuya Setoh[81], Albert V Smith[112,113], Lorraine Southam[114,115], Thomas Sparsø[48], Rona J Strawbridge[116,117], Fumihiko Takeuchi[118], Jingyi Tan[23], Stella Trompet[100,119], Erik van den Akker[120,121,122], Peter J Van der Most[36], Niek Verweij[123,124], Mandy Vogel[125], Heming Wang[54], Chaolong Wang[126,127], Nan Wang[128,129], Helen R Warren[52,53], Wanqing Wen[75], Tom Wilsgaard[130], Andrew Wong[131], Andrew R Wood[132], Tian Xie[36], Mohammad Hadi Zafarmand[133,134], Jing-Hua Zhao[135], Wei Zhao[49], Najaf Amin[63], Zorayr Arzumanyan[23], Arne Astrup[136], Stephan JL Bakker[137], Damiano Baldassarre[138,139], Marian Beekman[120], Richard N Bergman[140], Alain Bertoni[141], Matthias Blüher[142], Lori L. Bonnycastle[143], Stefan R Bornstein[144], Donald W Bowden[145], Qiuyin Cai[75], Archie Campbell[146,147], Harry Campbell[82], Yi Cheng Chang[148,149,150], Eco J.C. de Geus[27,28], Abbas Dehghan[63], Shufa Du[151], Gudny Eiriksdottir[113], Aliki Eleni Farmaki[152,153], Mattias Frånberg[154], Christian Fuchsberger[99], Yutang Gao[155], Anette P Gjesing[48], Anuj Goel[156,12], Sohee Han[80], Catharina A Hartman[157], Christian Herder[158,159,160], Andrew A. Hicks[99], Chang-Hsun Hsieh[161,162], Willa A. Hsueh[163], Sahoko Ichihara[164], Michiya Igase[165], M. Arfan Ikram[63], W. Craig Johnson[31], Marit E Jørgensen[17,166], Peter K Joshi[82], Rita R Kalyani[167], Fouad R. Kandeel[168], Tomohiro Katsuya[169,170], Chiea Chuen Khor[127], Wieland Kiess[125], Ivana Kolcic[171], Teemu Kuulasmaa[172], Johanna Kuusisto[173], Kristi Läll[15], Kelvin Lam[23], Deborah A Lawlor[7,8], Nanette R. Lee[174,175], Rozenn N. Lemaitre[50], Honglan Li[176], Lifelines Cohort Study[177], Shih-Yi Lin[178,179,180], Jaana Lindström[181], Allan Linneberg[182,183], Jianjun Liu[127,184], Carlos Lorenzo[185], Tatsuaki Matsubara[186], Fumihiko Matsuda[81], Geltrude Mingrone[187], Simon Mooijaart[100], Sanghoon Moon[80], Toru Nabika[188], Girish N. Nadkarni[39], Jerry L. Nadler[189], Mari Nelis[15], Matthew J Neville[11], Jill M Norris[190], Yasumasa Ohyagi[191], Annette Peters[192,96,193], Patricia A. Peyser[49], Ozren Polasek[171], Qibin Qi[194], Dennis Raven[157], Dermot F Reilly[195], Alex Reiner[196], Fernando Rivideneira[197], Kathryn Roll[23], Igor Rudan[198], Charumathi Sabanayagam[57,199], Kevin Sandow[23], Naveed Sattar[200], Annette Schürmann[201,202], Jinxiu Shi[203], Heather M Stringham[43,204], Kent D. Taylor[23], Tanya M. Teslovich[205], Betina Thuesen[182], Paul RHJ Timmers[82], Elena Tremoli[139], Michael Y Tsai[206], Andre Uitterlinden[197], Rob M van Dam[41,184,207], Diana van Heemst[100], Astrid van Hylckama Vlieg[19], Jana V Van Vliet-Ostaptchouk[36], Jagadish Vangipurapu[208], Henrik Vestergaard[48,17], Tao Wang[194], Ko Willems van Dijk[209,210,211], Tatijana Zemunik[212], Goncalo R Abecasis[43], Linda S. Adair[151,213], Carlos Alberto

Aguilar-Salinas[214,215,216], Marta E Alarcón-Riquelme[217,218], Ping An[219], Larissa Aviles-Santa[220], Diane M Becker[221], Lawrence J Beilin[222], Sven Bergmann[106,107,223], Hans Bisgaard[16], Corri Black[224], Michael Boehnke[225,226], Eric Boerwinkle[227,228], Bernhard O Böhm[229,230], Klaus Bønnelykke[16], D I. Boomsma[27,28], Erwin P. Bottinger[39,40,231], Thomas A Buchanan[232,233,129], Mickaël Canouil[234,235], Mark J Caulfield[52,53], John C. Chambers[90,45,84,236,237], Daniel I. Chasman[70,238], Yii-Der Ida Chen[23], Ching-Yu Cheng[57,199], Francis S. Collins[143], Adolfo Correa[239], Francesco Cucca[101], H. Janaka de Silva[240], George Dedoussis[241], Sölve Elmståhl[242], Michele K. Evans[61], Ele Ferranni[243], Luigi Ferrucci[244], Jose C Florez[245,246,247], Paul Franks[248], Timothy M Frayling[132], Philippe Froguel[234,235,249], Bruna Gigante[250], Mark O. Goodarzi[251], Penny Gordon-Larsen[151,213], Harald Grallert[95,96], Niels Grarup[48], Sameline Grimsgaard[130], Leif Groop[252,253], Vilmundur Gudnason[113,254], Xiuqing Guo[23], Anders Hamsten[117], Torben Hansen[48], Caroline Hayward[255], Susan R. Heckbert[256], Bernardo L Horta[73], Wei Huang[203], Erik Ingelsson[257], Pankow S James[258], Jost B Jonas[259,260], J. Wouter Jukema[119,261], Pontiano Kaleebu[262], Robert Kaplan[194,196], Sharon L.R. Kardia[49], Norihiro Kato[118], Sirkka M. Keinanen-Kiukaanniemi[263,264], Bong-Jo Kim[80], Mika Kivimaki[265], Heikki A. Koistinen[266,267,268], Jaspal S. Kooner[84,236,237,269], Antje Körner[125], Peter Kovacs[142,270], Diana Kuh[131], Meena Kumari[271], Zoltan Kutalik[272,107], Markku Laakso[173], Timo A. Lakka[273,274,275], Lenore J Launer[276], Karin Leander[277], Huaixing Li[29], Xu Lin[29], Lars Lind[278], Cecilia Lindgren[12,279,280], Simin Liu[85], Ruth J.F. Loos[39,102], Patrik Magnusson[281], Anubha Mahajan[12], Andres Metspalu[15], Dennis O Mook-Kanamori[19,282], Trevor A Mori[222], Patricia B Munroe[52,53], Inger Njølstad[130], Jeffrey R O'Connell[97], Albertine J Oldehinkel[157], Ken K Ong[9], Sandosh Padmanabhan[283], Colin N.A. Palmer[284], Nicholette D Palmer[145], Oluf Pedersen[48], Craig E Pennell[44], David J Porteous[146,285], Peter P. Pramstaller[99], Michael A. Province[219], Bruce M. Psaty[50,256,286,287], Lu Qi[76], Leslie J. Raffel[288], Rainer Rauramaa[275], Susan Redline[54,55], Paul M Ridker[70,238], Frits R. Rosendaal[19], Timo E. Saaristo[289,290], Manjinder Sandhu[291], Jouko Saramies[292], Neil Schneiderman[293], Peter Schwarz[144,294,202], Laura J. Scott[225,226], Elizabeth Selvin[38], Peter Sever[269], Xiao-ou Shu[75], P Eline Slagboom[120], Kerrin S Small[93], Blair H Smith[295], Harold Snieder[36], Tamar Sofer[296,246], Thorkild I.A. Sørensen[297,298,7,8], Tim D Spector[93], Alice Stanton[299], Claire J Steves[93,300], Michael Stumvoll[142], Liang Sun[29], Yasuharu Tabara[81], E Shyong Tai[184,301,302], Nicholas J Timpson[7,8], Anke Tönjes[142], Jaakko Tuomilehto[303,304,305], Teresa Tusie[79,306], Matti Uusitupa[307], Pim van der Harst[123,26], Cornelia van Duijn[308,63], Veronique Vitart[255], Peter Vollenweider[309], Tanja GM Vrijkotte[133], Lynne E Wagenknecht[310], Mark Walker[311], Ya X Wang[260], Nick J Wareham[9], Richard M Watanabe[128,233,129], Hugh Watkins[156,12], Wen B Wei[312], Ananda R Wickremasinghe[313], Gonneke Willemsen[27,28], James F Wilson[82,255], Tien-Yin Wong[57,199], Jer-Yuarn Wu[59], Anny H Xiang[314], Lisa R Yanek[221], Loïc Yengo[234,235,315], Mitsuhiro Yokota[316], Eleftheria Zeggini[317,318,319], Wei Zheng[320], Alan B Zonderman[61], Jerome I Rotter[23], Anna L Gloyn[11,12,321,322], Mark I. McCarthy[11,323,321,12], Josée Dupuis[18], James B Meigs[324,246,325], Robert Scott[9], Inga Prokopenko[30,24], Aaron Leong[326,327,328], Ching-Ti Liu[18], Stephen CJ Parker[6,329#], Karen L. Mohlke[3#], Claudia Langenberg[9#], Eleanor Wheeler[1,9#], Andrew P. Morris[330,331,12#], Inês Barroso[1,9,2#]

[1]Department of Human Genetics, Wellcome Sanger Institute, Hinxton, Cambridge, UK, [2]Exeter Centre of Excellence in Diabetes (ExCEeD), Exeter Medical School, University of Exeter, Exeter, UK, [3]Department of Genetics, University of North Carolina, Chapel Hill, NC, USA, [4]Department of Biostatistics and Epidemiology, University of Massachusetts, Amherst, MA, USA, [5]Inserm, Univ Brest, EFS, UMR 1078, GGB, Brest, France, [6]Department of Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, MI, USA, [7]MRC Integrative Epidemiology Unit, University of Bristol, Bristol, Bristol, UK, [8]Department of Population Health Sciences, Bristol Medical School, University of Bristol, Bristol, Bristol, UK, [9]MRC Epidemiology Unit, Institute of Metabolic Science, University of Cambridge, Cambridge, UK, [10]Department of Biostatistics, School of Public Health, Shandong University, Jinan, Shandong, China, [11]Oxford Centre for Diabetes, Endocrinology and Metabolism, Radcliffe Department of Medicine, University of Oxford, Oxford, UK, [12]Wellcome Centre for Human Genetics, University of Oxford, Oxford, UK, [13]Laboratory for Genomics of Diabetes and Metabolism, RIKEN Centre for Integrative Medical Sciences, Yokohama, Japan, [14]Medical Research Council Human Genetics Unit, Institute for Genetics and Molecular Medicine, Edinburgh, UK,

[15]Estonian Genome Center, Institute of Genomics, University of Tartu, Tartu, Estonia, [16]COPSAC, Copenhagen Prospective Studies on Asthma in Childhood, Herlev and Gentofte Hospital, University of Copenhagen, Copenhagen, Denmark, [17]Steno Diabetes Center Copenhagen, Gentofte, Denmark, [18]Department of Biostatistics, Boston University School of Public Health, Boston, MA, USA, [19]Department of Clinical Epidemiology, Leiden University Medical Center, Leiden, The Netherlands, [20]Department of Epidemiology, Brown University, Providence, RI, USA, [21]Department of Biomedical Sciences, City University of Hong Kong, Hong Kong SAR, China, [22]Department of Electrical Engineering, City University of Hong Kong, Hong Kong SAR, China, [23]The Institute for Translational Genomics and Population Sciences, `, The Lundquist Institute for Biomedical Innovation at Harbor-UCLA Medical Center, Torrance, CA, USA, [24]Department of Metabolism, Digestion, and Reproduction, Imperial College London, London, UK, [25]Division of Preventive Medicine, Brigham and Women's Hospital, Boston, MA, USA, [26]Department of Genetics, University Medical Center Groningen, University of Groningen, Groningen, The Netherlands, [27]Department of Biological Psychology, Faculty of Behaviour and Movement Sciences, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands, [28]Amsterdam Public Health Research Institute, Amsterdam Universities Medical Center, Amsterdam, The Netherlands, [29]CAS Key Laboratory of Nutrition, Metabolism and Food Safety, Shanghai Institute of Nutrition and Health, University of Chinese Academy of Sciences, Chinese Academy of Sciences, Shanghai, China, [30]Department of Clinical and Experimental Medicine, Section of Statistical Multi-Omics, University of Surrey, Guildford, UK, [31]Department of Biostatistics, University of Washington, Seattle, WA, USA, [32]SYNLAB Academy, SYNLAB Holding Deutschland GmbH, Mannheim, Germany, [33]Clinical Institute of Medical and Chemical Laboratory Diagnostics, Medical University Graz, Graz, Austria, [34]Vth Department of Medicine (Nephrology, Hypertensiology, Rheumatology, Endocrinology, Diabetology), Medical Faculty Mannheim, Heidelberg University, Mannheim, Baden-Württemberg, Germany, [35]Department of Economics, Metropolitan Autonomous University, Mexico City, Mexico, [36]Department of Epidemiology, University Medical Center Groningen, University of Groningen, Groningen, The Netherlands, [37]CVD Genetic Epidemiology Computational Laboratory, Gillings School of Global Public Health, University of North Carolina, Chapel Hill, NC, USA, [38]Department of Epidemiology, Johns Hopkins University, Baltimore, MD, USA, [39]The Charles Bronfman Institute for Personalized Medicine, Icahn School of Medicine at Mount Sinai, New York, NY, USA, [40]HPI Digital Health Center, Digital Health and Personalized Medicine, Hasso Plattner Institute, Potsdam, Germany, [41]Saw Swee Hock School of Public Health, National University of Singapore, Singapore, Singapore, [42]Center for Statistical Genetics, University of Michigan, Ann Arbor, MI, USA, [43]Department of Biostatistics, University of Michigan, Ann Arbor, MI, USA, [44]School of Medicine and Public Health, Faculty of Medicine and Health, The University of Newcastle, Newcastle, NSW, Australia, [45]Department of Epidemiology and Biostatistics, Imperial College London, London, UK, [46]Department of Cardiology, Ealing Hospital, London North West Healthcare NHS Trust, Middlesex, UK, [47]The Bioinformatics Centre, Department of Biology, University of Copenhagen, Copenhagen, Denmark, [48]Novo Nordisk Foundation Center for Basic Metabolic Research, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark, [49]Department of Epidemiology, School of Public Health, University of Michigan, Ann Arbor, MI, USA, [50]Department of Medicine, Cardiovascular Health Research Unit, University of Washington, Seattle, WA, USA, [51]Metabolism Program, Program in Medical and Population Genetics, Broad Institute, Cambridge, MA, USA, [52]Department of Clinical Pharmacology, William Harvey Research Institute Barts, The London School of Medicine and Dentistry,, Queen Mary University of London, London, UK, [53]NIHR Barts Cardiovascular Biomedical Research Centre, Queen Mary University of London, London, UK, [54]Department of Medicine, Sleep and Circadian Disorders, Brigham and Women's Hospital, Boston, MA, USA, [55]Department of Medicine, Sleep Medicine, Harvard Medical School, Boston, MA, USA, [56]Saw Swee Hock School of Public Health, National Univeristy of Singapore and National University Health System, Singapore, Singapore, [57]Ocular Epidemiology, Singapore Eye Research Institute, Singapore National Eye Centre, Singapore, Singapore, [58]Department of Ophthalmology, National University of Singapore and National University Health System, Singapore, Singapore,

[59]Institute of Biomedical Sciences, Academia Sinica, Taipei, Taiwan, Taiwan, [60]Department of Epidemiology, The Herbert Wertheim School of Public Health and Human Longevity Science, UC San Diego, La Jolla, CA, USA, [61]Laboratory of Epidemiology and Population Sciences, National Institute on Aging Intramural Research Program, National Institutes of Health, Baltimore, MD, USA, [62]Institute of Population Health Sciences, National Health Research Institutes, Miaoli, Taiwan, [63]Department of Epidemiology, Erasmus Medical Center, Rotterdam, The Netherlands, [64]Department of Statistics, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA, [65]Institute of Cardiovascular Science, UCL, London, UK, [66]Uganda Medical Informatics Centre (UMIC), MRC/UVRI and London School of Hygiene & Tropical Medicine (Uganda Research Unit), Entebbe, Uganda, [67]London School of Hygiene & Tropical Medicine, London, UK, [68]H3Africa Bioinformatics Network (H3ABioNet) Node, Centre for Genomics Research and Innovation, NABDA/FMST, Abuja, Nigeria, [69]Bioinfosol, Sevilla, Spain, [70]Division of Preventive Medicine, Brigham and Women's Hospital, Boston, MA, USA, [71]Department of Medical Sciences, Molecular Epidemiology and Science for Life Laboratory, Uppsala University, Uppsala, Sweden, [72]Department of Statistics, The University of Auckland, Science Center, Auckland, New Zealand, [73]Postgraduate Program in Epidemiology, Federal University of Pelotas, Pelotas, RS, Brazil, [74]MRC Integrative Epidemiology Unit, University of Bristol, Bristol, UK, [75]Department of Medicine, Epidemiology, Vanderbilt University Medical Center, Nashville, TN, USA, [76]Department of Epidemiology, Tulane University Obesity Research Center,, Tulane University, New Orleans, USA, [77]Department of Epidemiology and Biostatistics, School of Public Health, Peking University, Beijing, China, [78]Molecular Biology and Genomic Medicine Unit, National Council for Science and Technology, Mexico City, Mexico, [79]Molecular Biology and Genomic Medicine Unit, National Institute of Medical Sciences and Nutrition, Mexico City, Mexico, [80]Division of Genome Research, Center for Genome Research, Korea National Institute of Health, Cheongju-si, Chungcheongbuk-do, South Korea, [81]Center for Genomic Medicine, Kyoto Univsersity Graduate School of Medicine, Kyoto, Japan, [82]Centre for Global Health Research, Usher Institute of Population Health Sciences and Informatics, University of Edinburgh, Edinburgh, Scotland, [83]Centre for Cardiovascular Sciences, Queen's Medical Research Institute, University of Edinburgh, Edinburgh, Scotland, [84]Department of Cardiology, Ealing Hospital, London North West Healthcare NHS Trust, Middlesex, UK, [85]Department of Epidemiology, Brown University School of Public Health, Providence, RI, USA, [86]Department of Medicine, Divison of Biomedical Informatics and Personalized Medicine, University of Colorado Anschutz Medical Campus, Denver, CO, USA, [87]Department of Public Health Sciences, Biostatistics and Data Science, Wake Forest School of Medicine, Winston-Salem, NC, USA, [88]Department of Medicine, Division of Nephrology and Hypertension, University of Utah, Salt Lake City, UT, USA, [89]Nuffield Department of Population Health, University of Oxford, Oxford, UK, [90]Lee Kong Chian School of Medicine, Nanyang Technological University, Singapore, Singapore, [91]Department of Clinical Science, Center for Diabetes Research, University of Bergen, Bergen, Norway, [92]Department of Clinical Sciences, Lund University Diabetes Centre, Lund University, Malmo, Sweden, [93]Department of Twin Research and Genetic Epidemiology, School of Life Course Sciences, King's College London, London, UK, [94]NIHR Biomedical Research Centre, Guy's and St Thomas' Foundation Trust, London, UK, [95]Institute of Epidemiology, Research Unit of Molecular Epidemiology, Helmholtz Zentrum München Research Center for Environmental Health, Neuherberg, Bavaria, Germany, [96]German Center for Diabetes Research (DZD), Neuherberg, Bavaria, Germany, [97]Department of Medicine, Division of Endocrinology, Diabetes, and Nutrition, University of Maryland School of Medicine, Baltimore, MD, USA, [98]Public Health Informatics Unit, Department of Integrated Sciences, Nagoya University Graduate School of Medicine, Nagoya, Japan, [99]Institute for Biomedicine, Eurac Research, Bolzano, BZ, Italy, [100]Department of Internal Medicine, Section of Gerontology and Geriatrics, Leiden University Medical Center, Leiden, The Netherlands, [101]Istituto di Ricerca Genetica e Biomedica (IRGB), Consiglio Nazionale delle Ricerche (CNR), Monserrato, Italy, Italy, [102]The Mindich Child Health and Development Institute for Personalized Medicine, Icahn School of Medicine at Mount Sinai, New York, NY, USA, [103]Department of Preventive Medicine, Northwestern University Feinberg School of Medicine, Chicago, IL, USA, [104]Center for Public Health Genomics, University of Virginia,

Charlottesville, VA, USA, [105]Department of Public Health Sciences, University of Virginia, Charlottesville, VA, USA, [106]Department of Computational Biology, University of Lausanne, Lausanne, Switzerland, [107]Swiss Institute of Bioinformatics, Lausanne, Switzerland, [108]Center for Genomic Medicine, Massachusetts General Hospital, Harvard Medical School, Boston, MA, USA, [109]Department of Anesthesia, Critical Care and Pain Medicine, Massachusetts General Hospital, Boston, MA, USA, [110]Program in Medical and Population Genetics,, Broad Institute, Cambridge, MA, USA, [111]Department of Cell and Molecular Biology., National Bioinformatics Infrastructure Sweden,, Science for Life Laboratory, Uppsala University, Uppsala, Sweden, [112]Department of Biostatistics, University of Mitchican, Ann Arbor, MI, USA, [113]Icelandic Heart Association, Kopavogur, Iceland, [114]Institute of Translational Genomics, Helmholtz Zentrum München – German Research Center for Environmental Health, Institute of Translational Genomics, Helmholtz Zentrum München – German Research Center for Environmental Health, Neuherberg, Germany, [115]Wellcome Sanger Institute, Hinxton, Cambridge, UK, [116]Institute of Health and Wellbeing, University of Glasgow, Glasgow, Glasgow, UK, [117]Department of Medicine Solna, Cardiovascular medicine, Karolinska Institutet, Stockholm, Sweden, [118]National Center for Global Health and Medicine, Tokyo, Japan, [119]Department of Cardiology, Leiden University Medical Center, Leiden, The Netherlands, [120]Department of Biomedical Data Sciences, Molecular Epidemiology, Leiden University Medical Center, Leiden, The Netherlands, [121]Department of Pattern Recognition & Bioinformatics, Delft University of Technology, Delft, The Netherlands, [122]Department of Biomedical Data Sciences, Leiden Computational Biology Center, Leiden University Medical Center, Leiden, The Netherlands, [123]Department of Cardiology, University Medical Center Groningen, University of Groningen, Groningen, The Netherlands, [124]Genomics plc, Oxford, UK, [125]Center of Pediatric Research, University Children´s Hospital Leipzig, University of Leipzig Medical Center, Leipzig, Germany, [126]Department of Epidemiology and Biostatistics, School of Public Health, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China, [127]Genome Institute of Singapore, Agency for Science, Technology and Research, Singapore, Singapore, [128]Department of Preventive Medicine, Keck School of Medicine of USC, Los Angeles, CA, USA, [129]USC Diabetes and Obesity Research Institute, Keck School of Medicine of USC, Los Angeles, CA, USA, [130]Department of Community Medicine, Faculty of Health Sciences, UIT the Arctic University of Norway, Tromsø, Norway, [131]MRC Unit for Lifelong Health & Ageing at UCL, London, UK, [132]Department of Human Genetics, University of Exeter RILD Building Royal Devon and Exeter NHS trust Barrack Road, Exeter, UK, [133]Department of Public Health, Amsterdam Public Health Research Institute, Amsterdam Universities Medical Center, Amsterdam, The Netherlands, [134]Department of Clinical Epidemiology, Biostatistics, and Bioinformatics, Amsterdam Public Health Research Institute, Amsterdam Universities Medical Center, Amsterdam, The Netherlands, [135]Department of Public Health and Primary Care, School of Clinical Medicine, University of Cambridge, Cambridge, UK, [136]Department of Nutrition, Exercise, and Sports, Faculty of Science, University of Copenhagen, Copenhagen, Denmark, [137]Department of Internal Medicine, University Medical Center Groningen, University of Groningen, Groningen, The Netherlands, [138]Department of Medical Biotechnology and Translational Medicine, University of Milan, Milan, Milan, Italy, [139]Centro Cardiologico Monzino, IRCCS, Milan, Italy, [140]Department of Physiology and Biophysics, University of Southern California, Los Angeles, CA, USA, [141]Department of Epidemiology and Prevention, Division of Public Health Sciences, Wake Forest School of Medicine, Winston-Salem, NC, USA, [142]Medical Department III – Endocrinology, Nephrology, Rheumatology, University of Leipzig Medical Center, Leipzig, Germany, [143]Medical Genomics and Metabolic Genetics Branch, National Human Genome Research Institute, National Institues of Health, Bethesda, MD, USA, [144]Department for Prevention and Care of Diabetes, Faculty of Medicine Carl Gustav Carus, Technische Universität Dresden, Dresden, Germany, [145]Department of Biochemistry, Wake Forest School of Medicine, Winston-Salem, NC, USA, [146]Centre for Genomic and Experimental Medicine, Institute of Genetics & Molecular Medicine, University of Edinburgh, Western General Hospital, Edinburgh, UK, [147]Usher Institute for Population Health Sciences and Informatics, University of Edinburgh, Edinburgh, UK, [148]Internal Medicine, National Taiwan University Hospital, Taipei, Taiwan, [149]Medical Genomics and

Proteomics, National Taiwan University, Taipei, Taiwan, [150]Biomedical Sciences, Institute of Biomedical Sciences, Academia Sinica, Taiepi, Taiwan, [151]Department of Nutrition, Gillings School of Global Public Health, University of North Carolina, Chapel Hill, NC, USA, [152]Department of Population Science and Experimental Medicine, Institute of Cardiovascular Science, University College London, London, UK, [153]Department of Nutrition and Dietetics, School of Health Science and Education, Harokopio University of Athens, Athens, Greece, [154]Department of Medicine Solna, Cardiovascular medicine, Stockholm, Sweden, [155]Department of Epidemiology, Shanghai Cancer Institute, Shanghai, China, [156]Division of Cardiovascular Medicine, Radcliffe Department of Medicine, University of Oxford, Oxford, UK, [157]Department of Psychiatry, Interdisciplinary Center Psychopathology and Emotion Regulation, University Medical Center Groningen, University of Groningen, Groningen, The Netherlands, [158]Institute for Clinical Diabetology, German Diabetes Center, Leibniz Center for Diabetes Research at Heinrich Heine University Düsseldorf, Düsseldorf, Germany, [159]Division of Endocrinology and Diabetology, Medical Faculty, Heinrich Heine University Düsseldorf, Düsseldorf, Germany, [160]German Center for Diabetes Research (DZD), Düsseldorf, Germany, [161]Internal Medicine, Endocrine & Metabolism, Tri-Service General Hospital, Taipei, Taiwan, [162]School of Medicine, National Defense Medical Center, Taipei, Taiwan, [163]Internal Medicine, Endocrinology, Diabetes & Metabolism, Diabetes and Metabolism Research Center, The Ohio State University Wexner Medical Center, Columbus, OH, USA, [164]Department of Environmental and Preventive Medicine, Jichi Medical University School of Medicine, Shimotsuke, Japan, [165]Department of Anti-aging Medicine, Ehime University Graduate School of Medicine, Toon, Japan, [166]National Institute of Public Health, University of Southern Denmark, Odense, Denmark, [167]Department of Medicine, Endocrinology, Diabetes & Metabolism, Johns Hopkins University School of Medicine, Baltimore, MD, USA, [168]Clinical Diabetes, Endocrinology & Metabolism, Translational Research & Cellular Therapeutics, Beckman Research Institute of the City of Hope, Duarte, CA, USA, [169]Department of Clinical Gene Therapy, Osaka University Graduate School of Medicine, Suita, Japan, [170]Department of Geriatric and General Medicine, Osaka University Graduate School of Medicine, Suita, Japan, [171]Department of Public Health, University of Split School of Medicine, Split, Croatia, [172]Institute of Biomedicine, Bioinformatics Center, Univeristy of Eastern Finland, Kuopio, Finland, [173]Department of Medicine, University of Eastern Finland and Kuopio University Hospital, Kuopio, Finland, [174]USC-Office of Population Studies Foundation, University of San Carlos, Cebu City, Philippines, [175]Department of Anthropology, Sociology and History, University of San Carlos, Cebu City, Philippines, [176]State Key Laboratory of Oncogene and Related Genes & Department of Epidemiology, Shanghai Cancer Institute, Renji Hospital, Shanghai Jiaotong University School of Medicine, Shanghai, China, [177]Lifelines cohort, Groningen, The Netherlands, [178]Internal Medicine, Endocrine & Metabolism, Taichung Veterans General Hospital, Taichung, Taiwan, [179]Center for Geriatrics and Gerontology,, Taichung Veterans General Hospital, Taichung, Taiwan, [180]National Defense Medical Center, National Yang-Ming University, Taipei, Taiwan, [181]Diabetes Prevention Unit, National Institute for Health and Welfare, Helsinki, Finland, [182]Center for Clinical Research and Prevention, Bispebjerg and Frederiksberg Hospital, Copenhagen, Denmark, [183]Department of Clinical Medicine, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark, [184]Yong Loo Lin School of Medicine, National University of Singapore and National University Health System, Singapore, Singapore, [185]Department of Medicine, University of Texas Health Sciences Center, San Antonio, TX, USA, [186]Department of Internal Medicine, Aichi Gakuin University School of Dentistry, Nagoya, Japan, [187]Department of Diabetes, Diabetes, & Nutritional Sciences, James Black Centre, King's College London, London, UK, [188]Department of Functional Pathology, Shimane University School of Medicine, Izumo, Japan, [189]Department of Medicine and Pharmacology, New York Medical College School of Medicine, Valhalla, NY, USA, [190]Colorado School of Public Health, University of Colorado Anschutz Medical Campus, Aurora, CO, USA, [191]Department of Geriatric Medicine and Neurology, Ehime University Graduate School of Medicine, Toon, Japan, [192]Institute of Epidemiology, Helmholtz Zentrum München Research Center for Environmental Health, Neuherberg, Bavaria, Germany, [193]Institute for Medical Information Processing, Biometry, and Epidemiology, Ludwig-

Maximilians University Munich, Munich, Bavaria, Germany, [194]Department of Epidemiology and Population Health, Albert Einstein College of Medicine, Bronx, NY, USA, [195]Genetics and Pharmacogenomics, Merck Sharp & Dohme Corp., Kenilworth, NJ, USA, [196]Department of Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, WA, USA, [197]Department of Internal Medicine, Erasmus Medical Center, Rotterdam, The Netherlands, [198]Usher Institute of Population Health Sciences and Informatics, University of Edinburgh, Edinburgh, UK, [199]Ophthalmology & Visual Sciences Academic Clinical Program (Eye ACP), Duke-NUS Medical School, Singapore, Singapore, [200]BHF Glasgow Cardiovascular Research Centre, Institute of Cardiovascular and Medical Sciences, University of Glasgow, Glasgow, UK, [201]Department of Experimental Diabetology, German Institute of Human Nutrition Potsdam-Rehbruecke, Nuthetal, Germany, [202]German Center for Diabetes Research (DZD), Neuherberg, Germany, [203]Department of Genetics, Shanghai-MOST Key Laboratory of Health and Disease Genomics, Chinese National Human Genome Center at Shanghai (CHGC) and Shanghai Academy of Science & Technology (SAST), Shanghai, China, [204]Center for Statistical Genetics, Ann Arbor, MI, USA, [205]Sarepta Therapeutics, Cambridge, Massachusetts, USA, [206]Department of Laboratory Medicine and Pathology, University of Minnesota, Minneapolis, MN, USA, [207]Department of Nutrition, Harvard T.H. Chan School of Public Health, Boston, MA, USA, [208]Institute of Clinical Medicine, Internal Medicine, University of Eastern Finland, Kuopio, Finland, [209]Department of Internal Medicine, Division of Endocrinology, Leiden University Medical Center, Leiden, The Netherlands, [210]Laboratory for Experimental Vascular Medicine, Leiden University Medical Center, Leiden, The Netherlands, [211]Department of Human Genetics, Leiden University Medical Center, Leiden, The Netherlands, [212]Department of Human Biology, University of Split School of Medicine, Split, Croatia, [213]Carolina Population Center, University of North Carolina, Chapel Hill, NC, USA, [214]Department of Endocrinology and Metabolism, Instituto Nacional de Ciencias Medicas y Nutricion, Mexico City, Mexico, [215]225Unidad de Investigacion de Enfermedades Metabolicas. Tec Salud, Mexico City, Mexico, [216]Instituto Tecnológico y de Estudios Superiores de Monterrey Tec Salud, Mexico City, Mexico, [217]Department of Medical Genomics, Pfizer/University of Granada/Andalusian Government Center for Genomics and Oncological Research (GENYO), Granada, Spain, [218]Institute for Environmental Medicine, Chronic Inflammatory Diseases, Karolinska Institutet, Solna, Sweden, [219]Department of Genetics, Division of Statistical Genomics, Washington University School of Medicine, St. Louis, MO, USA, [220]Clinical and Health Services Research, National Institute on Minority Health and Health Disparities, Bethesda, MD, USA, [221]Department of Medicine, General Internal Medicine, Johns Hopkins University School of Medicine, Baltimore, MD, USA, [222]School of Medicine and Pharmacology, Royal Perth Hospital Unit, University of Western Australia, Perth, WA, Australia, [223]Department of Integrative Biomedical Sciences, University of Cape Town, Cape Town, South Africa, [224]Aberdeen Centre for Health Data Science, 1:042 Polwarth Building,, School of Medicine, Medical, Science and Nutrition, University of Aberdeen, Foresterhill, Aberdeen, UK, [225]Department of Biostatistics, School of Public Health, University of Michigan, Ann Arbor, MI, USA, [226]Center for Statistical Genetics, University of Michigan, Ann Arbor, MI, USA, [227]Human Genetics Center, School of Public Health, The University of Texas Health Science Center at Houston, Houston, TX, USA, [228]Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX, USA, [229]Division of Endocrinology and Diabetes, Graduate School of Molecular Endocrinology and Diabetes, University of Ulm, Ulm, Baden-Württemberg, Germany, [230]LKC School of Medicine, Nanyang Technological University, Singapore and Imperial College London, UK, Singapore, Singapore, [231]University of Potsdam, Berlin-Potsdam, Germany, [232]Department of Medicine, Keck School of Medicine of USC, Los Angeles, CA, USA, [233]Department of Physiology and Neuroscience, Keck School of Medicine of USC, Los Angeles, CA, USA, [234]INSERM UMR 1283 / CNRS UMR 8199, European Institute for Diabetes (EGID), Université de Lille, Lille, France, [235]INSERM UMR 1283 / CNRS UMR 8199, European Institute for Diabetes (EGID), Institut Pasteur de Lille, Lille, France, [236]Imperial College Healthcare NHS Trust, Imperial College London, London, UK, [237]MRC-PHE Centre for Environment and Health, Imperial College London, London, UK, [238]Havard Medical School, Boston, MA, USA, [239]Department of Medicine, Jackson Heart Study, University of Mississippi Medical Center,

Jackson, MS, USA, [240]Department of Medicine, Faculty of Medicine, University of Kelaniya, Ragama, Sri Lanka, [241]Department of Nutrition and Dietetics, School of Health Science and Education, Harokopio University of Athens, Kallithea, Greece, [242]Department of Clinical Sciences, Lund University, Malmö, Sweden, [243]CNR Institute of Clinical Physiology, Pisa, Italy, [244]Intramural Research Program, National Institute of Aging, Baltimore, MD, USA, [245]Diabetes Unit and Center for Genomic Medicine, Massachusetts General Hospital, Boston, MA, USA, [246]Department of Medicine, Harvard Medical School, Boston, MA, USA, [247]Programs in Metabolism and Medical & Population Genetics, Broad Institute, Cambridge, MA, USA, [248]Department of Clinical Sciences, Lund University Diabetes Centre, Malmö, Sweeden, [249]Department of Genomics of Common Disease, Imperial College London, London, UK, [250]Department of Medicine, Cardiovascular medicine, Karolinska Institutet, Stockholm, Sweden, [251]Department of Medicine, Division of Endocrinology, Diabetes & Metabolism, Cedars-Sinai Medical Center, Los Angeles, CA, USA, [252]Diabetes Centre, Lund University, Sweden, [253]Finnish Instituter of Molecular Medicine, Helsinki University, Helsinki, Finland, [254]Faculty of Medicine, School of health sciences, University of Iceland, Reykjavik, Iceland, [255]MRC Human Genetics Unit, Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh, Scotland, [256]Department of Epidemiology, Cardiovascular Health Research Unit, University of Washington, Seattle, WA, USA, [257]Department of Medicine, Division of Cardiovascular Medicine, Stanford University School of Medicine, Stanford University, Stanford, CA, USA, [258]Division of Epidemiology and Community Health, University of Minnesota, Minneapolis, MN, USA, [259]Department of Ophthalmology, Medical Faculty Mannheim, Heidelberg University, Mannheim, Germany, [260]Beijing Institute of Ophthalmology, Beijing Ophthalmology and Visual Science Key Lab, Beijing Tongren Eye Center, Beijing Tongren Hospital, Capital Medical University, Beijing, China, [261]Netherlands Heart Institute, Utrecht, The Netherlands, [262]MRC/UVRI and LSHTM (Uganda Research Unit), Entebbe, Uganda, [263]Faculty of Medicine, Institute of Health Sciences, University of Oulu, Oulu, Finland, [264]Unit of General Practice, Oulu University Hospital, Oulu, Finland, [265]Department of Epidemiology and Public Health, UCL, London, UK, [266]Department of Public Health Solutions, Finnish Institute for Health and Welfare, Helsinki, Finland, [267]Department of Medicine, University of Helsinki and Helsinki University Central Hospital, Helsinki, Finland, [268]Minerva Foundation Institute for Medical Research, Helsinki, Finland, [269]National Heart and Lung Institute, Imperial College London, London, UK, [270]IFB Adiposity Diseases, University of Leipzig Medical Center, Leipzig, Germany, [271]Institute for Social and Economic Research, University of Essex, Colchester, UK, [272]University Institute of Primary Care and Public Health, Division of Biostatistics, University of Lausanne, Lausanne, Switzerland, [273]Institute of Biomedicine, School of Medicine, University of Eastern Finland, Finland, [274]Department of Clinical Physiology and Nuclear Medicine, Kuopio University Hospital, Kuopio, Finland, [275]Foundation for Research in Health Exercise and Nutrition, Kuopio Research Institute of Exercise Medicine, Kuopio, Finland, [276]Laboratory of Epidemiology and Population Sciences, National Institute on Aging, National Institutes of Health, Baltimore, MD, USA, [277]Institute of Environmental Medicine, Cardiovascular and Nutritional Epidemiology, Karolinska Institutet, Stockholm, Sweden, [278]Department of Medical Sciences, Uppsala, Sweden, [279]Big Data Institute, Nuffield Department of Medicine, University of Oxford, Oxford, UK, [280]Nuffield Department of Women's and Reproductive Health, University of Oxford, Oxford, UK, [281]Department of Medical Epidemiology and Biostatistics and the Swedish Twin Registry, Karolinska Institute, Stockholm, Sweeden, [282]Department of Public Health and Primary Care, Leiden University Medical Center, Leiden, The Netherlands, [283]Institute of Cardiovascular and Medical Sciences, University of Glasgow, Glasgow, UK, [284]Division of Population Health and Genomics, School of Medicine, University of Dundee, Ninewells Hospital and Medical School, Dundee, UK, [285]Centre for Cognitive Ageing and Cognitive Epidemiology, University of Edinburgh, Edinburgh, UK, [286]Department of Health Services, Cardiovascular Health Research Unit, University of Washington, Seattle, WA, USA, [287]Kaiser Permanente Washington Health Research Institute, Seattle, WA, USA, [288]Department of Pediatrics, Genetic and Genomic medicine, University of California, Irvine, Irvine, CA, USA, [289]Tampere, Finnish Diabetes Association, Tampere, Finland, [290]Pirkanmaa Hospital District, Tampere, Finland, [291]Department of Medicine, University of

Cambridge, Cambridge, UK, [292]South Karelia Central Hospital, Lappeenranta, Finland, [293]Department of Psychology, University of Miami, Miami, FL, USA, [294]Paul Langerhans Institute Dresden of the Helmholtz Center Munich, University Hospital and Faculty of Medicine, Dresden, Germany, [295]Division of Population Health and Genomics, Ninewells Hospital and Medical School, University of Dundee, Dundee, UK, [296]Division of Sleep and Circadian Disorders, Brigham and Women's Hospital, Boston, MA, USA, [297]Novo Nordisk Foundation Center for Basic Metabolic Research, Section of Metabolic Genetics,, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark, [298]Department of Public Health, Section of Epidemiology, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark, [299]Departent of Molecular and Cellular Therapeutics, Royal College of Surgeons in Ireland, Dublin, Ireland, [300]Department of Aging and Health, Guy's and St Thomas' Foundation Trust, London, UK, [301]Saw Swee Hock School of Public Health, National University of Singapore and National University Health System, Singapore, Singapore, [302]Duke-NUS Medical School, Singapore, Singapore, [303]Department of Public Health Solutions, National Institute for Health and Welfare, Helsinki, Finland, [304]Department of Public Health, University of Helsinki, Helsinki, Finland, [305]Saudi Diabetes Research Group, King Abdulaziz University, Jeddah, Saudi Arabia, [306]Department of Genomic Medicine and Environmental Toxicology, Instituto de Investigaciones Biomedicas, Universidad Nacional Autonoma de Mexico, Mexico City, Mexico, [307]Department of Public Health and Clinical Nutrition, University of Eastern Finland, Finland, [308]Department of Epidemiology, University of Oxford, Oxford, UK, [309]Department of Medicine, Internal Medicine, Lausanne University Hospital (CHUV), Lausanne, Switzerland, [310]Department of Public Health Sciences, Wake Forest School of Medicine, Winston-Salem, NC, USA, [311]Faculty of Medical Sciences, Newcastle University, Newcastle upon Tyne, UK, [312]Beijing Tongren Eye Center, Beijing Key Laboratory of Intraocular Tumor Diagnosis and Treatment, Beijing Ophthalmology & Visual Sciences Key Lab, Beijing Tongren Hospital, Capital Medical University, Beijing, China, China, [313]Department of Public Health, Faculty of Medicine, University of Kelaniya, Ragama, Sri Lanka, [314]Department of Research and Evaluation, Kaiser Permanente of Southern California, Pasadena, CA, USA, [315]Institute for Molecular Bioscience, The University of Queensland, Queensland, Australia, [316]Kurume University School of Medicine, Japan, [317]Institute of Translational Genomics, Helmholtz Zentrum München – German Research Center for Environmental Health, Neuherberg, Germany, [318]Wellcome Sanger Institute, Hinxton, UK, [319]TUM School of Medicin, Technical University of Munich and Klinikum Rechts der Isar, Munich, Germany, [320]Department of Medicine, Division of Epidemiology, Vanderbilt Epidemiology Center, Vanderbilt University Medical Center, Nashville, TN, USA, [321]Oxford NIHR Biomedical Research Centre, Oxford University Hospitals NHS Foundation Trust, Oxford, UK, [322]Department of Pediatrics, Division of Endocrinology, Stanford School of Medicine, Stanford, CA, USA, [323]Wellcome Centre for Human Genetics, Nuffield Department of Medicine, University of Oxford, Oxford, UK, [324]Department of Medicine, Division of General Internal Medicine, Massachusetts General Hospital, Boston, MA, USA, [325]Program in Medical and Population Genetics, Broad Institute, Cambridge, MA, USA, [326]Department of Medicine, General Internal Medicine, Massachusetts General Hospital, Boston, MA, USA, [327]Department of Medicine, Diabetes Unit and Endocrine Unit, Massachusetts General Hospital, Boston, MA, USA, [328]Harvard Medical School, Boston, MA, USA, [329]Department of Human Genetics, University of Michigan, Ann Arbor, MI, USA, [330]Department of Musculoskeletal and Dermatological Sciences, Faculty of Medicine, Biology and Health, University of Manchester, Manchester, UK, [331]Department of Biostatistics, University of Liverpool, Liverpool, UK

[#] Denote shared authorship contributions

***Corresponding author:** Inês Barroso, Exeter Centre of Excellence in Diabetes (EXCEED), Exeter Medical School, University of Exeter, Exeter, UK, +44 1392 408221, ines.barroso@exeter.ac.uk

Abstract

Glycaemic traits are used to diagnose and monitor type 2 diabetes, and cardiometabolic health. To date, most genetic studies of glycaemic traits have focused on individuals of European ancestry. Here, we aggregated genome-wide association studies in up to 281,416 individuals without diabetes (30% non-European ancestry) with fasting glucose, 2h-glucose post-challenge, glycated haemoglobin, and fasting insulin data. Trans-ancestry and single-ancestry meta-analyses identified 242 loci (99 novel; $P<5\times10^{-8}$), 80% with no significant evidence of between-ancestry heterogeneity. Analyses restricted to European ancestry individuals with equivalent sample size would have led to 24 fewer new loci. Compared to single-ancestry, equivalent sized trans-ancestry fine-mapping reduced the number of estimated variants in 99% credible sets by a median of 37.5%. Genomic feature, gene-expression and gene-set analyses revealed distinct biological signatures for each trait, highlighting different underlying biological pathways. Our results increase understanding of diabetes pathophysiology by use of trans-ancestry studies for improved power and resolution.

Fasting glucose (FG), 2h-glucose post-challenge (2hGlu), and glycated haemoglobin (HbA1c) are glycaemic traits used to diagnose diabetes [1]. In addition, HbA1c is the most commonly used biomarker to monitor glucose control in patients with diabetes. Fasting insulin (FI) reflects a combination of insulin secretion and insulin resistance, both components of type 2 diabetes (T2D), and insulin clearance [2]. Collectively, all four of these glycaemic traits can be useful to better understand T2D pathophysiology [3-5] and are useful measures of cardiometabolic health as they are associated with cardiometabolic outcomes even within the non-diabetic range, albeit modestly so [6].

To date, genome-wide association studies (GWAS) and analysis of next-generation targeted arrays (Metabochip and exome array) have identified >120 loci associated with glycaemic traits in individuals without diabetes [7-15]. However, despite considerable differences in the prevalence of T2D risk factors across ancestries [16-18], most glycaemic trait GWAS in individuals without diabetes have insufficient representation of individuals of non-European ancestry and limited resolution for fine-mapping of causal variants and effector transcript identification. Here, we present large-scale trans-ancestry discovery meta-analyses of GWAS for four glycaemic traits (FG, 2hGlu, FI, and HbA1c) in individuals without diabetes with genotype imputation to the 1000 Genomes Project reference panel phase 1 version 3 [19]. Our aims were to identify additional glycaemic trait-associated loci; investigate the portability of loci and genetic scores across ancestries; leverage differences in effect allele frequency (EAF), effect size, and linkage disequilibrium (LD) across diverse populations to conduct fine-mapping and aid causal variant/effector transcript identification; and compare and contrast the genetic architecture of these four glycaemic traits to further elucidate their underlying biology and gain insights into pathophysiological pathways implicated in T2D.

Results

Study design, lead variant, index variant and trans-ancestry locus definitions

To identify loci associated with glycaemic traits FG, 2hGlu, FI, and HbA1c, we aggregated GWAS in up to 281,416 individuals without diabetes, ~30% of whom were of non-European ancestry [13% East Asian, 7% Hispanic, 6% African-American, 3% South Asian, and 2% sub-Saharan African (Ugandan - data only available for HbA1c)]. Prior to meta-analysis each contributing cohort imputed data to the 1000 Genomes Project reference panel (phase 1 v3, March 2012, or later; **Methods, Supplementary Table 1, Supplementary Figure 1).** In total, up to ~49.3 million variants were directly genotyped or imputed, with between 38.6 million (2hGlu) and 43.5 million variants (HbA1c) available for analysis after exclusions based on minor allele count (MAC < 3) and imputation quality (imputation r^2 or INFO score <0.40) in each cohort. As we had previously found adjusting for body mass index (BMI) provided similar results for FG and 2hGlu, but aided in new locus discovery for FI [15], here we conducted analyses for FG, 2hGlu and FI adjusted for BMI, but for simplicity these traits are abbreviated as FG, 2hGlu and FI (**Methods).**

We first performed trait-specific fixed-effect meta-analyses *within* each ancestry using METAL [20]. We defined "single-ancestry lead" variants as the strongest trait-associated variants ($P<5\times10^{-8}$) within a 1Mb region in a particular ancestry (**Glossary box**). Within each ancestry and each autosome, we used approximate conditional analyses in GCTA [21,22], to identify distinct "single-ancestry index variants" ($P<5\times10^{-8}$) that exert conditionally distinct effects on the trait (**Glossary Box**, **Methods**, **Supplementary Figure 2).** Overall, this approach identified 124 distinct FG, 15 2hGlu, 48 FI and 139 HbA1c variants that were significant in at least one ancestry **(Supplementary Table 2).**

Next, we conducted trait-specific *trans-ancestry* meta-analyses of ancestry-specific results using MANTRA (**Methods, Supplementary Table 1, Supplementary Figures 1 and 3)** to identify genome-wide significant "trans-ancestry lead variants", defined as the most significant trait-associated variant across all ancestries ($\log_{10}$ Bayes Factor [BF] >6, equivalent to $P < 5\times10^{-8}$ [23]) (**Glossary box**,

Methods). Here, we present trans-ancestry results based on data from all participating cohorts as our primary results (**Supplementary Table 2**).

Causal variants are expected to affect multiple related glycaemic traits and may be shared across ancestries. Therefore, we combined all single-ancestry lead variants, single-ancestry index variants, and/or trans-ancestry lead variants (for any trait) mapping within 500Kb of each other, into a single "trans-ancestry locus" that was bounded by a 500Kb flanking sequence (**Glossary Box).** As defined, a trans-ancestry locus may contain multiple causal variants affecting one or more glycaemic traits, exerting their effect in one or more ancestry.

Glycaemic trait locus discovery

In the trans-ancestry meta-analyses, we observed genome-wide significant associations at 235 trans-ancestry loci, of which 59 contained trans-ancestry lead variants for more than one trait. In addition, we identified seven "single-ancestry loci" that did not contain any trans-ancestry lead variants (**Glossary box**, **Supplementary Table 2).** Of the 242 trans-ancestry and single-ancestry loci, 99 (including 6 of the 7 single-ancestry) had not been previously associated with any of the four glycaemic traits or with T2D, at the time of analysis (**Figure 1**, **Supplementary Figures 1 and 3**, **Supplementary Table 3, Supplementary note**). However, based on the currently available largest East Asian ancestry and trans-ancestry T2D GWAS meta-analyses[23-27], the lead variants at 27/99 novel glycaemic trait loci have strong evidence of association with T2D ($P<10^{-4}$; 13 loci with $P<5x10^{-8}$), suggesting some of the novel loci are also important in diabetes pathophysiology (**Supplementary Tables 2 and 4)**.

Of the 99 novel loci, six were identified in a single ancestry (**Supplementary Table 3**). Three single-ancestry loci were associated in individuals of non-European ancestry: (i) an African American association for FG (lead variant rs61909476) near the gene *ETS1*, (ii) an African American association for FI (lead variant rs12056334) near the gene *LOC100128993* (an uncharacterised RNA gene; **Supplementary Note**), and (iii) a Hispanic association for FG (lead variant rs12315677) within the gene *PIK3C2G* (**Supplementary Table 3**). The associations of rs61909476 and rs12315677 with FG are noteworthy. The variant rs61909476 has an EAF of ~7% in African American, and 10-17% in all other ancestries (**Supplementary Table 2**), but the effect on FG is only detectable in African American individuals (b=0.0812 mmol/l, SE=0.01 mmol/l, $P=3.9\times10^{-8}$, all other ancestries b=0-0.002 mmol/l, se=0.003-0.017 mmol/l, p=0.44-0.95, **Supplementary table 2, Supplementary Figure 4, Supplementary note).** The nearest gene, *ETS1*, encodes a transcription factor which has been shown to localize to insulin-positive cells in mouse islets, and its overexpression was shown to decrease glucose-stimulated insulin secretion in mouse islets [28]. Located within the *PIK3C2G* gene, rs12315677 has an 84% EAF in Hispanic and ranges from 70-94% in other ancestry populations, but is significantly associated with FG only in our Hispanic GWAS (b=0.0387 mmol/l, SE=0.0075 mmol/l, $P=4.0\times10^{-8}$) compared with other ancestries (b=-0.0128-0.010 mmol/l, SE=0.003-0.018 mmol/l, P=0.14-0.76) (**Supplementary Figure 5, Supplementary note**). *PIK3C2G* has been shown to be a Rab5 effector which, when deleted in *Pik3c2g*$^{-/-}$mice, selectively inhibits *Akt2* activation and leads to a phenotype characterised by reduced glycogen storage in the liver, hyperlipidaemia, adiposity, and insulin resistance with increasing age, or after a high fat diet [29]. Instances where the EAFs are similar between populations, but the effect sizes differ, could be due to specific genotype-by-environment or other genotype epistatic effects that differ across ancestries, or lower imputation accuracy in ancestries with smaller sample sizes, although this would likely lead to deflated effect sizes and imputation quality is good for these variants (average r^2=0.81). It is also possible that the variants detected here are not themselves causal, but are in LD with ancestry-specific causal variants that are not directly interrogated in our meta-analysis and that differ in frequency across ancestries. We looked at data from 1000G in the cognate populations, but could not find evidence of rarer alleles in those ancestries that may themselves be driving the association signals (**Supplementary Table 5).**

However, this does not preclude the possibility that other rarer variants exist which are not represented in the 1000G populations. The final three single-ancestry loci were identified in individuals of European ancestry, but without any evidence of association in the other ancestries despite similar MAF, although this may be due to differences in power given the much smaller sample sizes in non-European ancestries (**Supplementary Figures 6-8**).

Next, to investigate the contribution of non-European ancestry data to novel trans-ancestry locus discovery, independent of the total sample size, we artificially boosted the sample size of the European meta-analysis to match that of trans-ancestry meta-analysis by rescaling the standard errors of allelic effect sizes (**Supplementary note**). Using this approach, we determined that 21 of the novel trans-ancestry loci would not have been discovered with an equivalent sample size comprised exclusively of European ancestry individuals (**Supplementary note**). Their discovery was due to the higher EAF and/or larger effect size in non-European ancestry populations. In particular, two loci (nearest genes *LINC0088*5 and *MIR4278*) contain East Asian and African American single-ancestry lead variants, respectively, suggesting that these specific ancestries may be driving the trans-ancestry discovery (**Supplementary Tables 2-3**). Combined with the three single-ancestry non-European loci described above, our results show that 24% (24/99) of novel loci were discovered due to the contribution of non-European ancestry participants, strengthening the argument for extending genetic studies to larger samples sizes in diverse populations.

Allelic architecture of glycaemic traits

Trans-ancestry and single-ancestry loci comprised a range of association patterns, with most loci harbouring one single-ancestry signal for any given trait (**Supplementary note**). However, 29 loci contained multiple distinct index variants that did not fully overlap between ancestries. The most complex locus we observed was in the region spanning *G6PC2*, which contained 14 distinct FG index variants in the European single-ancestry meta-analysis. Of these, four are shared ($P<5x10^{-8}$) with South Asian ancestry, two with East Asian ancestry, and two with Hispanic ancestry (**Supplementary Figure 9)**. The complexity of association signals at this locus is consistent with previous work that also reported common variant (MAF>5%) association signals and multiple rare variant (MAF≤1%) associations at this locus that influenced protein function by multiple mechanisms [30].

Combined, single-ancestry lead, single-ancestry index, and trans-ancestry lead variants increase the number of established loci for FG to 102 (182 signals, 53 novel loci), FI to 66 (95 signals, 49 novel loci), 2hGlu to 21 (28 signals, 11 novel loci), and HbA1c to 127 (218 signals, 62 novel loci) (**Supplementary Table 2**) and demonstrate significant overlap across glycaemic traits **(Supplementary Figure 10**). We also detected ($P<0.05$ or $\log_{10}BF>0$) the vast majority (~90%) of previously established glycaemic trait association signals in our data, 70-88% of which attained genome-wide significance in the current analyses (see further details in the **Supplementary Note** and **Supplementary Table 6**). Given that analyses for FG, FI, and 2hGlu were performed adjusted for BMI, we also confirmed that collider bias was not influencing discovery for more than 98% of our results (**Supplementary note**)[31].

Finally, as expected, given the greater power due to increased sample sizes, new association signals tended to have smaller effect sizes and/or EAFs in European ancestry individuals (in whom this analysis was conducted) compared to previously established signals (**Supplementary Figure 11**).

Characterisation of trans-ancestry lead variants and European index variants across ancestries

We next employed a series of complementary analyses to better understand the transferability of trans-ancestry lead variants across all ancestries. For each trans-ancestry lead variant, we investigated the pairwise EAF correlation between ancestries, as well as the pairwise summarised heterogeneity of effect sizes between ancestries [32] (**Methods** and **Supplementary Note**). In

agreement with population history and evolution, these results demonstrated considerable EAF correlation (ρ^2>0.70) between European and Hispanic populations, European and South Asian populations, and Hispanic and South Asian populations, consistent across all four traits, and between African Americans and Ugandans for HbA1c (**Supplementary Figure 12**). Despite significant EAF correlations, some pairwise comparisons exhibited strong evidence for effect size heterogeneity between ancestries that was less consistent between traits (**Supplementary Figure 12).** However, sensitivity analyses demonstrated that, across all comparisons, the evidence for heterogeneity is driven by a small number of variants, with between 81.5% (for HbA1c) and 85.7% of trans-ancestry lead variants (for FG) showing no evidence for trans-ancestry heterogeneity (*P*>0.05) (**Supplementary Note**).

We also took LD pruned European single-ancestry index variants and compared the direction of effect of these variants in European ancestry individuals with that in other ancestries (**Supplementary Note**). Consistent with the lack of heterogeneity in effect sizes, we saw >70% concordance in the direction of effect for all traits into all ancestries, with the exception of HbA1c into African Americans and Ugandans (**Supplementary Table 7**). Imperfect concordance between ancestries could reflect lower power in non-European ancestry groups due to sample size or variation in allele frequency, or could be explained by LD differences between index SNPs and causal variants. For HbA1c, we hypothesized that lower concordance might also be a reflection of the different pathways (glycaemic and non-glycaemic) through which variants can affect HbA1c levels, particularly effects mediated via the red blood cell (RBC) where balancing selection can lead to different associations in individuals of African ancestry [7] (**Supplementary Note** and below).

To further investigate the potential utility of trans-ancestry analyses, and to evaluate whether larger sample sizes might yield additional European ancestry signals that would be transferable across ancestries, we extended these concordance analyses to the entire genome, clumping variants mapping >1Mb apart (to eradicate the effect of LD in all ancestries) in different bins of association p-values obtained from the European ancestry meta-analysis (**Methods**). Aside from the bins with the weakest evidence for association in Europeans (i.e. in all bins with *P*≤0.05), we observed nominally significant concordance in the direction of effects between European and other ancestries for all traits except for 2hGlu, in which analyses were underpowered (**Supplementary Table 7**).

Trait variance explained by associated loci

The trait variance explained by genome-wide significant loci was assessed using the single-ancestry lead and index variants only or a combination of single-ancestry and trans-ancestry variants (**Supplementary Table 8)** with betas extracted from the relevant single-ancestry meta-analysis results (**Methods).** The variance explained was assessed by linear regression in a subset of the contributing cohorts (**Methods**, **Supplementary Tables 9-12**). In general, the optimal approach (i.e. that which explained the most variance) was to begin with the trans-ancestry lead variants (based on the MANTRA results) that have *P* < 0.1 in the relevant single-ancestry meta-analysis, then add in all single-ancestry lead and index variants that are not in LD with the trans-ancestry variants (LD r^2 < 0.1) (List C) (**Supplementary Tables 9-12**, **Figure 2**). However, in the European ancestry cohorts there was little gain from using trans-ancestry loci. Using this list of trans-ancestry lead variants supplemented with single-ancestry signals, the mean variance in the trait distribution explained was between 0.7% (2hGlu in EUR) and 6% (HbA1c in AA). In European ancestry studies, these estimates represent an improvement (i.e. more variance explained) relative to previous estimates of 2.8% for FG and 1.7% for HbA1c [33] (see further discussion in **Supplementary Note**).

Transferability of European ancestry-derived polygenic scores across ancestries

To investigate the transferability of polygenic scores across ancestries we used the PRS-CSauto software[34] to first build polygenic scores for each glycaemic trait (FG, FI, 2hGlu and HbA1c) based on

European ancestry data. However, the training set for 2hGlu was too small so this trait was excluded (**Methods**). We have used the term polygenic scores (PGS) as strictly speaking for continuous traits they are not risk scores. To build the PGS, for each trait we first removed five of the largest European cohorts contributing to the respective European ancestry meta-analysis (**Methods).** These five cohorts were meta-analysed and used as our European ancestry test dataset, for each trait. The remaining European ancestry cohorts were also meta-analysed and used as the training dataset from which we derived a PGS for each trait (**Methods**). We used PRS-CSauto to revise the effect size estimates for the variants in the score (obtained from the training European datasets) based on the LD of the test population (**Methods**). Unfortunately, PRS-CSauto does not have LD reference panels for South Asian or Hispanic ancestry and as such we were unable to test the transferability of the PGS into those populations. The "gtx" package[35] (**Methods**) was used to obtain the R^2 for each test population (**Figure 3, Supplementary Table 13**). In line with observations from other complex traits[36], the European ancestry-derived PGS had greater predictive power into test data of European ancestry than other ancestry groups.

Fine-mapping

Of the 242 identified loci, 231 were autosomal trans-ancestry loci and six were autosomal single-ancestry loci, which we took forward for fine-mapping (**Supplementary Table 2).** Due to the absence of LD maps from adequately sized populations, fine-mapping was not attempted for the 5 loci (4 trans-ancestry and 1 single-ancestry) mapping to the X chromosome. Using FINEMAP with ancestry-specific LD and an average LD matrix across ancestries, we conducted fine-mapping both within single-ancestries (161 autosomal loci with single-ancestry lead variants999) and across ancestries (231 autosomal trans-ancestry loci) for each trait (**Methods**). Because 59 of the 231 trans-ancestry loci were associated with more than one trait, we conducted trans-ancestry fine-mapping for a total of 305 locus-trait associations. Of these 305 locus-trait combinations, FINEMAP estimated the presence of a single causal variant responsible for the association at 186 loci (61%), while multiple distinct causal variants were implicated at 126 loci (39%), for a total of 464 causal variants (**Figure 4A)**.

Credible sets for causal variants

At each locus, we next constructed credible sets (CS) for each causal variant that account for >=99% of the posterior probability of association (PPA). We identified 21 locus-trait associations (at 19 loci) for which the 99% CS included a single variant, and we highlight five examples below. (**Methods, Supplementary Note**, **Figure 4B, Supplementary Table 14).**

We highlight two positive controls which provide confidence in the results. At one locus near *MTNR1B*, rs10830963 (PPA>0.999, for both HbA1c and FG), located in an *MTNR1B* intron, has shown allelic differences in enhancer activity and transcription factor binding [37]. An additional FG-associated locus near *SIX3,* rs12712928 (PPA=0.997) has shown allelic differences in transcriptional activity, transcription factor binding, and association with islet expression levels of nearby genes *SIX3* and *SIX2* [38,39]. The EAF and effect size of this variant is larger in EAS than in other ancestries (heterogeneity p-value=7.2×10^{-8}), which is driving the association at this locus.

Next, we highlight three novel findings. At a locus near *PFKM* associated with HbA1c, trans-ancestry fine-mapping identified rs12819124 (PPA>0.999) as the likely causal variant. This variant has been previously associated with mean corpuscular haemoglobin [40], suggesting an effect of this locus on HbA1c is via the RBC. We note that this locus also harbours an association with FI in European and trans-ancestry meta-analyses, although it appears to be distinct from the HbA1c signal based on distance and LD. Fine-mapping of the nearby FI signal in European ancestry populations identified rs111264094 (PPA=0.994) as the likely causal variant (**Supplementary Figures 13-14**). rs111264094 is a low frequency variant in Europeans (EAF=0.025) that is monomorphic or rare in other ancestries, is

located >600 kb from HbA1c-associated variant rs12819124, and is in low LD with rs12819124 in European ancestry populations (r^2<0.1), which supports the hypothesis of two distinct signals (one for FI and one HbA1c) at this locus.

At the *HBB* locus, we identify rs334 (PPA>0.999; Glu7Val) as the likely causal variant associated with HbA1c. rs334 is a causal variant of sickle cell anaemia [41], with previously reported associations with urinary albumin-to-creatinine ratio in Caribbean Hispanic individuals [42], severe malaria in a Tanzanian study population [43], haematocrit and mean corpuscular volume in Hispanic/Latino populations [44], and more recently with RBC distribution in Ugandan individuals [45], all of which point to an effect of this variant on HbA1c via non-glycaemic pathways.

Lastly, our credible set analysis identified rs1799815 (PPA=0.993) as the likely causal variant at the *INSR* locus associated with FI. rs1799815 is a synonymous variant (Tyr3033Tyr) within *INSR*, the well-known insulin receptor gene that regulates the insulin signalling pathway. *INSR* as a target gene for this locus is further supported by our finding that rs1799815 colocalizes as an eQTL for *INSR* expression in adipose tissue (details shown below). The remaining locus-trait associations with a single variant in the 99% CS (**Supplementary Table 14)** point to variants that could be prioritised for downstream functional follow-up to further elucidate their impact on glycaemic trait physiology.

In addition to identifying 99% CS with a single variant, trans-ancestry fine-mapping identified 99% CS with 50 or fewer variants at 156 locus-trait associations (**Figure 4B, Supplementary Table 14)**. Overall, 74 locus-trait associations contained 87 variants with PPA>0.90; that is, some locus-trait associations contain more than one variant with a high predicted probability of being causal as there can be more than one causal variant in a locus (**Supplementary Table 15)**. In addition to those already described above, the identified variants are strong candidate causal variants that merit prioritisation for future functional validation. For example, among the 87 variants, 10 are coding variants including several missense such as the *HBB* Glu7Val mentioned above, *GCKR* Leu446Pro, *RREB1* Asp1771Asn, *G6PC2* Pro324Ser, *GLP1R* Ala316Thr, and *TMPRSS6* Val736Ala**,** each of which have been proposed or shown to affect gene function [12,46-50]. We also additionally identify *AMPD3* Val311Leu (PPA=0.989) and *TMC6* Trp125Arg (PPA>0.999) variants associated with HbA1c which were previously detected in an exome array analysis but had not been fine-mapped with certainty due to the absence of backbone GWAS data [30]. Our current fine-mapping data now suggest these variants are likely to be causal and identify the cognate genes as the effector transcripts driving these associations.

Finally, we evaluated the resolution obtained in the trans-ancestry versus single-ancestry fine-mapping (**Methods, Supplementary Note**). To do this, we compared the number of variants in 99% CS across 98 locus-trait associations which, as suggested by FINEMAP, had a single causal variant in both trans-ancestry and single-ancestry analyses. Fine-mapping within and across ancestries was conducted using the same set of variants. At 8 of 98 locus-trait associations single-ancestry fine-mapping identified a single variant in the CS. In addition, at 72 of the 98 locus-trait associations, the number of variants in the 99% CS was smaller in trans-ancestry fine-mapping than in single-ancestry analyses (**Figure 4C**), which likely reflects the larger sample size and differences in LD structure, EAFs, and effect sizes across diverse populations. To quantify the estimated improvement in fine-mapping resolution attributable to the multi-ancestry GWAS, we then compared 99% CS sizes from the trans-ancestry fine-mapping to single-ancestry-specific data emulating the same total sample size by rescaling the standard errors (**Methods**). Of the 72 locus-trait associations with estimated improved fine-mapping in trans-ancestry analysis, resolution at 38 (53%) was improved because of the larger sample size in the trans-ancestry fine-mapping analysis (**Figure 4C)**, and this estimated improved resolution would likely have been obtained in a European-only fine-mapping effort with equivalent sample size. However, at 34 (47%) loci, the inclusion of samples from multiple diverse

populations yielded estimated improved resolution. On average, ancestry differences led to a reduction in the median number of variants in the 99% CS from 24 to 15 variants (37.5% median reduction; **Figure 4C),** demonstrating the value of conducting fine-mapping across ancestries.

HbA1c Signal Classification

We, and others, have previously suggested that HbA1c-associated variants appear to exert their effects on HbA1c levels through both glycaemic and non-glycaemic pathways [7,51]. Classification of loci into these pathways can have important implications for T2D diagnostic accuracy [7,52]. To further elucidate the biology of HbA1c-associated variants, we took advantage of prior association results for other glycaemic, RBC, and iron traits, and used a fuzzy clustering approach to classify variants into their most likely mode of action (**Methods**, **Supplementary note**). Of the 218 HbA1c-associated trans-ancestry lead variants and single-ancestry index variants, 27 (12%) could not be characterized due to missing summary statistics in the other datasets and 23 (11%) could not be classified into a "known" class (**Supplementary note**). The remaining signals were classified as principally: a) glycaemic (n=53; 24%), b) affecting iron levels/metabolism (n=12; 6%), or c) RBC traits (n=103; 47%). We found a genetic risk score (GRS) composed of all HbA1c-associated signals was strongly associated with T2D risk (OR=2.4, 95% CI 2.3-2.5, P=2.4x10^{-298}). However, when we tested partitioned GRSs composed of these different classes of variants (**Methods**), we found the T2D association was mainly driven by those variants influencing HbA1c through glycaemic pathways (OR=2.6, 95% CI 2.5-2.8, P=1.1x10^{-250}), with weaker evidence of association (despite the larger number of variants in the GRS) and a more modest risk (OR=1.4, 95% CI 1.2-1.7, P=4.7x10^{-4}) imparted by signals in the mature RBC cluster that were not glycaemic (i.e. where those specific variants had P>0.05 for FI, 2hGlu and FG) (**Supplementary Figure 15, Supplementary note**). This contrasts our previous finding where we found no significant association between a risk score of non-glycaemic variants and T2D [7]. Our current results could be partly driven by T2D cases being diagnosed based on HbA1c levels that may be influenced by the non-glycaemic signals, or by glycaemic effects not captured by FI, 2hGlu or FG measures.

Biological signatures of glycaemic trait associated loci

To better understand distinct and shared biological signatures underlying variant-trait associations, we conducted genomic feature enrichment, eQTL co-localisation, and tissue and gene-set enrichment analyses across all four traits.

Epigenomic landscape of trait-associated variants

We next explored the genomic context underlying glycaemic trait loci by computing overlap enrichment for static annotations such as coding, conserved regions, histone modification ChIP-seq peaks, and super enhancers, merged across various cell types [53-55] using the GREGOR tool [56]. We observed that FG, FI and HbA1c signals (**Supplementary Table 8**) were significantly (P<8.4x10^{-4}, Bonferroni threshold correcting for 59 total annotations) enriched in evolutionarily conserved regions, whereas 2hGlu signals were only nominally enriched (**Fig 5A, Supplementary Figure 16, Supplementary Table 16**).

We then focussed on the epigenomic landscapes defined in individual cell/tissue types. Previously, stretch enhancers (enhancer chromatin states ≥3kb in length) in pancreatic islets were shown to be highly cell-specific and strongly enriched with T2D risk signals [57]. We therefore calculated the enrichment of glycaemic trait-associated signals (**Supplementary Table 8**) in previously defined stretch enhancers [39] across a diverse panel of cell types and tissues most relevant to the traits of interest: pancreatic islets, skeletal muscle, adipose, and liver (**Methods**). These analyses strongly

suggest that variants associated with these glycaemic traits influence the function of tissue specific enhancers. Namely, FG- and 2hGlu-associated signals have the highest enrichment in islet stretch enhancers (FG: fold enrichment=4.70, P=2.7x10^{-24}; 2hGlu: fold enrichment=5.51, P=3.6x10^{-4} **Figure 5A, Supplementary Table 17**), which highlights the relevance of pancreatic islet tissue for the regulation of FG and 2hGlu. Interestingly, FI-associated variants are strongly enriched for overlap with stretch enhancers in skeletal muscle (fold enrichment=3.17, P=7.8x10^{-6}) and adipose tissue (fold enrichment=3.27, P=1.8x10^{-7}), which is consistent with these tissues being key targets of insulin action and their involvement in the insulin resistance phenotype (**Figure 5A**). We note that the high enrichment of stretch enhancers in individual cell types (see upper "stretch enhancer" labelled portion of **Figure 5A**) as compared to super enhancers merged across cell types (see lower "static annotations" labelled portion of **Figure 5A**) highlights the importance of using cell-specific annotations in enrichment analyses. HbA1c-associated signals are enriched in stretch enhancers of multiple cell types and tissues likely because of the complex nature of this trait, but have the strongest enrichment in stretch enhancers from the blood-derived leukaemia cell line K562 (fold enrichment=3.24, P=1.21x10^{-7}, **Figure 5A**). We next sought to identify potential cell specific epigenomic enrichments that are associated with the classified HbA1c-associated variants corresponding to the "hard" glycaemic and red blood cell clusters, the latter being the joint group of mature red blood cell and reticulocyte clusters. We found that these partitioned variants display expected cell type-specific enrichment trends with the HbA1c glycaemic variants significantly enriched in islet stretch enhancers (fold enrichment=3.96, P=3.69x10^{-16} **Figure 5B, Supplementary Table 18**) and not in K562. Conversely, the HbA1c red blood cell variants are significantly enriched in K562 stretch enhancers (fold enrichment=7.5, P=2.08x10^{-14}, **Figure 5B, Supplementary Table 18**) and not in islets.

To complement the overlap enrichment results from GREGOR, we computed enrichment with two additional approaches: fGWAS [58] and GARFIELD [59]. These independent analyses yielded consistent results (**Supplementary Figures 17-18, Supplementary Tables 16 and 19),** demonstrating reproducibility across different approaches. Notably, we also observed enrichment of FI-associated variants in liver stretch enhancers (odds ratio=1.92, P=1.7x10^{-4}) when considering a more lenient SNP significance threshold of P<10^{-5} with the GARFIELD approach (**Supplementary figure 18A**). This suggests that liver regulatory annotations are relevant for FI GWAS signals, but that we lack power to detect significant enrichment using the genome-wide significant loci and the current set of reference liver annotations.

Given the observed enrichment of FI loci with stretch enhancers from adipose and skeletal muscle tissue, we sought to explore these loci in more detail. We found that 11 of the 27 loci driving these enrichment signals include variants that overlap stretch enhancers in both adipose and skeletal muscle (**Figure 5C**). At the *COL4A2* locus, variants within an intronic region of the gene overlap stretch enhancer chromatin states in adipose tissue, skeletal muscle, and a human skeletal muscle myoblast (HSMM) cell line that are not shared across other cell types and tissues; among these variants, rs9555695 (in the 99% CS) also overlaps accessible chromatin regions in adipose (**Figure 5D**). At a narrow signal (no proxy variants with LD r^2>0.7 in Europeans, for the lead trans-ancestry rs62271373 variant), rs62271373 (PPA = 0.94) located in an intergenic region ~25kb from the *LINC01214* gene overlaps stretch enhancer chromatin states in adipose and HSMM and active enhancer chromatin states in skeletal muscle, but does not overlap any enhancer states in other tissues (**Figure 5E**). The lead rs62271373 variant also overlaps an ATAC-seq peak in adipose tissue. Collectively, the tissue-specific stretch enhancer epigenomic signatures at GWAS signals provide an opportunity to nominate tissues where these variants are likely to be active. Such a map will be helpful in future efforts to deconvolute GWAS signals into tissue-specific disease pathology.

Co-localisation of GWAS and eQTLs

Among the 99 novel glycaemic trait loci identified by this study, we identified co-localised eQTLs at 34 loci in blood, pancreatic islets, subcutaneous or visceral adipose, skeletal muscle, or liver, providing suggestive evidence of causal genes (**Supplementary Table 20**). The co-localised eQTLs include several genes previously reported at glycaemic trait loci: *ADCY5*, *CAMK1D, IRS1*, *JAZF1*, and *KLF14* [60-62]. For some additional loci, the co-localised genes have prior evidence for a role in glycaemic regulation. For example, the lead trans-ancestry variant and likely causal variant, rs1799815 (PPA=0.993, mentioned above), associated with FI is the strongest variant associated with expression of *INSR*, encoding the insulin receptor, in subcutaneous adipose from METSIM (P=2x10^{-9}) and GTEx (P=5x10^{-6}). The A allele at rs1799815 is associated with higher FI and lower expression of *INSR*, which is consistent with the well-established relationship in humans and model organisms between insulin resistance and reduced function of INSR protein [63]. In a second example, rs841572, the trans-ancestry lead variant associated with FG, is the variant with the highest PPA (PPA=0.535) among the 20 variants in the 99% CS and is in strong LD (r^2=0.87) with the lead eQTL variant (rs841576, also in the 99% CS) associated with expression of *SLC2A1* in blood from eQTLGen (P=1x10^{-8}). *SLC2A1*, also known as *GLUT1*, encodes the major glucose transporter in brain, placenta, and erythrocytes, and is responsible for glucose entry into the brain [64]. The A allele at rs841572 is associated with lower FG and lower *SLC2A1* expression. While rare missense variants in *SLC2A1* are an established cause of seizures and epilepsy [65], our data suggest that *SLC2A1* variants also affect plasma glucose levels within a healthy physiological range. These novel associations and co-localised eQTLs provide possible regulatory mechanisms for variant effects on genes to influence glycaemic traits.

The co-localised eQTLs also provide new insights into the mechanisms at glycaemic trait loci. For example, rs9884482 (a variant in the 99% CS) is associated with FI and expression of *TET2* in subcutaneous adipose (P=2x10^{-20}); rs9884482 is in high LD (r^2=0.96 in Europeans) with the lead *TET2* eQTL variant (rs974801). *TET2* encodes a DNA-demethylase through which *TET2* can affect transcriptional repression [66]. Adipose Tet2 expression is reduced in diet-induced insulin resistance in mice [67], and knockdown of Tet2 blocked adipogenesis by repressing *Pparg* expression [67,68]. Consistently, in human adipose tissue, rs9884482-C was associated with lower expression of *TET2* and higher FI. In a second example, HbA1c-associated variant rs617948 (a variant in the 99% CS) is the lead variant associated with expression of *C2CD2L* in blood from eQTLGen (P=3x10^{-96}). *C2CD2L*, also known as *TMEM24,* has been shown to regulate pulsatile insulin secretion and facilitate release of insulin pool reserves [69,70]. The G allele at rs617948 was associated with higher HbA1c and lower *C2CD2L*, providing evidence for a role of this insulin secretion protein in glucose homeostasis. Our HbA1c "soft" clustering classification assigns this signal to both the "unknown" (0.51 probability) and "reticulocyte" (0.42 probability) clusters, and this variant has no evidence for association with FG, FI or 2hGlu (P>0.05), but is strongly associated with HbA1c (P<6.8x10^{-8}), reticulocytes (RET; P<5x10^{-7}) and HbA1c adjusted for FG (P<6.12x10^{-7}; **Supplementary Table 21, Supplementary Note**). Together, these results would suggest a possible effect of this variant on reticulocyte biology, and an effect on insulin secretion (mediated through *C2CD2L*) which is not captured by any of our traits, both of which potentially influencing HbA1c levels through different tissues, and providing a plausible explanation for the classification as "unknown".

Tissue Expression

Consistent with results based on effector transcripts and expression analysis based on GTEx data [30], we found significant differences in tissue expression across the glycaemic trait-associated variants. FG-associated variants were enriched for genes expressed in the pancreas (at FDR<0.05), while there was insufficient power (insufficient number of genome-wide significant associations) in 2hGlu analysis to identify enrichment for any tissues or cell types at a more relaxed FDR<0.2 threshold. FI-

associated variants were enriched for connective tissue and cells (which includes adipose tissue), endocrine glands, blood cells, and muscles (at FDR<0.2) and HbA1c-associated variants were significantly enriched for genes expressed in the pancreas, hemic, and immune system (at FDR<0.05) (**Figure 6, Supplementary Table 22**). Consistent with our previous analysis [30], FI-enrichment for connective tissue was driven by adipose tissue (subcutaneous and visceral), while the newly described enrichment with endocrine glands was driven by the adrenal glands and cortex (**Supplementary Table 22**). Beyond enrichment for genes expressed in glycaemic-related tissues, the association of HbA1c-associated variants with genes expressed in blood is consistent with the role of RBC in this glycaemic measure and our previous results [30].

The association between FI-associated variants (a surrogate for insulin resistance) and genes expressed in adrenal glands is notable, suggesting a possible direct role for these genes in insulin resistance. One hypothesis is that these genes might influence cortisol levels, which could subsequently contribute to insulin resistance and FI levels through impairment of the insulin receptor signalling pathway in peripheral tissues, as well as influencing body fat distribution, stimulate lipolysis, and other indirect mechanisms [71,72].

Gene-set Analyses

Next, we performed gene-set analysis using DEPICT (**Methods)**. In keeping with previous results [30], we found distinct gene-sets enriched (FDR<0.05) for each glycaemic trait (except 2hGlu, for which genome-wide associations were insufficient to have power in this analysis). FG-associated variants highlighted gene-sets involved in metabolism in addition to gene-sets involved in more general cellular function such as “cytoplasmic vesicle membrane” and “circadian clock”” (**Figure 7A).** In contrast, in addition to metabolism related gene-sets FI-associated variants highlighted pathways related to growth, cancer and reproduction (**Figure 7B**). This is consistent with the role of insulin as a mitogenic hormone, and with epidemiological links between insulin and certain types of cancer [73] and reproductive disorders such as polycystic ovary syndrome [74]. HbA1c-associated variants highlighted a wide network of gene-sets (**Figure 7C**), including those linked to metabolism, as well as those linked to haematopoiesis, again recapitulating our postulated effects of variants on glucose and RBC biology. Additional pathways highlighted from HbA1c-associated variants also highlighted previous “CREBP PPi” and lipid biology related to T2D [75] and HbA1c [76], respectively, and potential new biology through which variants may influence HbA1c.

Discussion

Here we describe a large meta-analysis of GWAS of glycaemic traits for which 30% of the population was composed of East Asian, Hispanic, African-American, South Asian and sub-Saharan African participants, in addition to the European ancestry participants. Overall, this effort identified 242 loci (235 trans-ancestry and seven single-ancestry), which jointly explain between 0.7% (2hGlu in European ancestry individuals, SE=0.85% for 2hGlu) and 6% (HbA1c in African American ancestry, SE=1.2% for HbA1c) of the variance in glycaemic traits in any given ancestry. Of these 242 glycaemic trait loci, 114 have strong evidence of association with T2D ($P<10^{-4}$; 83 loci with $P<5x10^{-8}$, **Supplementary table 4**). Absence of strong evidence of association at the remaining loci (i.e. $P\geq10^{-4}$) suggests that for alleles more frequent than 5% we can exclude T2D ORs≥1.07 with 80% power (alpha=$5x10^{-8}$; and ORs≥1.05 for alpha=10^{-4}) given the current largest study which includes 228,499 T2D cases and 1,178,783 controls.[27] In total, we identified 486 signals associated with glycaemic traits (including all trans-ancestry and single-ancestry lead and index variants, **Supplementary table 2**). Of these 486 signals, eight have MAF<1%, and 45 have 1%<=MAF<5% in all ancestries, highlighting that 89% of signals identified are common in at least one of the ancestries studied.

A key aim of our study was to evaluate the added advantage of including population diversity into genetic discovery and fine-mapping efforts. Beyond the overall larger sample size included in the trans-ancestry meta-analysis, we were able to estimate the contribution of non-European ancestry data in locus discovery and fine-mapping resolution. We found that 24 of the 99 newly discovered loci owe their discovery to the inclusion of East Asian, Hispanic, African-American, South Asian and sub-Saharan African participant data, due to differences in EAF and effect sizes across ancestries.

Comparison of 295 trans-ancestry lead variants (315 locus-trait associations) across ancestries demonstrated that between 81.5% (for HbA1c) and 85.7% (for FG) of the trans-ancestry lead variants had no evidence of trans-ancestry heterogeneity in allelic effects (P>0.05). Expanded analyses including variants across the whole genome, demonstrated at least nominal concordance in the direction of effects between populations of European ancestry and other ancestries for all but the least significant association signals observed in European ancestry GWAS. These observations are consistent with a tail of variants with modest but homogenous effects on glycaemic traits across ancestries that would be amenable to discovery with even larger sample sizes in trans-ancestry meta-analysis.

Given sample size and power limitations, genome-wide significant trait associated variants in a single-ancestry (single-ancestry lead and index variants) explain only a modest proportion of trait variance in that ancestry (**Figure 2**). We demonstrate that trans-ancestry meta-GWAS identified loci (TA lead variants) provide additional information regarding trait variance explained above and beyond that contributed by the ancestry-specific meta-analysis results (**Figure 2**). This shows that even though not all TA lead variants are genome-wide significant in all ancestries they contribute to the genetic architecture of the trait in most ancestries.

We evaluated for the first time the transferability of European ancestry-derived glycaemic trait PGS into other ancestries. In agreement with results for other traits[36,77,78], we confirm that European ancestry-derived PGS perform much worse when the test dataset is from a different ancestry. We note that each trait-specific PGS improves trait variance explained by between 3.5-fold (HbA1c) and 6-fold (FG) in the European dataset (**Figure 3, Supplementary Table 12)** compared to using a score built only from TA lead variants and European index variants (**Figure 2, Supplementary tables 9-12**).

Despite development of novel approaches and software to derive polygenic risk scores[79], we note the difficulty in using summary level data to build a PGS in one ancestry and then apply it in test datasets of different ancestry. While PRS-CSauto[34] is able to use summary level data we noted that revision of the effect size estimates to account for LD required the use of reference panels that matched the ancestry of the test dataset. However, as the current version of the software lacks appropriate reference panels for many ancestries this precludes its broad application.

We further demonstrate that fine-mapping resolution is improved in trans-ancestry, compared to single-ancestry fine-mapping efforts. In ~50% of our loci, we were able to demonstrate the improvement is due to differences in EAF, effect size, or LD structure between ancestries, and not just due to the overall increased sample size available for trans-ancestry fine-mapping. By performing trans-ancestry fine-mapping, and co-localising GWAS signals with eQTL signals and coding variants, we identify new candidate causal genes. Altogether, these results provide additional strong motivation for continued expansion of genetic and genomic efforts in diverse populations, not least to improve understanding of these traits in diverse ancestries in whom individuals are often disproportionally affected by T2D.

Given data on four different glycaemic traits, and their utility to diagnose and monitor T2D and metabolic health, we also sought to characterise biological features underlying these traits. We

show that despite significant sharing of genetic loci across the four glycaemic traits, each trait is also characterised by a unique set of features based on stretch enhancer, gene expression and gene-set signatures. Combining genetic data from these traits with T2D data will further elucidate pathways driving normal physiology and pathophysiology, and help further develop useful predictive scores for disease classification and management [4,5].

Online Methods

Study design and participants

This study included trait data from four glycaemic traits: fasting glucose (FG), fasting insulin (FI), 2hr post-challenge glucose (2hGlu), and glycated haemoglobin (HbA1c). The total number of contributing cohorts ranged from 41 (2hGlu) to 131 (FG), and the maximum sample size for each trait ranged from 85,916 (2hGlu) to 281,416 (FG) (**Supplementary Table 1)**. Overall, European ancestry (EUR) participants dominated the sample size for all traits, representing between 68.0% (HbA1c) to 73.8% (2hGlu) of the overall sample size. African Americans (AA) represented between 1.7% (2hGlu) to 5.9% (FG) of participants; individuals of Hispanic ancestry (HISP) represented between 6.8% (FG) to 14.6% (2hGlu) of participants; individuals of East-Asian ancestry (EAS) represented between 9.9% (2hGlu) to 15.4% (HbA1c) of participants; and South-Asian ancestry (SAS) individuals represented between 0% (no contribution to 2hGlu) to 4.4% (HbA1c) of participants. Data from Ugandan participants were only available for the HbA1c analysis and represented 2% of participants.

Phenotypes

Analyses included data for FG and 2hGlu measured in mmol/l, FI measured in pmol/l, and HbA1c in % [where possible, studies reported HbA1c as a National Glycohemoglobin Standardization Program (NGSP) percent]. Similar to previous MAGIC efforts [7], individuals were excluded if they had type 1 or type 2 diabetes (defined by physician diagnosis); reported use of diabetes-relevant medication(s); or had a FG ≥7 mmol/L, 2hGlu ≥11.1mmol/L, or HbA1c ≥ 6.5%, as detailed in **Supplementary Table 1**. 2hGlu measures were obtained 120 minutes after a glucose challenge in an oral glucose tolerance test (OGTT). Measures for FG and FI taken from whole blood were corrected to plasma level using the correction factor 1.13 [80].

Genotyping, quality control, and imputation

Each participating cohort performed study-level quality control, imputation, and association analyses following a shared analysis plan. Cohorts were genotyped using commercially available genome-wide arrays or the Illumina CardioMetabochip (Metabochip) array (**Supplementary Table 1)** [81]. Prior to imputation, each cohort performed stringent sample and variant quality control (QC) to ensure only high-quality variants were kept in the genotype scaffold for imputation. Sample quality control checks included removing samples with low call rate < 95%, extreme heterozygosity, sex mismatch with X chromosome variants, duplicates, first- or second-degree relatives (unless by design), or ancestry outliers. Following sample QC, cohorts applied variant QC thresholds for call rate (< 95%), Hardy-Weinberg Equilibrium (HWE) $P < 1x10^{-6}$, and minor allele frequency (MAF). Full details of QC thresholds and exclusions by participating cohort are available in **Supplementary Table 1.**

Imputation was performed up to the 1000 Genomes Project phase 1 (v3) cosmopolitan reference panel [82], with a small number of cohorts imputing up to the 1000 Genomes phase 3 panel [19] or population-specific reference panels (**Supplementary Table 1).**

Study level association analyses

Each of the glycaemic traits (FG, natural log FI, and 2hGlu) were regressed on BMI (except HbA1c), study-specific covariates, and principal components (unless implementing a linear mixed model).

Analyses for FG, FI, and 2hGlu were adjusted for BMI as we had previously shown this did not materially affect results for FG and 2hGlu but improved our ability to detect FI-associated loci [15]. For simplicity, we refer to the traits as FG, FI and 2hGlu. For a discussion on collider bias see **Supplementary Note section 2c**. Both the raw and rank-based inverse normal transformed residuals from the regression were tested for association with genetic variants using SNPTEST [23] or Mach2Qtl [83,84]. Poorly imputed variants, defined as imputation $r^2 < 0.4$ or INFO score < 0.4, were excluded from downstream analyses (**Supplementary Table 1)**. Following study level QC, approximately 12,229,036 variants (GWAS cohorts) and 1,999,204 variants (Metabochip cohorts) were available for analysis (**Supplementary Table 1)**.

Centralised quality control

Each contributing cohort shared their summary statistic results with the central analysis group who performed additional QC using EasyQC [85]. Allele frequency estimates were compared to estimates from 1000Gp1 reference panel [82], and variants were excluded from downstream analyses if there was a minor allele frequency difference > 0.2 for AA, EUR, HISP, and EAS populations against AFR, EUR, MXL, and ASN populations from 1000 Genomes Phase 1, respectively, or a minor allele frequency difference > 0.4 for SAS against EUR populations. At this stage, additional variants were excluded from each cohort file if they met one of the following criteria: were tri-allelic; had a minor allele count (MAC) < 3; demonstrated a standard error of the effect size ≥ 10; or were missing an effect estimate, standard error, or imputation quality. All data that survived QC (approximately 12,186,053 variants from GWAS cohorts and 1,998,657 variants from Metabochip cohorts) were available for downstream meta-analyses.

Single-ancestry meta-analyses

Single-ancestry meta-analyses were performed within each ancestry group using the fixed-effects inverse variance meta-analysis implemented in METAL [20]. We applied a double-genomic control (GC) correction [15,86] to both the study-specific GWAS results and the single-ancestry meta-analysis results. Study-specific Metabochip results were GC-corrected using 4,973 SNPs included on the Metabochip array for replication of associations with QT-interval, a phenotype not correlated with our glycaemic traits [15].

Identification of single-ancestry index variants

To identify distinct association index variants across each chromosome within each ancestry (**Glossary box**), we performed approximate conditional analyses implemented in GCTA [21] using the --cojo-slct option (autosomes) and distance-based clumping (X chromosome). Linkage disequilibrium (LD) correlations for GCTA were estimated from a representative cohort from each ancestry: WGHS (EUR); CHNS (EAS); SINDI (SAS); BioMe (AA); SOL (HISP) and Uganda (for itself). The results from GCTA were comparable when using alternative cohorts for the LD reference. For any index variant with a QC flag which caused reason for concern, we performed manual inspection of forest plots to decide whether the signal was likely to be real (**Supplementary note**). Among 335 single-ancestry index variants across all traits, this manual inspection was done for 40 signals of which 32 passed and 8 failed after inspection. Thus, a total of 327 single-ancestry index variants passed and 8 failed.

Trans-ancestry meta-analyses

To leverage power across all ancestries, we also conducted trait-specific trans-ancestry meta-analysis by combining the single-ancestry meta-analysis results using MANTRA (**Supplementary Figure 3**) [87].We defined $\log_{10}$Bayes' Factor (BF) > 6 as genome-wide significant, approximately comparable to $P < 5 \times 10^{-8}$.

Manual curation of trans-ancestry lead variants

To ensure trans-ancestry lead variants were robust, we performed manual inspection of forest plots by at least two authors, for any variants with flags indicating possible QC issues (**Supplementary Note**). Of 463 trans-ancestry lead variants across all traits, 184 passed without inspection, 131 passed after inspection, and 148 failed after inspection.

Correlation in EAF and heterogeneity in effect sizes of TA lead variants across ancestries

For each pair of ancestries, we calculated Pearson's correlation in EAFs for each trans-ancestry lead variant. The pairwise summarised heterogeneity of effect sizes between ancestries was then tested using the joint F-test of heterogeneity [32]. The test statistic is the sum of Cochran Q-statistics for heterogeneity across all trans-ancestry signals. Under the null hypothesis, the statistics follows the χ^2 distribution with n degrees of freedom, where n is the number of the trans-ancestry lead variants.

Concordance analyses of LD pruned European single-ancestry index variants into other ancestries

We compared the direction of effect of variants on each trait separately. For each trait, we identified variants reported in the European ancestry meta-analysis and each non-European ancestry meta-analysis, in turn. These variants were assigned to *P*-value bins, according to the strength of the association with the trait in the European ancestry meta-analysis: $P < 5x10^{-8}$; $5x10^{-8} \leq P < 5x10^{-6}$; $5x10^{-6} \leq P < 5x10^{-4}$; $5x10^{-4} \leq P < 0.05$; and $P \geq 0.05$. Within each *P*-value bin, we selected a set of "independent" variants that were separated by 1 Mb. We defined independence using a distance-based threshold because of differences in patterns of LD between ancestry groups. For each *P*-value bin, the proportion of variants with the same direction of effect on the trait between the two ancestries was calculated along with a *P*-value from the binomial test to determine if the proportion of variants with the same direction of effect was greater than that expected by chance (50%, one sided).

LD-pruned variant lists

Several downstream analyses (for example, genomic feature enrichment, genetic scores, and estimation of variance explained by associated variants) require independent LD-pruned variants ($r^2 < 0.1$) to avoid double-counting variants which might otherwise be in LD with each other and that do not provide additional "independent" evidence. Therefore, for these analyses we generated different lists of either TA or single-ancestry LD pruned ($r^2 < 0.1$) variants, keeping in each case the variant with the strongest evidence of association (**Supplementary Table 8**). Subsequently, we combined TA and single-ancestry variant lists and conducted further LD pruning. For some analyses, we took the TA pruned variant list and added single-ancestry signals if the LD $r^2 < 0.1$, while for others we started with the single-ancestry pruned lists and supplemented with TA lead variants if the LD $r^2 < 0.1$. One exception was the list used for eQTL co-localisations, which included all single-ancestry European signals (without LD pruning) and supplemented with any additional TA lead variants (starting from the variants with the most significant P-values) in EUR LD $r^2 < 0.1$ with any of the variants already in list, and that reached at least $P < 1x10^{-5}$ in the European ancestry meta-analysis.

Trait variance explained by associated loci

To determine how much of the phenotypic variance of each trait could be explained by the corresponding trait-associated loci, variants were combined in a series of weighted genetic scores (GS). The analysis was performed in a subset of the cohorts included in the discovery GWAS (with representation from each ancestry) and in a smaller number of independent cohorts (European ancestry only). Up to three different GS were derived per trait (and for each ancestry) in order to evaluate the potential for the trans-ancestry meta-GWAS identified loci to provide additional information above and beyond that contributed by the ancestry-specific meta-analysis results. These GS comprised: List A - single-ancestry signals; List B - single-ancestry signals plus trans-ancestry signals; and List C - trans-ancestry signals plus single-ancestry signals (**Supplementary Table 8**). In

the case of the European ancestry cohorts that contributed to the GWAS, we employed the method of Nolte *et al.* [33] to adjust the effect sizes (betas) from the GWAS for the contribution of that cohort, providing sets of cohort-specific effect sizes that were then used to generate the GS. The association between each GS and its corresponding trait was tested by linear regression and the adjusted R^2 from the model extracted as an estimate of the variance explained.

Transferability of European ancestry-derived polygenic scores (PGS) across ancestries

We used the PRS-CSauto[34] software to first build European ancestry-derived PGS for each glycaemic trait (FG, FI, 2hGlu, HbA1c) on the basis of summary statistics. However, PRS-CSauto does not perform well when the training dataset is relatively small and the genetic architecture is sparse.[34] Consequently, 2hGlu was excluded from this analysis. For each trait, to obtain European ancestry training and test datasets, we first removed all cohorts only genotyped on the Metabochip which were not included in this analysis. From the remaining cohorts we then removed five of the largest European cohorts contributing to the respective European ancestry meta-analysis. For each trait, these five cohorts were meta-analysed and used as the European ancestry test dataset. Subsequently, the remaining European ancestry cohorts were also meta-analysed and used as the European ancestry training dataset. For each of the other ancestries, cohorts only genotyped on the Metabochip were also removed, and the remaining cohorts were meta-analysed, and used as the non-European ancestry test datasets. Variants with MAF<0.05 or missing in over half of the individuals in the training dataset were removed.[34,88] The PGS for each trait was built using PRS-CSauto with default settings[34] with the effect size estimates based on the European training dataset being revised based on an LD reference panel matching the test dataset. The proportion of the trait variance explained by the European ancestry-derived PGS (R^2) was estimated using the R package "gtx"[89] based on the revised effect sizes and summary statistics from the test dataset for each ancestry.

Fine-mapping

Of the 242 loci identified in this study, 237 were autosomal loci which we took forward for fine-mapping (**Supplementary Table 2)**. We used the Bayesian fine-mapping method FINEMAP [90] (version 1.1) to refine association signals and attempt to identify likely causal variants at each locus. FINEMAP estimates the maximum number of causal variants at each locus, calculates the posterior probability of each variant being causal, and proposes the most likely configuration of causal variants. The posterior probabilities of the configurations in each locus were used to construct 99% credible sets.

We performed both single-ancestry and trans-ancestry fine-mapping. In both analyses, only data from cohorts genotyped on GWAS arrays were used, and analyses were limited to trans-ancestry lead variants and other single-ancestry lead variants present in at least 90% of the samples for each trait. For the single-ancestry fine-mapping, FINEMAP estimates the number of causal variants in a region up to a maximum number, which we set to be two plus the number of distinct signals identified from the GCTA signal selection. FINEMAP uses single-ancestry and trait-specific z-scores from the fixed-effect meta-analysis in METAL [20] and an ancestry-specific LD reference, which we created from a subset of cohorts (combined sample size > 30% of the sample size for that ancestry), weighting each cohort by sample size. In the trans-ancestry fine-mapping, FINEMAP was similarly used to estimate the number of causal variants starting with two, and trait-specific z-scores and LD maps were generated from the sample size weighted average of those used in the single-ancestry fine-mapping. The maximum number of causal variants was iteratively increased by one until it was larger than the number of causal variants supported by data (Bayes factor), which was the estimated maximum number of causal variants used in the final run of fine-mapping analysis.

To compare fine-mapping results obtained from the single-ancestry and trans-ancestry efforts, analyses were limited to fine-mapping regions with evidence for a single likely causal variant in both, enabling a straightforward comparison of credible sets (**Supplementary note**). To ensure any difference in the fine-mapping results was not driven by different sets of variants being present in the different analyses, we repeated the single-ancestry fine-mapping limited to the same set of variants used in the trans-ancestry fine-mapping. The fine-mapping resolution was assessed based on comparisons of the 99% credible sets in terms of number of variants included in the set, and length of the region. To assess whether the improvement in the trans-ancestry fine-mapping was due to differences in LD, increased sample size, or both, we repeated the trans-ancestry fine-mapping mimicking the sample size present in the single-ancestry fine-mapping by dividing the standard errors by the square root of the sample size ratio and compared the results with those from the single-ancestry fine-mapping.

Functional Annotation of trait-associated variants

HbA1c signal classification

There were 218 HbA1c-associated signals from either the single-ancestry (i.e. all GCTA-signals from any ancestry) or trans-ancestry meta-analyses. To classify these signals in terms of their likely mode of action (i.e., glycaemic, erythrocytic, or other [7]), we examined association summary statistics for the lead variants at the 218 signals in other large European datasets for 19 additional traits: three glycaemic traits from this study (FG, 2hGlu and FI); seven mature red blood cell (RBC) traits[91,92] (red blood cell count, mean corpuscular volume, haematocrit, mean corpuscular haemoglobin, mean corpuscular haemoglobin concentration, haemoglobin concentration and red cell distribution width); five reticulocyte traits (reticulocyte count, reticulocyte fraction of red cells, immature fraction of reticulocytes, high light scatter reticulocyte count and high light scatter percentage of red cells)[91,92], and four iron traits (serum iron, transferrin, transferrin saturation and ferritin) [93]. Of the 218 HbA1c signals, data were available for the lead (n=183) or proxy (European LD $r^2 > 0.8$, n = 8) variants at 191 signals.

The additional traits were clustered using hierarchical clustering to ensure biologically related traits would cluster together (**Supplementary note**). We then used a non-negative matrix factorization (NMF) [94] process to cluster the HbA1c signals. Each cluster was labelled as glycaemic, reticulocyte, mature RBC, or iron related based on the strength of association of signals in the cluster to the glycaemic, reticulocyte, mature RBC and iron traits (**Supplementary note**). To verify that our cluster naming was correct, we used HbA1c association results conditioned on either FG or iron traits, or type 2 diabetes association results (**Supplementary note**).

HbA1c genetic risk scores (GRSs) and type 2 diabetes (T2D) risk

We constructed GRS for each cluster of HbA1c-associated signals (based on hard clustering) and tested the association of each cluster with T2D risk using samples from the UK Biobank. Pairs of HbA1c signals in LD (EUR $r^2 > 0.10$) were LD pruned by removing the signal with the less significant *P*-value of association with HbA1c. The GRS for each cluster was calculated based on the logarithm of odds ratios from the latest T2D study summary statistics [95] and UK Biobank genotypes imputed to the Haplotype Reference Consortium [19]. From 487,409 UK Biobank samples, we excluded participants for the following reasons: 373 with mismatched sex; 9 not used in the kinship calculation; 78,365 non-European ancestry individuals; and 138,504 with missing T2D status, age, or sex information. We further removed 26,896 related participants (kinship > 0.088, preferentially removing individuals with the largest number of relatives and controls where a T2D case was related to a control). T2D cases were defined by: (i) a history of diabetes without metformin or insulin treatment, (ii) self-reported diagnosis of T2D, or (iii) diagnosis of T2D in a national registry (N =

17,022). Controls were participants without a history of T2D (N = 226,240). We tested for association between each GRS and T2D using logistic regression including covariates for age, sex, and the first five principal components. Significance of association was evaluated by a bootstrap approach to incorporate the variance of each HbA1c associated signal in the T2D summary data. To do this, we generated the GRS of each cluster 200 times by resampling the logarithm of odds ratio of each signal with T2D. For each non-glycaemic class that had a GRS significantly associated with T2D, we performed sensitivity analyses to evaluate whether the association was driven from variants that also belonged to a glycaemic cluster when using a soft clustering approach (the signals were classified as also glycaemic in the soft clustering or had an association $P \leq 0.05$ with any of the three glycaemic traits).

Chromatin states

To identify genetic variants within association signals that overlapped predicted chromatin states, we used a previously published, 13 chromatin state model that included 31 diverse tissues, including pancreatic islets, skeletal muscle, adipose, and liver [39]. Briefly, this model was generated from cell/tissue ChIP-seq data for H3K27ac, H3K27me3, H3K36me3, H3K4me1, and H3K4me3, and input control from a diverse set of publicly available data [53,57,96,97] using the ChromHMM program [98]. As reported previously [39], stretch enhancers were defined as contiguous enhancer chromatin state (Active Enhancer 1 and 2, Genic Enhancer and Weak Enhancer) segments longer than 3kb [57].

Enrichment of genetic variants in genomic features

We used GREGOR (version 1.2.1) to calculate the enrichment of GWAS variants overlapping static and stretch enhancers [56]. For calculating the enrichment of glycaemic trait-associated variants in these annotations, we used the filtered list of trait-associated variants as described above (**Supplementary Table 8**) as input. For calculating the enrichment of sub-classified HbA1c variants, we included the list of loci characterized as Glycaemic, another list of loci characterized as Reticulocyte or mature Red Blood Cell, collectively representing the red blood cell fraction, along with lists of iron related or unclassified loci (**Supplementary Table 18**). We used the following parameters in GREGOR enrichment analyses: European r^2 threshold (for inclusion of variants in LD with the lead variant) = 0.8, LD window size = 1 Mb, and minimum neighbour number = 500.

We used fGWAS (version 0.3.6) [58] to calculate enrichment of glycaemic trait-associated variants in static and stretch enhancer annotations using summary level GWAS results. We used the default fGWAS parameters for enrichment analyses for individual annotations for each trait. For each annotation, the model provided the natural log of maximum likelihood estimate of the enrichment parameter. Annotations were considered as significantly enriched if the log2 (parameter estimate) and respective 95% confidence intervals were above zero or significantly depleted if the log2 (parameter estimate) and respective 95% confidence intervals were below zero.

We tested enrichment of trait-associated variants in static and stretch enhancer annotations with GARFIELD (v2) [59]. We formatted annotation overlap files as required by the tool; prepared input data at two GWAS thresholds - of $1x10^{-5}$ and a more stringent $1x10^{-8}$ by pruning and clumping with default parameters (garfield-prep-chr script). We calculated enrichment in each individual annotation using garfield-test.R with –c option set to 0. We also calculated the effective number of annotations using the garfield-Meff-Padj.R script. We used the effective number of annotations for each trait to obtain Bonferroni corrected significance thresholds for enrichment for each trait.

eQTL analyses

To aid in the identification of candidate casual genes at the European-only and trans-ancestry association signals, we examined whether any of the lead variants associated with glycaemic traits (**Supplementary Table 8**) were also associated with expression level (FDR < 5%) of nearby transcripts located within 1 Mb in existing eQTL data sets of blood, subcutaneous adipose, visceral adipose,

skeletal muscle, and pancreatic islet samples [60,61,99-102]. LD was estimated from the collected cohort pairwise LD information, where available, else from the European samples in 1000G phase 3. GWAS and eQTL signals likely co-localise when the GWAS variant and the variant most strongly associated with the expression level of the corresponding transcript (eSNP) exhibit high pairwise LD ($r^2 > 0.8$; 1000 Genomes Phase 3, EUR). At these signals, we conducted reciprocal conditional analyses to test association between the GWAS variant and transcript level when the eSNP was also included in the model, and vice versa. We report GWAS and eQTL signals as co-localised if the association for the eSNP was not significant (FDR ≥ 5%) when conditioned on the GWAS variant; we also report signals from the eQTLGen whole blood meta-analysis data that meet only the LD threshold because conditional analysis was not possible.

Tissue and gene-set analysis

We performed enrichment analysis using DEPICT (Data-driven Expression-Prioritized Integration for Complex Traits) version 3, specifically developed for 1000 Genomes Project imputed meta-analysis data [103] to identify cell types and tissues in which genes at trait-associated variants were strongly expressed, and to detect enrichment of gene-sets or pathways. DEPICT data included human gene expression data for 19,987 genes in 10,968 reconstituted gene sets, and 209 tissues/cell types. Because gene expression data in DEPICT is based on European samples and LD, we selected trait-associated variants with $P < 10^{-5}$ in the European meta-analysis and tested for enrichment of signals in each reconstituted gene-set, and each tissue or cell type. Enrichment results with a false discovery rate (FDR) < 0.05 were considered significant. We ran DEPICT based on association results for all traits among: (i) cohorts with genome-wide data, or (ii) all cohorts (genome-wide and Metabochip cohorts). Because results were broadly consistent between the two approaches, we present results from the analysis that contained all cohorts as it had greater statistical power.

References

1 in *Use of Glycated Haemoglobin (HbA1c) in the Diagnosis of Diabetes Mellitus: Abbreviated Report of a WHO Consultation* (World Health Organization Copyright © World Health Organization 2011., 2011).

2 Goodarzi, M. O. *et al.* Fasting insulin reflects heterogeneous physiological processes: role of insulin clearance. *American journal of physiology. Endocrinology and metabolism* **301**, E402-408, doi:10.1152/ajpendo.00013.2011 (2011).

3 Dimas, A. S. *et al.* Impact of type 2 diabetes susceptibility variants on quantitative glycemic traits reveals mechanistic heterogeneity. *Diabetes* **63**, 2158-2171, doi:10.2337/db13-0949 (2014).

4 Udler, M. S. *et al.* Type 2 diabetes genetic loci informed by multi-trait associations point to disease mechanisms and subtypes: A soft clustering analysis. *PLoS medicine* **15**, e1002654, doi:10.1371/journal.pmed.1002654 (2018).

5 Udler, M. S., McCarthy, M. I., Florez, J. C. & Mahajan, A. Genetic Risk Scores for Diabetes Diagnosis and Precision Medicine. *Endocrine reviews* **40**, 1500-1520, doi:10.1210/er.2019-00088 (2019).

6 Sarwar, N. *et al.* Diabetes mellitus, fasting blood glucose concentration, and risk of vascular disease: a collaborative meta-analysis of 102 prospective studies. *Lancet* **375**, 2215-2222, doi:10.1016/s0140-6736(10)60484-9 (2010).

7 Wheeler, E. *et al.* Impact of common genetic determinants of Hemoglobin A1c on type 2 diabetes risk and diagnosis in ancestrally diverse populations: A transethnic genome-wide meta-analysis. *PLoS medicine* **14**, e1002383, doi:10.1371/journal.pmed.1002383 (2017).

8 Dupuis, J. *et al.* New genetic loci implicated in fasting glucose homeostasis and their impact on type 2 diabetes risk. *Nature genetics* **42**, 105-116, doi:10.1038/ng.520 (2010).
9 Manning, A. K. *et al.* A genome-wide approach accounting for body mass index identifies genetic variants influencing fasting glycemic traits and insulin resistance. *Nature genetics* **44**, 659-669, doi:10.1038/ng.2274 (2012).
10 Walford, G. A. *et al.* Genome-Wide Association Study of the Modified Stumvoll Insulin Sensitivity Index Identifies BCL2 and FAM19A2 as Novel Insulin Sensitivity Loci. *Diabetes* **65**, 3200-3211, doi:10.2337/db16-0199 (2016).
11 Horikoshi, M. *et al.* Discovery and Fine-Mapping of Glycaemic and Obesity-Related Trait Loci Using High-Density Imputation. *PLoS genetics* **11**, e1005230, doi:10.1371/journal.pgen.1005230 (2015).
12 Mahajan, A. *et al.* Identification and functional characterization of G6PC2 coding variants influencing glycemic traits define an effector transcript at the G6PC2-ABCB11 locus. *PLoS genetics* **11**, e1004876, doi:10.1371/journal.pgen.1004876 (2015).
13 Hwang, J. Y. *et al.* Genome-wide association meta-analysis identifies novel variants associated with fasting plasma glucose in East Asians. *Diabetes* **64**, 291-298, doi:10.2337/db14-0563 (2015).
14 Chen, P. *et al.* Multiple nonglycemic genomic loci are newly associated with blood level of glycated hemoglobin in East Asians. *Diabetes* **63**, 2551-2562, doi:10.2337/db13-1815 (2014).
15 Scott, R. A. *et al.* Large-scale association analyses identify new loci influencing glycemic traits and provide insight into the underlying biological pathways. *Nature genetics* **44**, 991-1005, doi:10.1038/ng.2385 (2012).
16 Spanakis, E. K. & Golden, S. H. Race/ethnic difference in diabetes and diabetic complications. *Current diabetes reports* **13**, 814-823, doi:10.1007/s11892-013-0421-9 (2013).
17 Tillin, T. *et al.* Insulin resistance and truncal obesity as important determinants of the greater incidence of diabetes in Indian Asians and African Caribbeans compared with Europeans: the Southall And Brent REvisited (SABRE) cohort. *Diabetes care* **36**, 383-393, doi:10.2337/dc12-0544 (2013).
18 Whincup, P. H. *et al.* Early emergence of ethnic differences in type 2 diabetes precursors in the UK: the Child Heart and Health Study in England (CHASE Study). *PLoS medicine* **7**, e1000263, doi:10.1371/journal.pmed.1000263 (2010).
19 Auton, A. *et al.* A global reference for human genetic variation. *Nature* **526**, 68-74, doi:10.1038/nature15393 (2015).
20 Willer, C. J., Li, Y. & Abecasis, G. R. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics (Oxford, England)* **26**, 2190-2191, doi:10.1093/bioinformatics/btq340 (2010).
21 Yang, J., Lee, S. H., Goddard, M. E. & Visscher, P. M. GCTA: a tool for genome-wide complex trait analysis. *American journal of human genetics* **88**, 76-82, doi:10.1016/j.ajhg.2010.11.011 (2011).
22 Yang, J. *et al.* Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. *Nature genetics* **44**, 369-375, s361-363, doi:10.1038/ng.2213 (2012).

23 Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* **447**, 661-678, doi:10.1038/nature05911 (2007).
24 Mahajan, A. *et al.* Trans-ancestry genetic study of type 2 diabetes highlights the power of diverse populations for discovery and translation. *medRxiv*, 2020.2009.2022.20198937, doi:10.1101/2020.09.22.20198937 (2020).
25 Mahajan, A. *et al.* Fine-mapping type 2 diabetes loci to single-variant resolution using high-density imputation and islet-specific epigenome maps. *Nature genetics* **50**, 1505-1513, doi:10.1038/s41588-018-0241-6 (2018).
26 Spracklen, C. N. *et al.* Identification of type 2 diabetes loci in 433,540 East Asian individuals. *Nature* **582**, 240-245, doi:10.1038/s41586-020-2263-3 (2020).
27 Vujkovic, M. *et al.* Discovery of 318 new risk loci for type 2 diabetes and related vascular outcomes among 1.4 million participants in a multi-ancestry meta-analysis. *Nature genetics* **52**, 680-691, doi:10.1038/s41588-020-0637-y (2020).
28 Luo, Y. *et al.* Transcription factor Ets1 regulates expression of thioredoxin-interacting protein and inhibits insulin secretion in pancreatic beta-cells. *PloS one* **9**, e99049, doi:10.1371/journal.pone.0099049 (2014).
29 Braccini, L. *et al.* PI3K-C2gamma is a Rab5 effector selectively controlling endosomal Akt2 activation downstream of insulin signalling. *Nature communications* **6**, 7400, doi:10.1038/ncomms8400 (2015).
30 Ng, N. H. J. *et al.* Tissue-Specific Alteration of Metabolic Pathways Influences Glycemic Regulation. *bioRxiv*, 790618, doi:10.1101/790618 (2019).
31 Aschard, H., Vilhjalmsson, B. J., Joshi, A. D., Price, A. L. & Kraft, P. Adjusting for heritable covariates can bias effect estimates in genome-wide association studies. *American journal of human genetics* **96**, 329-339, doi:10.1016/j.ajhg.2014.12.021 (2015).
32 Lee, J. J. *et al.* Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals. *Nature genetics* **50**, 1112-1121, doi:10.1038/s41588-018-0147-3 (2018).
33 Nolte, I. M. *et al.* Missing heritability: is the gap closing? An analysis of 32 complex traits in the Lifelines Cohort Study. *European journal of human genetics : EJHG* **25**, 877-885, doi:10.1038/ejhg.2017.50 (2017).
34 Ge, T., Chen, C. Y., Ni, Y., Feng, Y. 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).
35 Dastani, Z. *et al.* Novel loci for adiponectin levels and their influence on type 2 diabetes and metabolic traits: a multi-ethnic meta-analysis of 45,891 individuals. *PLoS genetics* **8**, e1002607, doi:10.1371/journal.pgen.1002607 (2012).
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 Gaulton, K. J. *et al.* Genetic fine mapping and genomic annotation defines causal mechanisms at type 2 diabetes susceptibility loci. *Nature genetics* **47**, 1415-1425, doi:10.1038/ng.3437 (2015).
38 Spracklen, C. N. *et al.* Identification and functional analysis of glycemic trait loci in the China Health and Nutrition Survey. *PLoS genetics* **14**, e1007275, doi:10.1371/journal.pgen.1007275 (2018).

39 Varshney, A. *et al.* Genetic regulatory signatures underlying islet gene expression and type 2 diabetes. *Proceedings of the National Academy of Sciences of the United States of America* **114**, 2301-2306, doi:10.1073/pnas.1621192114 (2017).

40 Kichaev, G. *et al.* Leveraging Polygenic Functional Enrichment to Improve GWAS Power. *American journal of human genetics* **104**, 65-75, doi:10.1016/j.ajhg.2018.11.008 (2019).

41 Shriner, D. & Rotimi, C. N. Whole-Genome-Sequence-Based Haplotypes Reveal Single Origin of the Sickle Allele during the Holocene Wet Phase. *American journal of human genetics* **102**, 547-556, doi:10.1016/j.ajhg.2018.02.003 (2018).

42 Kramer, H. J. *et al.* African Ancestry-Specific Alleles and Kidney Disease Risk in Hispanics/Latinos. *Journal of the American Society of Nephrology : JASN* **28**, 915-922, doi:10.1681/asn.2016030357 (2017).

43 Ravenhall, M. *et al.* Novel genetic polymorphisms associated with severe malaria and under selective pressure in North-eastern Tanzania. *PLoS genetics* **14**, e1007172, doi:10.1371/journal.pgen.1007172 (2018).

44 Hodonsky, C. J. *et al.* Genome-wide association study of red blood cell traits in Hispanics/Latinos: The Hispanic Community Health Study/Study of Latinos. *PLoS genetics* **13**, e1006760, doi:10.1371/journal.pgen.1006760 (2017).

45 Gurdasani, D. *et al.* Uganda Genome Resource Enables Insights into Population History and Genomic Discovery in Africa. *Cell* **179**, 984-1002.e1036, doi:10.1016/j.cell.2019.10.004 (2019).

46 Rees, M. G. *et al.* Cellular characterisation of the GCKR P446L variant associated with type 2 diabetes risk. *Diabetologia* **55**, 114-122, doi:10.1007/s00125-011-2348-5 (2012).

47 Bonomo, J. A. *et al.* The ras responsive transcription factor RREB1 is a novel candidate gene for type 2 diabetes associated end-stage kidney disease. *Human molecular genetics* **23**, 6441-6447, doi:10.1093/hmg/ddu362 (2014).

48 Wessel, J. *et al.* Low-frequency and rare exome chip variants associate with fasting glucose and type 2 diabetes susceptibility. *Nature communications* **6**, 5897, doi:10.1038/ncomms6897 (2015).

49 Scott, R. A. *et al.* A genomic approach to therapeutic target validation identifies a glucose-lowering GLP1R variant protective for coronary heart disease. *Science translational medicine* **8**, 341ra376, doi:10.1126/scitranslmed.aad3744 (2016).

50 Nai, A. *et al.* TMPRSS6 rs855791 modulates hepcidin transcription in vitro and serum hepcidin levels in normal individuals. *Blood* **118**, 4459-4462, doi:10.1182/blood-2011-06-364034 (2011).

51 Soranzo, N. *et al.* Common variants at 10 genomic loci influence hemoglobin A(1)(C) levels via glycemic and nonglycemic pathways. *Diabetes* **59**, 3229-3239, doi:10.2337/db10-0502 (2010).

52 Sarnowski, C. *et al.* Impact of Rare and Common Genetic Variants on Diabetes Diagnosis by Hemoglobin A1c in Multi-Ancestry Cohorts: The Trans-Omics for Precision Medicine Program. *American journal of human genetics* **105**, 706-718, doi:10.1016/j.ajhg.2019.08.010 (2019).

53 Kundaje, A. *et al.* Integrative analysis of 111 reference human epigenomes. *Nature* **518**, 317-330, doi:10.1038/nature14248 (2015).

54 Nagel, M. *et al.* Meta-analysis of genome-wide association studies for neuroticism in 449,484 individuals identifies novel genetic loci and pathways. *Nature genetics* **50**, 920-927, doi:10.1038/s41588-018-0151-7 (2018).
55 Savage, J. E. *et al.* Genome-wide association meta-analysis in 269,867 individuals identifies new genetic and functional links to intelligence. *Nature genetics* **50**, 912-919, doi:10.1038/s41588-018-0152-6 (2018).
56 Schmidt, E. M. *et al.* GREGOR: evaluating global enrichment of trait-associated variants in epigenomic features using a systematic, data-driven approach. *Bioinformatics (Oxford, England)* **31**, 2601-2606, doi:10.1093/bioinformatics/btv201 (2015).
57 Parker, S. C. *et al.* Chromatin stretch enhancer states drive cell-specific gene regulation and harbor human disease risk variants. *Proceedings of the National Academy of Sciences of the United States of America* **110**, 17921-17926, doi:10.1073/pnas.1317023110 (2013).
58 Pickrell, J. K. Joint analysis of functional genomic data and genome-wide association studies of 18 human traits. *American journal of human genetics* **94**, 559-573, doi:10.1016/j.ajhg.2014.03.004 (2014).
59 Iotchkova, V. *et al.* GARFIELD classifies disease-relevant genomic features through integration of functional annotations with association signals. *Nature genetics* **51**, 343-353, doi:10.1038/s41588-018-0322-6 (2019).
60 van de Bunt, M. *et al.* Transcript Expression Data from Human Islets Links Regulatory Signals from Genome-Wide Association Studies for Type 2 Diabetes and Glycemic Traits to Their Downstream Effectors. *PLoS genetics* **11**, e1005694, doi:10.1371/journal.pgen.1005694 (2015).
61 Civelek, M. *et al.* Genetic Regulation of Adipose Gene Expression and Cardio-Metabolic Traits. *American journal of human genetics* **100**, 428-443, doi:10.1016/j.ajhg.2017.01.027 (2017).
62 Scott, L. J. *et al.* The genetic regulatory signature of type 2 diabetes in human skeletal muscle. *Nature communications* **7**, 11764, doi:10.1038/ncomms11764 (2016).
63 Ben Harouch, S., Klar, A. & Falik Zaccai, T. C. in *GeneReviews((R))* (eds M. P. Adam *et al.*) (University of Washington, Seattle
University of Washington, Seattle. GeneReviews is a registered trademark of the University of Washington, Seattle. All rights reserved., 1993).
64 Agus, D. B. *et al.* Vitamin C crosses the blood-brain barrier in the oxidized form through the glucose transporters. *The Journal of clinical investigation* **100**, 2842-2848, doi:10.1172/jci119832 (1997).
65 Wolking, S. *et al.* Focal epilepsy in glucose transporter type 1 (Glut1) defects: case reports and a review of literature. *Journal of neurology* **261**, 1881-1886, doi:10.1007/s00415-014-7433-5 (2014).
66 Guallar, D. *et al.* RNA-dependent chromatin targeting of TET2 for endogenous retrovirus control in pluripotent stem cells. *Nature genetics* **50**, 443-451, doi:10.1038/s41588-018-0060-9 (2018).
67 Bian, F. *et al.* TET2 facilitates PPARgamma agonist-mediated gene regulation and insulin sensitization in adipocytes. *Metabolism: clinical and experimental* **89**, 39-47, doi:10.1016/j.metabol.2018.08.006 (2018).

68 Yoo, Y. *et al.* TET-mediated hydroxymethylcytosine at the Ppargamma locus is required for initiation of adipogenic differentiation. *International journal of obesity (2005)* **41**, 652-659, doi:10.1038/ijo.2017.8 (2017).

69 Lees, J. A. *et al.* Lipid transport by TMEM24 at ER-plasma membrane contacts regulates pulsatile insulin secretion. *Science (New York, N.Y.)* **355**, doi:10.1126/science.aah6171 (2017).

70 Pottekat, A. *et al.* Insulin biosynthetic interaction network component, TMEM24, facilitates insulin reserve pool release. *Cell reports* **4**, 921-930, doi:10.1016/j.celrep.2013.07.050 (2013).

71 Androulakis, II *et al.* Patients with apparently nonfunctioning adrenal incidentalomas may be at increased cardiovascular risk due to excessive cortisol secretion. *The Journal of clinical endocrinology and metabolism* **99**, 2754-2762, doi:10.1210/jc.2013-4064 (2014).

72 Altieri, B. *et al.* Adrenocortical tumors and insulin resistance: What is the first step? *International journal of cancer* **138**, 2785-2794, doi:10.1002/ijc.29950 (2016).

73 Johansson, M. *et al.* The influence of obesity-related factors in the etiology of renal cell carcinoma-A mendelian randomization study. *PLoS medicine* **16**, e1002724, doi:10.1371/journal.pmed.1002724 (2019).

74 Diamanti-Kandarakis, E. & Dunaif, A. Insulin resistance and the polycystic ovary syndrome revisited: an update on mechanisms and implications. *Endocrine reviews* **33**, 981-1030, doi:10.1210/er.2011-1034 (2012).

75 Morris, A. P. *et al.* Large-scale association analysis provides insights into the genetic architecture and pathophysiology of type 2 diabetes. *Nature genetics* **44**, 981-990, doi:10.1038/ng.2383 (2012).

76 Leong, A. *et al.* Mendelian Randomization Analysis of Hemoglobin A(1c) as a Risk Factor for Coronary Artery Disease. *Diabetes care* **42**, 1202-1208, doi:10.2337/dc18-1712 (2019).

77 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).

78 Mostafavi, H. *et al.* Variable prediction accuracy of polygenic scores within an ancestry group. *eLife* **9**, doi:10.7554/eLife.48376 (2020).

79 Choi, S. W., Mak, T. S. & O'Reilly, P. F. Tutorial: a guide to performing polygenic risk score analyses. *Nature protocols* **15**, 2759-2772, doi:10.1038/s41596-020-0353-1 (2020).

80 D'Orazio, P. *et al.* Approved IFCC recommendation on reporting results for blood glucose (abbreviated). *Clinical chemistry* **51**, 1573-1576, doi:10.1373/clinchem.2005.051979 (2005).

81 Voight, B. F. *et al.* The metabochip, a custom genotyping array for genetic studies of metabolic, cardiovascular, and anthropometric traits. *PLoS genetics* **8**, e1002793, doi:10.1371/journal.pgen.1002793 (2012).

82 Abecasis, G. R. *et al.* An integrated map of genetic variation from 1,092 human genomes. *Nature* **491**, 56-65, doi:10.1038/nature11632 (2012).

83 Li, Y., Willer, C. J., Ding, J., Scheet, P. & Abecasis, G. R. MaCH: using sequence and genotype data to estimate haplotypes and unobserved genotypes. *Genetic epidemiology* **34**, 816-834, doi:10.1002/gepi.20533 (2010).

84 Pei, Y. F., Zhang, L., Li, J. & Deng, H. W. Analyses and comparison of imputation-based association methods. *PloS one* **5**, e10827, doi:10.1371/journal.pone.0010827 (2010).
85 Winkler, T. W. *et al.* Quality control and conduct of genome-wide association meta-analyses. *Nature protocols* **9**, 1192-1212, doi:10.1038/nprot.2014.071 (2014).
86 Devlin, B. & Roeder, K. Genomic control for association studies. *Biometrics* **55**, 997-1004 (1999).
87 Morris, A. P. Transethnic meta-analysis of genomewide association studies. *Genetic epidemiology* **35**, 809-822, doi:10.1002/gepi.20630 (2011).
88 Bulik-Sullivan, B. K. *et al.* LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nature genetics* **47**, 291-295, doi:10.1038/ng.3211 (2015).
89 Dastani, Z. *et al.* Novel loci for adiponectin levels and their influence on type 2 diabetes and metabolic traits: a multi-ethnic meta-analysis of 45,891 individuals. *PLoS genetics* **8**, e1002607, doi:10.1371/journal.pgen.1002607 (2012).
90 Benner, C. *et al.* FINEMAP: efficient variable selection using summary data from genome-wide association studies. *Bioinformatics (Oxford, England)* **32**, 1493-1501, doi:10.1093/bioinformatics/btw018 (2016).
91 Astle, W. J. *et al.* The Allelic Landscape of Human Blood Cell Trait Variation and Links to Common Complex Disease. *Cell* **167**, 1415-1429.e1419, doi:10.1016/j.cell.2016.10.042 (2016).
92 Canela-Xandri, O., Rawlik, K. & Tenesa, A. An atlas of genetic associations in UK Biobank. *Nature genetics* **50**, 1593-1599, doi:10.1038/s41588-018-0248-z (2018).
93 Benyamin, B. *et al.* Novel loci affecting iron homeostasis and their effects in individuals at risk for hemochromatosis. *Nature communications* **5**, 4926, doi:10.1038/ncomms5926 (2014).
94 Binesh, N. & Rezghi, M. Fuzzy clustering in community detection based on nonnegative matrix factoriztion with two novel evaluation criteria. *Applied Soft Computing* **69**, 689-703 (2018).
95 Scott, R. A. *et al.* An Expanded Genome-Wide Association Study of Type 2 Diabetes in Europeans. *Diabetes* **66**, 2888-2902, doi:10.2337/db16-1253 (2017).
96 Ernst, J. *et al.* Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature* **473**, 43-49, doi:10.1038/nature09906 (2011).
97 Mikkelsen, T. S. *et al.* Comparative epigenomic analysis of murine and human adipogenesis. *Cell* **143**, 156-169, doi:10.1016/j.cell.2010.09.006 (2010).
98 Ernst, J. & Kellis, M. ChromHMM: automating chromatin-state discovery and characterization. *Nature methods* **9**, 215-216, doi:10.1038/nmeth.1906 (2012).
99 Battle, A., Brown, C. D., Engelhardt, B. E. & Montgomery, S. B. Genetic effects on gene expression across human tissues. *Nature* **550**, 204-213, doi:10.1038/nature24277 (2017).
100 Zhernakova, D. V. *et al.* Identification of context-dependent expression quantitative trait loci in whole blood. *Nature genetics* **49**, 139-145, doi:10.1038/ng.3737 (2017).
101 Westra, H. J. *et al.* Systematic identification of trans eQTLs as putative drivers of known disease associations. *Nature genetics* **45**, 1238-1243, doi:10.1038/ng.2756 (2013).

102 Joehanes, R. *et al.* Integrated genome-wide analysis of expression quantitative trait loci aids interpretation of genomic association studies. *Genome biology* **18**, 16, doi:10.1186/s13059-016-1142-6 (2017).
103 Pers, T. H. *et al.* Biological interpretation of genome-wide association studies using predicted gene functions. *Nature communications* **6**, 5890, doi:10.1038/ncomms6890 (2015).

Acknowledgments
The authors thank all investigators, staff members, and study participants for their contribution to all participating studies. A full list of funding and individual and study acknowledgments appears in the **Supplementary Note**.

Competing interests statement
The authors declare the existence of a financial/non-financial competing interest. Full disclosures are included in the **Supplementary Note**.

Figure Legends

Figure 1 - Summary of all 242 loci identified in this study. 235 trans-ancestry loci are shown in orange (novel) or black (established) along with seven single-ancestry loci (blue) represented by nearest gene. Each locus is mapped to corresponding chromosome (outer segment). Each set of rows shows the results from the trans-ancestry analysis (orange) and each of the ancestries: European (purple), African American (tan), East Asian (grey), South Asian (green), Hispanic (yellow), sub-Saharan African (Ugandan-pink). Loci with a corresponding type 2 diabetes signal are represented by red circles in the middle of the plot.

Figure 2 – Trait variance explained by associated loci. The boxplots show the maximum, first quartile, median, third quartile and minimum of trait variance explained when using a genetic score with single-ancestry lead and index variants (EUR, AA, EAS, HISP and SAS) or a combination of individual trait trans-ancestry lead variants and single-ancestry lead and index variants (TA+EUR, TA+AA, TA+EAS, TA+HISP and TA+SAS). Variance explained for each trait (FG, FI and HbA1c) in each ancestry is shown on different panels and in different colours. R^2 was estimated in 1 to 11 cohorts with sample sizes ranging from 489 to 9,758 (**Supplementary Tables 9-12**).

Figure 3 – Transferability of PGS across ancestries. For each trait, the barplots represent trait variance explained when using a European ancestry-derived PGS in European, East Asian and African American test datasets. Variance explained (the height of each bar) for each trait (FG, FI and HbA1c) in each ancestry is shown on different panels and in different colours.

Figure 4 - Trans-ancestry fine-mapping. A) Number of plausible causal variants at each locus-trait association derived from FINEMAP. B) Number of variants within each 99% credible set. Twenty-one locus-trait associations at 19 loci were mapped to a single variant in the 99% credible set. C) Fine-mapping resolution. For each of the 98 locus-trait associations with a predicted single causal variant in both trans-ancestry and single-ancestry analyses, the number of variants included in the 99% credible set in the single-ancestry fine-mapping (x axis; logarithmic scale) is plotted against those in the trans-ancestry fine-mapping (y axis; logarithmic scale). Trans-ancestry and single-ancestry fine-mapping were based on the same set of variants. After removing eight locus-trait associations with

one variant in the 99% credible sets in both trans-ancestry and single-ancestry analyses, there were 18 locus-trait associations (in grey) where trans-ancestry fine-mapping did not improve the resolution of fine-mapping results (i.e. number of variants in the 99% credible set did not decrease). Of the 72 locus-trait associations with improved trans-ancestry fine-mapping resolution (blue and red) further analyses in European fine-mapping emulating the total sample size in trans-ancestry fine-mapping demonstrated that 34 locus-trait associations (in red) were improved because of both total sample size and differences across ancestries, while 38 locus-trait associations (in blue) were only improved due to increased sample size in the original trans-ancestry fine-mapping analysis.

Figure 5 - Epigenomic landscape of trait-associated variants. A: Enrichment of GWAS variants to overlap genomic regions including 'Static Annotations' which are common or 'static' across cell types and 'Stretch Enhancers' which are identified in each tissue/cell type. The numbers of signals for each trait are indicated in parentheses. Enrichment was calculated using GREGOR [56]. Significance (red) is determined after Bonferroni correction to account for 59 total annotations tested for each trait; nominal significance (P<0.05) is indicated in yellow. B: Enrichment for HbA1c GWAS signals partitioned into "hard" Glycaemic and Red Blood Cell cluster (signals from "hard" mature Red Blood Cell and reticulocyte clusters together) to overlap annotations including stretch enhancers in Islets and the blood-derived leukemia cell line K562, respectively (additional partitioned results in **Supplementary Table 18**). C: Individual FI GWAS signals that drive enrichment in Adipose and Skeletal Muscle stretch enhancers. D, E: Genome browser shots of FI GWAS signals – intronic region of the *COL4A2* gene (D) and an inter-genic region ~25kb from *LINC01214* gene (E) showing GWAS SNPs (lead and LD r^2>0.8 proxies), ATAC-seq signal tracks and chromatin state annotations in different tissues/cell types.

Figure 6 - Tissues and cell types significantly enriched for genes within glycaemic-associated loci. Top panel FG-associated loci, middle panel FI-associated loci, bottom panel Hba1c-associated loci. FDR thresholds are shown in red (q<0.05), orange (q<0.2), grey (q≥0.2).

Figure 7 - Gene-set enrichment analyses. Results from affinity-propagation clustering of significantly enriched gene sets (FDR<0.05) identified by DEPICT for A) FG, B) FI, and C) HbA1c. Each node is a cluster of gene-sets represented by an exemplar gene-set with similarities between the clusters represented by the Pearson correlation coefficients (r>0.3). The nodes are coloured according to the minimum gene-set enrichment p-value for gene-sets in that cluster. Example clusters are expanded to show the contributing gene-sets.

Tables

Table 1 – Glossary of terms

Orange (novel)

Blue (novel, single-ancestry)

Black (known)

TA
FG
FI
2hGlu
HbA1c

EUR
FG
FI
2hGlu
HbA1c

AA
FG
FI
2hGlu
HbA1c

EAS
FG
FI
2hGlu
HbA1c

SAS
FG
FI
2hGlu
HbA1c

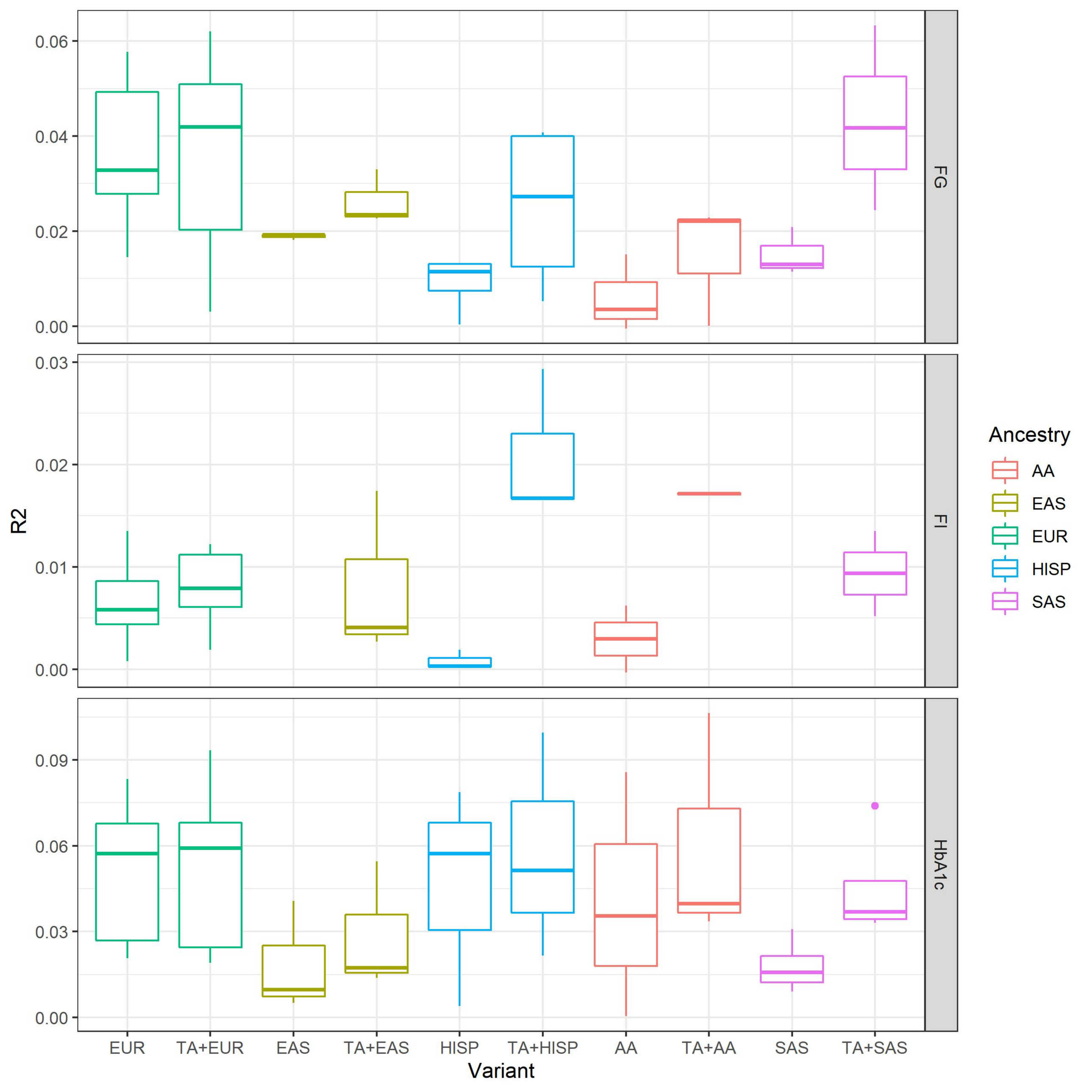
HISP
FG
FI
2hGlu
HbA1c

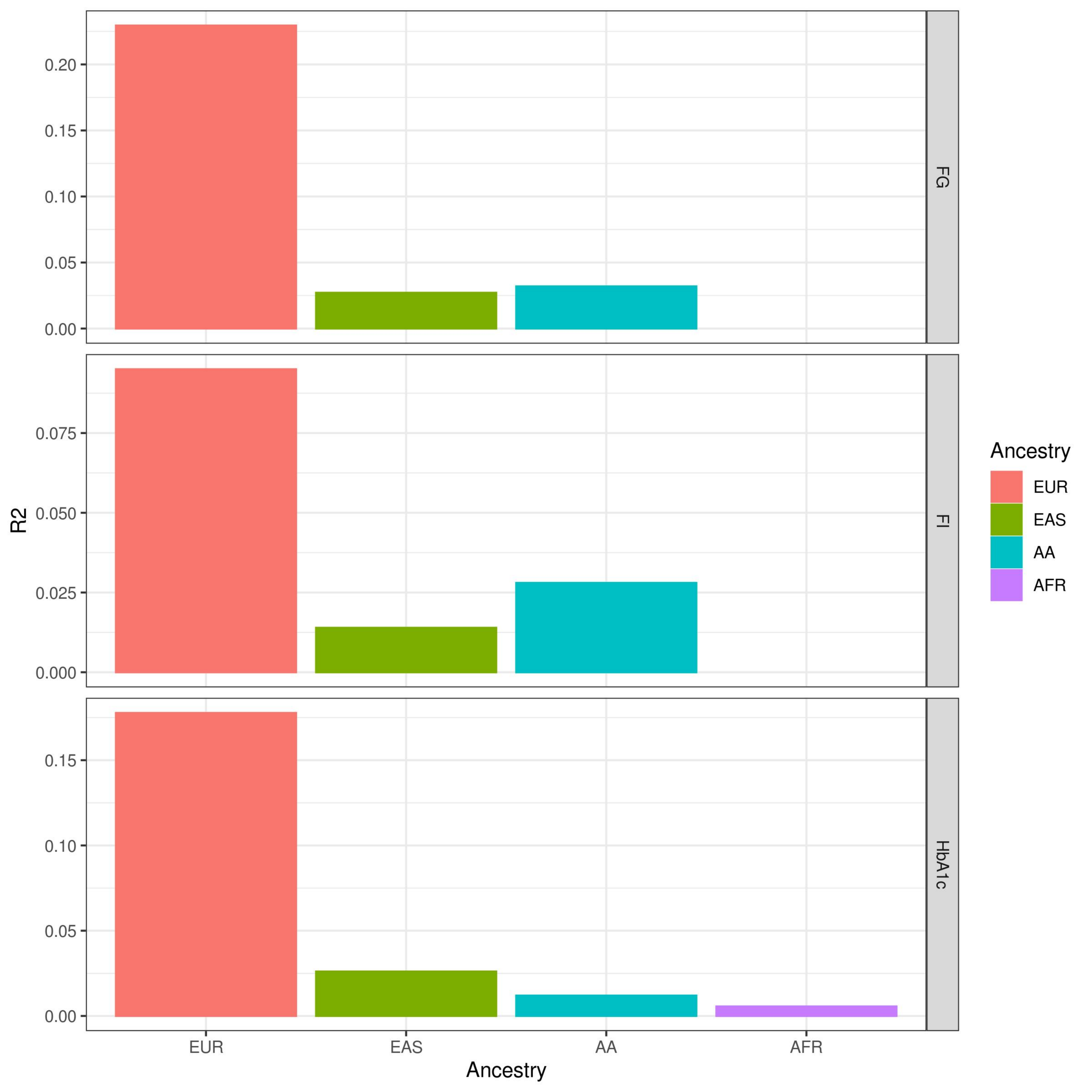
AFR
HbA1c

Type 2 diabetes

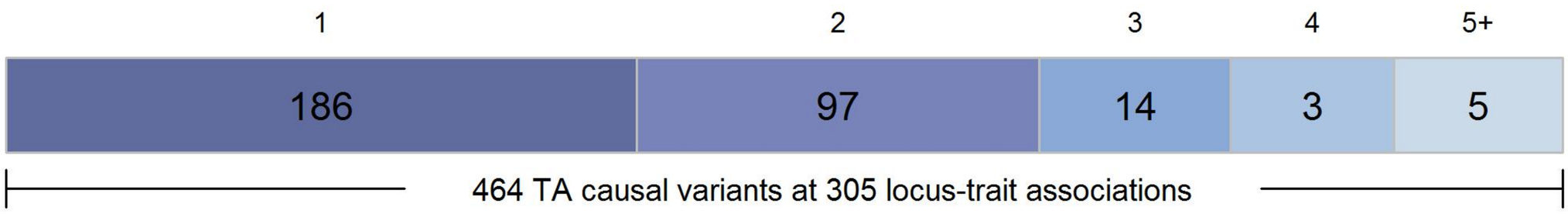
TA signal

Single-ancestry signal

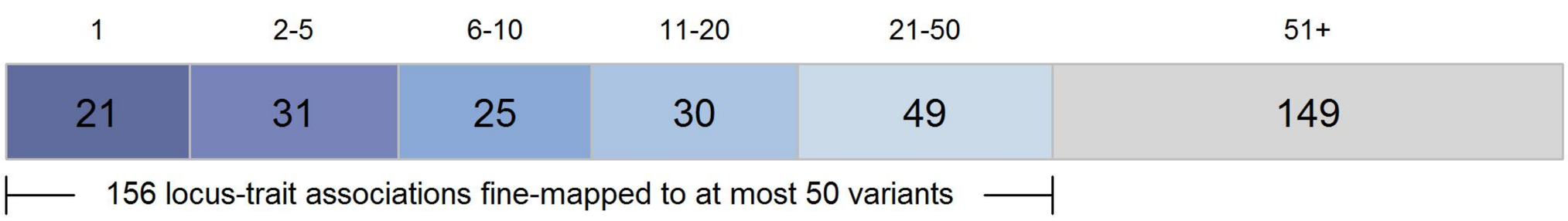

FG
FI
HbA1c
R2
0.06
0.04
0.02
0.00
0.03
0.01
0.09
EUR
TA+EUR
EAS
TA+EAS
HISP
TA+HISP
AA
TA+AA
SAS
TA+SAS
Variant
Ancestry
AA
EAS
EUR
HISP
SAS


FG
FI
HbA1c
R2
0.20
0.15
0.10
0.05
0.00
0.075
0.050
0.025
0.000
EUR
EAS
AA
AFR
Ancestry

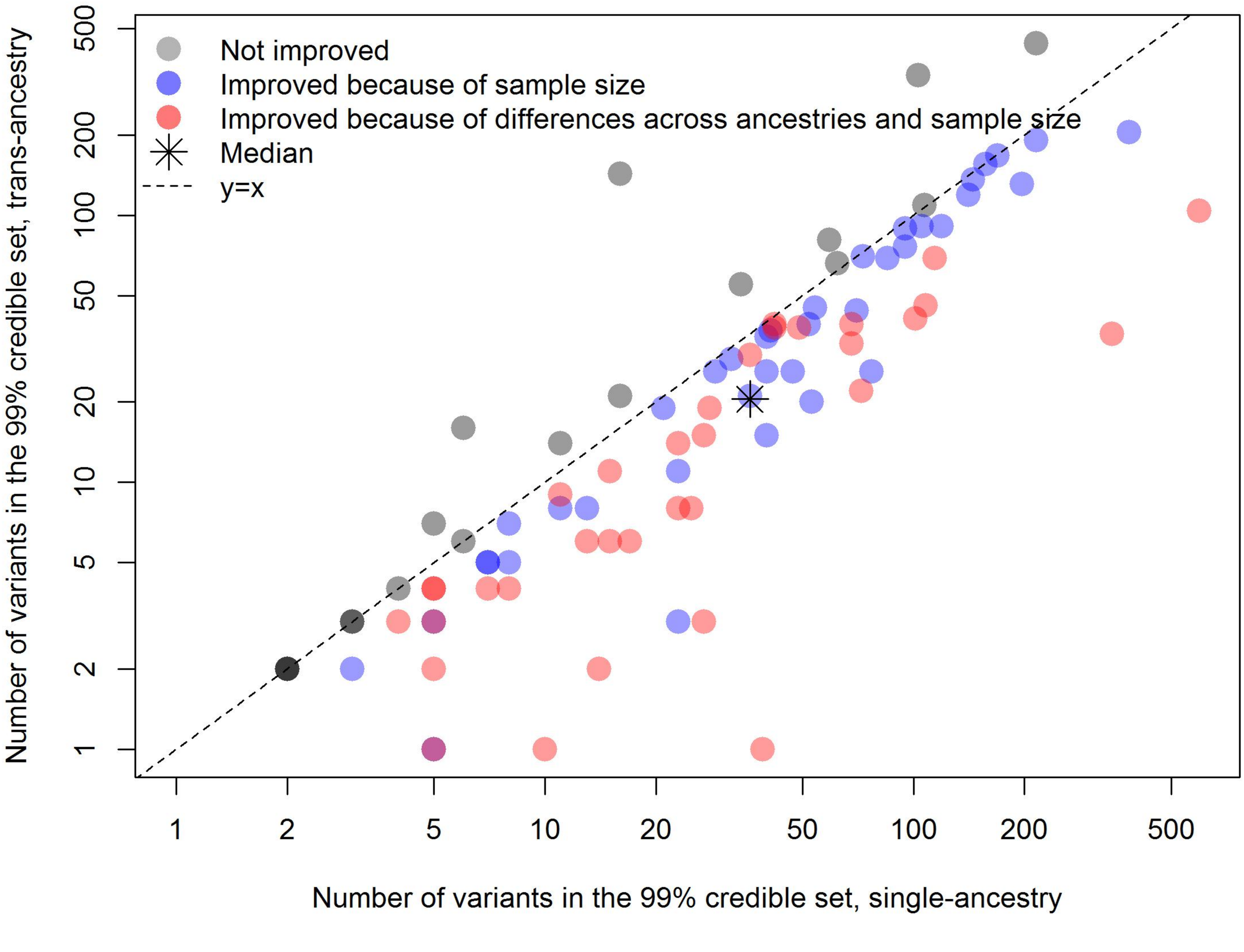
A. Estimated number of causal variants at each locus-trait association



B. Number of variants at each 99% credible set



C. Fine-mapping resolution



A

Fasting Glucose (BMI adjusted) (N=105)
HbA1c (N = 120)
2hr Glucose (BMI adjusted) (N = 17)
Fasting Insulin (BMI adjusted) (N = 61)

Stretch Enhancers in cell/tissues: Islets, Liver, HepG2, Anterior Caudate, Stomach Smooth Muscle, H1, Cingulate Gyrus, Huvec, Skeletal Muscle, Adipose, K562, GM12878, NHEK

Static Annotations: 5' UTRs, 3' UTRs, Coding, Promoters, Conserved, Super Enhancers, H3K4me3 peaks, Transcribed Enhancers, DHS peaks, TFBS, H3K4me1 peaks, H3K27ac peaks, Introns

Fold Enrichment

Significant (Bonferroni corrected) Nominally significant (P<0.05) Not significant

B

trait Glycemic Red Blood Cell

Fold Enrichment

Coding Conserved Islets K562

Static Annotations Stretch Enhancers

C

KLF14
LOC101927697
LOC101928448
LINC01214
PPARG
SYN2
IRS1
COBLL1
LYPLAL1
TCF7L2
QKI
RSPO3
VEGFA
FAM13A
ADCY5
TRAIP
PEPD
COL4A2
DNAH10
HDAC7
IGF1
PHTF2
GRB10
C6orf1
PDGFC
MACROD1
ITIH3

Locus name (closest gene)

Locus overlap with Stretch Enhancer
Overlap
No Overlap

Adipose Skeletal Muscle
Stretch Enhancer

D

Intronic region of COL4A2 gene on chr13

20 kb hg19

Fasting Insulin GWAS (lead + LD $r^2 > 0.8$)

ATAC-seq: Adipose, Skeletal Muscle, Islet, GM12878

Chromatin states: Adipose, Skeletal Muscle, HSMM, Islets, GM12878, Liver, HepG2, H1, K562, Huvec, Anterior Caudate

E

Inttergenic region on chr3, ~25kb from LINC01214

5 kb hg19

Fasting Insulin GWAS

Adipose, Skeletal Muscle, Islet, GM12878

Adipose, Skeletal Muscle, HSMM, Islets, GM12878, Liver, HepG2, H1, K562, Huvec, Anterior Caudate

Chromatin States

Active TSS
Weak/Flanking TSS
Bivalent/Poised TSS
Active Enhancer 1/2
Weak Enhancer
Genic Enhancer
Strong Transcription
Weak Transcription
Repressed Polycomb
Weak Repressed Polycomb
Quiescent/Low Signal

False discovery rate
< 0.05
< 0.20
≥ 0.20

Fasting Glucose

Fasting Insulin

HbA1c

-log10(p) for tissue enrichment

0
1
2
3
4
5

Antibody-Producing Cells
Antigen-Presenting Cells
Blood
Blood Cells
Blood Vessels
Cartilage
Central Nervous System
Connective Tissue
Connective Tissue Cells
Endocrine Cells
Endocrine Glands
Epithelial Cells
Epithelium
Erythroid Cells
Exocrine Glands
Eye
Gastrointestinal Tract
Genitalia
Germ Cells
Heart
Hematopoietic System
Immune System
Liver
Lung
Lymphoid Tissue
Membranes
Mouth
Muscle Cells
Muscles
Myeloid Cells
Neurosecretory Systems
Nose
Pancreas
Pharynx
Skeleton
Skin
Stem Cells
Urinary Tract

A

complete embryonic lethality during organogenesis
abnormal placenta labyrinth morphology
cytoplasmic vesicle membrane
Membrane trafficking
KIF5B PPI
TRIP11 PPI
Downstream signal transduction
Irs:related events
cellular response to peptide hormone stimulus
abnormal glucose homeostasis
Maturity onset diabetes of the young
regulation of peptide hormone secretion
response to glucose stimulus
regulation of cellular carbohydrate metabolic process
monosaccharide biosynthetic process
TSC22D1 PPI
SCIN PPI
monosaccharide transport
Toll like receptor 5 tlr5 cascade
posttranscriptional regulation of gene expression
PKN3 PPI
glycerophospholipid metabolic process
Circadian clock
response to inorganic substance
liver development
MS4A5 PPI
Diabetes pathways
BRF2 PPI
TSC22D1 PPI
PHB2 PPI
ILK PPI

B

transforming growth factor beta-activated receptor activity
Chronic myeloid leukemia
PDGFRA PPI
Irs:mediated signalling
SMAD6 PPI
SMAD3 PPI
kinase binding
SRMS PPI
cellular response to hormone stimulus
EP300 PPI
DVL1 PPI
complete embryonic lethality during organogenesis
sequence-specific DNA binding RNA polymerase II transcription factor activity
partial postnatal lethality
abnormal glucose homeostasis
WNT5B PPI
tube morphogenesis
mesenchymal cell development
persistent truncus arteriosis
ARNT2 PPI
positive regulation of kinase activity
urogenital system development
ossification
reproductive structure development
protein import into nucleus
female infertility
ERH PPI
leukemia
CDK6 PPI
positive regulation of rotein metabolic process
Acute myeloid leukemia
Signaling by erbb4
Colorectal cancer
Endometrial cancer
Activated tlr4 signalling
Prostate cancer
Non small cell lung cancer
CFLAR PPI
MAPK8 PPI
Chronic myeloid leukemia
Pancreatic cancer
Renal cell carcinoma
Pathways in cancer
Small cell lung cancer
decreased survivor rate
reduced fertility
protein heterodimerization activity
partial perinatal lethality
decreased birth weight
caveola
partial prenatal lethality
partial postnatal lethality
abnormal eye morphology
postnatal growth retardation
complete neonatal lethality

C

HSPA8 PPI
C12orf10 PPI
DDX3Y PPI
NCBP2 PPI
abnormal myocardial fiber morphology
ENSG00000137379 PPI
NCL PPI
cortical cytoskeleton
PRRC1 PPI
ENSG00000215292 PPI
CHERP PPI
CGN PPI
FLNA PPI
abnormal placenta labyrinth morphology
MYO1G PPI
actin filament bundle
MAPK1 PPI
FMNL1 PPI
cellular response to peptide hormone stimulus
abnormal glucose homeostasis
protein kinase binding
complete embryonic lethality during organogenesis
MARK2 PPI
focal adhesion
Integrin cell surface interactions
nuclear envelope
AP1S2 PPI
SRC PPI
Signaling by egfr in cancer
decreased circulating triglyceride level
regulation of lipid metabolic process
glucan metabolic process
RBM33 PPI
EZR PPI
CREBBP PPI
SHC1 PPI
Hemostasis
endocytosis
vacuole
CEP170 PPI
DNAJB12 PPI
E2F6 PPI
CASP8 PPI
Toll like receptor 10 tlr10 cascade
positive regulation of kinase activity
hydrolase activity, hydrolyzing O-glycosyl compounds
lipid catabolic process
Lipoprotein metabolism
FKBP5 PPI
Interleukin:3 5 and gm:csf signaling
STAT5B PPI
SYK PPI
binding, bridging
positive regulation of phosphorus metabolic process
monosaccharide binding
ENSG00000105216 PPI
abnormal macrophage physiology
JAK3 PPI
FCGR1A PPI
positive regulation of cellular protein metabolic process
decreased T cell proliferation
TBKBP1 PPI
induction of apoptosis
decreased IgG3 level
enlarged spleen
hemopoiesis
abnormal erythropoiesis
LIF PPI
porphyrin-containing compound metabolic process
erythrocyte homeostasis
decreased mean corpuscular volume

DEPICT GSE within-cluster minimum p-value FDR (<0.05)
0.00374
3.32×10^{-10}
Pearson correlation (r)
r>0.3
r>0.5
r>0.7

Glossary Box

This study combined analyses of trait-associations across multiple correlated glycaemic traits and across multiple ancestries, which has presented challenges in our ability to apply commonly used terms with clarity. For this reason, we define below terms often used in the field with variable meaning, as well as definitions of new terms used in this study.

EA – the effect allele was that defined by METAL based on trans-ancestry FG results and aligned such that the same allele was kept as the effect allele across all ancestries and traits, irrespective of its allele frequency or effect size for that particular ancestry and trait, in this way the effect allele is not necessarily the trait-increasing allele.

Single-ancestry lead variant – variant with the smallest p-value amongst all with $P < 5x10^{-8}$, within a 1Mb region, based on analysis of a single trait in a single ancestry.

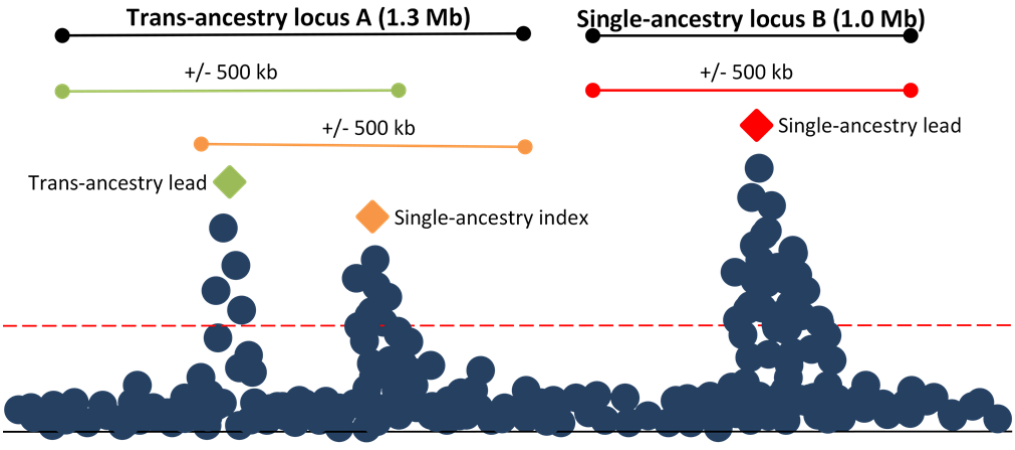
Single-ancestry index variants – variants identified by GCTA analysis of each autosome, and that appear to exert conditionally distinct effects on a given trait in a given ancestry ($P < 5x10^{-8}$). As defined, these include the single-ancestry lead variant.

Trans-ancestry lead variant – variant identified by trans-ethnic meta-analysis of a given trait that has the strongest association for that trait ($log_{10}BF > 6$, which is broadly equivalent to $P < 5x10^{-8}$) within a 1Mb region.

Single-ancestry locus – a 1Mb region centred on a single-ancestry lead variant which does not contain a lead variant identified in the trans-ancestry meta-analysis (i.e., does not contain a trans-ancestry lead variant).

Signal - a conditionally independent association between a trait and a set of variants in LD with each other and which is noted by the corresponding index variant.

Trans-ancestry locus – As we expected some genetic variants to influence multiple correlated traits and that functional variants would influence traits across multiple ancestries, we combined results across traits and across ancestries into multi-trait trans-ancestry loci. A **trans-ancestry locus** is a genomic interval that contains trans-ancestry trait-specific lead variants, with/out additional single-ancestry index variants, for one or more trait. This region is defined by starting at the telomere of each chromosome and selecting the first single-ancestry index variant or trans-ancestry lead variant for any trait. If other trans-ancestry lead variants or single-ancestry index variants mapped within 500kb of the first signal, then they were merged into the same locus. This process was repeated until there were no more signals within 500kb of the previous variant. A 500kb interval was added to the beginning of the first signal, and the end of the last signal to establish the final boundary of the trans-ancestry locus. As defined, a trans-ancestry locus may not have a single lead trans-ancestry variant, but may instead contain multiple trans-ancestry lead variants, one for each trait.



Locus diagram– In this diagram, trans-ancestry locus A contains a trans-ancestry lead variant for one glycaemic trait represented by the green diamond, and another single-ancestry index variant for another glycaemic trait represented by the orange triangle. Single-ancestry locus B contains a single-ancestry lead variant represented by the red square. The orange, green and red bars represent a +/- 500Kb window around the orange, green, and red variants, respectively. The black bars indicate the full locus window where trans-ancestry locus A contains trans-ancestry lead and single-ancestry index variants for two traits and single-ancestry locus B has a single-ancestry lead variant for a single trait.