Genetic drivers of heterogeneity in type 2 diabetes pathophysiology

Ken Suzuki[1,2,3,*], Konstantinos Hatzikotoulas[4,*,‡], Lorraine Southam[4,*], Henry J. Taylor[5,6,7,*], Xianyong Yin[8,9,*], Kim M. Lorenz[10,11,12,*], Ravi Mandla[13,14,*], Alicia Huerta-Chagoya[13], Giorgio E. M. Melloni[15], Stavroula Kanoni[16], Nigel W. Rayner[4], Ozvan Bocher[4], Ana Luiza Arruda[4], Kyuto Sonehara[3,17,18,19], Shinichi Namba[3], Simon S. K. Lee[20], Michael H. Preuss[20], Lauren E. Petty[21], Philip Schroeder[13,14], Brett Vanderwerff[9], Mart Kals[22], Fiona Bragg[23,24], Kuang Lin[23], Xiuqing Guo[25], Weihua Zhang[26,27], Jie Yao[25], Young Jin Kim[28], Mariaelisa Graff[29], Fumihiko Takeuchi[30], Jana Nano[31], Amel Lamri[32,33], Masahiro Nakatochi[34], Sanghoon Moon[28], Robert A. Scott[35], James P. Cook[36], Jung-Jin Lee[37], Ian Pan[38], Daniel Taliun[9], Esteban J. Parra[39], Jin-Fang Chai[40], Lawrence F. Bielak[41], Yasuharu Tabara[42], Yang Hai[25], Gudmar Thorleifsson[43], Niels Grarup[44], Tamar Sofer[45,46,47], Matthias Wuttke[48], Chloé Sarnowski[49], Christian Gieger[31,50,51], Darryl Nousome[52], Stella Trompet[53,54], Soo-Heon Kwak[55], Jirong Long[56], Meng Sun[57], Lin Tong[58], Wei-Min Chen[59], Suraj S. Nongmaithem[60], Raymond Noordam[54], Victor J. Y. Lim[40], Claudia H. T. Tam[61,62], Yoonjung Yoonie Joo[63,64], Chien-Hsiun Chen[65], Laura M. Raffield[66], Bram Peter Prins[67], Aude Nicolas[68], Lisa R. Yanek[69], Guanjie Chen[70], Jennifer A. Brody[71], Edmond Kabagambe[56,72], Ping An[73], Anny H. Xiang[74], Hyeok Sun Choi[75], Brian E. Cade[46,76], Jingyi Tan[25], K. Alaine Broadaway[66], Alice Williamson[35,77], Zoha Kamali[78,79], Jinrui Cui[80], Manonanthini Thangam[81], Linda S. Adair[82], Adebowale Adeyemo[70], Carlos A. Aguilar-Salinas[83], Tarunveer S. Ahluwalia[84,85], Sonia S. Anand[32,33,86], Alain Bertoni[87], Jette Bork-Jensen[44], Ivan Brandslund[88,89], Thomas A. Buchanan[90], Charles F. Burant[91], Adam S. Butterworth[6,7,92,93,94], Mickaël Canouil[95,96], Juliana C. N. Chan[61,62,97,98], Li-Ching Chang[65], Miao-Li Chee[99], Ji Chen[100,101], Shyh-Huei Chen[102], Yuan-Tsong Chen[65], Zhengming Chen[23,24], Lee-Ming Chuang[103,104], Mary Cushman[105], John Danesh[6,7,67,92,93,94], Swapan K. Das[106], H Janaka de Silva[107], George Dedoussis[108], Latchezar Dimitrov[109], Ayo P. Doumatey[70], Shufa Du[82,110], Qing Duan[66], Kai-Uwe Eckardt[111,112], Leslie S. Emery[113], Daniel S. Evans[114], Michele K. Evans[115], Krista Fischer[22,116], James S. Floyd[71], Ian Ford[117], Oscar H. Franco[118], Timothy M. Frayling[119], Barry I. Freedman[120], Pauline Genter[121], Hertzel C. Gerstein[32,33,86], Vilmantas Giedraitis[122], Clicerio González-Villalpando[123], Maria Elena González-Villalpando[123], Penny Gordon-Larsen[82,110], Myron Gross[124], Lindsay A. Guare[125], Sophie Hackinger[67], Liisa Hakaste[126,127], Sohee Han[28], Andrew T. Hattersley[128], Christian Herder[50,129,130], Momoko Horikoshi[131], Annie-Green Howard[110,132], Willa Hsueh[133], Mengna Huang[38,134], Wei Huang[135], Yi-Jen Hung[136,137], Mi Yeong Hwang[138], Chii-Min Hwu[139,140], Sahoko Ichihara[141], Mohammad Arfan Ikram[118], Martin Ingelsson[122], Md. Tariqul Islam[142], Masato Isono[30], Hye-Mi Jang[138], Farzana Jasmine[58], Guozhi Jiang[61,62], Jost B. Jonas[143], Torben Jørgensen[144,145,146], Frederick K. Kamanu[15], Fouad R. Kandeel[147], Anuradhani Kasturiratne[148], Tomohiro Katsuya[149,150], Varinderpal Kaur[14], Takahisa Kawaguchi[42], Jacob M. Keaton[5,56,109], Abel N. Kho[151,152], Chiea-Chuen Khor[153], Muhammad G. Kibriya[58], Duk-Hwan Kim[154], Florian Kronenberg[155], Johanna Kuusisto[156], Kristi Läll[22], Leslie A. Lange[157], Kyung Min Lee[158,159], Myung-Shik Lee[160,161], Nanette R. Lee[162], Aaron Leong[163,164], Liming Li[165,166], Yun Li[66], Ruifang Li-Gao[167], Symen Ligthart[118], Cecilia M. Lindgren[168,169,170], Allan Linneberg[144,171], Ching-Ti Liu[172], Jianjun Liu[153,173], Adam E. Locke[174,175,297], Tin Louie[113], Jian'an Luan[35], Andrea O. Luk[61,62], Xi Luo[176], Jun Lv[165,166], Julie A. Lynch[158,159], Valeriya Lyssenko[177,178], Shiro Maeda[131,179,180], Vasiliki Mamakou[181], Sohail Rafik Mansuri[60], Koichi Matsuda[182], Thomas Meitinger[183,184,185], Olle Melander[81], Andres Metspalu[22], Huan Mo[5], Andrew D. Morris[186], Filipe A. Moura[15], Jerry L.

47 Nadler[187], Michael A. Nalls[68,188,189], Uma Nayak[59], Ioanna Ntalla[16], Yukinori
48 Okada[3,17,18,19,190,191], Lorena Orozco[192], Sanjay R. Patel[193], Snehal Patil[9], Pei Pei[166], Mark A.
49 Pereira[194], Annette Peters[31,50,185,195], Fraser J. Pirie[196], Hannah G. Polikowsky[21], Bianca
50 Porneala[164], Gauri Prasad[197,198], Laura J. Rasmussen-Torvik[199], Alexander P. Reiner[200],
51 Michael Roden[50,129,130], Rebecca Rohde[29], Katheryn Roll[25], Charumathi Sabanayagam[99,201,202],
52 Kevin Sandow[25], Alagu Sankareswaran[60], Naveed Sattar[203], Sebastian Schönherr[155],
53 Mohammad Shahriar[58], Botong Shen[115], Jinxiu Shi[135], Dong Mun Shin[138], Nobuhiro Shojima[2],
54 Jennifer A. Smith[41,204], Wing Yee So[61,98], Alena Stančáková[156], Valgerdur Steinthorsdottir[43],
55 Adrienne M. Stilp[113], Konstantin Strauch[205,206,207], Kent D. Taylor[25], Barbara Thorand[31,50],
56 Unnur Thorsteinsdottir[43,208], Brian Tomlinson[61,209], Tam C. Tran[5], Fuu-Jen Tsai[210], Jaakko
57 Tuomilehto[211,212,213,214], Teresa Tusie-Luna[215,216], Miriam S Udler[13,14,163], Adan Valladares-
58 Salgado[217], Rob M. van Dam[40,173], Jan B. van Klinken[218,219,220], Rohit Varma[221], Niels Wacher-
59 Rodarte[222], Eleanor Wheeler[35], Ananda R. Wickremasinghe[148], Ko Willems van Dijk[218,219,223],
60 Daniel R. Witte[224,225], Chittaranjan S. Yajnik[226], Ken Yamamoto[227], Kenichi Yamamoto[3,190,228],
61 Kyungheon Yoon[138], Canqing Yu[165,166], Jian-Min Yuan[229,230], Salim Yusuf[32,33,86], Matthew
62 Zawistowski[9], Liang Zhang[99], Wei Zheng[56], VA Million Veteran Program, AMED GRIFIN
63 Diabetes Initiative Japan, Biobank Japan Project, Penn Medicine BioBank, Regeneron
64 Genetics Center, Genes & Health Research Team, eMERGE Consortium, International
65 Consortium of Blood Pressure (ICBP), Meta-Analyses of Glucose and Insulin-Related Traits
66 Consortium (MAGIC), Leslie J. Raffel[231], Michiya Igase[232], Eli Ipp[121], Susan Redline[46,76,233],
67 Yoon Shin Cho[75], Lars Lind[234], Michael A. Province[73], Myriam Fornage[235], Craig L. Hanis[236],
68 Erik Ingelsson[237,238], Alan B. Zonderman[115], Bruce M. Psaty[71,239,240], Ya-Xing Wang[241], Charles
69 N. Rotimi[70], Diane M. Becker[69], Fumihiko Matsuda[42], Yongmei Liu[87,242], Mitsuhiro Yokota[243],
70 Sharon L. R. Kardia[41], Patricia A. Peyser[41], James S. Pankow[194], James C. Engert[244,245], Amélie
71 Bonnefond[95,96,246], Philippe Froguel[95,96,246], James G. Wilson[247], Wayne H. H. Sheu[137,140,248]
72 Jer-Yuarn Wu[65], M Geoffrey Hayes[249,250,251], Ronald C. W. Ma[61,62,97,98], Tien-Yin Wong[99,201,202],
73 Dennis O. Mook-Kanamori[167], Tiinamaija Tuomi[81,126,127,252], Giriraj R. Chandak[60,298], Francis S.
74 Collins[5], Dwaipayan Bharadwaj[253], Guillaume Paré[33,254], Michèle M. Sale[59,300], Habibul
75 Ahsan[58], Ayesha A. Motala[196], Xiao-Ou Shu[56], Kyong-Soo Park[55,255], J. Wouter Jukema[53,256],
76 Miguel Cruz[217], Yii-Der Ida Chen[25], Stephen S. Rich[257], Roberta McKean-Cowdin[52], Harald
77 Grallert[31,50,258], Ching-Yu Cheng[99,201,202], Mohsen Ghanbari [118], E-Shyong Tai[40,173,259], Josee
78 Dupuis[172,260] Norihiro Kato[30], Markku Laakso[156], Anna Köttgen[48], Woon-Puay Koh[261,262],
79 Donald W Bowden[109,263,264], Colin N. A. Palmer[265], Jaspal S. Kooner[27,266,267,268], Charles
80 Kooperberg[200], Simin Liu[38,134,269], Kari E. North[29], Danish Saleheen[270,271,272], Torben Hansen[44],
81 Oluf Pedersen[44], Nicholas J. Wareham[186], Juyoung Lee[138], Bong-Jo Kim[138], Iona Y.
82 Millwood[23,24], Robin G. Walters[23,24], Kari Stefansson[43,208], Emma Ahlqvist[81], Mark O.
83 Goodarzi[80], Karen L. Mohlke[66], Claudia Langenberg[35,273,274], Christopher A. Haiman[275], Ruth J.
84 F. Loos[20,44,276], Jose C. Florez[13,14,163] Daniel J. Rader[12,277,278,279], Marylyn D. Ritchie[12,280,281],
85 Sebastian Zöllner[9,282], Reedik Mägi[22], Nicholas A. Marston[15], Christian T. Ruff[15], David A. van
86 Heel[283], Sarah Finer[284], Joshua C. Denny[5,285], Toshimasa Yamauchi[2], Takashi Kadowaki[2,286],
87 John C. Chambers[26,27,266,287], Maggie C. Y. Ng[109,264,288], Xueling Sim[40], Jennifer E. Below[21],
88 Philip S. Tsao[237,289,290], Kyong-Mi Chang[10,291], Mark I. McCarthy[168,292,293,299,†], James B.
89 Meigs[13,163,164,†], Anubha Mahajan[168,292,299,†], Cassandra N. Spracklen[294,†], Josep M.
90 Mercader[13,14,76,†], Michael Boehnke[9,†], Jerome I. Rotter[25,†], Marijana Vujkovic[10,291,295,†],
91 Benjamin F. Voight[10,11,12,278,†,‡], Andrew P. Morris[1,4,22,†,‡], Eleftheria Zeggini[4,296,†,‡].
92

93 [1]Centre for Genetics and Genomics Versus Arthritis, Centre for Musculoskeletal Research, Division of
94 Musculoskeletal and Dermatological Sciences, The University of Manchester, Manchester, UK. [2]Department of
95 Diabetes and Metabolic Diseases, Graduate School of Medicine, The University of Tokyo, Tokyo, Japan.
96 [3]Department of Statistical Genetics, Osaka University Graduate School of Medicine, Suita, Japan. [4]Institute of
97 Translational Genomics, Helmholtz Zentrum München, German Research Center for Environmental Health,
98 Neuherberg, Germany. [5]Center for Precision Health Research, National Human Genome Research Institute,
99 National Institutes of Health, Bethesda, MD, USA. [6]British Heart Foundation Cardiovascular Epidemiology Unit,
100 Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK. [7]Heart and Lung
101 Research Institute, University of Cambridge, Cambridge, UK. [8]Department of Epidemiology, School of Public
102 Health, Nanjing Medical University, Nanjing City, China. [9]Department of Biostatistics and Center for Statistical
103 Genetics, University of Michigan, Ann Arbor, MI, USA. [10]Corporal Michael J Crescenz VA Medical Center,
104 Philadelphia, PA, USA. [11]Department of Systems Pharmacology and Translational Therapeutics, University of
105 Pennsylvania Perelman School of Medicine, Philadelphia, PA, USA. [12]Department of Genetics, University of
106 Pennsylvania Perelman School of Medicine, Philadelphia, PA, USA. [13]Programs in Metabolism and Medical and
107 Population Genetics, Broad Institute of Harvard and MIT, Cambridge, MA, USA. [14]Diabetes Unit and Center for
108 Genomic Medicine, Massachusetts General Hospital, Boston, MA, USA. [15]TIMI Study Group, Division of
109 Cardiovascular Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA. [16]William
110 Harvey Research Institute, Barts and the London School of Medicine and Dentistry, Queen Mary University of
111 London, London, UK. [17]Department of Genome Informatics, Graduate School of Medicine, The University of
112 Tokyo, Tokyo, Japan. [18]Integrated Frontier Research for Medical Science Division, Institute for Open and
113 Transdisciplinary Research Initiatives, Osaka University, Suita, Japan. [19]Laboratory for Systems Genetics, RIKEN
114 Center for Integrative Medical Sciences, Kanagawa, Japan. [20]The Charles Bronfman Institute for Personalized
115 Medicine, Icahn School of Medicine at Mount Sinai, New York, NY, USA. [21]Department of Medicine, Vanderbilt
116 University Medical Center, Nashville, TN, USA. [22]Estonian Genome Centre, Institute of Genomics, University of
117 Tartu, Tartu, Estonia. [23]Nuffield Department of Population Health, University of Oxford, Oxford, UK. [24]Medical
118 Research Council Population Health Research Unit, University of Oxford, Oxford, UK. [25]The Institute for
119 Translational Genomics and Population Sciences, Department of Pediatrics, The Lundquist Institute for
120 Biomedical Innovation (formerly Los Angeles Biomedical Research Institute) at Harbor-UCLA Medical Center,
121 Torrance, CA, USA. [26]Department of Epidemiology and Biostatistics, Imperial College London, London, UK.
122 [27]Department of Cardiology, Ealing Hosptial, London NorthWest Healthcare NHS Trust, Middlesex, UK.
123 [28]Division of Genome Science, Department of Precision Medicine, National Institute of Health, Cheongju-si,
124 South Korea. [29]Department of Epidemiology, Gillings School of Global Public Health, University of North
125 Carolina at Chapel Hill, Chapel Hill, NC, USA. [30]Department of Gene Diagnostics and Therapeutics, Research
126 Institute, National Center for Global Health and Medicine, Tokyo, Japan. [31]Institute of Epidemiology, Helmholtz
127 Zentrum München, German Research Center for Environmental Health, Neuherberg, Germany. [32]Department
128 of Medicine, McMaster University, Hamilton, ON, Canada. [33]Population Health Research Institute, Hamilton
129 Health Sciences and McMaster University, Hamilton, ON, Canada. [34]Public Health Informatics Unit,
130 Department of Integrated Health Sciences, Nagoya University Graduate School of Medicine, Nagoya, Japan.
131 [35]MRC Epidemiology Unit, Institute of Metabolic Science, University of Cambridge School of Clinical Medicine,
132 Cambridge, UK. [36]Department of Health Data Science, University of Liverpool, Liverpool, UK. [37]Division of
133 Translational Medicine and Human Genetics, University of Pennsylvania, Philadelphia, PA, USA. [38]Department
134 of Epidemiology, Brown University School of Public Health, Providence, RI, USA. [39]Department of
135 Anthropology, University of Toronto at Mississauga, Mississauga, ON, Canada. [40]Saw Swee Hock School of
136 Public Health, National University of Singapore and National University Health System, Singapore, Singapore.
137 [41]Department of Epidemiology, School of Public Health, University of Michigan, Ann Arbor, MI, USA. [42]Center
138 for Genomic Medicine, Kyoto University Graduate School of Medicine, Kyoto, Japan. [43]deCODE Genetics,
139 Amgen Inc., Reykjavik, Iceland. [44]Novo Nordisk Foundation Center for Basic Metabolic Research, Faculty of
140 Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark. [45]Department of Biostatistics,
141 Harvard University, Boston, MA, USA. [46]Division of Sleep and Circadian Disorders, Brigham and Women's
142 Hospital, Boston, MA, USA. [47]Department of Medicine, Harvard University, Boston, MA, USA. [48]Institute of
143 Genetic Epidemiology, Department of Data Driven Medicine, Faculty of Medicine and Medical Center,
144 University of Freiburg, Freiburg, Germany. [49]Department of Epidemiology, Human Genetics, and
145 Environmental Sciences, The University of Texas Health Science Center at Houston School of Public Health,
146 Houston, TX, USA. [50]German Center for Diabetes Research (DZD), Neuherberg, Germany. [51]Research Unit of
147 Molecular Epidemiology, Helmholtz Zentrum München, German Research Center for Environmental Health,
148 Neuherberg, Germany. [52]Department of Population and Public Health Sciences, Keck School of Medicine of
149 USC, Los Angeles, CA, USA. [53]Department of Cardiology, Leiden University Medical Center, Leiden, The

150 Netherlands. [54]Section of Gerontology and Geriatrics, Department of Internal Medicine, Leiden University
151 Medical Center, Leiden, The Netherlands. [55]Department of Internal Medicine, Seoul National University
152 Hospital, Seoul, South Korea. [56]Division of Epidemiology, Department of Medicine, Institute for Medicine and
153 Public Health, Vanderbilt Genetics Institute, Vanderbilt University Medical Center, Nashville, TN, USA.
154 [57]Nuffield Department of Surgical Sciences, University of Oxford, Oxford, UK. [58]Institute for Population and
155 Precision Health (IPPH), Biological Sciences Division, The University of Chicago, Chicago, IL, USA. [59]Department
156 of Public Health Sciences and Center for Public Health Genomics, University of Virginia School of Medicine,
157 Charlottesville, VA, USA. [60]Genomic Research on Complex Diseases (GRC-Group), CSIR-Centre for Cellular and
158 Molecular Biology (CSIR-CCMB), Hyderabad, India. [61]Department of Medicine and Therapeutics, The Chinese
159 University of Hong Kong, Hong Kong, China. [62]Chinese University of Hong Kong-Shanghai Jiao Tong University
160 Joint Research Centre in Diabetes Genomics and Precision Medicine, The Chinese University of Hong Kong,
161 Hong Kong, China. [63]Samsung Advanced Institute for Health Sciences & Technology (SAIHST), Sungkyunkwan
162 University, Samsung Medical Center, Seoul, South Korea. [64]Department of Preventive Medicine, Northwestern
163 University Feinberg School of Medicine, Chicago, IL, USA. [65]Institute of Biomedical Sciences, Academia Sinica,
164 Taipei, Taiwan. [66]Department of Genetics, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA.
165 [67]Department of Human Genetics, Wellcome Sanger Institute, Wellcome Genome Campus, Hinxton, UK.
166 [68]Laboratory of Neurogenetics, National Institute on Aging, National Institutes of Health, Bethesda, MD, USA.
167 [69]Department of Medicine, Johns Hopkins University School of Medicine, Baltimore, MD, USA. [70]Center for
168 Research on Genomics and Global Health, National Human Genome Research Institute, National Institutes of
169 Health, Bethesda, MD, USA. [71]Cardiovascular Health Research Unit, Department of Medicine, University of
170 Washington, Seattle, WA, USA. [72]Division of Academics, Ochsner Health, New Orleans, LA, USA. [73]Division of
171 Statistical Genomics, Washington University School of Medicine, St. Louis, MO, USA. [74]Department of Research
172 & Evaluation, Division of Biostatistics Research, Kaiser Permanente of Southern California, Pasadena, CA, USA.
173 [75]Department of Biomedical Science, Hallym University, Chuncheon, South Korea. [76]Harvard Medical School,
174 Boston, MA, USA. [77]Metabolic Research Laboratories, Wellcome Trust-Medical Research Council Institute of
175 Metabolic Science, Department of Clinical Biochemistry, University of Cambridge, Cambridge, UK.
176 [78]Department of Epidemiology, University of Groningen, University Medical Centre Groningen, Groningen, The
177 Netherlands. [79]Department of Bioinformatics, Isfahan University of Medical Sciences, Isfahan, Iran.
178 [80]Department of Medicine, Division of Endocrinology, Diabetes and Metabolism, Cedars-Sinai Medical Center,
179 Los Angeles, CA, USA. [81]Lund University Diabetes Centre, Department of Clinical Sciences, Lund University,
180 Skåne University Hospital, Malmö, Sweden. [82]Department of Nutrition, Gillings School of Global Public Health,
181 University of North Carolina at Chapel Hill, Chapel Hill, NC, USA. [83]Unidad de Investigación en Enfermedades
182 Metabólicas and Departamento de Endocrinología y Metabolismo, Instituto Nacional de Ciencias Médicas y
183 Nutrición Salvador Zubirán, Mexico City, Mexico. [84]Steno Diabetes Center Copenhagen, Herlev, Denmark. [85]The
184 Bioinformatics Center, Department of Biology, University of Copenhagen, Copenhagen, Denmark.
185 [86]Department of Health Research Methods, Evidence, and Impact, McMaster University, Hamilton, ON,
186 Canada. [87]Department of Epidemiology and Prevention, Division of Public Health Sciences, Wake Forest School
187 of Medicine, Winston-Salem, NC, USA. [88]Institute of Regional Health Research, University of Southern
188 Denmark, Odense, Denmark. [89]Department of Clinical Biochemistry, Vejle Hospital, Vejle, Denmark.
189 [90]Department of Medicine, Division of Endocrinology and Diabetes, Keck School of Medicine of USC, Los
190 Angeles, CA, USA. [91]Department of Internal Medicine, University of Michigan, Ann Arbor, MI, USA. [92]British
191 Heart Foundation Centre of Research Excellence, School of Clinical Medicine, Addenbrooke's Hospital,
192 University of Cambridge, Cambridge, UK. [93]Health Data Research UK Cambridge, Wellcome Genome Campus
193 and University of Cambridge, Hinxton, UK. [94]National Institute for Health and Care Research (NIHR) Blood and
194 Transplant Unit (BTRU) in Donor Health and Behaviour, Heart and Lung Research Institute, University of
195 Cambridge, Cambridge, UK. [95]Inserm U1283, CNRS UMR 8199, European Genomic Institute for Diabetes (EGID),
196 Institut Pasteur de Lille, Lille University Hospital, Lille, France. [96]University of Lille, Lille, France. [97]Li Ka Shing
197 Institute of Health Sciences, The Chinese University of Hong Kong, Hong Kong, China. [98]Hong Kong Institute of
198 Diabetes and Obesity, The Chinese University of Hong Kong, Hong Kong, China. [99]Singapore Eye Research
199 Institute, Singapore National Eye Centre, Singapore, Singapore. [100]Exeter Centre of Excellence in Diabetes
200 (ExCEeD), Exeter Medical School, University of Exeter, Exeter, UK. [101]Wellcome Sanger Institute, Wellcome
201 Genome Campus, Hinxton, UK. [102]Department of Biostatistics and Data Science, Wake Forest School of
202 Medicine, Winston-Salem, NC, USA. [103]Division of Endocrinology and Metabolism, Department of Internal
203 Medicine, National Taiwan University Hospital, Taipei, Taiwan. [104]Institute of Epidemiology and Preventive
204 Medicine, National Taiwan University, Taipei, Taiwan. [105]Department of Medicine, University of Vermont,
205 Colchester, VT, USA. [106]Section on Endocrinology and Metabolism, Department of Internal Medicine, Wake
206 Forest School of Medicine, Winston-Salem, NC, USA. [107]Department of Medicine, Faculty of Medicine,

207 University of Kelaniya, Ragama, Sri Lanka. [108]Department of Nutrition and Dietetics, Harokopio University of
208 Athens, Athens, Greece. [109]Center for Genomics and Personalized Medicine Research, Wake Forest School of
209 Medicine, Winston-Salem, NC, USA. [110]Carolina Population Center, University of North Carolina at Chapel Hill,
210 Chapel Hill, NC, USA. [111]Department of Nephrology and Medical Intensive Care Medicine, Charité
211 Universitätsmedizin Berlin, Berlin, Germany. [112]Department of Nephrology and Hypertension, Friedrich-
212 Alexander-Universität Erlangen-Nürnberg, Erlangen, Germany. [113]Department of Biostatistics, University of
213 Washington, Seattle, WA, USA. [114]California Pacific Medical Center Research Institute, San Francisco, CA, USA.
214 [115]Laboratory of Epidemiology and Population Sciences, National Institute on Aging, National Institutes of
215 Health, Baltimore, MD, USA. [116]Institute of Mathematics and Statistics, University of Tartu, Tartu, Estonia.
216 [117]Robertson Centre for Biostatistics, University of Glasgow, Glasgow, UK. [118]Department of Epidemiology,
217 Erasmus MC University Medical Center, Rotterdam, The Netherlands. [119]Genetics of Complex Traits, University
218 of Exeter Medical School, University of Exeter, Exeter, UK. [120]Department of Internal Medicine, Wake Forest
219 School of Medicine, Winston-Salem, NC, USA. [121]Department of Medicine, Division of Endocrinology and
220 Metabolism, Lundquist Research Institute at Harbor-UCLA Medical Center, Torrance, CA, USA. [122]Department
221 of Public Health and Caring Sciences, Uppsala University, Uppsala, Sweden. [123]Centro de Estudios en Diabetes,
222 Unidad de Investigacion en Diabetes y Riesgo Cardiovascular, Centro de Investigacion en Salud Poblacional,
223 Instituto Nacional de Salud Publica, Mexico City, Mexico. [124]Department of Laboratory Medicine and
224 Pathology, University of Minnesota, Minneapolis, MN, USA. [125]Genomics and Computational Biology Graduate
225 Group, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA, USA. [126]Institute for
226 Molecular Medicine Finland (FIMM), University of Helsinki, Helsinki, Finland. [127]Folkhalsan Research Center,
227 Helsinki, Finland. [128]University of Exeter Medical School, University of Exeter, Exeter, UK. [129]Institute for Clinical
228 Diabetology, German Diabetes Center, Leibniz Center for Diabetes Research at Heinrich Heine University
229 Düsseldorf, Düsseldorf, Germany. [130]Department of Endocrinology and Diabetology, Medical Faculty and
230 University Hospital Düsseldorf, Heinrich Heine University Düsseldorf, Düsseldorf, Germany. [131]Laboratory for
231 Genomics of Diabetes and Metabolism, RIKEN Center for Integrative Medical Sciences, Kanagawa, Japan.
232 [132]Department of Biostatistics, Gillings School of Global Public Health, University of North Carolina at Chapel
233 Hill, Chapel Hill, NC, USA. [133]Department of Internal Medicine, Diabetes and Metabolism Research Center, The
234 Ohio State University Wexner Medical Center, Columbus, OH, USA. [134]Center for Global Cardiometabolic
235 Health, Brown University, Providence, RI, USA. [135]Shanghai-MOST Key Laboratory of Health and Disease
236 Genomics, Shanghai Institute for Biomedical and Pharmaceutical Technologies, Shanghai, China. [136]Division of
237 Endocrine and Metabolism, Tri-Service General Hospital Songshan Branch, Taipei, Taiwan. [137]School of
238 Medicine, National Defense Medical Center, Taipei, Taiwan. [138]Division of Genome Science, Department of
239 Precision Medicine, National Institute of Health, Cheongju-si, Korea. [139]Section of Endocrinology and
240 Metabolism, Department of Medicine, Taipei Veterans General Hospital, Taipei, Taiwan. [140]School of Medicine,
241 National Yang Ming Chiao Tung University, Taipei, Taiwan. [141]Department of Environmental and Preventive
242 Medicine, Jichi Medical University School of Medicine, Shimotsuke, Japan. [142]University of Chicago Research
243 Bangladesh, Dhaka, Bangladesh. [143]Institute of Molecular and Clinical Ophthalmology Basel, Basel, Switzerland.
244 [144]Center for Clinical Research and Prevention, Bispebjerg and Frederiksberg Hospital, Frederiksberg, Denmark.
245 [145]Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark. [146]Faculty of
246 Medicine, Aalborg University, Aalborg, Denmark. [147]Department of Clinical Diabetes, Endocrinology and
247 Metabolism, Department of Translational Research and Cellular Therapeutics, City of Hope, Duarte, CA, USA.
248 [148]Department of Public Health, Faculty of Medicine, University of Kelaniya, Ragama, Sri Lanka. [149]Department
249 of Clinical Gene Therapy, Osaka University Graduate School of Medicine, Osaka, Japan. [150]Department of
250 Geriatric and General Medicine, Graduate School of Medicine, Osaka University, Osaka, Japan. [151]Division of
251 General Internal Medicine and Geriatrics, Department of Medicine, Northwestern University Feinberg School
252 of Medicine, Chicago, IL, USA. [152]Center for Health Information Partnerships, Institute for Public Health and
253 Medicine, Northwestern University Feinberg School of Medicine, Chicago, IL, USA. [153]Genome Institute of
254 Singapore, Agency for Science, Technology and Research, Singapore, Singapore. [154]Department of Molecular
255 Cell Biology, Sungkyunkwan University School of Medicine, Suwon, South Korea. [155]Institute of Genetic
256 Epidemiology, Medical University of Innsbruck, Innsbruck, Austria. [156]Institute of Clinical Medicine, Internal
257 Medicine, University of Eastern Finland and Kuopio University Hospital, Kuopio, Finland. [157]Department of
258 Medicine, University of Colorado Denver, Anschutz Medical Campus, Aurora, CO, USA. [158]VA Salt Lake City
259 Health Care System, Salt Lake City, UT, USA. [159]Department of Internal Medicine, University of Utah School of
260 Medicine, Salt Lake City, UT, USA. [160]Soochunhyang Institute of Medi-bio Science and Division of
261 Endocrinology, Department of Internal Medicine, Soochunhyang University College of Medicine, Cheonan,
262 South Korea. [161]Department of Medicine, Samsung Medical Center, Sungkyunkwan University School of
263 Medicine, Seoul, South Korea. [162]USC-Office of Population Studies Foundation Inc., University of San Carlos,

264 Cebu City, Philippines. [163]Department of Medicine, Harvard Medical School, Boston, MA, USA. [164]Division of
265 General Internal Medicine, Massachusetts General Hospital, Boston, MA, USA. [165]Department of Epidemiology
266 and Biostatistics, School of Public Health, Peking University, Beijing, China. [166]Peking University Center for
267 Public Health and Epidemic Preparedness and Response, Beijing, China. [167]Department of Clinical
268 Epidemiology, Leiden University Medical Center, Leiden, The Netherlands. [168]Wellcome Centre for Human
269 Genetics, Nuffield Department of Medicine, University of Oxford, Oxford, UK. [169]Program in Medical and
270 Population Genetics, Broad Institute, Cambridge, MA, USA. [170]Big Data Institute, Li Ka Shing Centre for Health
271 Information and Discovery, University of Oxford, Oxford, UK. [171]Department of Clinical Medicine, Faculty of
272 Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark. [172]Department of Biostatistics,
273 Boston University School of Public Health, Boston, MA, USA. [173]Department of Medicine, Yong Loo Lin School
274 of Medicine, National University of Singapore and National University Health System, Singapore, Singapore.
275 [174]McDonnell Genome Institute, Washington University School of Medicine, St. Louis, MO, USA. [175]Department
276 of Medicine, Division of Genomics and Bioinformatics, Washington University School of Medicine, St. Louis,
277 MO, USA. [176]Department of Biostatistics and Data Science, The University of Texas Health Science Center at
278 Houston School of Public Health, Houston, TX, USA. [177]Department of Clinical Sciences, Diabetes and
279 Endocrinology, Lund University Diabetes Centre, Malmö, Sweden. [178]Department of Clinical Science, Center for
280 Diabetes Research, University of Bergen, Bergen, Norway. [179]Department of Advanced Genomic and
281 Laboratory Medicine, Graduate School of Medicine, University of the Ryukyus, Okinawa, Japan. [180]Division of
282 Clinical Laboratory and Blood Transfusion, University of the Ryukyus Hospital, Okinawa, Japan. [181]Dromokaiteio
283 Psychiatric Hospital, National and Kapodistrian University of Athens, Athens, Greece. [182]Computational Biology
284 and Medical Sciences, Graduate School of Frontier Sciences, The University of Tokyo, Tokyo, Japan. [183]Institute
285 of Human Genetics, Helmholtz Zentrum München, German Research Center for Environmental Health,
286 Neuherberg, Germany. [184]Institute of Human Genetics, Technical University Munich, Munich, Germany.
287 [185]German Centre for Cardiovascular Research (DZHK), Partner Site Munich Heart Alliance, Munich, Germany.
288 [186]The Usher Institute to the Population Health Sciences and Informatics, University of Edinburgh, Edinburgh,
289 UK. [187]Department of Medicine and Pharmacology, New York Medical College, Valhalla, NY, USA. [188]Data
290 Tecnica International LLC, Glen Echo, MD, USA. [189]Center for Alzheimer's and Related Dementias, National
291 Institutes of Health, Bethesda, MD, USA. [190]Laboratory of Statistical Immunology, Immunology Frontier
292 Research Center (WPI-IFReC), Osaka University, Suita, Japan. [191]Premium Research Institute for Human
293 Metaverse Medicine (WPI-PRIMe), Osaka University, Suita, Japan. [192]Instituto Nacional de Medicina Genómica,
294 Mexico City, Mexico. [193]Division of Pulmonary, Allergy, and Critical Care Medicine, Department of Medicine,
295 University of Pittsburgh, Pittsburgh, PA, USA. [194]Division of Epidemiology and Community Health, School of
296 Public Health, University of Minnesota, Minneapolis, MN, USA. [195]Institute for Medical Information Processing,
297 Biometry and Epidemiology, Ludwig-Maximilians-Universität München, Munich, Germany. [196]Department of
298 Diabetes and Endocrinology, Nelson R Mandela School of Medicine, College of Health Sciences, University of
299 KwaZulu-Natal, Durban, South Africa. [197]Academy of Scientific and Innovative Research, CSIR-Human Resource
300 Development Campus, Ghaziabad, India. [198]Genomics and Molecular Medicine Unit, CSIR-Institute of Genomics
301 and Integrative Biology, New Delhi, India. [199]Department of Preventive Medicine, Northwestern University
302 Feinberg School of Medicine, Chicago, IL, USA. [200]Fred Hutchinson Cancer Research Center, Seattle, WA, USA.
303 [201]Ophthalmology and Visual Sciences Academic Clinical Program (Eye ACP), Duke-NUS Medical School,
304 Singapore, Singapore. [202]Department of Ophthalmology, Yong Loo Lin School of Medicine, National University
305 of Singapore and National University Health System, Singapore, Singapore. [203]Institute of Cardiovascular and
306 Medical Sciences, University of Glasgow, Glasgow, UK. [204]Survey Research Center, Institute for Social Research,
307 University of Michigan, Ann Arbor, MI, USA. [205]Institute of Genetic Epidemiology, Helmholtz Zentrum
308 Munchen, German Research Center for Environmental Health, Neuherberg, Germany. [206]Institute for Medical
309 Biostatistics, Epidemiology, and Informatics (IMBEI), University Medical Center, Johannes Gutenberg
310 University, Mainz, Germany. [207]Chair of Genetic Epidemiology, Institute of Medical Information Processing,
311 Biometry, and Epidemiology, Faculty of Medicine, Ludwig-Maximilians-Universität München, Munich,
312 Germany. [208]Faculty of Medicine, University of Iceland, Reykjavik, Iceland. [209]Faculty of Medicine, Macau
313 University of Science and Technology, Macau, China. [210]Department of Medical Genetics and Medical
314 Research, China Medical University Hospital, Taichung, Taiwan. [211]Population Health Unit, Finnish Institute for
315 Health and Welfare, Helsinki, Finland, Finnish Institute for Health and Welfare, Helsinki, Finland. [212]National
316 School of Public Health, Madrid, Spain. [213]Department of Public Health, University of Helsinki, Helsinki, Finland.
317 [214]Diabetes Research Group, King Abdulaziz University, Jeddah, Saudi Arabia. [215]Unidad de Biología Molecular y
318 Medicina Genómica, Instituto Nacional de Ciencias Médicas y Nutrición Salvador Zubirán, Mexico City, Mexico.
319 [216]Departamento de Medicina Genómica y Toxiología Ambiental, Instituto de Investigaciones Biomédicas,
320 UNAM, Mexico City, Mexico. [217]Unidad de Investigacion Medica en Bioquimica, Hospital de Especialidades,

Centro Medico Nacional Siglo XXI, Instituto Mexicano del Seguro Social, Mexico City, Mexico. [218]Einthoven Laboratory for Experimental Vascular Medicine, Leiden University Medical Center, Leiden, The Netherlands. [219]Department of Human Genetics, Leiden University Medical Center, Leiden, The Netherlands. [220]Department of Clinical Chemistry, Laboratory of Genetic Metabolic Disease, Amsterdam University Medical Center, Amsterdam, The Netherlands. [221]Southern California Eye Institute, CHA Hollywood Presbyterian Hospital, Los Angeles, CA, USA. [222]Unidad de Investigación Médica en Epidemiologia Clinica, Hospital de Especialidades, Centro Medico Nacional Siglo XXI, Instituto Mexicano del Seguro Social, Mexico City, Mexico. [223]Department of Internal Medicine, Division of Endocrinology, Leiden University Medical Center, Leiden, The Netherlands. [224]Department of Public Health, Aarhus University, Aarhus, Denmark. [225]Danish Diabetes Academy, Odense, Denmark. [226]Diabetology Research Centre, King Edward Memorial Hospital and Research Centre, Pune, India. [227]Department of Medical Biochemistry, Kurume University School of Medicine, Kurume, Japan. [228]Department of Pediatrics, Osaka University Graduate School of Medicine, Suita, Japan. [229]Division of Cancer Control and Population Sciences, UPMC Hillman Cancer Center, University of Pittsburgh, Pittsburgh, PA, USA. [230]Department of Epidemiology, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, PA, USA. [231]Department of Pediatrics, Division of Genetic and Genomic Medicine, UCI Irvine School of Medicine, Irvine, CA, USA. [232]Department of Anti-Aging Medicine, Ehime University Graduate School of Medicine, Ehime, Japan. [233]Division of Pulmonary, Critical Care, and Sleep Medicine, Beth Israel Deaconess Medical Center, Boston, MA, USA. [234]Department of Medical Sciences, Uppsala University, Uppsala, Sweden. [235]Institute of Molecular Medicine, The University of Texas Health Science Center at Houston School of Public Health, Houston, TX, USA. [236]Human Genetics Center, University of Texas Health Science Center at Houston, Houston, TX, USA. [237]Department of Medicine, Stanford University School of Medicine, Stanford, CA, USA. [238]Department of Medical Sciences, Molecular Epidemiology and Science for Life Laboratory, Uppsala University, Uppsala, Sweden. [239]Department of Epidemiology, University of Washington, Seattle, WA, USA. [240]Department of Health Systems and Population Health, University of Washington, Seattle, WA, USA. [241]Beijing Institute of Ophthalmology, Ophthalmology and Visual Sciences Key Laboratory, Beijing Tongren Hospital, Capital Medical University, Beijing, China. [242]Department of Medicine, Division of Cardiology, Duke University School of Medicine, Durham, NC, USA. [243]Kurume University School of Medicine, Kurume, Japan. [244]Department of Medicine, McGill University, Montreal, QC, Canada. [245]Department of Human Genetics, McGill University, Montreal, QC, Canada. [246]Department of Metabolism, Digestion and Reproduction, Imperial College London, London, UK. [247]Division of Cardiovascular Medicine, Beth Israel Deaconess Medical Center, Boston, MA, USA. [248]Division of Endocrinology and Metabolism, Department of Medicine, Taichung Veterans General Hospital, Taichung, Taiwan. [249]Division of Endocrinology, Metabolism, and Molecular Medicine, Department of Medicine, Northwestern University Feinberg School of Medicine, Chicago, IL, USA. [250]Center for Genetic Medicine, Northwestern University Feinberg School of Medicine, Chicago, IL, USA. [251]Department of Anthropology, Northwestern University, Evanston, IL, USA. [252]Department of Endocrinology, Helsinki University Hospital, Helsinki, Finland. [253]Systems Genomics Laboratory, School of Biotechnology, Jawaharlal Nehru University, New Delhi, India. [254]Department of Pathology and Molecular Medicine, McMaster University, Hamilton, ON, Canada. [255]Department of Molecular Medicine and Biopharmaceutical Sciences, Graduate School of Convergence Science and Technology, Seoul National University, Seoul, South Korea. [256]Netherlands Heart Institute, Utrecht, The Netherlands. [257]Center for Public Health Genomics, University of Virginia School of Medicine, Charlottesville, VA, USA. [258]Research Unit of Molecular Epidemiology, Helmholtz Zentrum München, German Research Center for Environmental Health, Munich, Germany. [259]Duke-NUS Medical School, Singapore, Singapore. [260]Department of Epidemiology, Biostatistics and Occupational Health, McGill University, Montreal, QC, Canada. [261]Singapore Institute for Clinical Sciences, Agency for Science Technology and Research (A*STAR), Singapore, Singapore. [262]Healthy Longevity Translational Research Programme, Yong Loo Lin School of Medicine, National University of Singapore, Singapore, Singapore. [263]Center for Diabetes Research, Wake Forest School of Medicine, Winston-Salem, NC, USA. [264]Department of Biochemistry, Wake Forest School of Medicine, Winston-Salem, NC, USA. [265]Pat Macpherson Centre for Pharmacogenetics and Pharmacogenomics, University of Dundee, Dundee, UK. [266]Imperial College Healthcare NHS Trust, Imperial College London, London, UK. [267]MRC-PHE Centre for Environment and Health, Imperial College London, London, UK. [268]National Heart and Lung Institute, Imperial College London, London, UK. [269]Department of Medicine, Brown University Alpert School of Medicine, Providence, RI, USA. [270]Department of Medicine, Columbia University Irving Medical Center, New York, NY, USA. [271]Department of Cardiology, Columbia University Irving Medical Center, New York, NY, USA. [272]Center for Non-Communicable Diseases, Karachi, Pakistan. [273]Computational Medicine, Berlin Institute of Health at Charité–Universitätsmedizin, Berlin, Germany. [274]Precision Healthcare University Research Institute, Queen Mary University of London, London, UK. [275]Department of Preventive Medicine, Keck School of Medicine of USC, Los Angeles, CA, USA. [276]The Mindich

378 Child Health and Development Institute, Ichan School of Medicine at Mount Sinai, New York, NY, USA.
379 [277]Division of Translational Medicine and Therapeutics, Department of Medicine, University of Pennsylvania
380 Perelman School of Medicine, Philadelphia, PA, USA. [278]Institute for Translational Medicine and Therapeutics,
381 University of Pennsylvania Perelman School of Medicine, Philadelphia, PA, USA. [279]Department of Pediatrics,
382 University of Pennsylvania Perelman School of Medicine, Philadelphia, PA, USA. [280]Center for Precision
383 Medicine, University of Pennsylvania - Perelman School of Medicine, Philadelphia, PA, USA. [281]Institute for
384 Biomedical Informatics, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA, USA.
385 [282]Department of Psychiatry, University of Michigan, Ann Arbor, MI, USA. [283]Blizard Institute, Queen Mary
386 University of London, London, UK. [284]Institute for Population Health Sciences, Barts and the London School of
387 Medicine and Dentistry, Queen Mary University of London, London, UK. [285]All of Us Research Program,
388 National Institutes of Health, Bethesda, MD, USA. [286]Toranomon Hospital, Tokyo, Japan. [287]Lee Kong Chian
389 School of Medicine, Nanyang Technological University, Singapore, Singapore. [288]Vanderbilt Genetics Institute,
390 Division of Genetic Medicine, Vanderbilt University Medical Center, Nashville, TN, USA. [289]VA Palo Alto Health
391 Care System, Palo Alto, CA, USA. [290]Stanford Cardiovascular Institute, Stanford University School of Medicine,
392 Stanford, CA, USA. [291]Department of Medicine, University of Pennsylvania Perelman School of Medicine,
393 Philadelphia, PA, USA. [292]Oxford Centre for Diabetes, Endocrinology and Metabolism, Radcliffe Department of
394 Medicine, University of Oxford, Oxford, UK. [293]Oxford NIHR Biomedical Research Centre, Churchill Hospital,
395 Oxford University Hospitals NHS Foundation Trust, Oxford, UK. [294]Department of Biostatistics and
396 Epidemiology, University of Massachusetts Amherst, Amherst, MA, USA. [295]Department of Biostatistics,
397 Epidemiology and Informatics, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA, USA.
398 [296]TUM School of Medicine, Technical University of Munich and Klinikum Rechts der Isar, Munich, Germany.
399 [297]Present address: Regeneron Genetics Center, Tarrytown, NY, USA. [298]Present address: Sir J C Bose Fellow,
400 Science and Engineering Research Board, Department of Science and Technology, Ministry of Science and
401 Technology, Government of India, New Delhi, India. [299]Present address: Genentech, South San Francisco, CA,
402 USA. [300]Deceased. [*]These authors contributed equally to this work. [†]These authors supervised this work.
403 [‡]Corresponding authors.
404

405 **Type 2 diabetes (T2D) is a heterogeneous disease that develops through diverse**
406 **pathophysiological processes[1,2] and molecular mechanisms that are often cell-type**
407 **specific[3,4]. To characterise the genetic contribution to these processes across ancestry**
408 **groups, we aggregate genome-wide association study (GWAS) data from 2,535,601**
409 **individuals (39.7% non-European ancestry), including 428,452 T2D cases. We identify**
410 **1,289 independent association signals at genome-wide significance (P<5x10^{-8}) that map to**
411 **611 loci, of which 145 loci are previously unreported. We define eight non-overlapping**
412 **clusters of T2D signals characterised by distinct profiles of cardiometabolic trait**
413 **associations. These clusters are differentially enriched for cell-type specific regions of**
414 **open chromatin, including pancreatic islets, adipocytes, endothelial, and enteroendocrine**
415 **cells. We build cluster-specific partitioned polygenic scores (PS)[5] in an additional 279,552**
416 **individuals of diverse ancestry, including 30,288 T2D cases, and test their association with**
417 **T2D-related vascular outcomes. Cluster-specific partitioned PS are associated with**
418 **coronary artery disease (CAD), peripheral artery disease and end-stage diabetic**
419 **nephropathy (ESDN) across ancestry groups, highlighting the importance of obesity-**
420 **related processes in the development of vascular outcomes. Our findings demonstrate the**
421 **value of integrating multi-ancestry GWAS with single-cell epigenomics to disentangle the**
422 **aetiological heterogeneity driving the development and progression of T2D, which may**
423 **offer a route to optimise global access to genetically-informed diabetes care.**
424

425 Diabetes mellitus is a huge public health burden, with an estimated prevalence of 537
426 million adults worldwide in 2021, of whom >90% are affected by T2D[6]. The biological
427 processes through which T2D develops are diverse and include impaired insulin secretion
428 and insulin resistance. This aetiological heterogeneity leads to substantial variability in
429 patient phenotypes, including age of disease onset, manifestation of disease complications,
430 and response to management strategies[1,2]. Whilst environment and lifestyle are well-
431 established risk factors for T2D, heritability has been estimated to be 69% amongst
432 individuals of 35 to 60 years[7]. Previous GWAS of T2D have identified >500 risk loci[8,9], which
433 demonstrated variable patterns of association with clinical features mediated via effector
434 genes acting through distinct molecular mechanisms that are often cell-type specific[3,4].
435 Through the newly-established Type 2 Diabetes Global Genomics Initiative, we present
436 findings from the largest T2D GWAS meta-analysis to date, comprising >2.5 million
437 individuals of diverse ancestry and a near three-fold increase in effective sample size over
438 previous efforts[8,9]. We leverage the power afforded by this increased sample size with
439 emerging single-cell functional genomics data derived from disease-relevant tissues to
440 disentangle the aetiological heterogeneity of T2D. For the first time across multiple ancestry
441 groups, we construct partitioned PS[5] and assess their association with T2D-related
442 macrovascular outcomes and progression to microvascular complications.
443

444 **Study overview.** We assembled GWAS including 428,452 T2D cases and 2,107,149 controls
445 (**Supplementary Figure 1, Supplementary Tables 1 and 2**). We organised these GWAS into
446 six subsets of genetically similar studies, which we refer to as "ancestry groups" (**Extended**
447 **Data Figure 1**). Specifically, we considered: a European ancestry group (EUR, 60.3% of the
448 effective sample size); an East Asian ancestry group (EAS, 19.8%); an admixed African
449 American group with ancestry predominantly from West Africa and Europe (AFA, 10.5%); an
450 admixed Hispanic group with ancestry predominantly from the Americas, West Africa, and
451 Europe (HIS, 5.9%); a South Asian ancestry group (SAS, 3.3%); and a South African ancestry

452 group (SAF, 0.2%). Association analyses accounted for study-level population structure and
453 relatedness, and adjusted for age and sex, where appropriate, and additional study-specific
454 covariates (**Supplementary Table 3, Methods**).
455
456 **Discovery of T2D loci.** We aggregated association summary statistics across GWAS via multi-
457 ancestry meta-regression, implemented in MR-MEGA[10], which allows for allelic effect
458 heterogeneity that is correlated with ancestry. We included three axes of genetic variation
459 as covariates in the meta-regression model that separated GWAS from different ancestry
460 groups (**Extended Data Figure 1, Methods**), which resulted in lower genomic control
461 inflation than a fixed-effects meta-analysis (λ_{GC}=1.120 and λ_{GC}=1.396, respectively).
462 The DIAMANTE Consortium previously advocated the use of a multi-ancestry
463 genome-wide significance threshold ($P<5\times10^{-9}$) to define loci, which takes account of the
464 weaker linkage disequilibrium (LD) between SNVs expected after multi-ancestry meta-
465 analysis[9]. To gain insight into true positive signals meeting conventional genome-wide
466 significance ($P<5\times10^{-8}$) that would be overlooked at this more stringent threshold, we
467 considered loci reported by the DIAMANTE Consortium, which contributed 39.5% of the
468 effective sample size of the current study. Of 39 loci with association signals meeting
469 $5\times10^{-9}\leq P<5\times10^{-8}$ in the DIAMANTE Consortium analysis, 36 (92.3%) attained multi-ancestry
470 genome-wide significance with the larger sample size available to us in the current study
471 (**Supplementary Text**). We therefore focussed our downstream analyses on SNVs meeting
472 the conventional genome-wide significance threshold.
473 We identified at total of 1,289 distinct T2D association signals ($P<5\times10^{-8}$) that were
474 represented by independent ($r^2<0.05$) index SNVs (**Supplementary Figure 2, Supplementary**
475 **Table 4, Methods**). The 1,289 association signals mapped to 611 loci, of which 145 loci
476 (23.7%) have not been previously reported in T2D GWAS. At association signals mapping to
477 loci not previously reported for T2D, index SNVs were predominantly common (MAF >5% in
478 at least one ancestry group) with odds-ratios (ORs) <1.05 (**Supplementary Figure 3**).
479
480 **Mechanistic clusters of T2D index SNVs.** To understand the genetic contribution to
481 phenotypic heterogeneity in T2D, we classified the 1,289 index SNVs according to their
482 profile of associations (aligned to the T2D risk allele) with 37 cardiometabolic phenotypes.
483 These included glycaemic traits, anthropometric measures, body fat and adipose tissue
484 volume, blood pressure, circulating plasma lipid levels, and biomarkers of liver function and
485 lipid metabolism[11-19] (**Supplementary Table 5**). We applied an unsupervised "hard
486 clustering" approach with imputation of missing phenotype associations, which identified
487 eight non-overlapping but exhaustive subsets of index SNVs with similar cardiometabolic
488 profiles (**Figure 1, Table 1, Extended Data Figure 2, Supplementary Figure 4,**
489 **Supplementary Tables 6 and 7, Methods**).
490 We observed overlap in the cardiometabolic features and loci of five of the identified
491 clusters with those reported in previous efforts[3,4,20,21], representing beta-cell dysfunction
492 with positive or negative association with proinsulin (PI), and insulin resistance mediated via
493 obesity, lipodystrophy, and liver/lipid metabolism (**Supplementary Table 8**). T2D risk alleles
494 at index SNVs in the two beta-cell dysfunction clusters are associated with increased fasting
495 glucose (FG), two-hour glucose, and glycated haemoglobin (HbA1c), and with decreased
496 fasting insulin (FI). Index SNVs in both clusters are also associated with PI, but with opposite
497 directions of effect for the T2D risk allele. The clusters reflecting mechanisms of insulin
498 resistance mediated via obesity, lipodystrophy, and liver/lipid metabolism, include index

499 SNVs that are associated with anthropometric measures and circulating plasma lipid levels.
500 T2D risk alleles at index SNVs in the obesity cluster are associated with increased body mass
501 index (BMI), waist-hip ratio (WHR), body fat percentage, and basal metabolic rate, and with
502 decreased high-density lipoprotein (HDL) cholesterol. The lipodystrophy cluster comprises
503 index SNVs for which T2D risk alleles are associated with increased FI, WHR, blood pressure,
504 and triglycerides, and with decreased body fat percentage, gluteofemoral adipose tissue
505 (GFAT) volume, and HDL cholesterol. T2D risk alleles at index SNVs assigned to the liver/lipid
506 metabolism cluster are associated with increased liver fat and liver-related biomarkers, and
507 with decreased low-density lipoprotein cholesterol and total cholesterol.
508 By increasing the number of index SNVs in the clustering by nearly four-fold over
509 previous efforts, we provide a more granular view of the biological processes through which
510 T2D associations impact disease, and highlight three previously-unreported clusters of
511 signals with cardiometabolic profiles that are representative of the metabolic syndrome,
512 body fat, and residual glycaemic effects. We observed significantly weaker allelic effects on
513 T2D in these three clusters than in those previously reported (mean OR of 1.028 versus
514 1.033, P=2.2x10^{-7}), but there was no noticeable difference in disparity around the centroid
515 between clusters (**Extended Data Figure 3, Supplementary Table 9, Supplementary Figure**
516 **5**). T2D risk alleles at index SNVs assigned to the metabolic syndrome cluster are associated
517 with increased FG, WHR, triglycerides, and blood pressure, and with decreased HDL
518 cholesterol, which together are used to define metabolic syndrome. T2D risk alleles in this
519 cluster are also associated with increased FI, and with accumulations of unhealthy fat
520 depots: increased visceral adipose tissue (VAT) volume and liver fat, and with decreased
521 GFAT volume. Observationally, individuals with the metabolic syndrome are at increased
522 risk of T2D[22], although Mendelian randomisation studies indicate that a causal effect is
523 driven by increased waist circumference and increased FG[23]. T2D risk alleles at index SNVs
524 assigned to the body fat cluster are associated with increased abdominal subcutaneous
525 adipose tissue volume, VAT volume, and body fat percentage. Whilst the body fat cluster
526 profile of associations with cardiometabolic phenotypes shares these features in common
527 with obesity-mediated insulin resistance, index SNVs in the body fat cluster are not strongly
528 associated with BMI, lipid levels, or basal metabolic rate. Previous investigations have
529 highlighted that body fat percentage is predictive of abnormal blood glucose in individuals
530 with a healthy BMI[24]. Finally, T2D risk alleles at index SNVs assigned to the residual
531 glycaemic cluster are most strongly associated with increased FG and HbA1c, but unlike the
532 two beta-cell dysfunction clusters, are not associated with PI or decreased FI.
533 Clustering provides a framework to better understand the diverse physiological
534 processes through which T2D develops and shared biological pathways driving genetic
535 correlations with other insulin resistance-related disorders, including gestational diabetes
536 mellitus (GDM) and polycystic ovary syndrome (PCOS). T2D risk alleles at index SNVs
537 showed a gradient of effects on these correlated insulin-related endophenotypes across
538 clusters (**Supplementary Text, Extended Data Figure 4, Supplementary Tables 10 and 11**),
539 representing a cline from insulin production and processing in the two beta-cell dysfunction
540 clusters through to insulin resistance that was most extreme in the lipodystrophy cluster.
541 Index SNVs in the beta cell +PI cluster showed the strongest associations with GDM, whilst
542 those in the obesity cluster were most strongly associated with PCOS (**Supplementary Text,**
543 **Extended Data Figure 5, Supplementary Table 12**).
544

545 **Regulatory processes underlying clusters.** To gain insight into tissue-specific regulatory
546 processes underpinning mechanistic clusters, we integrated T2D association signals with
547 Assay for Transposase-Accessible Chromatin using sequencing (ATAC-seq) peaks from single-
548 cell atlases of chromatin accessibility (CATLAS and DESCARTES) for 222 cell types derived
549 from 30 human adult and 15 human foetal tissues[25,26] and an additional 106 cell types from
550 human brain[27] (**Figure 2, Supplementary Tables 13 and 14, Methods**).
551 We observed significant enrichment for regions of open chromatin in foetal islets
552 and adult neuroendocrine cells in pancreatic islets (alpha, beta, gamma, and delta) in the
553 beta cell +PI, beta cell -PI, and residual glycaemic clusters. In addition, the residual glycaemic
554 cluster was enriched in foetal and adult pancreatic ductal cells, whilst the beta cell -PI
555 cluster was enriched in adult enterochromaffin cells, a type of enteroendocrine cell that
556 plays an essential role in regulating intestinal motility and secretion in the gastrointestinal
557 tract[28]. Enterochromaffin cells are a major target for GLP-1 and highly express GLP-1
558 receptor, whose agonists are widely used as medications for T2D[29] (**Supplementary Text**).
559 The obesity cluster was also significantly enriched for regions of open chromatin in
560 adult pancreatic islets, although not as strongly as the beta-cell dysfunction clusters.
561 Enrichment was observed only for alpha, gamma, and delta cells, suggesting potential
562 alternative pathways through which islets impact the development of T2D than through
563 insulin secretion from beta cells. The obesity cluster was further enriched in foetal adrenal
564 gland (chromaffin cells and adrenal neurons), foetal heart (ventricular cardiomyocytes), and
565 foetal kidney (metanephric cells). Previous studies have reported enrichment of BMI
566 loci/heritability for epigenomic annotations in pancreatic islets and adrenal gland[30,31],
567 consistent with our findings. In human brain, the obesity cluster was significantly enriched
568 for regions of open chromatin in cell types including intratelencephalic (IT) projecting
569 neurons, somatostatin positive (SST+) GABAergic inhibitory neurons, and D1 medium spiny
570 neurons. SST+ GABAergic neurons exist in the hypothalamus and regulate food intake[32]. D1
571 medium spiny neurons are a type of GABAergic neuron in the human striatum that express
572 D1-type dopamine receptors, and have been implicated in food motivation and the
573 development of diet-induced obesity in mice[33].
574 The remaining four clusters (lipodystrophy, metabolic syndrome, body fat, and
575 liver/lipid metabolism) were not significantly enriched for regions of open chromatin in
576 pancreatic islets. The lipodystrophy cluster was enriched only in adult adipocytes, which
577 confirms previous reports in bulk adipose tissue[4,20]. Consistent with these results,
578 association signals for WHR, triglycerides, and HDL cholesterol, which are strongly impacted
579 by index SNVs in the lipodystrophy cluster, have been shown to be enriched in candidate *cis*-
580 regulatory elements in adipocytes[26]. The metabolic syndrome cluster was enriched in cells
581 that reside in the walls of blood vessels (adult pericytes and foetal endothelial cells), foetal
582 kidney (mesangial cells), and foetal fibroblasts. Association signals for systolic and diastolic
583 blood pressure, a key component of the metabolic syndrome, have been shown to be
584 enriched in candidate *cis*-regulatory elements in these cell types[26]. Endothelial dysfunction
585 is not only a consequence of insulin resistance, but impairs insulin signalling to further
586 reduce insulin sensitivity, thereby providing a pathophysiological mechanism that links
587 metabolic and cardiovascular components of metabolic syndrome[34]. In human brain, the
588 metabolic syndrome cluster was significantly enriched for regions of open chromatin in cell
589 types including IT projecting neurons and SST+ GABAergic inhibitory neurons. IT projecting
590 neurons are a type of glutamatergic excitatory pyramidal neuron in the cerebral cortex, and
591 metabolic syndrome has been associated with pyramidal neurons and GABAergic neurons in

592 cell-type specificity analysis in metabolic syndrome factor GWAS[35]. We observed no
593 significant enrichments in the body fat cluster or liver/lipid metabolism cluster.
594
595 **Ancestry-correlated heterogeneity.** Previous multi-ancestry GWAS have demonstrated
596 widespread heterogeneity in allelic effects at T2D association signals across ancestry
597 groups[9,36]. We took advantage of the meta-regression model to partition heterogeneity into
598 an ancestry-correlated component explained by three axes of genetic variation, and a
599 residual component reflecting differences in environmental exposures (that are not
600 correlated with ancestry) and/or study design (**Supplementary Table 15**). We observed 127
601 (9.9%) independent T2D association signals with significant evidence for ancestry-correlated
602 heterogeneity ($P_{HET}<3.9x10^{-5}$, Bonferroni correction for 1,289 signals). We would expect <1
603 signal to meet this threshold of significance, highlighting that ancestry-correlated
604 heterogeneity is strongly enriched at T2D associations (one-sided binomial test $P<2.2x10^{-16}$).
605 In contrast, we observed significant evidence of residual heterogeneity at only 4 (0.3%)
606 association signals (one-sided binomial test $P=0.031$). These results therefore suggest that
607 differences in allelic effects at index SNVs are more strongly correlated with genetic
608 ancestry than other factors that vary between GWAS.
609 We next sought to understand better the impact of genetic diversity on differences
610 in allelic effects between GWAS at the 127 association signals with significant evidence of
611 ancestry-correlated heterogeneity (**Methods**). For 118 (92.9%) signals, allelic effect sizes
612 were most strongly associated with the first two axes of genetic variation, which reflect
613 differences between AFA/EUR and EAS GWAS (AFA-EAS axis), and between AFA/EAS and
614 EUR GWAS (AFA-EUR axis), respectively (**Supplementary Text, Extended Data Figure 1,**
615 **Extended Data Figure 6, Supplementary Table 16**).
616 We observed significant differences in mean Z-scores for association between
617 clusters for both the AFA-EAS axis ($P=4.1x10^{-6}$) and the AFA-EUR axis ($P=1.5x10^{-6}$). Index
618 SNVs in the two beta-cell dysfunction clusters were most positively associated with the AFR-
619 EAS axis, indicating allelic effects on T2D that were greater in EAS than in AFA and EUR
620 GWAS (**Extended Data Figure 7, Supplementary Table 17**). In contrast, index SNVs in the
621 lipodystrophy and obesity clusters were most positively associated with the AFA-EUR axis,
622 indicating allelic effects on T2D that were greater in EUR GWAS than in EAS/AFA GWAS.
623 These results indicate that ancestry-correlated heterogeneity varies between mechanistic
624 clusters, with allelic effects greatest for EAS GWAS at association signals assigned to clusters
625 acting via beta-cell dysfunction and greatest for EUR GWAS at those assigned to clusters
626 operating through insulin resistance.
627 Ancestry-correlated heterogeneity in allelic effects between GWAS is not driven by
628 differences in allele frequency between ancestry groups but can occur because of
629 interaction between index SNVs and environmental/lifestyle factors, if not accounted for in
630 the association analysis[37]. We observed substantial variation in the distribution of study-
631 level mean BMI in T2D cases and controls across ancestry groups (**Supplementary Figure 6**).
632 Such variation could impact on ancestry-correlated heterogeneity because, when cases and
633 controls are selected from the extremes of the BMI distribution, the magnitude of allelic
634 effect estimates at T2D signals acting via beta-cell dysfunction can be inflated[38]. We
635 therefore extended the MR-MEGA meta-regression model to allow for allelic effect
636 heterogeneity at index SNVs due to mean BMI in T2D cases and controls, in addition to axes
637 of genetic variation (**Methods**).

638 After adjustment for study-level mean BMI in T2D cases and in controls, only 24
639 association signals retained significant evidence of ancestry-correlated heterogeneity
640 (P<3.9x10^{-5}), compared with 127 signals without adjustment (**Supplementary Text,**
641 **Supplementary Table 18**). After adjustment for BMI, significant differences in mean Z-scores
642 for association between clusters for the AFA-EUR axis were maintained (P=3.2x10^{-5} versus
643 P=1.5x10^{-6} without adjustment), whilst those for the AFA-EAS axis were not (P=0.18 versus
644 P=4.1x10^{-6} without adjustment). Furthermore, after adjustment for BMI, the two beta-cell
645 dysfunction clusters were no longer strongly positively associated with the AFA-EAS axis
646 (**Extended Data Figure 7, Supplementary Table 19**). Taken together, these results suggest
647 that heterogeneity in allelic effects between EAS GWAS and EUR/AFA GWAS, which occur
648 most often at association signals assigned to the beta-cell dysfunction clusters, can be
649 mostly accounted for by differences in the distributions of mean BMI in T2D cases and in
650 controls between these ancestry groups.
651
652 **Partitioned PS association with outcomes.** The major complications in individuals with T2D
653 are macrovascular outcomes including CAD, ischemic stroke, and peripheral artery disease,
654 and microvascular outcomes, including ESDN and proliferative diabetic retinopathy. We
655 tested for association of a cluster-specific partitioned PS with these vascular outcomes in up
656 to 279,552 individuals (including 30,288 T2D cases) across five ancestry groups (AFA, EAS,
657 EUR, HIS, and SAS) from the All of Us Research Program, Biobank Japan, and Genes & Health
658 (**Methods**). These individuals were not included in the multi-ancestry meta-analysis,
659 avoiding potential inflated type I error rates due to overlap between the discovery and
660 testing datasets. To maximise sample size, we tested macrovascular outcomes in all
661 individuals, adjusted for T2D status, and microvascular complications only in individuals with
662 T2D (**Methods, Supplementary Table 20**). To assess the additional information afforded by
663 the partitioned PS over an overall T2D PS, agnostic to cluster membership, we tested for
664 association of each cluster-specific component of the partitioned PS after adjustment for
665 the overall PS. **Figure 3** provides an overview of the associations of each cluster-specific
666 component of the partitioned PS with the five vascular outcomes across ancestry groups.
667
668 We observed significant association (P<0.0063, Bonferroni correction for eight
669 clusters) of two components of the partitioned PS with CAD: a negative association with the
670 beta cell +PI cluster (OR=0.96 per standard deviation of the PS, P=1.3x10^{-6}) and a positive
671 association with the obesity cluster (OR=1.04, P=0.00019). There was no evidence of
672 heterogeneity in the effects of these two clusters on CAD across ancestry groups
673 (**Supplementary Figure 7, Supplementary Table 21**). Importantly, after adjustment for a
674 CAD PS derived from a recently published multi-ancestry meta-analysis of CAD GWAS[39], the
675 positive CAD association with both components of the partitioned PS remained significant
676 (**Extended Data Figure 8, Supplementary Table 22**): beta cell +PI cluster (OR=0.96,
677 P=4.4x10^{-5}) and obesity cluster (OR=1.04, P=0.00065). We also observed significant positive
678 association of the obesity cluster from the partitioned PS with peripheral artery disease
679 (OR=1.05, P=0.00045), with no evidence of heterogeneity in effects across ancestry groups
680 (**Supplementary Figure 8, Supplementary Table 21**). Across all three macrovascular
681 outcomes, there was a general trend of negative association with the beta cell +PI cluster
682 and positive association with the obesity cluster, although no cluster-specific components of
683 the partitioned PS attained significance for ischemic stroke (**Supplementary Figure 9,**
684 **Supplementary Table 21**). There was no strong association of the overall T2D PS with CAD

685 (P=0.17), ischemic stroke (P=0.022), or peripheral artery disease (P=0.77) after meta-
686 analysis across ancestry groups. Taken together, these results highlight the advantages of
687 the partitioned PS over an overall T2D PS for detecting association with macrovascular
688 outcomes and provide insight into the biological processes that lead to their development.
689 We observed significant association of two components of the partitioned PS with
690 ESDN: a negative association with the beta cell +PI cluster (OR=0.83, P=0.00024) and a
691 positive association with the obesity cluster (OR=1.19, P=0.00050). There was no evidence
692 of heterogeneity in the effects of these two clusters across ancestry groups,
693 (**Supplementary Figure 10, Supplementary Table 21**), and the overall PS was not strongly
694 associated with ESDN (P=0.048). In contrast, none of the cluster-specific components of the
695 partitioned PS were associated with proliferative diabetic retinopathy. However, there was
696 a strong positive association of the overall PS with this microvascular outcome (OR=1.32,
697 P=1.1x10^{-9}), with no evidence of heterogeneity in effects across ancestry groups
698 (**Supplementary Figure 11, Supplementary Table 21**). Taken together, these results suggest
699 that ESDN is associated with obesity and beta-cell dysfunction with opposite direction of
700 effect and confirm previous reports that proliferative diabetic retinopathy is driven by
701 hyperglycaemia[40] and therefore strongly associated with the overall burden of T2D risk
702 variants.
703 Finally, we tested for association of the cluster-specific components of the
704 partitioned PS and the overall T2D PS with age of onset of T2D (**Extended Data Figure 9,**
705 **Methods**). The overall PS was strongly associated with earlier age of onset (1.15 years per
706 standard deviation of the PS, P=5.1x10^{-8}), although the effects were highly heterogeneous
707 across ancestry groups (**Supplementary Figure 12, Supplementary Table 23**). However,
708 even after adjustment for the overall PS, the obesity cluster was significantly associated
709 with earlier age of onset (0.38 years, P=1.4x10^{-7}) with no evidence of heterogeneity across
710 ancestry groups. These findings highlight the importance of obesity-related processes for
711 T2D onset, in addition to the development of vascular complications.
712
713 **Obesity and beta cell +PI clusters are consistently associated with vascular outcomes in**
714 **clinical trials.** To gain insight into the associations of the obesity and beta cell +PI clusters
715 with a broader range of vascular outcomes, we assessed the performance of the partitioned
716 PS (after adjustment for the overall PS) in prospective GWAS in up to 29,827 EUR individuals
717 with T2D from six clinical trials from the Thrombolysis in Myocardial Infarction (TIMI) Study
718 Group (**Methods, Supplementary Table 24**). We observed the strongest associations of
719 cluster-specific components of the partitioned PS with risk of hospitalization for heart
720 failure: positive with the obesity cluster (hazard-ratio [HR]=1.15 per standard deviation of
721 the PS, P=4.8x10^{-6}) and negative with the beta cell +PI cluster (HR=0.90, P=0.00092).
722 Amongst macrovascular outcomes, the beta cell +PI cluster was also negatively associated
723 with cardiovascular death (HR=0.90, P=0.0020), major cardiovascular events (HR=0.94,
724 P=0.0050), and myocardial infarction (HR=0.94, P=0.027). For microvascular outcomes, the
725 two clusters showed associations with opposite directions of effect for albuminuria: obesity
726 cluster (HR=1.06, P=0.012) and beta cell +PI cluster (HR=0.95, P=0.047). Across all outcomes,
727 there was a general trend of positive association with the obesity cluster and negative
728 association with the beta cell +PI cluster (**Extended Data Figure 10**), consistent with the
729 associations observed from our analyses of retrospective GWAS across ancestry groups.
730

731 **DISCUSSION.** To better understand the aetiological heterogeneity of T2D across diverse
732 populations, we assembled the largest collection of T2D GWAS to date for five ancestry
733 groups through the Type 2 Diabetes Global Genomics Initiative. By increasing the effective
734 sample size by almost three-fold compared to previous efforts, we identified a total of 611
735 loci attaining the conventional threshold of genome-wide significance ($P<5 \times 10^{-8}$), 145
736 (23.7%) of which have not been previously reported. This conventional threshold is
737 equivalent to a Bonferroni correction for the effective number of independent SNVs in EUR
738 reference data[41]. Using empirical data from the 1000 Genomes Project, the DIAMANTE
739 Consortium and others have advocated more stringent thresholds for multi-ancestry meta-
740 analysis because the structure of LD is broken down across ancestry groups and the
741 effective number of independent SNVs is increased[9,42]. In fact, our analyses suggest that loci
742 meeting conventional genome-wide significance are unlikely to be false positive association
743 signals, but instead are driven by index SNVs that have modest effects that require larger
744 sample sizes to meet more stringent thresholds. We therefore recommend the use of this
745 conventional threshold but advocate careful review of reported signals to ensure that
746 associations are not driven by single studies or poorly imputed variants to protect against
747 false positives.

748 Multi-ancestry meta-regression maximizes power to detect associations that are
749 shared across ancestry groups by allowing for heterogeneity in allelic effects at index SNVs.
750 MR-MEGA is not restricted to broad continental ancestry labels that can be used to
751 reinforce the concept of fundamental genetic differences between groups[43], but instead
752 represents ancestry as continuous axes of genetic variation, which better reflect the
753 continuum of human genetic diversity and demographic history[44]. Still, it is important to
754 emphasize that our meta-analysis does not fully capture global genetic diversity, in
755 particular under-represented populations across Africa, South and Central America, the
756 Middle East, and Oceania. For example, 98.2% of the total effective sample size of
757 individuals with the highest proportion of ancestry from Africa are African Americans. The
758 ancestry of these individuals represents a cline of admixture that is predominantly from
759 West Africa and is therefore not representative of other regions in Africa, where the level of
760 genetic variation is equivalent to the differences observed between other continental
761 groups[43]. Bolstering GWAS collections in these under-represented populations remains an
762 urgent priority for the human genetics research community and highlights the need for
763 careful interpretation of results that do not generalise findings across ancestry groups that
764 are sensitive to biased representation.

765 Within the landscape of the genetic architecture of T2D, we identified eight clusters
766 of index SNVs with distinct profiles of associations with 37 cardiometabolic phenotypes,
767 which defined pathophysiologic-relevant groupings. The addition of previously-unreported
768 T2D signals identified through the multi-ancestry meta-analysis helped define three clusters
769 that were not detected in previous clustering efforts[3,4,20,21], with cardiometabolic profiles
770 that are consistent with residual glycaemic effects, accumulations of body fat, and the
771 metabolic syndrome. These previous efforts have implemented "soft clustering"
772 approaches, such as Bayesian non-negative matrix factorisation, that generate weights for
773 cluster membership for each index SNV[4]. Assignment of index SNVs to clusters is then
774 determined given a threshold weight for cluster membership, allowing for the possibility
775 that a T2D association signal impacts on disease through multiple pathophysiological
776 pathways. However, depending on the threshold for cluster membership, some index SNVs
777 will be unassigned. Bayesian non-negative matrix factorisation also considers positive and

778 negative associations with the same phenotype as independent variables, and most
779 clustering methods cannot directly accommodate missing phenotype associations. To
780 address these potential limitations, we implemented methodology that jointly conducts K-
781 means clustering of index SNVs with powerful iterative multiple imputation of missing
782 phenotype associations. In this "hard clustering" approach, each index SNV is assigned to
783 exactly one cluster. This has the potential disadvantage, therefore, that index SNVs with
784 outlying or intermediate profiles of trait associations are "forced" into a cluster that does
785 not fit well. However, the previously-unreported clusters we identified in our hard clustering
786 were not noticeably more disparate than the clusters reported previously, suggesting that
787 we have not introduced substantial noise by forcing all SNVs into exactly one cluster.
788 Ultimately, the choice of clustering approach may depend on the objectives of any
789 downstream investigations.
790 Our analyses highlighted a significant excess of T2D association signals with ancestry-
791 correlated heterogeneity, which is primarily driven by differences in allelic effects between
792 AFA, EAS, and EUR GWAS. The two beta-cell dysfunction clusters are most strongly
793 associated with the AFA-EAS axis, where effects are typically larger in EAS GWAS than in
794 other ancestry groups. These two clusters are also most strongly associated with reduced
795 insulin secretion and lower insulin resistance. In contrast, the lipodystrophy and obesity
796 clusters, which are characterised by reduced insulin sensitivity and higher insulin resistance,
797 are most strongly associated with the AFA-EUR axis, where effects are typically larger in EUR
798 than other ancestry groups. These observations are consistent with studies reporting
799 differences in the pathogenesis of T2D between ancestry groups, where T2D is primarily
800 initiated through increased insulin resistance in EUR individuals, but is characterised by
801 reduced insulin secretion with lower insulin resistance in EAS individuals[45,46]. For the first
802 time, we have demonstrated that most signals with ancestry-correlated heterogeneity can
803 be explained by differences in the distribution of BMI in T2D cases and controls between
804 ancestry groups. Furthermore, after adjustment for study-level mean BMI, we observe no
805 difference in allelic effects between clusters along the AFA-EAS axis. This is consistent with
806 previous reports that body composition is the main determinant of variation in T2D
807 pathogenesis between EAS and EUR individuals, because insulin sensitivity and beta-cell
808 response are similar in the two ancestry groups after accounting for differences in BMI[45,47].
809 For the first time across multiple ancestry groups, we demonstrated significant
810 associations of vascular outcomes with cluster-specific components of the partitioned PS
811 after adjustment for the overall PS, suggesting that disease trajectories may be associated
812 with genetic burden in particular biological pathways that are consistent across diverse
813 populations. Whilst the effect sizes of the cluster-specific components of the partitioned PS
814 were small, they motivate future work to strengthen these effects through identification of
815 additional T2D associations in larger sample sizes. Through integration with single-cell
816 chromatin accessibility data across diverse cell types, they also enhance understanding of
817 key biological processes driving heterogeneity in the clinical features of T2D phenotypes. For
818 example, the obesity cluster-specific component of the PS was positively associated with
819 CAD and ESDN, and included index SNVs that were enriched for regions of open chromatin
820 in foetal ventricular cardiomyocytes, foetal adrenal neuron, adult chromaffin cells in the
821 adrenal gland, and foetal metanephric cells. These findings are in line with the reported
822 enrichments of CAD association signals for transcriptomic/epigenomic annotations in bulk
823 tissues including aorta/arteries, heart, and adrenal gland[39,48,49] and of renal function
824 association signals in kidney tissue-specific regulatory annotations[50]. Taken together, these

825 findings provide a clear link to shared biological mechanisms driving development of T2D
826 and other vascular diseases.
827 In conclusion, our findings demonstrate the value of integrating multi-ancestry
828 GWAS of T2D and cardiometabolic traits with single-cell epigenomics across diverse tissues
829 to disentangle the aetiological heterogeneity driving the development and progression of
830 T2D across population groups. Improved understanding of the varied pathophysiological
831 processes that link T2D to vascular outcomes may offer a route to genetically-informed
832 diabetes care and global opportunities for the clinical translation of T2D GWAS findings.
833

834 **FIGURE LEGENDS**
835
836 **Figure 1. Heatmap of associations of 37 cardiometabolic phenotypes with eight**
837 **mechanistic clusters of index SNVs for T2D association signals.** Each column corresponds to
838 a cluster. Each row corresponds to a cardiometabolic phenotype. The "temperature" of each
839 cell represents the Z-score (aligned to the T2D risk allele) of association of the phenotype
840 with index SNVs assigned to the cluster. *Phenotype is adjusted for body mass index.
841
842 **Figure 2. Heatmap of cluster-specific enrichments of T2D associations for cell type-specific**
843 **regions of open chromatin derived from single-cell ATAC-seq peaks in adult and foetal**
844 **tissue.** (A) 222 cell types from 30 human adult tissues and 15 human foetal tissues. (B) 106
845 cell types from human brain. In each panel, columns represent mechanistic clusters. Each
846 row represents a cell type that was significantly enriched (Bonferroni correction for the
847 number of cell types) for T2D associations in at least one cluster (indicated by an asterisk).
848 The "temperature" of each cell defines the magnitude of the log-fold enrichment. The
849 liver/lipid metabolism cluster is not presented because it includes only three T2D
850 association signals and the model parameter estimates were unstable.
851
852 **Figure 3. Associations of cluster-specific components of the partitioned PS with five T2D-**
853 **related vascular outcomes in up to 279,552 individuals from multiple ancestry groups.**
854 Each of the panels summarise the associations each cluster-specific component of the
855 partitioned PS with coronary artery disease (CAD), peripheral artery disease (PAD), ischemic
856 stroke (IS), end-stage diabetic nephropathy (ESDN), and proliferative diabetic retinopathy
857 (PDR). The height of each bar corresponds to the log-odds ratio (beta) per standard
858 deviation of the PS, and the grey bar shows the 95% confidence interval. Analyses of T2D-
859 related macrovascular complications (CAD, PAD, and IS) were undertaken in all individuals,
860 with adjustment for T2D status. Analysis of microvascular complications were undertaken in
861 individuals with T2D only. *P<0.05, nominal association. **P<0.0063, Bonferroni correction
862 for eight clusters. Exact P-values are presented in **Supplementary Table 21**.
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879

880 **Table 1. Cardiometabolic profile, exemplar loci, and physiological impact of index SNVs at**
881 **T2D association signals allocated to eight mechanistic clusters.**
882

Mechanistic cluster	Cardiometabolic profile	Number of T2D associations	Exemplar loci	Physiological impact	
				Insulin secretion	Insulin sensitivity
Beta cell +PI	+FG*, +2hG*, +HbA1c, +PI*	91	*TCF7L2*, *KCNQ1*, *CDKAL1*, *CDKN2A-CDKN2B*, *SLC30A8*	-	+
Beta cell -PI	+FG*, +2hG*, +HbA1c, -PI*	89	*CDC123-CAMK1D*, *HNF1B*, *KCNJ11-ABCC8*, *HNF4A*, *HNF1A*	-	+
Residual glycaemic	+FG*, +HbA1c	389	*GCC1-PAX4-LEP*, *ANKRD55*, *GCKR*, *UBE2E2*	-	-
Body fat	+body fat, +ASAT*	273	*ZMIZ1*, *HMGA2*, *CTBP1*	+	-
Metabolic syndrome	+FG*, +FI*, +WHR, +VAT*, -GFAT*, +TG, -HDL, +BP	166	*IGF2BP2*, *CCND2*, *HHEX-IDE*, *JAZF1*, *GPSM1*	+	-
Obesity	+BMI, +WHR, +body fat, +BMR, +TG, -HDL	233	*FTO*, *MC4R*, *MACF1*, *TMEM18*	+	-
Lipodystrophy	+FI*, +WHR, -body fat, -GFAT*, +TG, -HDL, +BP	45	*IRS1*, *GRB14-COBBL1*, *PPARG*	+	-
Liver/lipid metabolism	-LDL, -TC, +liver fat, +liver biomarkers	3	*TOMM40-APOE-GIPR*, *TM6SF2*, *PNPLA3*	-	-

883
884 +/-: T2D risk alleles associated with increased/decreased phenotype values.
885 FG: fasting glucose. FI: fasting insulin. 2hG: two-hour glucose. HbA1c: glycated haemoglobin. PI: proinsulin.
886 BMI: body mass index. WHR: waist-hip ratio. VAT: visceral adipose tissue volume. ASAT: abdominal
887 subcutaneous adipose tissue volume. GFAT: gluteofemoral adipose tissue volume. LDL: low-density lipoprotein
888 cholesterol. HDL: high-density lipoprotein cholesterol. TC: total cholesterol. TG: triglycerides. BMR: basal
889 metabolic rate. BP: blood pressure.
890 *adjusted for BMI.

891 **REFERENCES**
892

893 1. McCarthy, M. I. Painting a new picture of personalised medicine for diabetes.
894 *Diabetologia* **60**, 793-799 (2017).
895 2. Pearson, E. R. Type 2 diabetes: a multifaceted disease. *Diabetologia* **62**, 1107-1112
896 (2019).
897 3. Mahajan, A. *et al.* Refining the accuracy of validated target identification through
898 coding variant fine-mapping in type 2 diabetes. *Nat Genet* **50**, 559-571 (2018).
899 4. Udler, M. S. *et al.* Type 2 diabetes genetic loci informed by multi-trait associations
900 point to disease mechanisms and subtypes: A soft clustering analysis. *PLoS Med* **15**,
901 e1002654 (2018).
902 5. Udler, M. S., McCarthy, M. I., Florez, J. C. & Mahajan, A. Genetic risk scores for
903 diabetes diagnosis and precision medicine. *Endocr Rev* **40**, 1500-1520 (2019).
904 6. International Diabetes Federation. IDF Diabetes Atlas. 10th edn., Brussels, Belgium
905 (2021). https://diabetesatlas.org/.
906 7. Almgren, P. *et al.* Heritability and familiality of type 2 diabetes and related
907 quantitative traits in the Botnia Study. *Diabetologia* **54**, 2811-2819 (2011).
908 8. Vujkovic, M. *et al.* Discovery of 318 new risk loci for type 2 diabetes and related
909 vascular outcomes among 1.4 million participants in a multi-ancestry meta-analysis.
910 *Nat Genet* **52**, 680-691 (2020).
911 9. Mahajan, A. *et al.* Multi-ancestry genetic study of type 2 diabetes highlights the
912 power of diverse populations for discovery and translation. *Nat Genet* **54**, 560-572
913 (2022).
914 10. Mägi, R. *et al.* Trans-ethnic meta-regression of genome-wide association studies
915 accounting for ancestry increases power for discovery and improves fine-mapping
916 resolution. *Hum Mol Genet* **26**, 3639-3650 (2017).
917 11. Warrington, N. M. *et al.* Maternal and fetal genetic effects on birth weight and their
918 relevance to cardio-metabolic risk factors. *Nat Genet* **51**, 804-814 (2019).
919 12. Chen, J. *et al.* The trans-ancestral genomic architecture of glycemic traits. *Nat Genet*
920 **53**, 840-860 (2021).
921 13. Liu, Y. *et al.* Genetic architecture of 11 organ traits derived from abdominal MRI
922 using deep learning. *Elife* **10** (2021).
923 14. Sulc, J. *et al.* Composite trait Mendelian randomization reveals distinct metabolic and
924 lifestyle consequences of differences in body shape. *Commun Biol* **4**, 1064 (2021).
925 15. Sakaue, S. *et al.* A cross-population atlas of genetic associations for 220 human
926 phenotypes. *Nat Genet* **53**, 1415-1424 (2021).
927 16. Graham, S. E. *et al.* The power of genetic diversity in genome-wide association
928 studies of lipids. *Nature* **600**, 675-679 (2021).
929 17. Agrawal, S. *et al.* Inherited basis of visceral, abdominal subcutaneous and
930 gluteofemoral fat depots. *Nat Commun* **13**, 3771 (2022).
931 18. Warren, H. *et al*. Genome-wide analysis in over 1 million individuals reveals over
932 2,000 independent genetic signals for blood pressure. Preprint at:
933 https://www.researchsquare.com/article/rs-1409164/v1 (2022).
934 19. Broadaway, K. A. *et al.* Loci for insulin processing and secretion provide insight into
935 type 2 diabetes risk. *Am J Hum Genet* **110**, 284-299 (2023).

936 20. Kim, H. *et al.* High-throughput genetic clustering of type 2 diabetes loci reveals
937 heterogeneous mechanistic pathways of metabolic disease. *Diabetologia* **66**, 495-
938 507 (2023).
939 21. Smith, K. et al. Multi-ancestry mechanisms of type 2 diabetes elucidate disease
940 processes and clinical heterogeneity. Preprint at:
941 https://www.medrxiv.org/content/10.1101/2023.09.28.23296294v1 (2023).
942 22. Ford, E. S., Li, C. & Sattar, N. Metabolic syndrome and incident diabetes: current
943 state of the evidence. *Diabetes Care* **31**, 1898-1904 (2008).
944 23. Marott, S. C., Nordestgaard, B. G., Tybjærg-Hansen, A. & Benn, M. Components of
945 the metabolic syndrome and risk of type 2 diabetes. *J Clin Endocrinol Metab* **101**,
946 3212-3221 (2016).
947 24. Jo, A. & Mainous, A. G., 3rd. Informational value of percent body fat with body mass
948 index for the risk of abnormal blood glucose: a nationally representative cross-
949 sectional study. *BMJ Open* **8**, e019200 (2018).
950 25. Domcke, S. *et al.* A human cell atlas of fetal chromatin accessibility. *Science* **370**,
951 eaba7612 (2020).
952 26. Zhang, K. *et al.* A single-cell atlas of chromatin accessibility in the human genome.
953 *Cell* **184**, 5985-6001.e5919 (2021).
954 27. Li, Y. E. *et al.* A comparative atlas of single-cell chromatin accessibility in the human
955 brain. *Science* **382**, eadf7044 (2023).
956 28. Bertrand, P. P. & Bertrand, R. L. Serotonin release and uptake in the gastrointestinal
957 tract. *Auton Neurosci* **153**, 47-57 (2010).
958 29. Lund, M. L. *et al.* Enterochromaffin 5-HT cells - A major target for GLP-1 and gut
959 microbial metabolites. *Mol Metab* **11**, 70-83 (2018).
960 30. Finucane, H. K. *et al.* Partitioning heritability by functional annotation using genome-
961 wide association summary statistics. *Nat Genet* **47**, 1228-1235 (2015).
962 31. Akiyama, M. *et al.* Genome-wide association study identifies 112 new loci for body
963 mass index in the Japanese population. *Nat Genet* **49**, 1458-1467 (2017).
964 32. Suyama, S. & Yada, T. New insight into GABAergic neurons in the hypothalamic
965 feeding regulation. *J Physiol Sci* **68**, 717-722 (2018).
966 33. Matikainen-Ankney, B. A. et al. Nucleus Accumbens D1 Receptor-Expressing Spiny
967 Projection Neurons Control Food Motivation and Obesity. *Biol Psychiatry* **93**, 512-523
968 (2023).
969 34. Kim, J. A., Montagnani, M., Koh, K. K. & Quon, M. J. Reciprocal relationships between
970 insulin resistance and endothelial dysfunction: molecular and pathophysiological
971 mechanisms. *Circulation* **113**, 1888-1904 (2006).
972 35. van Walree, E. S. *et al*. Disentangling Genetic Risks for Metabolic Syndrome. *Diabetes*
973 **71**, 2447-2457 (2022).
974 36. Wojcik, G. L. *et al.* Genetic analyses of diverse populations improves discovery for
975 complex traits. *Nature* **570**, 514-518 (2019).
976 37. Petersen, R. E. *et al*. Genome-wide Association Studies in Ancestrally Diverse
977 Populations: Opportunities, Methods, Pitfalls, and Recommendations. *Cell* **179**, 589-
978 603 (2019).
979 38. Perry, J. R. B. *et al.* Stratifying type 2 diabetes cases by BMI identifies genetic risk
980 variants in LAMA1 and enrichment for risk variants in lean compared to obese cases.
981 *PLoS Genet* **8**, e1002741 (2012).

982 39. Aragam, K. G. *et al.* Discovery and systematic characterization of risk variants and
983 genes for coronary artery disease in over a million participants. *Nat Genet* **54**, 1803-
984 1815 (2022).
985 40. UK Prospective Diabetes Study (UKPDS) Group. Intensive blood-glucose control with
986 sulphonylureas or insulin compared with conventional treatment and risk of
987 complications in patients with type 2 diabetes (UKPDS 33). UK Prospective Diabetes
988 Study (UKPDS) Group. *Lancet* **352**, 837-853 (1998).
989 41. Pe'er, I., Yelensky, R., Altshuler, D. & Daly, M. J. Estimation of the multiple testing
990 burden for genomewide association studies of nearly all common variants. *Genet*
991 *Epidemiol* **32**, 381-385 (2008).
992 42. Kanai, M., Tanaka, T. & Okada, Y. Empirical estimation of genome-wide significance
993 thresholds based on the 1000 Genomes Project data set. *J Hum Genet* **61**, 861-866
994 (2016).
995 43. Carlson, J., Henn, B. M., Al-Hindi, R. R. & Ramachandran, S. Counter the
996 weaponization of genetics research by extremists. *Nature* **610**, 444-447 (2022).
997 44. Lewis, A. C. F. *et al.* Getting genetic ancestry right for science and society. *Science*
998 **376**, 250-252 (2022).
999 45. Moller, J. B. *et al*. Ethnic differences in insulin sensitivity, β-cell function, and hepatic
1000 extraction between Japanese and Caucasians: a minimal model analysis. *J Clin*
1001 *Endocrinol Metab* **99**, 4273-4280 (2014).
1002 46. Yabe, D., Seino, Y., Fukushima, M. & Seino, S. β cell dysfunction versus insulin
1003 resistance in the pathogenesis of type 2 diabetes in East Asians. *Curr Diab Rep* **15**,
1004 602 (2015).
1005 47. Moller, J. B. *et al*. Body composition is the main determinant for the difference in
1006 type 2 diabetes pathophysiology between Japanese and Caucasians. *Diabetes Care*
1007 **37**, 796-804 (2014).
1008 48. Matsunaga, H. *et al.* Transethnic meta-analysis of genome-wide association studies
1009 identifies three new loci and characterizes population-specific differences for
1010 coronary artery disease. *Circ Genom Precis Med* **13**, e002670 (2020).
1011 49. Koyama, S. *et al.* Population-specific and trans-ancestry genome-wide analyses
1012 identify distinct and shared genetic risk loci for coronary artery disease. *Nat Genet*
1013 **52**, 1169-1177 (2020).
1014 50. Morris, A. P. *et al.* Trans-ethnic kidney function association study reveals putative
1015 causal genes and effects on kidney-specific disease aetiologies. *Nat Commun* **10**, 29
1016 (2019).

METHODS

Ethics statement. Study-level ethics statements are provided in the **Supplementary Note**.

Study-level analyses. Within each study, we assigned individuals to ancestry groups using self-report and genetic background (**Supplementary Tables 1 and 2**). Any individuals not assigned to an ancestry group were excluded as population outliers. Within each ancestry group-specific GWAS, we conducted quality control of genotype data and imputed up to reference panels from the Trans-Omics for Precision Medicine Program[51], Haplotype Reference Consortium[52], 1000 Genomes Project (phase 1, March 2012 release; phase 3, October 2014 release)[53,54], or population-specific whole-genome sequencing[55-61] (**Supplementary Table 3**). Studies imputed to reference panels mapped to GRCh38/hg38 were lifted back to hg19 using the UCSC liftOver tool (https://genome.ucsc.edu/cgi-bin/hgLiftOver). We excluded SNVs with poor imputation quality and/or minor allele count (MAC) <5 (**Supplementary Table 3**).

Within each ancestry group-specific GWAS, we tested for association of each SNV with T2D via generalised linear (mixed) modelling, under an additive dosage of the minor allele, with adjustment for age and sex (where appropriate), and additional study-specific covariates (**Supplementary Table 3**). We employed different strategies to account for population stratification and/or kinship: (i) exclude closely related individuals and adjust for principal components derived from a genetic relatedness matrix (GRM) as additional covariates; or (ii) incorporate a random effect for the GRM (**Supplementary Table 3**). Allelic effects and corresponding standard errors that were estimated from a linear mixed model were converted to the log-odds scale[62]. We corrected study-level association summary statistics for residual structure by the LD-score regression intercept[63] (**Supplementary Table 3**) using an LD reference that we derived from ancestry-matched haplotypes from continental groups in the 1000 Genomes Project (phase 3, October 2014 release)[54]. We matched AFA GWAS to the “African” continental group and HIS GWAS to the “American” continental group.

Multi-ancestry meta-analyses. We analysed autosomal bi-allelic SNVs that overlap reference panels from the 1000 Genomes Project (phase 3, October 2014 release)[54] and the Haplotype Reference Consortium[52]. We considered SNVs with MAF >0.5% in at least one of the five continental groups in the 1000 Genomes Project (phase 3, October 2014 release)[54]. We excluded SNVs that differed in allele frequency by >20% when comparing reference panels in the same subsets of haplotypes.

We used meta-regression, implemented in MR-MEGA[10], to aggregate association summary statistics across GWAS. MR-MEGA models allelic effect heterogeneity that is correlated with genetic ancestry by including axes of genetic variation as covariates in the meta-regression model to capture diversity between GWAS. We used SNVs reported in all studies to construct a distance matrix of mean effect allele frequency differences between each pair of GWAS. We implemented multi-dimensional scaling of the distance matrix to obtain three principal components that represent axes of genetic variation to separate GWAS across ancestry groups (**Extended Data Figure 1**).

For each SNV, we aggregated inverse-variance weighted allelic effects across GWAS via linear regression, including three axes of genetic variation as covariates. We tested for: (i) association with T2D allowing for ancestry-correlated allelic effect heterogeneity

1064 between GWAS; (ii) ancestry-correlated allelic effect heterogeneity between GWAS (defined
1065 by the axes of genetic variation); and (iii) residual allelic effect heterogeneity between
1066 GWAS. MR-MEGA is a meta-regression approach, and therefore does not produce an allelic
1067 effect estimate because this is allowed to vary with the axes of genetic variation.
1068 Consequently, we also aggregated association summary statistics across GWAS via fixed-
1069 effects meta-analysis (inverse-variance weighting of allelic effects) using METAL[64]. To assess
1070 the extent of residual structure between GWAS, we calculated the genomic control inflation
1071 factor[65] for the multi-ancestry meta-regression and fixed-effects meta-analysis. We
1072 considered only those SNVs reported in at least five GWAS for downstream interrogation.
1073
1074 **Defining T2D signals and loci.** We identified all SNVs attaining genome-wide significance
1075 (P<5x10^{-8}) for association with T2D from the multi-ancestry meta-regression. Clumps were
1076 formed around index variants, which were selected using a greedy algorithm in PLINKv1.9[66],
1077 after ranking SNVs by ascending P-value. SNVs <5Mb from an index SNV were assigned to
1078 the clump if r^2>0.05 in at least one of the five continental groups from the 1000 Genomes
1079 Project (phase 3, October 2014 release)[54]. Index SNVs separated by <1Mb were assigned to
1080 the same locus. Each locus was then defined as mapping 500kb up- and down-stream of
1081 index SNVs contained within it. We considered the locus to have been previously reported if
1082 it contained variants discovered in published large-scale T2D GWAS at genome-wide
1083 significance.
1084
1085 **Ancestry group-specific meta-analyses.** We aggregated association summary statistics
1086 across GWAS from the same ancestry group via fixed-effects meta-analysis (inverse-variance
1087 weighting of allelic effects) using METAL[64]. We estimated the mean effect allele frequency
1088 across GWAS from each ancestry group, weighted by the effective sample size of the study.
1089 We generated forest plots of association summary statistics of index SNVs across ancestry
1090 groups using the R package meta (https://cran.r-project.org/package=meta/).
1091
1092 **Defining clusters of T2D index SNVs with distinct cardiometabolic profiles.** We considered
1093 cardiometabolic-related quantitative phenotypes that are used to define T2D status and/or
1094 are associated with risk of T2D or complications. We excluded phenotypes for which GWAS
1095 summary statistics were available only after imputation to reference panels from the
1096 International HapMap Project[67] because they did not provide sufficient coverage of SNVs
1097 included in the multi-ancestry meta-analysis. We considered the largest available GWAS
1098 meta-analysis (ancestry-specific or multi-ancestry) that provided the following association
1099 summary statistics for each SNV: effect allele, other allele, allelic effect, and corresponding
1100 standard error (**Supplementary Table 5**). We re-aligned the effect estimate to the T2D risk
1101 allele from the fixed-effects multi-ancestry meta-analysis, denoted β_{ij} for the jth index SNV
1102 and the ith phenotype. We then calculated a sample size corrected Z-score, given by $Z_{ij} =$
1103 $\beta_{ij}/\left(\sqrt{N_i}s_{ij}\right)$, where s_{ij} is the standard error of the effect estimate of the jth index SNV and
1104 the ith phenotype, and N_i is the maximum sample size reported for the ith phenotype.
1105 Where association summary statistics were not reported, the Z-score was set as “missing”.
1106 We conducted K-means clustering of index SNVs with imputation of missing Z-scores
1107 using the R package ClustImpute (https://cran.r-project.org/package=ClustImpute). For a
1108 pre-defined number of clusters, ClustImpute replaces missing Z-scores at random from the
1109 marginal distribution for the phenotype in the first iteration and performs K-means
1110 clustering. In subsequent iterations, missing Z-scores are updated, conditional on the

1111 current cluster assignment, so that correlations between phenotypes are considered. At
1112 each iteration, penalizing weights are imposed on imputed values and successively
1113 decreased (to zero) as the missing data imputation improves. Finally, we determined the
1114 "optimal" number of clusters according to the majority rule across 27 indices of cluster
1115 performance[68], implemented in the R package NbClust (https://cran.r-
1116 project.org/package=NbClust).

1117 We tested for association of the ith phenotype with index SNVs across clusters in a
1118 linear regression model, given by $E(Z_{ij}) = \sum_k \gamma_{ik} C_{jk}$, where C_{jk} is an indicator variable that
1119 takes the value "1" if the jth index SNV was assigned to the kth cluster and "0" otherwise.
1120 The strength/direction of the association of each phenotype with each cluster was then
1121 presented in a heatmap, where the "temperature" was defined by the direction of the
1122 regression coefficient γ_{ik} and the corresponding -$\log_{10}$ *P*-value. Regression models were
1123 fitted using the glm function in R.

1124 We extracted cardiometabolic phenotype Z-scores from the final imputed dataset
1125 from ClustImpute. We calculated the Euclidean distance between the jth SNV and kth
1126 cluster centroid as
1127

1128 $$\delta_{jk} = \sqrt{\sum_i (Z_{ij} - \mu_{ik})^2},$$

1129

1130 where Z_{ij} and μ_{ik} are the Z-score of the jth SNV and the location of the kth cluster centroid
1131 for the ith cardiometabolic phenotype. To assess cluster disparity, we also performed
1132 principal components analysis of cardiometabolic phenotype Z-scores from the final
1133 imputed dataset using the R package factoextra (https://cran.r-
1134 project.org/package=factoextra).
1135

1136 **Cluster-specific associations of index SNVs with T2D.** We tested for association of T2D with
1137 index SNVs across clusters in a linear regression model, given by $E(\beta_j) = \sum_k \gamma_k C_{jk}$, where
1138 C_{jk} is an indicator variable that takes the value "1" if the jth index SNV was assigned to the
1139 kth cluster and "0" otherwise, and weighted by the inverse of the variance of the allelic
1140 effect. We tested for heterogeneity in cluster effects on T2D by comparing the deviance of
1141 this model with that of $E(\beta_j) = \gamma_0$, again weighted by the inverse of the variance of the
1142 allelic effect. To compare associations between previously-reported clusters and previously-
1143 unreported clusters, we replaced C_{jk} by an indicator variable that takes the value "1" if the
1144 jth index SNV was assigned to a previously-reported cluster and "0" otherwise. Regression
1145 models were fitted using the glm function in R.
1146

1147 **Enrichment of T2D associations for cell type-specific regions of open chromatin within**
1148 **clusters.** For each T2D association signal, we defined "null" SNVs that mapped within 50kb
1149 of the index SNV and were not in LD (r^2>0.05) with the index SNV in any of the five
1150 continental groups from the 1000 Genomes Project (phase 3, October 2014 release)[54]. We
1151 defined an indicator variable, Y_j, taking the value "1" if the jth SNV is an index SNV and "0"
1152 if the jth SNV is a null SNV. We mapped index SNVs and null SNVs to genic regions defined
1153 by the Ensembl Project (release 104)[69], including protein-coding exons, and 3' UTRs and 5'
1154 UTRs. We defined indicator variables, G_j^{EXON}, G_j^{3UTR}, and G_j^{5UTR}, that each take the value
1155 "1" if the jth SNV mapped to the respective genic annotation and "0" otherwise. We also

1156 mapped index SNVs and null SNVs to ATAC-seq peaks from single-cell atlases of chromatin
1157 accessibility (CATLAS and DESCARTES) for: 222 cell types derived from 30 human adult and
1158 15 human foetal tissues[25,26]; and 106 cell types derived from human brain[27]. We defined an
1159 indicator variable, X_{ij}, that takes the value "1" if the jth SNV mapped to an ATAC-seq peak
1160 for the ith cell type and "0" otherwise.

1161 Within each cluster, we modelled enrichment of T2D associations for ATAC-seq
1162 peaks in the ith cell type, after accounting for genic annotations, in a Firth bias-reduced
1163 logistic regression, given by
1164
1165 $$f^{-1}(Y_j) = \alpha_0 + \alpha_{\mathrm{EXON}} G_j^{\mathrm{EXON}} + \alpha_{\mathrm{3UTR}} G_j^{\mathrm{3UTR}} + \alpha_{\mathrm{5UTR}} G_j^{\mathrm{5UTR}} + \theta_i X_{ij},$$
1166
1167 where f is the logit link function. In this expression, α_0 is an intercept, α_{EXON}, α_{3UTR}, and
1168 α_{5UTR} are log-fold enrichments of genic annotations, and θ_i is the log-fold enrichment of
1169 ATAC-seq peaks in the ith cell type. We conducted a test of enrichment of the ith cell type
1170 by comparing the deviances of models in which $\theta_i = 0$ and θ_i is unconstrained. We
1171 identified cell types with significant evidence of enrichment (*P*<0.00023, Bonferroni
1172 correction for 222 cell types in adult/foetal tissues; *P*<0.00047, Bonferroni correction for
1173 106 cell types in brain). All models were fitted using the R package logistf (https://cran.r-
1174 project.org/package=logistf).
1175
1176 **Contribution of each axis of genetic variation to ancestry-correlated heterogeneity.** For
1177 each index SNV, we calculated a Z-score (beta/SE) for association with each axis of variation
1178 by aligning the effect from the meta-regression model to the T2D-risk allele. For each index
1179 SNV, we identified the axis of genetic variation with the strongest association (greatest
1180 magnitude Z-score).
1181
1182 **Difference in ancestry-correlated heterogeneity between mechanistic clusters.** We tested
1183 for differences in Z-scores (beta/SE) for association of index SNVs in each cluster with the
1184 ith axis of genetic variation by comparing two linear models via ANOVA: (i) $f^{-1}(Z_{ij}) = \tau_{0i}$;
1185 and (ii) $f^{-1}(Z_{ij}) = \sum_k \tau_{ki} C_{jk}$. In these expressions: f is the identity link function; Z_{ij} is Z-
1186 score for the jth index SNV; C_{jk} is an indicator variable that takes the value "1" if the jth
1187 index SNV was assigned to the kth cluster and "0" otherwise; and τ_{0i} and τ_{ki} are regression
1188 coefficients. Regression models were fitted using the glm function in R.
1189
1190 **Impact of BMI on ancestry-correlated and residual heterogeneity in allelic effects between**
1191 **GWAS.** For each index SNV, we aggregated inverse-variance weighted allelic effects across
1192 GWAS via linear regression, implemented in MR-MEGA[10], including as covariates: (i) three
1193 axes of genetic variation; (ii) mean BMI in controls; and (iii) mean BMI in T2D cases. After
1194 adjustment for BMI, we tested for: (i) ancestry-correlated allelic effect heterogeneity
1195 between GWAS; and (ii) residual allelic effect heterogeneity between GWAS. After
1196 adjustment, as outlined above, we re-assessed: (i) the contribution of each axis of genetic
1197 variation to ancestry-correlated heterogeneity; and (ii) the difference in ancestry-correlated
1198 heterogeneity between mechanistic clusters.
1199
1200 **Cluster-specific partitioned PS analyses of vascular outcomes and age of T2D onset.** We
1201 tested for association of cluster-specific components of the partitioned PS and an overall PS
1202 with T2D-related macrovascular outcomes (CAD, ischemic stroke, and peripheral artery

1203 disease), microvascular complications (ESDN and proliferative diabetic retinopathy), and age
1204 of T2D onset in participants from the All of Us Research Program (AoURP; AFA, EUR, and HIS
1205 ancestry groups), Biobank Japan (BBJ; EAS ancestry group), and Genes & Health (G&H; SAS
1206 ancestry group). Cohort descriptions and details of sequencing/genotyping, quality control,
1207 and phenotype derivation are provided in the **Supplementary Methods**.
1208 We conducted analyses separately for each ancestry group in AoURP, BBJ, and G&H.
1209 For each ancestry, we performed analyses for macrovascular outcomes using all individuals,
1210 irrespective of T2D status, and for microvascular complications in individuals with T2D only.
1211 For each analysis, we calculated the overall PS and cluster-specific partitioned PS for each
1212 individual, with each index SNV weighted by the allelic log-OR from the ancestry-specific
1213 meta-analyses. We did not include index SNVs with MAF <1% in the PS. We also excluded
1214 index SNVs with poor imputation quality (r^2<0.7) in BBJ and G&H, and those with extreme
1215 deviation from Hardy-Weinberg equilibrium (P<10^{-6}) in AoURP. We standardised the overall
1216 PS and each cluster-specific component of the partitioned PS to have mean zero and unit
1217 variance. We tested for association with each vascular outcome via generalised linear
1218 regression and with age of T2D onset via linear regression. For each outcome, we
1219 considered a model including the overall PS and then each cluster-specific component the
1220 partitioned PS adjusted for the overall PS. All association analyses were conducted using the
1221 glm function in R.
1222 We adjusted association analyses with vascular outcomes for age, sex, and the first
1223 20 principal components. In BBJ, we also adjusted for recruitment phase and status of the
1224 registered common diseases (other than T2D) to account for ascertainment. We further
1225 adjusted analyses of macrovascular outcomes for T2D status. We also further adjusted
1226 analyses of microvascular complications for duration of T2D. In AoURP, we defined age as
1227 age at last hospital visit. In BBJ, we defined age as age at first record. In G&H, we defined
1228 age as age at diagnosis for T2D cases and age at last follow-up for controls. For CAD, we also
1229 conducted sensitivity analyses by including, as an additional covariate, a CAD PS from the
1230 largest published multi-ancestry CAD GWAS[39]. The PS was constructed from index SNVs for
1231 241 conditionally independent CAD associations, weighted by the multi-ancestry allelic log-
1232 OR (ancestry-specific effects were not available), and standardised to have mean zero and
1233 unit variance. We adjusted association analyses with age of T2D onset for sex and the first
1234 20 principal components. In BBJ, we also adjusted for recruitment phase and status of the
1235 registered common diseases (other than T2D) to account for ascertainment.
1236 For each outcome, we aggregated association summary statistics from each cluster-
1237 specific component of the partitioned PS and the overall PS across ancestries via random-
1238 effects meta-analysis. All meta-analyses were conducted using the R package meta
1239 (https://cran.r-project.org/package=meta).
1240
1241 **Cluster-specific partitioned PS analyses of clinical outcomes.** We tested for association of
1242 cardiovascular and kidney-related clinical outcomes in European ancestry individuals with
1243 T2D in prospective GWAS from six clinical trials from the Thrombolysis in Myocardial
1244 Infarction (TIMI) Study Group (https://timi.org/). Trial descriptions and details of genotyping
1245 and quality control are provided in the **Supplementary Methods**.
1246 Within each trial, we calculated the overall PS and cluster-specific components of the
1247 partitioned PS for each individual, with each index SNV weighted by the allelic log-OR from
1248 the European ancestry-specific meta-analysis. We standardised the overall PS and each
1249 cluster-specific component of the partitioned PS to have mean zero and unit variance. Data

from the six trials were subsequently pooled, and we considered the following clinical outcomes in T2D patients only: myocardial infarction, ischemic stroke, cardiovascular death, hospitalization for heart failure, atrial fibrillation, acute limb ischemia, peripheral revascularization, end-stage renal disease or renal death, and albuminuria. We tested for association of each cluster-specific component of the partitioned PS with each clinical outcome under a Cox proportional hazards model, including age, sex, the first ten principal components, and the overall PS as covariates. All association analyses were conducted using the coxph function with Efron ties handling from the R package survival (https://cran.r-project.org/package=survival).

Data availability. Genome-wide association summary statistics from the multi-ancestry meta-analysis and ancestry-specific meta-analyses reported in this study are available through the DIAGRAM Consortium website (http://www.diagram-consortium.org/downloads.html).

Code availability. Analyses were conducted using publicly available software: UCSC liftOver tool (https://genome.ucsc.edu/cgi-bin/hgLiftOver), MR-MEGA v0.2 (https://genomics.ut.ee/en/tools), METAL v2011-03-25 (https://genome.sph.umich.edu/wiki/METAL), PLINKv1.9 (https://www.cog-genomics.org/plink/1.9/), Beagle 4.1 (https://faculty.washington.edu/browning/beagle/b4_1.html), SNPTEST v2.5.6 (https://www.well.ox.ac.uk/~gav/snptest/), GWAMA v2.2.2 (https://genomics.ut.ee/en/tools), EIGENSOFT v7.2.1 (https://www.hsph.harvard.edu/alkes-price/software/), PLINKv2.0 (https://www.cog-genomics.org/plink/2.0/), SHAPEIT4 (https://odelaneau.github.io/shapeit4/), Minimac4 (https://genome.sph.umich.edu/wiki/Minimac4), KING v2.3 (https://www.kingrelatedness.com/), and EAGLE v2.4 (https://alkesgroup.broadinstitute.org/Eagle/#Xeagle2). Analyses were also conducted using the following R packages: meta (https://cran.r-project.org/package=meta), ClustImpute (https://cran.r-project.org/package=ClustImpute), NbClust (https://cran.r-project.org/package=NbClust), factoextra (https://cran.r-project.org/package=factoextra), and logistf (https://cran.r-project.org/package=logistf).

REFERENCES

51. Taliun, D. *et al.* Sequencing of 53,831 diverse genomes from the NHLBI TOPMed Program. *Nature* **590**, 290-299 (2021).
52. McCarthy, S. *et al.* A reference panel of 64,976 haplotypes for genotype imputation. *Nat Genet* **48**, 1279-1283 (2016).
53. 1000 Genomes Project Consortium *et al.* An integrated map of genetic variation from 1,092 human genomes. *Nature* **491**, 56-65 (2012).
54. 1000 Genomes Project Consortium *et al.* A global reference for human genetic variation. *Nature* **526**, 68-74 (2015).
55. Mitt, M. *et al.* Improved imputation accuracy of rare and low-frequency variants using population-specific high-coverage WGS-based imputation reference panel. *Eur J Hum Genet* **25**, 869-876 (2017).

1297 56. Jónsson, H. *et al.* Whole genome characterization of sequence diversity of 15,220
1298 Icelanders. *Sci Data* **4**, 170115 (2017).
1299 57. Moon, S. *et al.* The Korea Biobank Array: Design and identification of coding variants
1300 associated with blood biochemical traits. *Sci Rep* **9**, 1382 (2019).
1301 58. Akiyama, M. *et al.* Characterizing rare and low-frequency height-associated variants
1302 in the Japanese population. *Nat Commun* **10**, 4393 (2019).
1303 59. Gurdasani, D. *et al.* The African Genome Variation Project shapes medical genetics in
1304 Africa. *Nature* **517**, 327-332 (2015).
1305 60. Gurdasani, D. *et al.* Uganda genome resource enables insights into population
1306 history and genomic discovery in Africa. *Cell* **179**, 984-1002.e1036 (2019).
1307 61. Kurki, M. I. *et al.* FinnGen provides genetic insights from a well-phenotyped isolated
1308 population. *Nature* **613**, 508-518 (2023).
1309 62. Cook, J. P., Mahajan, A. & Morris, A. P. Guidance for the utility of linear models in
1310 meta-analysis of genetic association studies of binary phenotypes. *Eur J Hum Genet*
1311 **25**, 240-245 (2017).
1312 63. Bulik-Sullivan, B. K. *et al.* LD Score regression distinguishes confounding from
1313 polygenicity in genome-wide association studies. *Nat Genet* **47**, 291-295 (2015).
1314 64. Willer, C. J., Li, Y. & Abecasis, G. R. METAL: fast and efficient meta-analysis of
1315 genomewide association scans. *Bioinformatics* **26**, 2190-2191 (2010).
1316 65. Devlin, B. & Roeder, K. Genomic control for association studies. *Biometrics* **55**, 997-
1317 1004 (1999).
1318 66. Chang, C. C. *et al.* Second-generation PLINK: rising to the challenge of larger and
1319 richer datasets. *Gigascience* **4**, 7 (2015).
1320 67. International HapMap Consortium. The International HapMap Project. *Nature* **426**,
1321 789-796 (2003).
1322 68. Charrad, M., Ghazzali, N., Boiteau, V. & Niknafs, A. NbClust: an R package for
1323 determining the relevant number of clusters in a data set. *J Stat Softw* **61**, 1-36
1324 (2014).
1325 69. Cunningham, F. *et al.* Ensembl 2022. *Nucleic Acids Res* **50**, D988-D995 (2022).

1326 **CONSORTIA**
1327
1328 **VA Million Veteran Program.** Kim M. Lorenz[10,11,12], Kyung Min Lee[158,159], Julie A. Lynch[158,159],
1329 Philip S. Tsao[237,289,290], Kyong-Mi Chang[10,291], Marijana Vujkovic[10,291,295], and Benjamin F.
1330 Voight[10,11,12,278]. Contributors to the VA Million Veteran Program are listed in the
1331 **Supplementary Note**.
1332
1333 **AMED GRIFIN Diabetes Initiative Japan.** Ken Suzuki[1,2,3], Kyuto Sonehara[3,17,18,19], Shinichi
1334 Namba[3], Momoko Horikoshi[131], Shiro Maeda[131,179,180], Koichi Matsuda[182], Yukinori
1335 Okada[3,17,18,19,190,191], Nobuhiro Shojima[2], Kenichi Yamamoto[3,190,228], Toshimasa Yamauchi[2],
1336 and Takashi Kadowaki[2,286]. Contributors to AMED GRIFIN Diabetes Initiative Japan are listed
1337 in the **Supplementary Note**.
1338
1339 **Biobank Japan Project.** Koichi Matsuda[182]. Contributors to the Biobank Japan Project are
1340 listed in the **Supplementary Note**.
1341
1342 **Penn Medicine BioBank.** Lindsay A. Guare[125], Daniel J. Rader[12,277,278,279], Marylyn D.
1343 Ritchie[12,280,281], Marijana Vujkovic[10,291,295], and Benjamin F. Voight[10,11,12,278]. Contributors to
1344 the Penn Medicine BioBank are listed in the **Supplementary Note**.
1345
1346 **Regeneron Genetics Center.** Contributors to the Regeneron Genetics Center are listed in the
1347 **Supplementary Note**.
1348
1349 **Genes & Health Research Team.** Stavroula Kanona[16], David A. van Heel[283], and Sarah
1350 Finer[284]. Contributors to the Genes & Health Research Team are listed in the **Supplementary**
1351 **Note**.
1352
1353 **eMERGE Consortium.** Yoonjung Yoonie Joo[63,64,], Abel N. Kho[151,152], and M. Geoffrey
1354 Hayes[249,250,251]. Contributors to the eMERGE Consortium are listed in the **Supplementary**
1355 **Note**.
1356
1357 **International Consortium of Blood Pressure (ICBP).** Zoha Kamali[78,79]. Contributors to ICBP
1358 are listed in the **Supplementary Note.**
1359
1360 **Meta-Analyses of Glucose and Insulin-Related Traits Consortium (MAGIC).** K Alaine
1361 Broadaway[66], Alice Williamson[35,77], Karen L. Mohlke[66], and Claudia Langenberg[35,273,274].
1362 Contributors to MAGIC are listed in the **Supplementary Note.**
1363
1364
1365 **ACKNOWLEDGEMENTS**
1366
1367 Central analyses were supported by Japan Agency for Medical Research and Development
1368 (JP21km0405213, JP20km0405202, and JP21tm0424218), NHGRI (HG011723), American
1369 Diabetes Association Innovative and Clinical Translational Award (1-19-ICTS-068), European
1370 Union's Horizon 2020 research and innovation programme under Grant Agreement No
1371 101017802 (OPTOMICS), American Heart Association Postdoctoral Fellowship
1372 (15POST24470131 and 17POST33650016), American Diabetes Association (11-22-JDFPM-

06), Corporal Michael J Crescenz VA Medical Center Research, NIDDK (DK126194 and DK105535), Versus Arthritis (21754), NIHR Manchester Biomedical Research Centre (NIHR203308), and MRC (MR/W029626/1). A complete list of acknowledgments and funding appears in the **Supplementary Note**.

We thank the International Consortium of Blood Pressure (ICBP) and the Meta-Analyses of Glucose and Insulin-related traits Consortium (MAGIC) for providing pre-publication access to GWAS summary statistics for blood pressure, proinsulin, and post-challenge insulin resistance measures.

The views expressed in this article are those of the authors and do not necessarily represent those of: the UK National Health Service, the UK National Institute for Health Research, or the UK Department of Health and Social Care; the US National Heart, Lung, and Blood Institute, the US National Institute of Neurological Disorders and Stroke, the US National Institute on Aging, the US National Institutes of Health, the US Department of Health and Human Services, the US Department of Veterans Affairs, the US Food and Drug Administration, or the US Government.

AUTHOR CONTRIBUTIONS

Central analysis group: K.Suzuki, K.H., L.S., H.J.T., X.Y., K.M.Lorenz, R.Mandla, A.H.-C., N.W.R., O.B., A.L.A., A.Mahajan, C.N.S., J.M.M., M.B., J.I.R., M.V., B.F.V., A.P.M., E.Z. **Polygenic score analysis:** K.Suzuki, K.H., L.S., H.J.T., G.E.M.M., S.K., A.P.M. **Study-level primary analysis:** K.Suzuki, H.J.T., X.Y., K.M.Lorenz, R.Mandla, A.H.-C., K.Sonehara, S.N., S.S.K.L., M.H.P., L.E.P., P.S., B.V., M.K., F.B., K.Lin, X.G., W.Zhang, J.Y., Y.J.K., M.Graff, F.T., J.N., A.Lamri, M.N., S.Moon, R.A.S., J.P.C., J.-J.L., I.P., D.T., E.J.P., J.-F.C., L.F.B., Y.T., Y.H., G.T., N.G., T.S., M.W., C.Sarnowski, C.G., D.N., S.T., S.-H.K., J.Long, M.Sun, L.T., W.-M.C., S.S.N., R.N., V.J.Y.L., C.H.T.T., Y.Y.J., C.-H.C., L.M.R., B.P.P., A.N., L.R.Y., G.C., J.A.B., E.K., P.A., A.H.X., H.S.C., B.E.C., J.Tan, M.C.Y.N., X.S., A.Mahajan, C.N.S., A.P.M. **Study-level sample collection, phenotyping, genotyping, and additional analysis:** L.S.A., A.A., C.A.A.-S., T.S.A., S.S.A., A.Bertoni, J.B.-J., I.B., T.A.B., C.F.B., A.S.B., M.Canouil, J.C.N.C., L.-C.C., M.-L.C., J.Chen, S.-H.C., Y.-T.C., Z.C., L.-M.C., M.Cushman, J.Danesh, S.K.D., H.J.d.S., G.D., L.D., A.P.D., S.D., Q.D., K.-U.E., L.S.E., D.S.E., M.K.E., K.F., J.S.F., I.F., O.H.F., T.M.F., B.I.F., P.G., H.C.G., V.G., C.G.-V., M.E.G.-V., P.G.-L., M.Gross, L.A.G., S.Hackinger, S.Han, A.T.H., C.H., M.Horikoshi, A.-G.H., W.Hsueh, M.Huang, W.Huang, Y.-J.H., M.Y.H., C.-M.H., S.I., M.A.I., M.Ingelsson, M.T.I., M.Isono, H.-M.J., F.J., G.J., J.B.J., T.J., F.K.K., F.R.K., A.Kasturiratne, T.Katsuya, V.K., T.Kawaguchi, J.M.K., A.N.K., C.-C.K., M.G.K., D.-H.K., F.K., J.K., K.Läll, L.A.L., K.M.Lee, M.-S.L., N.R.L., A.Leong, L.Li, Y.Liu, R.L.-G., S.Lithgart, C.M.L., A.Linneberg, C.-T.L., J.Liu, A.E.L., T.L., J.Luan, A.O.L., X.L., J.Lv, J.A.L., V.L., S.Maeda, V.M., S.R.M., K.M., T.M., A.Metspalu, H.M., A.D.M., F.A.M., J.L.N., M.A.N., U.N., I.N., Y.O., L.O., S.R.P., S.P., P.P., M.A.Pereira, A.P., F.J.P., H.G.P., B.P., G.Prasad, L.J.R.-T., A.P.R., M.R., R.R., K.R., C.Sabanayagam, K.Sandow, A.Sankareswaran, N.Sattar, S.S., M.Shahriar, B.S., J.S., D.M.S., N.Shojima, J.A.S., W.Y.S., A.Stančáková, V.S., A.M.S., K.Strauch, K.D.T., B.Thorand, U.T., B.Tomlinson, T.C.T., F.-J.T., J.Tuomilehto, T.T.-L., M.S.U., A.V., R.M.v.D., J.B.v.K., R.V., N.W.-R., E.W., A.R.W., K.W.v.D., D.R.W., C.S.Y., Ken.Yamamoto, Kenichi.Yamamoto, K.Yoon, C.Y., J.-M.Y., S.Y., M.Z., L.Z., W.Zheng, J.S.P., T.-Y.W., F.S.C., M.O.G., K.L.M., J.C.F., J.E.B., M.I.M., M.V. **Study-level principal investigator:** L.J.R., M.Igase, E.Ipp, S.R., Y.S.C., L.Lind, M.A.Province, M.F., C.L.H., E.Ingelsson, A.B.Z., B.M.P., Y.-X.W., C.N.R., D.M.B., F.M., Y.Liu, M.Y., S.L.R.K., P.A.P., J.S.P.,

1420 J.C.E., A.Bonnefond, P.F., J.G.W., W.H.H.S., J.-Y.W., M.G.H., R.C.W.M., T.-Y.W., D.O.M.-K.,
1421 T.T., G.R.C., F.S.C., D.B., G.Paré, M.M.S., H.A., A.A.M., X.-O.S., K.-S.P., J.W.J., M.Cruz, Y.-D.I.C.,
1422 S.S.R., R.M.-C., H.G., C.-Y.C., M.G., E.-S.T., J.Dupuis, N.K., M.L., A.Köttgen, W.-P.K., D.W.B.,
1423 C.N.A.P., J.S.K., C.K., S.Liu, K.E.N., D.S., T.H., O.P., N.J.W., J.Lee, B.-J.K., I.Y.M., R.G.W.,
1424 K.Stefansson, K.L.M., C.L., C.A.H., R.J.F.L., J.C.F., D.J.R., M.D.R., S.Z., R.Mägi, N.A.M., C.T.R.,
1425 D.A.v.H., S.F., J.C.D., T.Y., T.Kadowaki, J.C.C., J.E.B., P.S.T., K.-M.C., M.I.M., J.B.M., J.M.M.,
1426 M.B., J.I.R., M.V., B.F.V., A.P.M., E.Z. **MAGIC analysis:** K.A.B., A.W., K.L.M., C.L. **ICBP analysis:**
1427 Z.K. **GUARDIAN Consortium analysis:** J.Cui, M.O.G. **GLP-1 GWAS analysis:** M.T., L.H., O.M.,
1428 T.T., E.A. **DIAMANTE Consortium ancestry-specific data co-ordination:** J.C.C., M.C.Y.N., X.S.,
1429 J.E.B., A.Mahajan. **Manuscript preparation:** K.Suzuki, K.H., L.S., H.J.T., X.Y., K.M.Lorenz,
1430 R.Mandla, M.I.M., J.B.M., A.Mahajan, C.N.S., J.M.M., M.B., J.I.R., M.V., B.F.V., A.P.M., E.Z.
1431 **T2DGGI steering group:** M.I.M., J.B.M., A.Mahajan, C.N.S., J.M.M., M.B., J.I.R., M.V., B.F.V.,
1432 A.P.M., E.Z.
1433
1434
1435 **COMPETING INTERESTS**
1436
1437 R.A.S. is now an employee of GlaxoSmithKline. G.T. is an employee of deCODE
1438 genetics/Amgen Inc. A.S.B. reports institutional grants from AstraZeneca, Bayer, Biogen,
1439 BioMarin, Bioverativ, Novartis, Regeneron and Sanofi. J.Danesh serves on scientific advisory
1440 boards for AstraZeneca, Novartis, and UK Biobank, and has received multiple grants from
1441 academic, charitable and industry sources outside of the submitted work. L.S.E. is now an
1442 employee of Bristol Myers Squibb. J.S.F. has consulted for Shionogi Inc. T.M.F. has consulted
1443 for Sanofi, Boehringer Ingelheim, and received funding from GlaxoSmithKline. H.C.G. holds
1444 the McMaster-Sanofi Population Health Institute Chair in Diabetes Research and Care;
1445 reports research grants from Eli Lilly, AstraZeneca, Merck, Novo Nordisk and Sanofi;
1446 honoraria for speaking from AstraZeneca, Boehringer Ingelheim, Eli Lilly, Novo Nordisk,
1447 DKSH, Zuellig, Roche, and Sanofi; and consulting fees from Abbott, AstraZeneca, Boehringer
1448 Ingelheim, Eli Lilly, Merck, Novo Nordisk, Pfizer, Sanofi, Kowa and Hanmi.lth Institute Chair
1449 in Diabetes Research and Care; reports research grants from Eli Lilly, AstraZeneca, Merck,
1450 Novo Nordisk and Sanofi; honoraria for speaking from AstraZeneca, Boehringer Ingelheim,
1451 Eli Lilly, Novo Nordisk, and Sanofi; and consulting fees from Abbott, AstraZeneca, Boehringer
1452 Ingelheim, Eli Lilly, Merck, Novo Nordisk, Janssen, Sanofi, and Kowa. M.Ingelsson is a paid
1453 consultant to BioArctic AB. R.L.-G. is a part-time consultant of Metabolon Inc. A.E.L. is now
1454 an employee of Regeneron Genetics Center LLC and holds shares in Regeneron
1455 Pharmaceuticals. M.A.N. currently serves on the scientific advisory board for Clover
1456 Therapeutics and is an advisor to Neuron23 Inc. S.R.P. has received grant funding from
1457 Bayer Pharmaceuticals, Philips Respironics and Respicardia. N.Sattar has consulted for or
1458 been on speakers bureau for Abbott, Amgen, Astrazeneca, Boehringer Ingelheim, Eli Lilly,
1459 Hanmi, Novartis, Novo Nordisk, Sanofi and Pfizer and has received grant funding from
1460 Astrazeneca, Boehringer Ingelheim, Novartis and Roche Diagnostics. V.S. is now an
1461 employee of deCODE genetics/Amgen Inc. A.M.S. receives funding from Seven Bridges
1462 Genomics to develop tools for the NHLBI BioData Catalyst consortium. U.T. is an employee
1463 of deCODE genetics/Amgen Inc. E.Ingelsson is now an employee of GlaxoSmithKline. B.M.P.
1464 serves on the Steering Committee of the Yale Open Data Access Project funded by Johnson
1465 & Johnson. R.C.W.M. reports research funding from AstraZeneca, Bayer, Novo Nordisk,
1466 Pfizer, Tricida Inc. and Sanofi, and has consulted for or received speaker's fees from

AstraZeneca, Bayer, Boehringer Ingelheim, all of which have been donated to the Chinese University of Hong Kong to support diabetes research. D.O.M.-K. is a part-time clinical research consultant for Metabolon Inc. S.Liu reports consulting payments and honoraria or promises of the same for scientific presentations or reviews at numerous venues, including but not limited to Barilla, by-Health Inc, Ausa Pharmed Co.LTD, Fred Hutchinson Cancer Center, Harvard University, University of Buffalo, Guangdong General Hospital and Academy of Medical Sciences, Consulting member for Novo Nordisk, Inc; member of the Data Safety and Monitoring Board for a trial of pulmonary hypertension in diabetes patients at Massachusetts General Hospital; receives royalties from UpToDate; receives an honorarium from the American Society for Nutrition for his duties as Associate Editor. K.Stefansson is an employee of deCODE genetics/Amgen Inc. M.I.M. has served on advisory panels for Pfizer, NovoNordisk and Zoe Global, has received honoraria from Merck, Pfizer, Novo Nordisk and Eli Lilly, and research funding from Abbvie, Astra Zeneca, Boehringer Ingelheim, Eli Lilly, Janssen, Merck, NovoNordisk, Pfizer, Roche, Sanofi Aventis, Servier, and Takeda; and is now an employee of Genentech and a holder of Roche stock. J.B.M. is an Academic Associate for Quest Diagnostics R&D. A.Mahajan is an employee of Genentech, and a holder of Roche stock. All other authors declare no competing interests.

ADDITIONAL INFORMATION

Supplementary Information is available for this paper.

Correspondence and requests for materials should be addressed to: Andrew P. Morris (andrew.morris-5@manchester.ac.uk), Konstantinos Hatzikotoulas (konstantinos.hatzikotoulas@helmholtz-munich.de), Benjamin F. Voight (bvoight28@gmail.com), and Eleftheria Zeggini (eleftheria.zeggini@helmholtz-munich.de).

Reprints and permissions information is available at www.nature.com/reprints.

EXTENDED DATA

Extended Data Figure 1. Axes of genetic variation separating GWAS of T2D across ancestry groups. We used SNVs that were reported in all studies to construct a distance matrix of mean effect allele frequency differences between each pair of GWAS. We implemented multi-dimensional scaling of the distance matrix to principal components that represent axes of genetic variation. The first three axes of genetic variation (PC1, PC2 and PC3) from multi-dimensional scaling of the Euclidean distance matrix between populations are sufficient to separate GWAS from six ancestry groups: African American (AFA), East Asian (EAS), European (EUR), Hispanic (HIS), South African (SAF), and South Asian (SAS). Variance explained by each axis: PC1 90.7%; PC2 6.5%; PC3 1.0%.

Extended Data Figure 2. Cluster-specific associations of index SNVs with defining cardiometabolic phenotypes. Each bar presents $-\log_{10}$ *P*-value for association, with effect direction aligned to the T2D risk allele. FG: fasting glucose. FI: fasting insulin. PI: proinsulin. BMI: body mass index. WHR: waist-hip ratio. LDL: low-density lipoprotein cholesterol. HDL: high-density lipoprotein cholesterol. TG: triglycerides. *Trait adjusted for BMI.

Extended Data Figure 3. Cluster-specific associations of index SNVs with T2D. The height of each bar corresponds to the log-odds ratio (beta), and the grey bar shows the 95% confidence interval. *$P<0.05$, nominal association. **$P<0.0063$, Bonferroni correction for eight clusters. Exact *P*-values are presented in **Supplementary Table 9**.

Extended Data Figure 4. Cluster-specific associations of T2D risk alleles at index SNVs with insulin-related endophenotypes. Measures of insulin secretion and insulin sensitivity were derived from hyperinsulinemic-euglycemic clamp assessments and oral glucose tolerance tests in up to 1,316 Mexican American participants without diabetes. Homeostatic model assessment measures of beta-cell function (HOMA-B) and insulin resistance (HOMA-IR) were obtained from 36,466 non-diabetic individuals of European ancestry. Each point corresponds to the cluster-specific mean Z-score for each trait, and grey bars represent 95% confidence intervals. The liver/lipid metabolism cluster has been removed for ease of presentation.

Extended Data Figure 5. Cluster-specific associations of T2D risk alleles at index SNVs with insulin resistance-related disorders. Association with gestational diabetes mellitus (GDM) was assessed in 5,485 cases and 347,856 female controls of diverse ancestry. Association with polycystic ovary syndrome (PCOS) was assessed in 10,074 cases and 103,164 female controls of European ancestry. The height of each bar corresponds to the mean Z-score, and the grey bar shows the 95% confidence interval. The liver/lipid metabolism cluster has been removed for ease of presentation. *$P<0.05$, nominal association. **$P<0.0063$, Bonferroni correction for eight clusters. Exact *P*-values are presented in **Supplementary Table 12**.

Extended Data Figure 6. Ancestry-correlated heterogeneity is driven by differences in allelic effect sizes between African American, East Asian, and European ancestry groups. In the scatterplot, index SNVs with significant evidence ($P_{HET}<3.9\times10^{-5}$, Bonferroni correction for 1,289 signals) for ancestry-correlated heterogeneity are plotted according to their association (Z-score) with the first two axes of genetic variation. The first axis represents

differences in allelic effect sizes between AFA/EUR and EAS GWAS (AFA-EAS axis), whilst the second axis represents differences in effect size between AFA/EAS and EUR GWAS (AFA-EUR axis). The forest plots present examples of ancestry-correlated heterogeneity at index SNVs. In each forest plot, the allelic log-odds ratio (OR) from each ancestry group-specific fixed-effects meta-analysis is given by the black tick mark, the 95% confidence interval is given by the horizontal line, and the weight (inverse-variance) of each ancestry group by the grey box. AFA: African American ancestry group. EAS: East Asian ancestry group. EUR: European ancestry group. HIS: Hispanic ancestry group. SAF: South African ancestry group. SAS: South Asian ancestry group.

Extended Data Figure 7. Cluster-specific associations of index SNVs with the first two axes of genetic variation: (a) unadjusted for BMI; and (b) adjusted for study-level mean BMI in T2D cases and controls. Each point corresponds to a cluster, plotted according to the mean Z-score for association with the first two axes of genetic variation (PC1 and PC2) on the x-axis and y-axis, respectively. Grey bars correspond to 95% confidence intervals. The liver/lipid metabolism cluster has been removed for ease of presentation.

Extended Data Figure 8. Associations of cluster-specific components of the partitioned PS with CAD in up to 279,552 individuals across diverse ancestry groups. The panel summarizes the associations of each cluster-specific component of the partitioned PS with CAD, with and without adjustment for a previously-published multi-ancestry CAD PS. The height of each bar corresponds to the log-OR (beta) per standard deviation of the PS, and the grey bar shows the 95% confidence interval. Analyses were undertaken in all individuals, with adjustment for T2D status. **P*<0.05, nominal association. ***P*<0.0063, Bonferroni correction for eight clusters. Exact *P*-values are presented in **Supplementary Tables 21 and 22**.

Extended Data Figure 9. Associations of cluster-specific components of the partitioned PS with T2D age of onset in up to 30,288 individuals across diverse ancestry groups. The panel summarizes the associations of each cluster-specific component of the partitioned PS with age of onset. The height of each bar corresponds to years (beta) per standard deviation of the PS, and the grey bar shows the 95% confidence interval. A negative effect corresponds to earlier age of onset. **P*<0.05, nominal association. ***P*<0.0063, Bonferroni correction for eight clusters. Exact *P*-values are presented in **Supplementary Table 23**.

Extended Data Figure 10. Associations of the beta cell +PI and obesity cluster-specific components of the partitioned PS with vascular outcomes in up to 29,827 EUR individuals with T2D from six clinical trials from the Thrombolysis in Myocardial Infarction (TIMI) Study Group. Major cardiovascular event is defined as myocardial infarction, ischaemic stroke, or cardiovascular death. Major limb event is defined as acute limb ischaemia or peripheral revascularization. The height of each bar corresponds to the log-hazard ratio per standard deviation of the PS, and the grey bar shows the 95% confidence interval. **P*<0.05, nominal association. ***P*<0.0063, Bonferroni correction for eight clusters. Exact *P*-values are presented in **Supplementary Table 24**.