

Tissue-Specific Alteration of Metabolic Pathways Influences Glycemic Regulation

Natasha H. J. Ng^{1,2*}, Sara M. Willems^{3*}, Juan Fernandez⁴, Rebecca S. Fine^{5,6,7}, Eleanor Wheeler^{8,3}, Jennifer Wessel^{9,10}, Hidetoshi Kitajima⁴, Gaelle Marenne⁸, Jana K. Rundle¹, Xueling Sim^{11,12}, Hanieh Yeghootkar¹³, Nicola L. Beer¹, Anne Raimondo¹, Andrei I. Tarasov¹, Soren K. Thomsen^{1,14}, Martijn van de Bunt^{4,1}, Shuai Wang¹⁵, Sai Chen¹², Yuning Chen¹⁵, Yii-Der Ida Chen¹⁶, Hugoline G. de Haan¹⁷, Niels Grarup¹⁸, Ruifang Li-Gao¹⁷, Tibor V. Varga¹⁹, Jennifer L Asimit^{8,20}, Shuang Feng²¹, Rona J. Strawbridge^{22,23}, Erica L. Kleinbrink²⁴, Tarunveer S. Ahluwalia^{25,26}, Ping An²⁷, Emil V. Appel¹⁸, Dan E Arking²⁸, Juha Auvinen^{29,30}, Lawrence F. Bielak³¹, Nathan A. Bihlmeyer²⁸, Jette Bork-Jensen¹⁸, Jennifer A. Brody^{32,33}, Archie Campbell³⁴, Audrey Y Chu³⁵, Gail Davies^{36,37}, Ayse Demirkan³⁸, James S. Floyd^{32,33}, Franco Giulianini³⁵, Xiuqing Guo¹⁶, Stefan Gustafsson³⁹, Benoit Hastoy¹, Anne U. Jackson¹², Johanna Jakobsdottir⁴⁰, Marjo-Riitta Jarvelin^{41,29,42}, Richard A. Jensen^{32,33}, Stavroula Kanoni⁴³, Sirkka Keinanen-Kiukaanniemi^{44,45}, Jin Li⁴⁶, Man Li^{47,48}, Kurt Lohman⁴⁹, Yingchang Lu^{50,51}, Jian'an Luan³, Alisa K. Manning^{52,53}, Jonathan Marten⁵⁴, Carola Marzi^{55,56}, Karina Meidtner^{57,56}, Dennis O. Mook-Kanamori^{17,58}, Taulant Muka^{59,38}, Giorgio Pistis^{60,61}, Bram Prins⁸, Kenneth M. Rice^{32,62}, Neil Robertson⁴, Serena Sanna^{60,63}, Yuan Shi⁶⁴, Albert Vernon Smith^{40,65}, Jennifer A. Smith^{31,66}, Lorraine Southam^{8,4,67}, Heather M. Stringham¹², Salman M. Tajuddin⁶⁸, Vinicius Tragante⁶⁹, Sander W. van der Laan^{70,71}, Helen R. Warren^{43,72}, Jie Yao¹⁶, Andrianos M. Yiorkas^{73,74}, Weihua Zhang^{75,76}, Wei Zhao³¹, Emma Ahlqvist⁷⁷, Mariaelisa Graff⁷⁸, Heather M. Highland^{78,79}, Anne E. Justice⁷⁸, Ken Sin Lo⁸⁰, Eirini Marouli⁴³, Carolina Medina-Gomez^{38,81}, Saima Afaq⁴¹, Wesam A Alhejily^{43,82}, Najaf Amin³⁸, Folkert W. Asselbergs^{69,83,84}, Lori L. Bonnycastle⁸⁵, Michiel L. Bots⁸⁶, Ivan Brandslund⁸⁷, Ji Chen⁸, Cramer Christensen⁸⁸, John Danesh⁸⁹, Renée de Mutsert¹⁷, Abbas Dehghan^{38,90,91}, Tapani Ebeling⁹², Paul Elliott^{41,93,94}, EPIC-InterAct Consortium, Aliko-Eleni Farmaki^{95,96}, Jessica D. Faul⁶⁶, Paul W. Franks^{19,97,6}, Steve Franks⁹⁸, Andreas Fritsche^{99,56}, Anette P. Gjesing¹⁸, Mark O. Goodarzi¹⁰⁰, Vilmundur Gudnason^{40,65}, Göran Hallmans¹⁰¹, Tamara B. Harris⁴⁰, Karl-Heinz Herzig^{102,103}, Marie-France Hivert^{104,105}, Jan-Håkan Jansson¹⁰⁶, Min A Jhun^{31,107}, Torben Jørgensen^{108,109,110}, Marit E. Jørgensen^{26,111}, Pekka Jousilahti¹¹², Eero Kajantie^{113,112,114,115}, Maria Karaleftheri¹¹⁶, Sharon L.R. Kardia³¹, Leena Kinnunen¹¹², Heikki A. Koistinen^{112,117,118}, Pirjo Komulainen¹¹⁹, Peter Kovacs^{120,121}, Johanna Kuusisto¹²², Markku Laakso¹²², Leslie A. Lange¹²³, Lenore J. Launer⁴⁰, Jung-Jin Lee¹²⁴, Aaron Leong¹²⁵, Jaana Lindström¹¹², Jocelyn E. Manning Fox^{126,127}, Satu Männistö¹¹², Nisa M Maruthur^{128,129,48}, Leena Moilanen¹³⁰, Antonella Mulas^{60,131}, Mike A. Nalls^{132,133}, Matthew Neville¹, James S. Pankow¹³⁴, Alison Pattie³⁷, Eva R.B. Petersen¹³⁵, Hannu Puolijoki¹³⁶, Asif Rasheed¹³⁷, Paul Redmond³⁷, Frida Renström^{19,101}, Michael Roden^{138,139,56}, Danish Saleheen^{124,137}, Juha Saltevo¹⁴⁰, Kai Savonen^{119,141}, Sylvain Sebert^{44,42}, Tea Skaaby¹⁰⁸, Kerrin S Small¹⁴², Alena Stančáková¹²², Jakob Stokholm²⁵, Konstantin Strauch¹⁴³, E-Shyong Tai^{144,11,145}, Kent D. Taylor¹⁶, Betina H. Thuesen¹⁰⁸, Anke Tönjes¹²¹, Emmanouil Tsafantakis¹⁴⁶, Tiinamaija Tuomi^{147,148,149}, Jaakko Tuomilehto^{112,150,151,152}, Understanding Society Scientific Group, Matti Uusitupa¹⁵³, Marja Vääräsmäki^{154,155}, Ilonca Vaartjes⁸⁶, Magdalena Zoledziewska⁶⁰, Goncalo Abecasis⁶¹, Beverley Balkau¹⁵⁶, Hans Bisgaard²⁵, Alexandra I. Blakemore^{74,73}, Matthias Blüher^{120,121}, Heiner Boeing¹⁵⁷, Eric Boerwinkle¹⁵⁸, Klaus Bønnelykke²⁵, Erwin P. Bottinger⁵⁰,

Mark J. Caulfield^{43,72}, John C Chambers^{41,76,159}, Daniel I Chasman^{35,160,161}, Ching-Yu Cheng^{64,162,163}, Anne Clark¹, Francis S. Collins⁸⁵, Josef Coresh^{48,129}, Francesco Cucca^{60,131}, Gert J. de Borst¹⁶⁴, Ian J. Deary^{36,37}, George Dedoussis⁹⁵, Panos Deloukas^{43,165}, Hester M. den Ruijter⁷⁰, Josée Dupuis^{15,166}, Michele K. Evans⁶⁸, Ele Ferrannini¹⁶⁷, Oscar H Franco^{38,59}, Harald Grallert^{55,56}, Leif Groop^{168,149}, Torben Hansen^{18,169}, Andrew T. Hattersley¹⁷⁰, Caroline Hayward⁵⁴, Joel N. Hirschhorn^{171,6,7}, Arfan Ikram³⁸, Erik Ingelsson^{46,39,172,173}, Fredrik Karpe^{1,174}, Kay-Tee Kaw¹⁷⁵, Wieland Kiess¹⁷⁶, Jaspal S Kooner^{177,76,159}, Antje Körner¹⁷⁶, Timo Lakka^{178,119,141}, Claudia Langenberg³, Lars Lind¹⁷⁹, Cecilia M Lindgren^{4,180}, Allan Linneberg^{108,181}, Leonard Lipovich^{24,182}, Ching-Ti Liu¹⁵, Jun Liu³⁸, Yongmei Liu¹⁸³, Ruth J.F. Loos^{50,184}, Patrick E. MacDonald^{126,127}, Karen L. Mohlke¹⁸⁵, Andrew D Morris¹⁸⁶, Patricia B. Munroe^{43,72}, Alison Murray¹⁸⁷, Sandosh Padmanabhan¹⁸⁸, Colin N A Palmer¹⁸⁹, Gerard Pasterkamp^{70,71}, Oluf Pedersen¹⁸, Patricia A. Peyser³¹, Ozren Polasek¹⁹⁰, David Porteous³⁴, Michael A. Province²⁷, Bruce M Psaty^{32,33,191,192}, Rainer Rauramaa¹¹⁹, Paul M Ridker^{35,160,193}, Olov Rolandsson¹⁹⁴, Patrik Rorsman^{1,174}, Frits R. Rosendaal¹⁷, Igor Rudan¹⁸⁶, Veikko Salomaa¹¹², Matthias B. Schulze^{57,56}, Robert Sladek^{195,196}, Blair H Smith¹⁸⁹, Timothy D Spector¹⁴², John M. Starr^{197,36}, Michael Stumvoll¹²¹, Cornelia M van Duijn³⁸, Mark Walker¹⁹⁸, Nick J. Wareham³, David R. Weir⁶⁶, James G. Wilson¹⁹⁹, Tien Yin Wong^{64,162,163}, Eleftheria Zeggini^{8,67}, Alan B. Zonderman⁶⁸, Jerome I. Rotter¹⁶, Andrew P. Morris^{200,4}, Michael Boehnke¹², Jose Florez^{201,202}, Mark I McCarthy^{4,1,174}, James B Meigs^{125,202}, Anubha Mahajan⁴, Robert A. Scott³, Anna L Gloyn^{1,4,174*+}, Inês Barroso^{8,3*+}

* Authors contributed equally

+Corresponding authors:

Anna L Gloyn - anna.gloyn@drl.ox.ac.uk

Inês Barroso - ines.barroso@mrc-epid.cam.ac.uk

1. Oxford Centre for Diabetes, Endocrinology and Metabolism, University of Oxford, Oxford, UK, OX3 7LE.
2. Stem Cells and Diabetes Laboratory, Institute of Molecular and Cell Biology, Agency for Science, Technology and Research (A*STAR), Singapore 138673, Singapore
3. MRC Epidemiology Unit, University of Cambridge School of Clinical Medicine, Institute of Metabolic Science, Cambridge Biomedical Campus, Cambridge, CB2 0QQ, UK.
4. Wellcome Centre for Human Genetics, University of Oxford, Oxford, UK, OX3 7BN
5. Department of Genetics, Harvard Medical School, Boston, MA, 02115
6. Division of Endocrinology and Center for Basic and Translational Obesity Research, Boston Children's Hospital, Boston, MA 02115
7. Broad Institute of MIT and Harvard, Cambridge, MA, 02142
8. Department of Human Genetics, Wellcome Sanger Institute, Genome Campus, Hinxton, Cambridge, CB10 1SA, UK
9. Departments of Epidemiology & Medicine, Schools of Public Health & Medicine, Indiana University, Indianapolis, IN 46202
10. Diabetes Translational Research Center, Indiana University School of Medicine, Indianapolis, IN 46202

11. Saw Swee Hock School of Public Health, National University Health System, National University of Singapore, Singapore 117549, Singapore.
12. Department of Biostatistics and Center for Statistical Genetics, University of Michigan, Ann Arbor, MI 48109, USA
13. Genetics of Complex Traits, University of Exeter Medical School, University of Exeter, Exeter, EX2 5DW, UK
14. Present address: Vertex Pharmaceuticals Europe Ltd, Milton Park, Abingdon, OX14 4RW, UK
15. Department of Biostatistics, Boston University School of Public Health, Boston, MA, USA.
16. Institute for Translational Genomics and Population Sciences, Departments of Pediatrics and Medicine, LABiomed at Harbor-UCLA Medical Center, Torrance, CA, USA
17. Department of Clinical Epidemiology, Leiden University Medical Center, Leiden, 2333 ZA, the Netherlands
18. Novo Nordisk Foundation Center for Basic Metabolic Research, Faculty of Health and Medical Sciences, University of Copenhagen, 2200 Copenhagen, Denmark
19. Department of Clinical Sciences, Genetic and Molecular Epidemiology Unit, Lund University, Malmö, SE-205 02, Sweden
20. MRC Biostatistics Unit, Cambridge Biomedical Campus, Cambridge Institute of Public Health, Forvie Site, Robinson Way, Cambridge CB2 0SR, UK.
21. Department of Biostatistics, University of Michigan School of Public Health, Ann Arbor, Michigan, USA
22. UKRI Innovation Fellow at HDR UK, Mental Health and Wellbeing, Institute of Health and Wellbeing, College of Medical, Veterinary and Life Sciences, University of Glasgow, Room 113, 1 Lilybank Gardens, Glasgow, G12 8RZ
23. Cardiovascular Medicine Unit, Department of Medicine Solna, Karolinska Institutet, Stockholm, 171 76, Sweden
24. Center for Molecular Medicine and Genetics, Wayne State University, 540 E. Canfield St., 3208 Scott Hall, Detroit, Michigan 48201-1928, U.S.A.
25. COPSAC, Copenhagen Prospective Studies on Asthma in Childhood, Herlev and Gentofte Hospital 2820, University of Copenhagen, Copenhagen, Denmark
26. Steno Diabetes Center Copenhagen, 2820 Gentofte, Denmark
27. Department of Genetics, Division of Statistical Genomics, Washington University School of Medicine, St. Louis, Missouri 63108, USA
28. McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, MD, USA.
29. Center for Life Course Health Research, University of Oulu, PO Box 8000, 90014 Oulu, Finland
30. Unit of Primary Care, Oulu University Hospital, Oulu, Finland
31. Department of Epidemiology, School of Public Health, University of Michigan, 1415 Washington Hgts, Ann Arbor, MI, 48109, USA
32. Cardiovascular Health Research Unit, University of Washington, 1730 Minor Ave. Suite 1360, Seattle, WA 98101, USA.
33. Department of Medicine, University of Washington, Seattle, WA, USA.

34. Medical Genetics Section, Centre for Genomic and Experimental Medicine, Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh EH4 2XU, UK
35. Division of Preventive Medicine, Brigham and Women's Hospital, Boston MA, 02215, USA.
36. Centre for Cognitive Ageing and Cognitive Epidemiology, University of Edinburgh, Edinburgh EH8 9JZ
37. Department of Psychology, University of Edinburgh, Edinburgh, EH8 9JZ
38. Department of Epidemiology, Erasmus University Medical Center, Rotterdam, 3015 GE, The Netherlands.
39. Department of Medical Sciences, Molecular Epidemiology and Science for Life Laboratory, Uppsala University, Uppsala, 75237, Sweden
40. Icelandic Heart Association, Hotlasmari 1, 201 Kopavogur, Iceland.
41. Department of Epidemiology and Biostatistics, MRC-PHE Centre for Environment & Health, School of Public Health, Imperial College London, London W2 1PG, United Kingdom.
42. Biocenter Oulu, University of Oulu, Oulu, Finland
43. William Harvey Research Institute, Barts and The London School of Medicine and Dentistry, Queen Mary University of London, UK.
44. Faculty of Medicine, Center for Life Course Health Research, University of Oulu, Oulu, Finland
45. MRC and Unit of Primary Care, Oulu University Hospital, Oulu, Finland
46. Department of Medicine, Division of Cardiovascular Medicine, Stanford University School of Medicine, Stanford, CA 94305, USA
47. Division of Nephrology, Internal Medicine, School of Medicine, University of Utah, Salt Lake City, Utah
48. Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, Maryland, USA.
49. Department of Biostatistical Sciences, Public Health Sciences, Wake Forest School of Medicine, Winston-Salem, NC 27157
50. The Charles Bronfman Institute for Personalized Medicine, The Icahn School of Medicine at Mount Sinai, New York, NY, 10069, USA.
51. Department of Medicine, Division of Epidemiology, Vanderbilt-Ingram Cancer Center, Vanderbilt Epidemiology Center, Vanderbilt University School of Medicine Nashville, TN 37203 USA
52. Center for Human Genetics Research, Massachusetts General Hospital, Boston, MA, 02114, USA
53. Department of Medicine, Harvard Medical School, Boston, MA, USA.
54. Medical Research Council Human Genetics Unit, Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh EH4 2XU, UK.
55. Institute of Epidemiology II, Research Unit of Molecular Epidemiology, Helmholtz Zentrum München, German
56. German Center for Diabetes Research (DZD), 85764 München-Neuherberg, Germany.
57. Department of Molecular Epidemiology, German Institute of Human Nutrition Potsdam-Rehbruecke (DIfE), 14558 Nuthetal, Germany.

58. Department of Public Health and Primary Care, Leiden University Medical Center, Leiden, 2333 ZA, the Netherlands
59. Institute of Social and Preventive Medicine (ISPM), University of Bern, Bern, Switzerland
60. Italian National Research Council, Institute of Genetics and Biomedic Research, Cittadella Universitaria, Monserrato (CA) 09042, Italy
61. Center for Statistical Genetics, University of Michigan, 1415 Washington Heights, 48109 Ann Arbor, MI, USA
62. Department of Biostatistics, University of Washington, Seattle, WA, USA.
63. University of Groningen, University Medical Center Groningen, Department of Genetics, 9700 RB Groningen, the Netherlands.
64. Singapore Eye Research Institute, Singapore National Eye Centre, The Academia, 20 College Road, Singapore 169856.
65. Faculty of Medicine, University of Iceland, 101 Reykjavik, Iceland
66. Survey Research Center, Institute for Social Research, University of Michigan, 426 Thompson St, Ann Arbor, MI 48104
67. Institute of Translational Genomics, Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg, Germany
68. Laboratory of Epidemiology and Population Sciences, National Institute on Aging, National Institutes of Health, Baltimore, MD 21224, USA.
69. Department of Cardiology, Division Heart & Lungs, University Medical Center Utrecht, Utrecht University, 3584CX Utrecht, The Netherlands
70. Experimental Cardiology Laboratory, Division Heart and Lungs, University Medical Center Utrecht, 3584 CX Utrecht, Utrecht, the Netherlands
71. Laboratory of Clinical Chemistry and Hematology, University Medical Center Utrecht, 3584 CX Utrecht, Utrecht, the Netherlands
72. NIHR Barts Cardiovascular Research Unit, Barts and The London School of Medicine & Dentistry, Queen Mary University, London, London, EC1M 6BQ, UK
73. Section of Investigative Medicine, Department of Medicine, Imperial College London, London, W12 0NN, UK
74. Department of Life Sciences, Brunel University London, London, UB8 3PH, UK
75. Department of Epidemiology and Biostatistics, Imperial College London, London W2 1PG, UK
76. Ealing Hospital, London North West Healthcare NHS Trust, Middlesex UB1 3HW, UK
77. Lund University Diabetes Centre, Department of Clinical Sciences Malmö, Lund University, Skåne University Hospital, SE-20502 Malmö, Sweden
78. Department of Epidemiology, University of North Carolina at Chapel Hill, Chapel Hill, NC, 27514, USA
79. Human Genetics Center, The University of Texas School of Public Health; The University of Texas Graduate School of Biomedical Sciences at Houston; The University of Texas Health Science Center at Houston, Houston, TX 77030, USA
80. Montreal Heart Institute, Université de Montréal, Montreal, Quebec, H1T 1C8, Canada
81. Department of Internal Medicine, Erasmus University Medical Center, Rotterdam, 3015 GE, The Netherlands
82. Department of Medicine, King Abdulaziz University, Jeddah, 21589, Saudi Arabia

83. Durrer Center for Cardiovascular Research, Netherlands Heart Institute, 3501DG Utrecht, Netherlands
84. Institute of Cardiovascular Science, Faculty of Population Health Sciences, University College London, WC1E 6BT London, United Kingdom
85. Medical Genomics and Metabolic Genetics Branch, National Human Genome Research Institute, NIH, Bethesda, MD 20892, USA
86. Center for Circulatory Health, University Medical Center Utrecht, 3508GA Utrecht, the Netherlands
87. Department of Clinical Biochemistry, Lillebaelt Hospital Vejle, 7100 Vejle, Denmark
88. Medical Department, Lillebaelt Hospital Vejle, 7100 Vejle, Denmark
89. Department of Public Health and Primary Care, University of Cambridge, Cambridge, CB18RN, UK.
90. Department of Biostatistics and Epidemiology, MRC-PHE Centre for Environment and Health, School of Public Health, Imperial College London
91. UK Dementia Research Institute at Imperial College London, London, United Kingdom
92. Oulu University Hospital, 90220 Oulu, Finland
93. Imperial College NIHR Biomedical Research Centre, London, United Kingdom
94. Health Data Research UK – London at Imperial College London, United Kingdom
95. Department of Nutrition and Dietetics, School of Health Science and Education, Harokopio University, Athens, 17671, Greece
96. Department of Population Science and Experimental Medicine, Institute of Cardiovascular Science, University College London, London, UK
97. Department of Nutrition, Harvard School of Public Health, Boston, USA.
98. Institute of Reproductive and Developmental Biology, Imperial College London, W12 0NN London, UK
99. Department of Internal Medicine, Division of Endocrinology, Diabetology, Vascular Medicine, Nephrology, and Clinical Chemistry, University Hospital of Tübingen, Tübingen Germany.
100. Division of Endocrinology, Diabetes and Metabolism, Cedars-Sinai Medical Center, Los Angeles, CA, 90048, USA.
101. Department of Biobank Research, Umeå University, Umeå, SE-901 87, Sweden.
102. Institute of Biomedicine and Biocenter of Oulu, Faculty of Medicine, Medical Research Center Oulu and Oulu University Hospital (K.-H.H.), Oulu, Finland.
103. Department of Gastroenterology and Metabolism, Poznan University of Medical Sciences, 60-572 Poznan, Poland
104. Department of Population Medicine, Harvard Medical School, Harvard Pilgrim Health Care Institute, Boston, MA, USA
105. Diabetes Unit, Massachusetts General Hospital, Boston, MA, USA
106. Department of Public Health and Clinical Medicine, Umeå University, Umeå, SE-901 85, Sweden.
107. Division of Public Health Sciences, Fred Hutchison Cancer Research Center, Seattle, Washington, 98109, USA
108. Center for Clinical Research and Prevention, Bispebjerg and Frederiksberg Hospital, 2000 Frederiksberg, Denmark

109. Department of Public Health, Faculty of Health and Medical Sciences, University of Copenhagen, 2200 Copenhagen, Denmark
110. Faculty of Medicine, University of Aalborg, 9100 Aalborg, Denmark
111. National Institute of Public Health, Southern Denmark University, 5000 Odense, Denmark
112. Department of Public Health Solutions, National Institute for Health and Welfare, FI-00271 Helsinki, Finland
113. PEDEGO Research Unit, MRC Oulu, Oulu University Hospital and University of Oulu, Oulu, Finland
114. Norwegian University of Science and Technology, Department of Clinical and Molecular Medicine, Trondheim, Norway
115. Children's Hospital, Helsinki University Hospital and University of Helsinki
116. Echinops Medical Centre, Echinops, Greece.
117. University of Helsinki and Department of Medicine, Helsinki University Central Hospital, P.O.Box 340, Haartmaninkatu 4, Helsinki, FI-00029, Finland
118. Minerva Foundation Institute for Medical Research, Biomedicum 2U Helsinki, Tukholmankatu 8, FI-00290 Helsinki, Finland
119. Foundation for Research in Health Exercise and Nutrition, Kuopio Research Institute of Exercise Medicine, Kuopio, 70100 Kuopio, Finland
120. Integrated Research and Treatment (IFB) Center AdiposityDiseases, University of Leipzig, Liebigstrasse 19-21, 04103 Leipzig, Germany
121. Department of Medicine, University of Leipzig, Liebigstrasse 18, 04103 Leipzig, Germany
122. Institute of Clinical Medicine, Internal Medicine, University of Eastern Finland, 70210 Kuopio, Finland
123. Department of Medicine, Division of Bioinformatics and Personalized Medicine, University of Colorado Denver, Aurora, CO, USA
124. Department of Biostatistics and Epidemiology, University of Pennsylvania, USA, 19104
125. Division of General Internal Medicine, Massachusetts General Hospital, Department of Medicine, Harvard Medical School, Boston, MA, USA
126. Alberta Diabetes Institute IsletCore, University of Alberta, Edmonton, Canada, T6G 2E1
127. Department of Pharmacology, University of Alberta, Edmonton, Canada, T6G 2E1
128. Department of Medicine, Division of General Internal Medicine, The Johns Hopkins University School of Medicine, Baltimore, Maryland
129. Welch Center for Prevention, Epidemiology, and Clinical Research, Johns Hopkins University, Baltimore, MD, USA.
130. Kuopio University Hospital, 70210 Kuopio, Finland
131. Dipartimento di Scienze Biomediche, Università degli Studi di Sassari, 07100 Sassari, Italy.
132. Laboratory of Neurogenetics, National Institute on Aging, Bethesda, MD, 20892, USA.
133. Data Tecnica International, LLC, Glen Echo, Maryland 20812, USA
134. Division of Epidemiology and Community Health, School of Public Health, University of Minnesota, Minneapolis, MN, 55455, USA.
135. Department of Clinical Biochemistry and Pharmacology, Odense University Hospital, 5000 Odense, Denmark

136. South Ostbothnia Central Hospital, 60220 Seinajoki, Finland
137. Center for Non-Communicable Diseases, Karachi, Pakistan
138. Institute for Clinical Diabetology, German Diabetes Center, Leibniz Institute for Diabetes Research at Heinrich Heine University Düsseldorf, Düsseldorf, Germany
139. Division of Endocrinology and Diabetology, Medical Faculty, University Hospital Düsseldorf, Düsseldorf, Germany
140. Central Finland Central Hospital, 40620 Jyväskylä, Finland
141. Department of Clinical Physiology and Nuclear Medicine, Kuopio University Hospital, 70029 Kuopio, Finland
142. Department of Twin Research and Genetic Epidemiology, King's College London, London, , SE1 7EH, UK
143. Institute of Genetic Epidemiology, Helmholtz Center Munich, German Research Center for Environmental Health, Neuherberg, Germany, German Center for Diabetes Research (DZD e.V.), Neuherberg, Germany
144. Department of Medicine, Yong Loo Lin School of Medicine, National University of Singapore, 1E Kent Ridge Road, Singapore 119228.
145. Duke-NUS Medical School, Singapore, 8 College Road, Singapore 169857
146. Anogia Medical Centre, Anogia, Greece.
147. Folkhälsan Research Centre, Helsinki, Finland
148. Department of Endocrinology, Helsinki University Central Hospital, Helsinki, Finland.
149. Institute for Molecular Medicine Finland FIMM, University of Helsinki, Helsinki, Finland
150. Department of Public Health, University of Helsinki, Helsinki, Finland
151. Department of Neuroscience and Preventive Medicine, Danube-University Krems, 3500 Krems, Austria
152. Saudi Diabetes Research Group, King Abdulaziz University, 21589 Jeddah, Saudi Arabia
153. Department of Public Health and Clinical Nutrition, University of Eastern Finland, 70210 Kuopio, Finland
154. PEDEGO Research Unit (Research Unit for Pediatrics, Dermatology, Clinical Genetics, Obstetrics and Gynecology), Medical Research Center Oulu (MRC Oulu), Oulu University Hospital and University of Oulu, Oulu, Finland.
155. Department of Welfare, Children, Adolescents and Families Unit, National Institute for Health and Welfare, Oulu, Finland
156. INSERM U1018, Centre de recherche en Épidémiologie et Santé des Populations (CESP), Villejuif, France
157. Department of Epidemiology, German Institute of Human Nutrition Potsdam-Rehbrücke (DIfE), 14558 Nuthetal, Germany.
158. The Human Genetics Center and Institute of Molecular Medicine, University of Texas Health Science Center, Houston, Texas, 77030, USA.
159. Imperial College Healthcare NHS Trust, London W12 0HS, UK
160. Harvard School of Medicine, Boston MA, 02115, USA
161. Division of Genetics, Brigham and Women's Hospital and Harvard Medical School, Boston MA, USA.
162. Ophthalmology & Visual Sciences Academic Clinical Program (Eye ACP), Duke-NUS Medical School, Singapore, 8 College Road, Singapore 169857

163. Department of Ophthalmology, Yong Loo Lin School of Medicine, National University of Singapore, Singapore, 1E Kent Ridge Rd, Singapore 119228.
164. Department of Vascular Surgery; Division of Surgical Specialties; University Medical Center Utrecht, 3584 CX Utrecht, the Netherlands
165. Princess Al-Jawhara Al-Brahim Centre of Excellence in Research of Hereditary Disorders (PACER-HD), King Abdulaziz University, Jeddah, Saudi Arabia
166. National Heart, Lung, and Blood Institute (NHLBI) Framingham Heart Study, Framingham, MA, USA.
167. CNR Institute of Clinical Physiology, Pisa, Italy, Department of Clinical & Experimental Medicine, University of Pisa, Italy
168. Department of Clinical Sciences, Diabetes and Endocrinology, Lund University Diabetes Centre, Malmö, Sweden.
169. Faculty of Health Sciences, University of Southern Denmark, 5000 Odense, Denmark
170. University of Exeter Medical School, University of Exeter, Exeter, EX2 5DW, UK
171. Departments of Pediatrics and Genetics, Harvard Medical School, Boston, MA 02115
172. Stanford Cardiovascular Institute, Stanford University, Stanford, CA 94305
173. Stanford Diabetes Research Center, Stanford University, Stanford, CA 94305
174. Oxford NIHR Biomedical Research Centre, Churchill Hospital, Oxford, UK, OX3 7LE
175. Department of Public Health and Primary Care, Institute of Public Health, University of Cambridge, Strangeways Research Laboratory, Cambridge, CB1 8RN, UK
176. Pediatric Research Center, Department of Women & Child Health, University of Leipzig, Leipzig, Germany
177. National Heart and Lung Institute, Imperial College London, London W12 0NN, UK
178. Institute of Biomedicine, School of Medicine, University of Eastern Finland, Kuopio Campus, 70211 Kuopio, Finland
179. Department of Medical Sciences, Molecular Epidemiology; EpiHealth, Uppsala University, Uppsala, 75185, Sweden
180. The Big Data Institute, Li Ka Shing Centre for Health Information and Discovery, University of Oxford, Oxford OX3 7BN, UK
181. Department of Clinical Medicine, Faculty of Health and Medical Sciences, University of Copenhagen, 2200 Copenhagen, Denmark
182. Department of Neurology, Wayne State University School of Medicine, Detroit, MI, USA
183. Department of Epidemiology & Prevention, Division of Public Health Sciences, Wake Forest University, Winston-Salem, NC, 27157, USA.
184. The Mindich Child Health and Development Institute, The Icahn School of Medicine at Mount Sinai, New York, NY, 10069, USA.
185. Department of Genetics, University of North Carolina, Chapel Hill, NC, 27599, USA.
186. Usher Institute of Population Health Sciences and Informatics, University of Edinburgh, Edinburgh, EH16 4UX, UK
187. Aberdeen Biomedical Imaging Centre, University of Aberdeen, Foresterhill Health Campus, Aberdeen, AB25 2ZD
188. British Heart Foundation Glasgow Cardiovascular Research Centre, Institute of Cardiovascular and Medical Sciences, College of Medical, Veterinary and Life Sciences, University of Glasgow, Glasgow G12 8TA, UK

189. Division of Population Health and Genomics, School of Medicine, University of Dundee, DUNEE DD2 4BF
190. Faculty of Medicine, University of Split, Split, Croatia
191. Departments of Epidemiology and Health Services, University of Washington, Seattle, WA, USA
192. Kaiser Permanent Washington Health Research Institute, Seattle, WA, USA
193. Division of Cardiovascular Medicine, Brigham and Women's Hospital, Boston MA 02115
194. Department of Public Health & Clinical Medicine, Section for Family Medicine, Umeå University, Umeå, SE-901 85, Sweden
195. Department of Medicine, McGill University, Montreal (Quebec) Canada H4A 3J1
196. Department of Human Genetics, McGill University, Montreal (Quebec) Canada H3A 1B1
197. Alzheimer Scotland Dementia Research Centre, University of Edinburgh, Edinburgh, EH8 9JZ
198. Institute of Cellular Medicine, The Medical School, Newcastle University, Newcastle, NE2 4HH, UK
199. Department of Physiology and Biophysics, University of Mississippi Medical Center, Jackson, MS, USA.
200. Department of Biostatistics, University of Liverpool, Liverpool L69 3GL, UK
201. Diabetes Unit, Department of Medicine, Massachusetts General, Boston, MA, USA.
202. Program in Medical Genetics, Broad Institute, Cambridge, MA, USA.

1 Tissue-Specific Alteration of Metabolic Pathways Influences Glycemic 2 Regulation

3 4 Highlights

- 5
- 6 • 23 novel coding variant associations (single-point and gene-based) for glycemic traits
- 7 • 51 effector transcripts highlighted different pathway/tissue signatures for each trait
- 8 • The exocrine pancreas and gut influence fasting and 2h glucose, respectively
- 9 • Multiple variants in liver-enriched *G6PC* and islet-specific *G6PC2* influence glycemia

10

11 Summary

12

13 Metabolic dysregulation in multiple tissues alters glucose homeostasis and influences risk for type 2
14 diabetes (T2D). To identify pathways and tissues influencing T2D-relevant glycemic traits (fasting glucose
15 [FG], fasting insulin [FI], two-hour glucose [2hGlu] and glycated hemoglobin [HbA1c]), we investigated
16 associations of exome-array variants in up to 144,060 individuals without diabetes of multiple
17 ancestries. Single-variant analyses identified novel associations at 21 coding variants in 18 novel loci,
18 whilst gene-based tests revealed signals at two genes, *TF* (HbA1c) and *G6PC* (FG, FI). Pathway and tissue
19 enrichment analyses of trait-associated transcripts confirmed the importance of liver and kidney for FI
20 and pancreatic islets for FG regulation, implicated adipose tissue in FI and the gut in 2hGlu, and
21 suggested a role for the non-endocrine pancreas in glucose homeostasis. Functional studies
22 demonstrated that a novel FG/FI association at the liver-enriched *G6PC* transcript was driven by multiple
23 rare loss-of-function variants. The FG/HbA1c-associated, islet-specific *G6PC2* transcript also contained
24 multiple rare functional variants, including two alleles within the same codon with divergent effects on
25 glucose levels. Our findings highlight the value of integrating genomic and functional data to maximize
26 biological inference.

27 Introduction

28

29 It has long been recognized that rare and penetrant disease-causing mutations can pinpoint key proteins
30 and pathways involved in human metabolism (Froguel et al., 1992; Gloyn et al., 2004; Montague et al.,
31 1997). Type 2 diabetes (T2D) results from an inability of the pancreatic islet beta cells to produce and
32 secrete sufficient insulin, compounded by the failure of metabolic tissues to respond to insulin and store
33 glucose appropriately. Blood glucose levels are regulated by the co-ordination of homeostatic pathways
34 operating across multiple tissues that control metabolism, therefore a clearer understanding of their
35 relative roles is critical in guiding efforts to modulate them pharmacologically to treat T2D and pre-
36 diabetes. In recent years, technological advances have made it possible to assay genetic variation
37 genome-wide and at scale. These provide tremendous opportunities to understand metabolic
38 differences within the physiological range through the study of quantitative fasting and post-challenge
39 glycemic measures (Mahajan et al., 2015; Scott et al., 2012; Wessel et al., 2015; Wheeler et al., 2017a).
40 These measures can influence the risk of developing pathophysiological conditions such as T2D and
41 cardiovascular disease. However, as in all genome-wide association studies (GWAS), it has proven
42 challenging to translate the associated genetic signals into biological pathways, as the vast majority of
43 association signals lie within non-coding regions, and connecting them to their respective effector genes
44 is less straightforward. There are to date over 97 loci reported to be associated with glycemic traits,
45 across different genetic approaches (Wheeler et al., 2017b). One approach to facilitate identification of
46 likely causal variants and transcripts is to focus on coding variation, whose effects on protein sequence
47 can be predicted and functionally tested, facilitating identification of likely causal genes and the ensuing
48 biological insights. This strategy has been successfully used to establish not only the effector genes but
49 also the direction of effect of T2D risk alleles on protein function such as in the case of SLC30A8
50 (Flannick et al., 2014) and PAM (Steinthorsdottir et al., 2014; Thomsen et al., 2018).

51

52 Here, we describe the largest exome-array study to date across four commonly-used glycemic traits
53 (fasting glucose [FG], fasting insulin [FI], glycated hemoglobin [HbA1c], and two-hour glucose [2hGlu]) in
54 up to 144,060 non-diabetic individuals from multiple ancestries, to discover variants and loci influencing
55 these traits within the physiological range. We sought to identify causal variants and putative effector
56 transcripts in known and novel loci, and subsequently highlight pathways and tissues that are enriched
57 for these glycemic trait associations. We further complemented our analyses with functional validation
58 of selected effector transcripts, focusing on novel FG/FI locus *G6PC* and FG/HbA1c locus *G6PC2*, to

59 establish functional links between the associated rare coding variants in those loci and glucose
60 regulation through different metabolic tissues. Together, our findings provide valuable insight into the
61 biology underlying glycemic traits, and build on the knowledge required for validating candidate genes
62 for therapeutic targeting in diabetes.

63

64 **Results**

65

66 **Identification of coding variant and gene-based glycemic trait associations**

67 We focused on coding variants on the exome chip as these could point more directly to their potential
68 effector transcripts (i.e. likely causal gene[s]). Single-variant and gene-based association analyses with
69 FG, FI, HbA1c, and 2hGlu levels were performed on exome-array coding variants in up to 144,060
70 individuals without diabetes of European (85%), African-American (6%), South Asian (5%), East Asian
71 (2%), and Hispanic (2%) ancestry from up to 64 cohorts (**Table S1, Methods**).

72 We performed single-variant analyses in each individual cohort using a linear mixed model and
73 combined results by fixed-effect meta-analyses within and across ancestries. As body mass index (BMI)
74 is a major risk factor for T2D and is correlated with glycemic traits, all analyses were adjusted for BMI
75 (**Methods**) to identify loci influencing glycemia independently from their effects on overall adiposity. We
76 used distance-based clumping and considered signals to be novel, if they were located more than 500 kb
77 from a variant with an established association with any of the glycemic traits or T2D at the time of the
78 study (**Methods**).

79

80 Based on the above definition, we found 21 coding variants (in 18 genes) which were not previously
81 associated with any other glycemic trait or T2D risk, that are now associated at exome-wide significance
82 (defined as $P < 2.2 \times 10^{-7}$) (Mahajan et al., 2018b; Sveinbjornsson et al., 2016) with their respective
83 glycemic trait(s) (**Table 1, Methods**). Among these novel loci were a missense (p.E1365D) and splice
84 region variant in *OBSL1* associated with FI, another missense variant (p.L300P) in *RAPGEF3* associated
85 with FI, a missense variant (p.S439N) in *SPTB* associated with HbA1c, and missense variants p.R187Q in
86 *ANKH* and p.R456Q in *STEAP2* that are associated with FG (**Table 1**). *OBSL1* encodes a cytoskeletal
87 protein related to obscurin, mutations in which have been shown to lead to an autosomal recessive
88 primordial growth disorder (OMIM: 612921). Loss of *OBSL1* leads to downregulation of *CUL7*, a protein
89 known to interact with *IRS-1*, downstream of the insulin receptor signaling pathway (Hanson et al.,

90 2009). *RAPGEF3* encodes a cAMP-regulated guanine nucleotide exchange factor and is part of a cAMP-
91 responsive signalling complex. The gene has been shown to be involved in cAMP-dependent
92 adipogenesis (Jia et al., 2012), and investigation of associations in other related traits showed that the
93 same *RAPGEF3* variant is also associated with BMI, waist-hip ratio and height (all $P < 1 \times 10^{-4}$; **Tables 1**
94 **and S2**). This suggests that its role in adiposity and obesity is likely to link it to FI regulation. Since the
95 directions of effect of the variant are opposite for FI and BMI, the observed association could however
96 be due to collider bias and should thus be interpreted with caution. *SPTB* encodes the protein spectrin
97 beta, which is a major constituent of the cytoskeletal network underlying the erythrocyte plasma
98 membrane. Mutations in this gene underlie a range of hematological disorders such as hemolytic
99 anemias (OMIM: 617948, 616649). Given that red blood cell disorders can interfere with HbA1c levels
100 (Wheeler et al., 2017a), this missense variant identifies *SPTB* as the likely effector transcript at this locus.
101 *ANKH* encodes a transmembrane protein likely acting as a transporter. Recently, the FG-lowering allele
102 reported here was shown to associate with decreased T2D risk in Europeans (OR=0.78 [0.69-0.87],
103 $P_{EUR}=2.0 \times 10^{-7}$), and had a >97% posterior probability of being causal, suggesting that this gene is the
104 effector transcript at this locus (Mahajan et al., 2018b). In the final example, *STEAP2* encodes a six
105 transmembrane protein localized both intracellularly and on the plasma membrane, and is suggested to
106 have roles in the regulation of iron transport (Sikkeland et al., 2016). A closely-related member of the
107 *STEAP* family, *STEAP4*, has been reported to mediate cellular response to inflammatory stress through
108 its role as a metalloredutase mediating iron and copper homeostasis (Scarl et al., 2017). Though little is
109 known about *STEAP2* function, in the recent T2D analysis, the FG-associated variant in *STEAP2* was also
110 found to be nominally associated with T2D risk ($P < 1 \times 10^{-4}$) (Mahajan et al., 2018b) (**Tables 1 and S2**). In
111 addition to the novel loci, 53 other significant coding variant associations (in 40 genes) were detected
112 that were within 500 kb of an established glycemic GWAS locus. These were of interest as they could
113 point to a causal gene (**Tables 1 and S3**).

114
115 To increase power to detect rare variant associations, we additionally performed gene-burden and
116 sequence kernel association (SKAT) tests for gene-level analyses (**Methods**). We identified six genes with
117 significant evidence of association ($P < 2.5 \times 10^{-6}$), of which two – *G6PC* (for FG and FI) and *TF* (for HbA1c)
118 – represented novel associations (**Tables 2 and S4**).

119

120 **Identification of effector transcripts**

121 To establish whether the associated coding variants (both novel and those at established loci) were
122 likely to be causal, and/or likely to pinpoint an effector transcript, we first integrated these results with
123 published data with higher density GWAS coverage (Manning et al., 2012; Wheeler et al., 2017a). This is
124 important because coding variants can sometimes erroneously point to the wrong effector transcript, as
125 they can “piggy-back” on non-coding alleles that drive the association, and by virtue of having a
126 predicted effect on protein sequence they may falsely implicate the gene in which they reside as the
127 causal one. For example, the coding variant rs56200889 (p.Q802E) at *ARAP1* is strongly associated with
128 FG ($\beta=-0.016$, $P=1.8 \times 10^{-14}$, **Table 1**), and when considered in isolation might have suggested *ARAP1* as
129 the effector transcript. However, T2D fine-mapping efforts showed this association to be secondary to a
130 much stronger non-coding signal (Mahajan et al., 2018b), and recent data integrating human islet and
131 mouse knockout information has established neighbouring gene *STARD10* as the most likely gene
132 mediating the GWAS signal at this locus (Carrat et al., 2017). Therefore, we conditioned the coding
133 variants identified here on existing non-coding GWAS index variants at established loci from two
134 previously published GWAS datasets (Manning et al., 2012; Wheeler et al., 2017a), and also performed
135 the reciprocal analysis (**Table S3, Methods**). At novel loci, we also assessed whether the coding variant
136 identified here was being driven by association of a sub-threshold (i.e. non genome-wide significant in
137 smaller sample size) non-coding variant based on published GWAS results with higher density coverage
138 (Manning et al., 2012; Wheeler et al., 2017a) (**Methods**). As reciprocal conditional analysis was not
139 always possible, or was not informative, we also used additional published data, including fine-mapping
140 results from comparable T2D efforts (Mahajan et al., 2018b), results for associations with blood cell
141 traits (Astle et al., 2016; Soranzo et al., 2010) (**Table S5**), as well as a body of literature establishing the
142 role of certain genes (mapping within our loci) in glucose metabolism, or red blood cell biology (for
143 HbA1c) to inform effector transcript classification. We further considered significant gene-based
144 associations driven by multiple coding variants within a single gene as strong evidence for the
145 determination of effector transcripts (**Methods**).

146
147 Combining the above approaches, we curated the 74 coding variant associations (in 58 genes) displayed
148 in Table 1, and where possible identified and classified effector transcripts into “gold”, “silver” and
149 “bronze” categories, depending on the strength of evidence (**Table S6, Methods**). Loci with strong
150 evidence from reciprocal conditional analysis or from published data that supported the relevance of
151 the identified effector transcript to the glycemic trait were labelled “gold” (e.g. *GLP1R*, *SLC30A8*, *G6PD*,
152 *PPARG*, *ANK1*); those where an effector transcript could not be defined by conditional analysis (either

153 because it was inconclusive or due to lack of data) but where there was strong biological plausibility for
154 a given gene at the locus were labelled “silver” (e.g. *MADD*, *MLXIPL*, *FN3K/FN3KRP*, *HK1*, *VPS13C*); those
155 where we had some evidence but that was not as strong as “silver” were labelled “bronze” (e.g. *DCAF12*,
156 *OBSL1*, *STEAP2*, *RAPGEF3*); the remaining were left with an undetermined effector transcript (**Figure 1**,
157 **Table S6**). Effector transcript classification into the three categories was undertaken independently by
158 four of the authors and the consensus was used as the final classification for effector transcripts. From
159 74 single variant and six gene-based signals, we identified 51 unique effector transcripts (24 gold, 11
160 silver, 16 bronze), with many of them shared across traits (**Figure 1**). One case in point pertains to
161 *VPS13C*, which harboured a missense variant (p.R974K) associated with 2hGlu (labelled “bronze”) at
162 exome-wide significance ($\beta=-0.069$, $P=6.4 \times 10^{-10}$; **Table 1**), and also exhibited a significant gene-based
163 association with FG (labelled “silver”; $P_{SKAT}=3.7 \times 10^{-7}$; **Table S4**). *VPS13C* belongs to the previously-
164 established *VPS13C/C2CD4A/C2CD4B* glycemic trait and T2D risk locus, and recent follow-up studies
165 have with varying levels of evidence suggested *C2CD4A*, encoding a calcium-dependent nuclear protein,
166 as the causal gene for T2D through its potential role in the pancreatic islets (Kycia et al., 2018; Mehta et
167 al., 2016; O'Hare et al., 2016). In our data, it is however not possible to rule out *VPS13C* as a potential
168 effector transcript at this locus, warranting further functional studies for *VPS13C*, which encodes a
169 protein reported to be necessary for proper mitochondrial function (Lesage et al., 2016).

170

171 **Pathway analyses identifies relevant gene sets regulating glycemia**

172 To identify pathways enriched for glycemic trait associations, and to subsequently determine the extent
173 to which associations within the same trait implicate the same or similar pathways (as indicated by the
174 functional connectivity of the network), we used GeneMANIA network analysis (Franz et al., 2018).
175 GeneMANIA takes a query list of genes and finds functionally-similar genes based on large, publicly-
176 available biological datasets. We analysed all loci harbouring non-synonymous variants that reached $P < 1$
177 $\times 10^{-5}$ for any of the four glycemic traits in our study (totaling 121 associations). A high degree of
178 connectivity was observed within the HbA1c network, with enrichment of processes related to blood cell
179 biology such as porphyrin metabolism, erythrocyte homeostasis and iron transport (**Figures 2 and S1**,
180 **Table S7**). In comparison, the network generated from FG-associated genes captured several processes
181 known to contribute to glucose regulation and islet function, including insulin secretion, zinc transport
182 and fatty acid metabolism (**Figure 2, Table S7**). The FG network further revealed linking nodes (that are
183 not among the association signals) with known links to glucose homeostasis and diabetes, such as *GCK*
184 (encoding the beta cell glucose sensor glucokinase), *GCG* (encoding the peptide hormone glucagon

185 secreted by the alpha cells of the pancreas) and *GIP* (encoding the incretin hormone gastric inhibitory
186 polypeptide). One gene within the FG cluster for lipid-related pathways is *CERS2*, which encodes
187 ceramide synthase 2, an enzyme known to be associated with the sphingolipid biosynthetic process
188 (**Figure 2, Table S7**). Although *CERS2* is only nominally associated with FG and is significantly associated
189 with HbA1c, it does not cluster together with any HbA1c-enriched pathway, suggesting that *CERS2* is
190 regulating FG and HbA1c indirectly through its role in lipid metabolism. Given that there were fewer
191 genes associated with FI and 2hGlu, we were less powered to draw meaningful insights from the
192 enriched pathways in those traits (**Figure S1, Table S7**).

193

194 We also performed gene set enrichment analysis (GSEA) using EC-DEPICT (Marouli et al., 2017; Turcot et
195 al., 2018) (**Methods**). The primary innovation of EC-DEPICT is the use of 14,462 gene sets extended
196 based on large-scale co-expression data (Fehrmann et al., 2015; Pers et al., 2015). These gene sets take
197 the form of z-scores, where higher z-scores indicate a stronger prediction that a given gene is a member
198 of a gene set. To reduce some of the redundancy in the gene sets (many of which are strongly correlated
199 with one another), we clustered them into 1,396 “meta-gene sets” using affinity propagation clustering
200 (Frey and Dueck, 2007). These meta-gene sets are used to simplify visualizations and aid interpretation
201 of results. Here, we combined and analyzed all variants that reached $P < 1 \times 10^{-5}$ for any of the four
202 glycemic traits (**Methods**). We found 234 significant gene sets in 86 meta-gene sets with false discovery
203 rate (FDR) of < 0.05 (**Table S8**). As expected, we observed a strong enrichment of insulin- and glucose-
204 related gene sets, as well as exocytosis biology (in keeping with insulin vesicle release). In agreement
205 with the GeneMANIA network analyses, we also noted a strong enrichment for blood-related pathways,
206 which was primarily driven by HbA1c-associated variants. This was likely because HbA1c levels are
207 influenced not only by glycation but also by blood cell turnover rate (Cohen et al., 2008; Wheeler et al.,
208 2017a). To disentangle blood cell turnover from effects due to glycation, we repeated the analysis
209 excluding variants that were significantly associated with HbA1c only and found 128 significant gene sets
210 in 53 meta-gene sets (FDR < 0.05) (**Table S8**). We also analyzed each of the four traits separately (**Table**
211 **S8, Methods**).

212 To identify additional candidate genes, we then performed heat map visualization with unsupervised
213 clustering of the membership predictions (z-scores) of trait-associated genes for each significant gene
214 set (**Figures 2, S2 and S3**). This strategy has previously been effective for gene prioritization for
215 downstream analyses (Marouli et al., 2017; Turcot et al., 2018), as it becomes visually apparent which

216 genes are the strongest drivers of the significant gene sets and thus are natural targets for follow-up.
217 This can be particularly helpful for prioritizing genes that are not well-characterized, as it leverages
218 DEPICT's prediction of gene function. For the analysis of all traits except HbA1c, one cluster showed
219 particularly strong predicted membership for highly relevant gene sets, including "abnormal glucose
220 homeostasis", "peptide hormone secretion", "Maturity Onset Diabetes of the Young", and multiple
221 pathways involved in the regulation of glycogen, incretin, and carbohydrate metabolism (**Figure 2C**).
222 Strikingly, this cluster of six genes (*PCSK1*, *GLP1R*, *GIPR*, *G6PC2*, *SLC30A8* and *CTRB2*) contained five of
223 the genes that had independently been assigned to "gold" status during effector transcript identification
224 (**Table S6**). Therefore, the sixth gene, *CTRB2*, represents a novel gene for prioritization, since it showed
225 strong similarity to other genes for which there was already substantial biological evidence. *CTRB2*
226 encodes chymotrypsinogen B2, a digestive enzyme that is expressed in the exocrine pancreas, and
227 subsequently secreted into the gut. The gene contains a borderline significant variant for 2hGlu
228 (rs147238447; p.L6V; $P=1.9 \times 10^{-6}$). Another variant at this locus, rs7202877 (6.2kb downstream of
229 *CTRB2*, $r^2=0.0006$, $D'=1$ with rs147238447 in European populations), has previously been shown to be an
230 eQTL for *CTRB1* and *CTRB2*, with the minor G allele (MAF=11%) associated with increased expression (t
231 Hart et al., 2013). In the same study, the rs7202877-G allele was associated with increased glucagon-like
232 peptide 1 (GLP-1)-stimulated insulin secretion ($P=8.8 \times 10^{-7}$, $N=196$). In our data, rs7202877-G was
233 nominally associated with lower 2hGlu ($P=6.3 \times 10^{-3}$) and lower FG ($P=2.8 \times 10^{-3}$) levels. Multiple distinct
234 signals in this region (previously referred to as the *BCAR1* locus) have also been associated with T2D risk,
235 including rs7202877 (where the G allele is protective), rs72802342 ($r^2=0.65$ with rs7202877 in European
236 populations) and rs3115960, although the coding variant rs147238447 described here is not (Mahajan et
237 al., 2018a; Mahajan et al., 2018b; Morris et al., 2012; Zhao et al., 2017). This can potentially be
238 explained by limited power to identify a significant association given the low MAF (~0.5%) of the coding
239 variant. In contrast to its effect on T2D, the rs7202877-G allele has been associated with increased risk
240 of type 1 diabetes (OR=1.28, $P=3.1 \times 10^{-15}$, $N=21,293$) (Barrett et al., 2009). Other variants at this locus
241 are associated with risk of chronic pancreatitis (rs8055167, $r^2=0.0021$ with rs147238447 and $r^2=0.12$
242 with rs7202877 in European populations, in LD with an inversion that changes the expression ratio of
243 *CTRB1* and *CTRB2* isoforms) (Rosendahl et al., 2017) and pancreatic cancer (rs7190458, $r^2=0.0002$ with
244 rs147238447 and $r^2=0.31$ with rs7202877 in European populations) (Wolpin et al., 2014). The
245 prioritization of *CTRB2* is intriguing as it supports an emerging hypothesis that the exocrine pancreas
246 contributes to complex mechanisms influencing 2hGlu levels and diabetes risk (Esteghamat et al., 2019;
247 Hart et al., 2018; Woodmansey et al., 2017). Given the earlier associations with GLP-1 stimulated insulin

248 secretion, we investigated whether this effect could be mediated by incretin levels. However, we found
249 no associations at rs147238447 for GLP-1 levels in the largest available dataset (fasting GLP-1, N=4170:
250 MAF=0.00457, $P=0.495$; 2h GLP-1, N=3839: MAF=0.00464, $P=0.076$) (Almgren et al., 2017), though this
251 might again be explained by limited power. Although additional validation of the rare coding variant
252 rs147238447 (p.L6V) as a potential causal variant is required given the absence of clear associations with
253 T2D risk and other glycemic traits, the results discussed above suggest a role of *CTRB2* in glycemic
254 regulation.

255 We also noted a small but distinct cluster in the FG-only analysis indicating the role of the
256 cilium/axoneme, pointing to novel biology relating to sensing and signaling in response to the
257 extracellular environment (**Figure 2D**). Two genes were the main drivers of this association: *WDR78* and
258 *AGBL2*. These represent potentially interesting candidates for follow-up, although we note that the
259 *AGBL2* signal may be driven through effects of the nearby *MADD* gene, which harbors a FG-associated
260 coding variant in our study and is labelled “silver” in our effector transcript classification (**Tables 1 and**
261 **S6**). Overall, our network and pathway analyses highlighted several trait-associated genes that do not
262 reach exome-wide significance in conventional single variant or gene-based tests, but show evidence of
263 contribution to glycemic regulation.

264

265 **Tissue enrichment analysis reveals shared roles of key tissues in the regulation of glycemic** 266 **traits**

267 In addition to identifying key metabolic pathways involved in glucose regulation, we sought to establish
268 the relative importance of particular tissues in the regulation of the different glycemic phenotypes. This
269 time, we assessed the tissues that are most highly enriched for the expression of the 51 effector
270 transcripts we have curated at the associated loci identified in this study, to highlight specific tissues
271 that contribute critically to the regulation of each glycemic trait. Using publicly-available tissue
272 expression data from GTEx (Battle et al., 2017) and human islets (van de Bunt et al., 2015), we noted
273 clear differences in tissue enrichment patterns as well as tissues shared between traits (**Figure 3**).

274 Comparisons between analyses of FG- and FI-associated effector transcripts underscored the relative
275 roles of the liver in both traits ($P<0.05$), whereas pancreatic islets were enriched in associations for FG
276 ($P=9.99 \times 10^{-5}$) but not FI ($P=0.75$). In contrast, adipose ($P=0.01$) and kidney tissues ($P=0.01$) were
277 enriched in FI but not FG ($P>0.05$). These results not only highlight the established role of pancreatic
278 islets in influencing FG levels, but also the under-appreciated role of insulin clearance in the kidney and

279 likely the liver, in addition to insulin action in liver and adipose tissue, in influencing FI levels (Goodarzi
280 et al., 2011). Consistent with the EC-DEPICT GSEA, there was also support for the role of the exocrine
281 pancreas (which typically represents >95% of whole pancreas tissue) in addition to the endocrine
282 pancreas (islets) in FG ($P=9.99 \times 10^{-5}$) and 2hGlu ($P=2.99 \times 10^{-4}$) associations. We also observed
283 enrichment for genes expressed in stomach for 2hGlu ($P=1.99 \times 10^{-4}$) but not for FG ($P=0.16$). HbA1c
284 analysis revealed enrichment in “metabolic” tissues reflecting insulin secretion (islets, $P=1.59 \times 10^{-2}$ and
285 pancreas, $P=0.01$), insulin action (muscle, $P=1.50 \times 10^{-2}$), insulin clearance (liver, $P=0.03$), as well as
286 strong enrichment for whole blood ($P=3.99 \times 10^{-3}$). These indicate key factors relating to hemoglobin
287 glycation and blood cell function in influencing overall HbA1c levels (**Figure 3**).

288

289 Our results from the pathway and tissue enrichment analyses demonstrate the role of specific tissues
290 with known functions in blood glucose regulation in particular glycemic traits. These observations add
291 further support to emerging reports of an underappreciated role for the exocrine pancreas in FG and
292 2hGlu regulation, the stomach-incretin axis in 2hGlu, and the importance of insulin clearance through
293 the kidney and liver in FI.

294

295 **Novel glycemic trait associations in liver-enriched *G6PC* are driven by functional coding** 296 **variants**

297 To delve deeper into tissue-specific gene effects, we focused on two homologues, *G6PC* and *G6PC2*,
298 with contrasting tissue expression profiles where we identified gene-based association signals for FG/FI
299 and FG/HbA1c, respectively (**Tables 2 and S4**). Both genes encode gluconeogenic enzymes that catalyze
300 the same biochemical pathway but are known to have distinct tissue expression profiles. *G6PC2* is
301 largely expressed in pancreatic islets whereas *G6PC* is highly expressed in the liver, kidney, and small
302 intestine (Foster et al., 1997; Mithieux, 1997). Our gene-based analyses highlighted *G6PC* through novel
303 associations with FG and FI, driven primarily by rare missense variants p.A204S (rs201961848) and
304 p.R83C (rs1801175), and protein-truncating variant (PTV) p.Q347X (rs80356487), none of which
305 achieved exome-wide significance at single-variant level (**Table S4**). Homozygous inactivating alleles in
306 *G6PC*, which include both p.R83C and p.Q347X, are known to give rise to glycogen storage disease type
307 Ia (GSD1a), a rare autosomal recessive metabolic disorder (Chou and Mansfield, 2008; Lei et al., 1995),
308 but this is the first time that rare coding variants in *G6PC* have been shown to influence FG and FI levels
309 in normoglycemic individuals.

310

311 Given the well-known role of *G6PC* in hepatic glucose homeostasis, we were interested in elucidating
312 the molecular impact of rare heterozygous *G6PC* coding variants highlighted in our exome-array
313 analysis, in particular novel variant p.A204S, one of the statistical drivers of the gene-based *G6PC* signal
314 (**Table S4**). In transient protein overexpression assays, p.R83C and p.A204S resulted in significantly
315 reduced protein levels compared to wild type (WT) *G6PC* in both Huh7 (human hepatoma) and HEK293
316 (human embryonic kidney) cell lines (**Figure 4A-D**). The PTV p.Q347X, which in our *in vitro* system
317 generated a smaller molecular weight protein, exhibited markedly lower protein expression levels in
318 Huh7 cells but not HEK293 cells. However, in both cell types, the cellular localization pattern of p.Q347X
319 appears to be largely diffuse and did not co-localize with the Golgi apparatus, which is important for
320 post-translational modification of *G6PC* protein (**Figures 4E and S4A**). Further functional characterization
321 of glucose-6-phosphatase (*G6Pase*) activity revealed that both p.R83C and p.Q347X variants lead to
322 proteins lacking any detectable phosphatase activity (**Figure S4B-C**), consistent with previous
323 observations of several *GSD1a*-causing coding variants (Shieh et al., 2002). As we observed that the
324 p.R83C variant resulted in complete loss of glycosylation, we determined if glycosylation is essential for
325 *G6Pase* activity by treating cells with tunicamycin to inhibit N-linked glycosylation. The ability of
326 unglycosylated *G6PC* to catalyze glucose-6-phosphate (*G6P*) was found to be downregulated by up to
327 14%, although this difference was not statistically significant (**Figure 4F**). We therefore concluded that
328 whilst glycosylation contributes to overall functional activity, it may not be a requisite for *G6P*
329 hydrolysis. Finally, we were unable to accurately assess p.A204S-*G6PC* phosphatase activity as the level
330 of expression in the microsomes was reduced by 41% relative to WT, supporting the hypothesis that
331 p.A204S-*G6PC* exhibits partial loss-of-function (LOF) most likely due to loss of protein expression.

332
333 Together, our functional studies support p.A204S, p.R83C, and p.Q347X as functional LOF variants due
334 to loss of *G6Pase* protein expression and/or activity. This results in a reduced potential to hydrolyze *G6P*
335 to glucose in gluconeogenic tissues (such as in the liver and kidney), thus directly reducing *FG* levels and
336 consequently lowering circulating *FI* levels in the plasma. Our data suggest that rare inactivating
337 mutations in *G6PC* (such as p.R83C and p.Q347X) that cause the autosomal recessive disorder *GSD1a*
338 can also modulate fasting glycemic traits within a normoglycemic range in asymptomatic heterozygous
339 variant carriers.

340

341 ***G6PC2* alleles influence protein function by multiple mechanisms**

342 *G6PC2*, a gene homolog of *G6PC*, is an established effector transcript at a GWAS locus which contains
343 multiple coding variants known to influence FG and HbA1c but not FI levels (Bouatia-Naji et al., 2008;
344 Chen et al., 2008; Mahajan et al., 2015; Soranzo et al., 2010; Wessel et al., 2015). In this current study,
345 gene-based association signals for both FG and HbA1c were observed at the *G6PC2* locus, primarily
346 driven by multiple coding variants (p.H177Y, p.Y207S, p.R283X, and p.S324P) (**Table S4**). We aimed to
347 extend the investigation of coding variation in this gene, which is likely to harbor a series of functional
348 alleles, by characterizing the four *G6PC2* coding variants above and six others, across the allelic
349 frequency spectrum (all with single-variant $P < 0.05$ for FG or HbA1c in our analyses) (**Table S4; Figure**
350 **S5A**). Protein overexpression studies in the rat insulinoma cell line INS-1 832/13 and HEK293 cells
351 revealed that seven of the *G6PC2* variants characterized (including PTV p.R283X) resulted in significantly
352 reduced protein levels (**Figures 5A and S5B-C**). In INS-1 832/13 cells, this effect was largely due to partial
353 or total loss of the glycosylated form of the protein. In HEK293 cells, the reduction in total protein levels
354 could be rescued when the proteasomal pathway (but not the lysosomal pathway) was inhibited,
355 consistent with an earlier study involving a smaller subset of variants (Mahajan et al., 2015), confirming
356 proteasome-mediated protein turnover.

357
358 As three variants (p.I171T, p.I171V, and p.F256L) appeared to be stably expressed and processed like WT
359 *G6PC2* protein, we hypothesized that these alleles could be influencing glycemic levels through effects
360 on protein activity. As there is a high level of conservation between the catalytic domains in *G6PC* and
361 *G6PC2*, we adapted the *G6Pase* assay used earlier, to indirectly analyse the effect of the *G6PC2* variants
362 on *G6Pase* enzymatic activity. We assumed that the *G6PC2* alleles of interest, which mapped to the
363 conserved regions, will give rise to the same consequence in the *G6PC* backbone due to the strong
364 homology and preserved topology of both proteins. The adaptation was necessary as we were unable to
365 detect *G6PC2* activity using the same experimental conditions. First, we generated variants that mapped
366 to equivalent sites within the *G6PC* protein (*G6PC*-p.L173T, p.L173V, and p.F258L correspond to *G6PC2*-
367 p.I171T, p.I171V, and p.F256L, respectively), and then performed the enzymatic studies. Two alleles,
368 p.L173T, p.L173V, affected the same codon and were each genetically associated with FG levels but with
369 opposite directions of effect (**Table S4**). We found that *G6PC*-p.L173T exhibited ~20% decreased activity
370 compared to WT based on assessment of V_{max} (maximal rate of reaction), a measure of enzymatic
371 activity (**Figure 5B**). In contrast, *G6PC*-p.L173V had enhanced activity through both increased V_{max} and
372 lowered K_m (Michaelis constant, whereby a lower K_m indicates higher substrate affinity) (**Figure 5B**).
373 Importantly, our *in vitro* observations mirrored the genetic effects on FG ($\beta_{171T} = -0.084$ mmol/l;

374 $\beta_{1171V}=+0.131$ mmol/l) and HbA1c levels ($\beta_{1171T}=-0.007\%$; $\beta_{1171V}=+0.093\%$) (**Table S4**). The G6PC-p.F258L
375 variant also displayed impaired phosphatase activity due to reduced V_{max} and a tendency towards higher
376 K_m relative to WT (**Figure 5C**), consistent with the observed glucose-lowering effects of G6PC2-p.F256L.
377 To ensure that the observed effects of the rare variants on FG were not influenced by the common
378 *G6PC2* variant rs560887, as was the case for a common variant V219L shown in an earlier study
379 (Mahajan et al., 2015) which we confirm here, conditional analyses were performed conditioning on
380 rs560887 (**Table S9**). Conditional results for p.I171T, p.I171V and p.F256L confirmed that the directions
381 of effect for the variants remain unchanged, making it unlikely that the regulatory variant rs560887 is
382 regulating these effects (**Table S9**). These results provided the first example of an activating allele in
383 *G6PC2* (p.I171V) and highlighted the unique protein changes at a single codon that can give rise to a
384 corresponding loss or gain of functional activity. These data therefore show that variations in *G6PC2*
385 may influence FG levels through their impact on protein expression or activity.

386

387 To further characterize these variants, we set out to determine the effect of the G6PC2 LOF variants on
388 ER integrity, given that G6PC2 is an ER-resident protein and that beta cells, which are highly-specialized
389 secretory cells, are highly sensitive to ER stress. Specifically, we evaluated the expression of G6PC2
390 variant proteins on the canonical ER stress response (ERSE) and unfolded protein response (UPRE)
391 pathways. The three G6PC2 variants which displayed relatively severe effects on protein stability
392 (p.H177Y, p.Y207S, p.S324P) in our study were found to activate ERSE and UPRE reporter activities by
393 ~3-fold, in contrast to the variants p.I171T and p.F256L which exert their effects primarily on enzymatic
394 function (**Figure 5D**). The common p.V219L variant, which reduces protein expression by approximately
395 50%, displayed an intermediary effect (**Figure 5D**). These results suggest that G6PC2 variant proteins,
396 especially those that result in severe LOF due to protein instability, may also influence beta cell ER
397 homeostasis.

398

399 In previous studies, the G6PC2-p.R283X variant has shown inconsistencies in terms of their associations
400 with FG levels (Mahajan et al., 2015; Wessel et al., 2015). With a larger dataset we have now confirmed
401 that this variant influences both FG and HbA1c levels (**Tables 1 and S3**). As the nonsense p.283X allele is
402 located in the last exon of the gene and may evade NMD, we queried RNA sequencing data from human
403 islets and observed an allelic balance in heterozygous carriers, indicating that variant transcripts are
404 indeed likely to escape NMD and be translated (**Figure S6A**). Based on our pipeline of *in vitro* assays, we

405 confirmed G6PC2-p.R283X loss-of-function due to reduced protein expression, failure to localize to the
406 Golgi network, and a high likelihood of complete loss of phosphatase activity (**Figures 5A and S5D**).

407
408 In contrast to the mechanisms in play for the coding variants in *G6PC2*, the non-coding GWAS index
409 variant at the *G6PC2* locus (rs560887) is suggested to influence expression of *G6PC2* splice variants
410 based on previous minigene analyses in HeLa cells (Baerenwald et al., 2013; Bouatia-Naji et al., 2010). To
411 establish whether this variant indeed influences *G6PC2* regulation in human islets, we determined its
412 effect on *G6PC2* isoform expression. We found that in human islets, the presence of the rs560887-G
413 allele is associated with increased expression of the full-length *G6PC2* isoform as compared with the
414 shorter isoform lacking exon 4 (**Figure S6B**). This observation supports the hypothesis that rs560887 may
415 alter splicing and is consistent with the association between rs560887-G and elevated FG and HbA1c
416 levels due to increased *G6PC2* function. As the phenotypic consequence of rare coding variants can be
417 influenced by regulatory variants on the same haplotype, we therefore performed conditional analyses
418 to explore the relationship between rs560887 and the rare coding variants. We showed that the
419 direction of effects of all the rare alleles in our study remained the same after conditioning on rs560887,
420 though it is notable that the variants p.Y207S and p.R283X showed some reduction in strength of
421 association after conditioning (**Table S9**).

422

423 **Functional assessment of *G6PC2* variants improves gene-based association analysis**

424 We next evaluated the utility of our functional data to enhance gene-based association analyses. We
425 showed that the gene-based signals were strengthened when the tests were informed by *in vitro*
426 functional validation of the variants (as determined in this study) as opposed to the predictive *in silico*
427 annotations based on the NSbroad and NSstrict masks (**Table S9, Methods**). In fact, in line with
428 expectation, flipping the alleles in the gain-of-function variant p.I171V (which we now know acts in the
429 opposite direction compared to other rare variants in the test), to align all alleles with the same
430 direction of effect, augmented the strength of association for both FG (from $P=4.34 \times 10^{-71}$ to $P=6.47 \times$
431 10^{-78}) and HbA1c ($P=6.37 \times 10^{-30}$ to $P=6.37 \times 10^{-33}$) in the gene burden test (**Table S9**). Improved methods
432 of filtering variants will enhance the performance of gene-based tests and increase the likelihood of
433 identifying true association signals, especially for those that are of borderline significance or that initially
434 fall below the significance threshold.

435

436 **G6PC2 regulates basal insulin secretion in human beta cells**

437 Although G6PC2 is known to be specifically enriched in pancreatic islet beta cells, its role in the
438 regulation of human beta cell function has not been shown. Using gene knockdown studies in the
439 human EndoC- β H1 beta cell line, we found that G6PC2-deficient cells exhibited significantly (but
440 modestly) increased insulin secretion at low glucose (1 mM) and a trend towards increased insulin
441 secretion at sub-maximal glucose (6 mM) levels (**Figures 5F and 5E**). When expressed as a fraction of
442 insulin content (**Figure 5F**), insulin secretion was significantly increased across multiple glucose
443 conditions, although this was primarily driven by reduced total insulin content in G6PC2-deficient cells
444 by ~15%. Overall, *G6PC2* knockdown increases glucose responsiveness at sub-threshold levels of glucose
445 but not at maximal glucose concentration in EndoC- β H1 cells, suggesting enhancement of basal glucose
446 sensitivity by promoting glycolytic flux at sub-stimulatory glucose concentrations, and warranting more
447 in-depth characterization experiments.

448 Discussion

449 We have identified novel coding variant associations with FG, FI, 2hGlu and HbA1c, across the allele
450 frequency spectrum, and assigned these variants to their effector transcripts using available genetic and
451 biological evidence. We further pinpointed novel loci and effector transcripts that have now been
452 associated with T2D and other related metabolic traits since the time of our analysis. Our results
453 revealed that 15 out of 58 glycemic trait-associated loci have evidence of association with T2D risk
454 (**Table S2**) (Hara et al., 2014; Mahajan et al., 2018b; Williams et al., 2014). For instance, FG-associated
455 loci *ANKH* and *STEAP2*, and HbA1c-associated *DCAF12* all associate with T2D risk (**Table S2**), providing
456 opportunities to investigate the mechanisms through which associated variants influence both glycemic
457 regulation within the physiological range as well as T2D pathophysiology. The FI-associated *RAPGEF3*
458 locus is also associated with various obesity-related measures including BMI and WHR, potentially
459 supporting our tissue enrichment analyses linking FI with adiposity.

460 We used this work to explore the pathways and metabolic tissues through which the associated genes
461 influence variation in glycemic traits and highlighted those with key roles in glucose regulation and traits
462 that act through multiple metabolic tissues, including islets, liver, fat, and in addition, exocrine pancreas,
463 gut and kidney. Our GSEA enabled us to identify additional genes (e.g. *CTRB2*) within these tissues and
464 pathways which were below the threshold for statistical significance in our initial discovery effort and
465 that merit follow-up. We report an emerging role for the gut and exocrine pancreas for 2hGlu levels and
466 potentially T2D risk through multiple analyses, consistent with current understanding that both incretins
467 and digestive enzymes are important in controlling postprandial glucose levels (Esteghamat et al., 2019;
468 Hart et al., 2018; Woodmansey et al., 2017). We also show that different traits are influenced by
469 pathways operating in distinct tissues: FG is almost exclusively influenced by pathways in the endocrine
470 and exocrine pancreas and liver, whilst FI is mediated by the insulin-sensitive tissues such as liver,
471 kidney, and adipose tissue, indicating the importance of both insulin action and insulin clearance
472 mechanisms. Genes expressed in muscle, also an insulin-sensitive tissue, were enriched in HbA1c-
473 associated effector loci but not FI, though this could be due to differences in power between the two
474 analyses. We see evidence of multiple metabolic tissues being important for HbA1c regulation, and note
475 that the HbA1c-associated set of effector transcripts appear enriched for those that influence blood cell
476 biology.

477 We have also shown for the first time that genetic variation in *G6PC*, a gene implicated in GSD1a,
478 influences glycemic traits within the normal physiological range in heterozygote carriers. *In vitro* follow-

479 up of the variants driving the gene-based association – p.A204S, p.R83C, and p.Q347X – confirmed that
480 these were indeed causal LOF variants at this locus that contribute to modulation of FG and FI levels. We
481 then reported novel rare coding variant associations for FG and HbA1c within a member of the same
482 gene family, *G6PC2*, and expanded the allelic spectrum of reported variants to include variants affecting
483 the same codon with both loss and gain of function alleles. Our comprehensive analysis of this locus
484 demonstrates multiple molecular mechanisms by which variants influence protein function, including
485 evidence from human islets that the common regulatory variant rs560887 influences *G6PC2* isoform
486 expression, and that a rare PTV (p.R283X) evades NMD and results in a catalytically-null enzyme. Given
487 the possibility that the effects of any coding variants in exon 4 which are carried in *cis* with the
488 rs560887-A allele could potentially be “diluted” due to the splicing effect, we checked whether the
489 observed rare variant effects could be driven by rs560887 in LD by repeating the single-variant
490 association tests with conditional analyses (**Table S9**). In our analysis, the directions of effect of the rare
491 coding alleles do not appear to be influenced by the non-coding regulatory allele. We then used our *in*
492 *vitro* data to refine existing methods for gene-based association analysis to demonstrate the value of
493 functional data in improving their sensitivity. New developments in high-throughput functional
494 annotation that can overcome the time-consuming nature of functional experiments will greatly
495 facilitate such efforts (Liu et al., 2017; Tewhey et al., 2016; Ulirsch et al., 2016). Finally, to understand
496 how loss of *G6PC2* influences FG levels, we silenced it in a human beta cell model and demonstrated
497 increased insulin secretion at low glucose levels, in line with the genetic observations.

498

499 It has long been suspected that particular metabolic tissues are key to governing specific processes of
500 glucose metabolism. Using human genetics, our study has explored this within an unbiased approach
501 and has illustrated the impact of altered glycolysis in multiple metabolic tissues on various glycemic
502 phenotypes. Uniquely, our parallel studies of *G6PC* and *G6PC2* highlighted two homologous proteins
503 that act through different tissues to influence glycemic traits. As *G6PC* is involved in hepatic glucose
504 production it influences both FG and FI levels. Previous studies have also established a potential role for
505 *G6PC* in influencing lipid and urate levels (Dewey et al., 2016; Sever et al., 2012). In contrast, due to its
506 restricted expression in the islet beta cell, variants in *G6PC2* only influence FG and HbA1c due to a beta
507 cell-driven effect. There are also notable differences in the molecular mechanisms underlying protein
508 dysfunction: for *G6PC* variants the effect is primarily on enzymatic activity, whilst *G6PC2* variants largely
509 cause protein instability.

510

511 A limitation of the present study is that we were not able to fine-map association signals, being
512 restricted to variants captured on the exome array, leaving many associated loci with unknown effector
513 transcripts. Additional large-scale studies, with higher density GWAS arrays and imputation to dense
514 reference panels, will be required for fine-mapping and further effector transcript identification.

515

516 In conclusion, we have combined human genetic discovery with pathway analysis and functional studies
517 to uncover tissue-specific effects in common pathways that influence glycemetic traits. Our findings will
518 inform efforts to target these pathways therapeutically to modulate metabolic function.

519 **Figure Legends**

520

521 **Figure 1. Effector transcript classification into “gold”, “silver” and “bronze” categories based on**
522 **strength of genetic and biological evidence.** A total of 51 effector transcripts from 74 single variant and
523 six gene-based signals were identified, with many of them shared across traits. The classification was
524 undertaken independently by four of the authors and the consensus was used as the final classification
525 for effector transcripts (see **Methods**). *Asterisk indicates “silver” for FG, “bronze” for 2hGlu.

526

527 **Figure 2. Network and pathway analyses identify relevant gene sets regulating glycemia using two**
528 **different methods for variant associations with $P < 1 \times 10^{-5}$.** (A-B) The networks represent composite
529 networks for (A) HbA1c and (B) FG, from the GeneMANIA analysis using genes with variant associations
530 at $P < 1 \times 10^{-5}$ for each trait as input. Nodes outlined in red correspond to genes from the input list. Other
531 nodes correspond to related genes based on 50 default databases. Based on the network, GO terms and
532 Reactome pathways that were significantly enriched are depicted. To summarize these results, the most
533 significant term of all calculated terms within the same group is represented. Barplots with the
534 Bonferroni-adjusted $-\log_{10}(\text{p-values})$ of the most significant terms within each group are shown.
535 Each group was assigned a specific color; if a gene is present in more than one term, it is displayed in
536 more than one color.

537 (C-D) Heatmaps showing EC-DEPICT results from analysis of (C) all traits except HbA1c and (D) FG. The
538 columns represent the input genes for the analysis. In (C), these are genes with variant associations of
539 $P < 1 \times 10^{-5}$ for FG, FI, and/or 2hGlu, and in (D) these are genes with variant associations of $P < 1 \times 10^{-5}$ for
540 FG. Rows in the heatmap represent significant meta-gene sets (FDR < 0.05). The color of each square
541 indicates DEPICT’s z-score for membership of that gene in that gene set, where dark red means “very
542 likely a member” and dark blue means “very unlikely a member.” The gene set annotations indicate
543 whether that meta-gene set was significant at FDR < 0.05 or not significant (n.s.) for each of the other EC-
544 DEPICT analyses. For heatmap intensity and EC-DEPICT P -values, the meta-gene set values are taken
545 from the most significantly enriched member gene set. The gene variant annotations are as follows: (1)
546 the European minor allele frequency (MAF) of the input variant, where rare is MAF < 1%, low-frequency
547 is MAF 1-5%, and common is MAF > 5%, (2) whether the gene has an Online Mendelian Inheritance in
548 Man (OMIM) annotation as causal for a diabetes/glycemic-relevant syndrome or blood disorder, (3) the
549 effector transcript classification for that variant: gold, silver, bronze, or NA (note that only array-wide
550 significant variants were classified, so suggestively-significant variants are by default classified as “NA”),

551 (4-7) whether each variant was significant ($P < 2 \times 10^{-7}$), suggestively significant ($P < 1 \times 10^{-5}$), or not
552 significant in Europeans for each of the four traits, and (8) whether each variant was included in the
553 analysis or excluded by filters (see **Methods**). AWS: array-wide significant. Related to Figures S1 to S3.

554

555 **Figure 3. Tissue enrichment analysis reveals the key tissues involved in the regulation of glycemic**
556 **traits.** The figures display expression enrichment of genes from all of the golden, silver, and bronze gene
557 set lists for (A) HbA1c, (B) FG, (C) FI and (D) 2hGlu in GTEx tissue samples plus islet data. Enrichment *P*-
558 values were assessed empirically for each tissue using a permutation procedure (10,000 iterations), and
559 the red vertical line shows the significance threshold (empirical $P < 0.05$).

560

561 **Figure 4. Functional characterisation of G6PC variant proteins.** Related to Figure S4.

562 (A) Protein expression levels of missense G6PC variants were determined in Huh7 cells (n=4-5) and (B)
563 HEK293 cells (n=5) by western blot densitometric analysis of FLAG-tagged G6PC constructs relative to
564 tubulin control, with representative blots shown.

565 (C) Protein expression levels of PTV Q347X were determined in Huh7 cells (n=3) and (D) HEK293 cells
566 (n=4) by western blot densitometric analysis of V5-tagged G6PC constructs relative to tubulin control,
567 with representative blots shown. Bars in red indicate variants that are statistical drivers of the gene-
568 based signal.

569 (E) Cellular localisation of V5-tagged G6PC-Q347X was assessed in Huh7 cells and overlaid with markers
570 for the ER (calreticulin) and the trans-Golgi network (TGN46). White arrows point to positions of the
571 Golgi apparatus. Scale bar indicates 10µm.

572 (F) Glucose-6-phosphatase activity of unglycosylated WT G6PC protein obtained from tunicamycin-
573 treated (Tuni) HEK293 microsomes (n=2), with representative western blot of microsomal protein
574 shown. All data presented as mean \pm SEM. * $p=0.01-0.05$; ** $p=0.001-0.01$; *** $p < 0.001$.

575

576 **Figure 5. Functional characterisation of G6PC2 variant proteins and the role of G6PC2 in human beta**
577 **cells.** Related to Figure S5.

578 (A) Expression levels of the glycosylated forms (upper bands only) of G6PC2 variant proteins were
579 determined in INS-1 832/13 cells by western blot densitometric analysis of Myc-tagged G6PC2
580 constructs relative to tubulin control (n=5). Representative blots are shown for untreated cells together
581 with cells treated with proteasomal inhibitor MG-132 or lysosomal inhibitor chloroquine.

582 (B) Glucose-6-phosphatase activity of L173T and L173V variants in G6PC (proxy for I171T and I171V in
583 G6PC2 respectively) in HEK293 against increasing glucose-6-phosphate concentrations (n=4), with mean
584 $V_{max} \pm SEM$ and $K_m \pm SEM$ values shown for WT and each variant.

585 (C) Glucose-6-phosphatase activity of F258L variant in G6PC (proxy for F256L in G6PC2) in HEK293
586 against increasing glucose-6-phosphate concentrations (n=3), with mean $V_{max} \pm SEM$ and $K_m \pm SEM$
587 values shown. V_{max} and K_m results were computed based on the Michaelis-Menten kinetic model.

588 (D) Effect of G6PC2 WT and variant protein expression on luciferase activity driven by ER stress response
589 elements in HEK293 cells. Relative luciferase units corrected for background activity were normalised to
590 WT for each reporter, from n=6 across two independent experiments (except for F256L, n=3 in one
591 experiment) using two-way ANOVA with Fisher's LSD test comparing each variant to WT.

592 (E) Cellular localisation of R283X in EndoC- β H1 overlaid with markers for the ER (calreticulin) and the
593 trans-Golgi network (TGN46). White arrows point to positions of the Golgi apparatus. Scale bar indicates
594 $10\mu m$.

595 (F) Insulin secretion normalised to total content at basal and high glucose conditions (with and without
596 drug treatments) following 96-120h *G6PC2* knockdown in EndoC- β H1. Unpaired two-tailed Students' t
597 tests were used to compare *G6PC2* knockdown to control for each condition, from n=16 across 4
598 independent experiments. Tol: tolbutamide; Diaz: diazoxide. All data presented as mean $\pm SEM$. *
599 $p=0.01-0.05$; ** $p=0.001-0.01$; *** $p<0.001$.

600

601

602 **Table Legends**

603

604 **Table 1. Single-point coding variant associations meeting the significant threshold for coding variants**
605 **of $P < 2.2 \times 10^{-7}$.** This table includes all novel coding variants meeting this threshold, irrespective of
606 whether they fall in completely new loci or in previously-established loci, provided that the association
607 at the established locus was not shown to be due to a non-coding variant (Table S3) or another coding
608 variant at the same locus. Novel loci are highlighted in bold. HbA1c: glycated haemoglobin; FG: fasting
609 glucose; FI: fasting insulin; 2hGlu: 2h glucose; Alleles E/O: effect allele/other allele; Freq. Effect Allele:
610 frequency of effect allele; Effect (SE): effect size (standard error); *P*: p-value; N: number of samples in
611 the analysis; Novel/previous glycemic trait association: Novel corresponds to a new association result;
612 Locus name of previous association – name used for previously-reported locus. ¹Significant in the
613 European-only analysis in our study. ²Genome-wide significant association with T2D since date of
614 analysis (Mahajan et al., 2018b). ³Association with T2D at $P < 1 \times 10^{-4}$ since date of analysis (Mahajan et al.,
615 2018b). ⁴T2D locus identified in Japanese (Hara et al., 2014) and Mexican (Williams et al., 2014)
616 populations only. The date of our exomes analysis is May 2015. Related to Table S3.

617

618 **Table 2. Gene-based results from broad (NSbroad mask) and strict (NSstrict mask) analyses.** Genes in
619 bold are newly discovered from this effort. N var: total number of variants in that gene-based analysis;
620 P_{burden} : p-value from burden test which assumes all variants have the same direction of effect; P_{SKAT} : p-
621 value from SKAT test which allows for different directions of effect between variants. The lowest p-value
622 is highlighted in bold. Related to Table S4.

623 **Methods**

624

625 **LEAD CONTACT AND MATERIALS AVAILABILITY**

626

627 Further information and requests for resources and reagents should be directed to and will be fulfilled
628 by the Lead Contacts, Inês Barroso (ines.barroso@mrc-epid.cam.ac.uk) and Anna L Gloyn
629 (anna.gloyn@drl.ox.ac.uk).

630

631 **EXPERIMENTAL MODEL AND SUBJECT DETAILS**

632

633 **Studies in humans**

634 MAGIC (Meta-Analysis of Glucose and Insulin-related traits Consortium) was established to focus on the
635 genetic analysis of glycemic traits in individuals without diabetes. In this MAGIC effort, non-diabetic
636 individuals of European (85%), African-American (6%), South Asian (5%), East Asian (2%) and Hispanic
637 (2%) ancestry from up to 64 cohorts participated. Sample sizes were up to 144,060 for HbA1c, 129,665
638 for FG, 104,140 for FI and 57,878 for 2hGlu. Participating cohorts and their characteristics are detailed in
639 **Table S1**.

640

641 **Studies in cellular models**

642 HEK293 cells were cultured in Dulbecco's Modified Eagle's Medium (DMEM) (D6429, Sigma Aldrich),
643 10% (v/v) foetal bovine serum (FBS) (10500-064, Life Technologies), 100 U/ml penicillin and 100 µg/ml
644 streptomycin (15140122, Life Technologies). Huh7 cells were cultured in DMEM (31885, Life
645 Technologies), 10% (v/v) FBS, 100 U/ml penicillin and 100 µg/ml streptomycin. INS-1 832/13 cells were
646 cultured in Roswell Park Memorial Institute-1640 (RPMI-1640) media (R0883, Sigma Aldrich), 10% (v/v)
647 FBS, 100 U/ml penicillin and 100 µg/ml streptomycin, 2 mM L-glutamine (25030081, Life Technologies),
648 1 mM sodium pyruvate (S8636, Sigma Aldrich), 10 mM HEPES (H3537, Sigma Aldrich), 50 µM 2-
649 mercaptoethanol (Life Technologies). EndoC-βH1 cells were cultured in DMEM (31885, Life
650 Technologies), Bovine Serum Albumin (BSA) fraction V (10775835001, Roche), 100 U/ml penicillin and
651 100 µg/ml streptomycin, 2 mM L-glutamine, 50 µM 2-mercaptoethanol, 10 mM nicotinamide (Sigma
652 Aldrich), 5.5 µg/ml transferrin (Sigma Aldrich) and 6.6 ng/ml, sodium selenite (Sigma Aldrich). All cell

653 lines were tested negative for mycoplasma contamination using the MycoAlert Assay kit (Lonza). Cells
654 were maintained at 37°C and 5% CO₂.

655 **METHOD DETAILS**

656

657 **Studies in humans**

658 **Phenotypes**

659 Studied outcomes were FG (mmol/L), Ln-transformed FI (pmol/L), 2hGlu (mmol/L) and HbA1c (% of
660 hemoglobin). Glycemic measurements are described in detail for each contributing cohort in **Table S1**.
661 Individuals with diagnosed or treated diabetes, or those with diabetes on the basis of FG (≥ 7 mmol/L),
662 2hGlu (≥ 11.1 mmol/L) and/or HbA1c ($\geq 6.5\%$) were excluded from analyses.

663

664 **Genotyping and QC**

665 The Illumina HumanExome BeadChip is a genotyping array containing variants that have been observed
666 in sequencing data of ~12,000 individuals. Non-synonymous variants seen at least three times across at
667 least two datasets were included on the exome chip. More lenient criteria were used for splice and
668 nonsense variants. Besides the core content of protein-altering variants, the exome chip contains
669 additional variants including common variants identified in GWAS, ancestry informative markers,
670 mitochondrial variants, randomly selected synonymous variants, HLA tag variants and Y chromosome
671 variants. In this study we analysed association with glycemic traits of 247,470 autosomal and X
672 chromosome variants present on the exome chip. Genotype calling and quality control were performed
673 following protocols developed by the UK Exome Chip or CHARGE consortium (Grove et al., 2013). The
674 exact genotyping array, calling algorithm and QC procedure used by each cohort are depicted in **Table**
675 **S1**.

676

677 **Annotation and functional prediction of variants**

678 Annotation of the exome chip variants was performed using the Ensembl Variant Effect Predictor v78. *In*
679 *silico* functional prediction from SIFT, Polyphen HumDiv, Polyphen HumVar, LRT and MutationTaster was
680 added using dbNSFP v2.9 (Liu et al., 2013; Yourshaw et al., 2015).

681

682 **Statistical analyses**

683 *Single variant analyses*

684 Individual cohorts ran linear mixed models using the raremetalworker (v 4.13.2) or rvtests (v20140723)
685 software (**Table S1**). For each glycemetic outcome, analyses were performed using an additive model for
686 the raw and the inverse normal transformed trait. In the manuscript and in all tables and figures effect
687 estimates and standard errors are for the raw trait, while the p-values are from the inverse normal
688 transformed trait analyses. Analyses were adjusted for age, sex, BMI, study-specific number of PCs and
689 other study-specific covariates (**Table S1**). Raremetal (v4.13.7 or higher) was used to combine results by
690 fixed-effect meta-analyses. Variants with $P < 10^{-4}$ for deviation from Hardy-Weinberg equilibrium or with
691 call rate < 0.99 in individual cohorts were excluded from meta-analyses. In single variant analyses, the
692 threshold for significance was $P < 2.2 \times 10^{-7}$ for coding variants (stop-gained, stop lost, frameshift, splice
693 donor, splice acceptor, initiator codon, missense, in-frame indel and splice region variants). These P -
694 value thresholds were based on a Bonferroni correction weighted by the enrichment for complex trait
695 associations among the different functional annotation categories (Mahajan et al., 2018b;
696 Sveinbjornsson et al., 2016). Significant association signals located more than 500 kb from any variant
697 already known to be associated with the trait at the time of analysis (May 2015) were considered novel
698 for the trait.

699

700 *Gene-based analyses*

701 In addition, raremetal was used to perform gene-based burden and sequence kernel association (SKAT)
702 tests. For both burden and SKAT tests, two *in silico* masks for inclusion of variants in the test were used:
703 NSstrict and NSbroad. The NSstrict mask includes PTVs (splice donor, splice acceptor, stop gained,
704 frameshift, stop lost or initiator codon variant) OR variants that are missense and predicted to be
705 damaging by five prediction algorithms (SIFT, Polyphen HumDiv, Polyphen HumVar, LRT,
706 MutationTaster). The NSbroad mask additionally includes missense variants predicted to be damaging
707 by at least one of the five prediction algorithms AND that have a MAF $<1\%$ in each ancestry group. These
708 MAFs were derived from our single variant HbA1c meta-analyses results (N up to 144,060). For *G6PC2*,
709 we also used masks filtering on functional variants that have been determined *in vitro* to influence
710 protein expression or function. The P -value threshold for significance in gene-based analyses was
711 2.5×10^{-6} (Bonferroni correction for 20,000 genes).

712

713 *Conditional analyses*

714 Approximate conditional analyses were performed using Raremetal v 4.13.8. At known glycemetic trait
715 loci, if previously known GWAS index variants (or good proxies) were present on the exome chip,

716 significant lead coding variants were conditioned on these known index variants and vice versa to
717 identify distinct coding variant signals. At novel loci, to identify additional distinct associated variants,
718 analyses were performed conditioning on the most significant variant at the locus. These analyses were
719 repeated by including the next most significant and distinct associated variant until no exome- or
720 genome-wide significantly-associated variants were left at the locus. For gene-based signals, to identify
721 the variants driving the signal, analyses were performed conditioning on the variant with the most
722 significant p-value that was included in the mask. These analyses were repeated including the next most
723 significant variant until association at the gene was attenuated ($P > 0.05$). If there were both gene-level
724 and known or novel single variant associations at the same locus (within 500 kb), we additionally
725 conditioned on the associated single variant to assess whether the gene-based association was distinct
726 from the single variant association.

727

728 **Putative effector transcript identification**

729 To identify putative effector transcripts, at known glycemic trait loci we considered the transcript a
730 putative effector transcript if there was a distinct coding variant signal (still meeting the threshold for
731 significance of $P < 2.2 \times 10^{-7}$ after conditioning on the non-coding GWAS index variant, for details on
732 these conditional analyses methods refer to the *conditional analyses* methods section above). Coding
733 variant associations at novel loci were followed up on in published GWAS results with higher density
734 coverage (Manning et al., 2012; Wheeler et al., 2017a). If the coding variant was present in the GWAS
735 results, approximate conditional analyses were performed using GCTA (Yang et al., 2012). If the GWAS
736 index variant signal was abolished by conditioning on the coding variant, we considered this as evidence
737 supporting the transcript as a putative effector transcript. If the both the GWAS index variant and the
738 coding variant signals were attenuated, the results were considered uninformative and we considered
739 the transcript in light of other data. We additionally utilized published data to classify effector
740 transcripts, including (1) fine-mapping results from comparable T2D efforts (Mahajan et al., 2018b) and
741 (2) a body of literature establishing a role in glucose metabolism or red blood cell biology (for HbA1c) for
742 certain genes that mapped within our loci. Significant gene-based associations driven by multiple coding
743 variants within a single gene, in particular where an impact on protein expression or function could be
744 demonstrated, were considered strong evidence for the determination of effector transcripts.
745 Combining these approaches, we attempted to identify effector transcripts at each locus, and we
746 classified their likelihood of being correct depending on the strength of the evidence. Those effector
747 transcripts where there was strong evidence from reciprocal conditional analysis or support from

748 published data for the relevant glycemetic trait or phenotype were labelled “gold”; those where the
749 effector transcript could not be defined by conditional analysis (either because it was inconclusive or
750 due to lack of data) but where there was strong biological plausibility for a given gene at the locus were
751 labelled “silver”; those where we had some tentative evidence but that was not strong enough to
752 warrant a “silver” classification were labelled “bronze”, and the remainder were left with an unknown
753 effector transcript. Effector transcript classification into “gold”, “silver” and “bronze” was undertaken
754 independently by four of the authors and the highly concordant consensus score was given (**Table S6**).

755

756 **GeneMANIA network analysis**

757 For network analyses, we used GeneMANIA (v3.5.1), a network approach that searches many large,
758 publicly-available biological datasets to find related genes. These include protein-protein, protein-DNA
759 and genetic interactions, pathways, reactions, gene and protein expression data, protein domains and
760 phenotypic screening profiles. Briefly, GeneMANIA uses a label propagation algorithm for predicting
761 gene function given the composite functional association network (calculated from the databases
762 selected). The weights needed for the label propagation method to work are selected at the beginning
763 of the process. In our case, and according to the defaults, we weighted the network using linear
764 regression, to make genes in the input list interact as much as possible with each other. We analyzed all
765 non-synonymous variants for each locus with a cut-off of association $P < 1 \times 10^{-5}$ with any trait (input
766 genes). We performed four network analyses: (1) HbA1c-associated variants only, (2) FI-associated
767 variants only, (3) FG-associated variants only, and (4) 2hGlu-associated variants only.

768 We selected the 50 default databases to create the composite network, and we allowed the method to
769 find at most 50 genes that are related to our query input list. The resultant networks were investigated
770 to find enriched Gene Ontology (GO) terms and Reactome Pathways. Gene Set Enrichment (GSE) of
771 networks and sub-networks were assessed with ClueGO (Bindea et al., 2009) using GO terms and
772 Reactome gene sets (Croft et al., 2014). The enrichment results were grouped using a Cohen’s Kappa
773 score of 0.4, and terms were considered significant with a Bonferroni-adjusted p-value < 0.05 , provided
774 that there was an overlap of at least three network genes in the relevant GO gene set when calculating
775 GO enrichment. For the pathway selection (Reactome), we set a threshold that the network genes
776 should represent at least 4% of the pathway. These values were applied given the recommended
777 defaults when running ClueGO (Bindea et al., 2009). Cohen’s Kappa statistic was used to measure the
778 gene-set similarity of GO terms and Reactome pathways and allowed us to group enriched terms into

779 functional groups to improve visualization of enriched pathways. We used all genes with GO annotations
780 and at least one interaction in our network database as the background set.

781

782 **Gene set enrichment analysis (GSEA)**

783 For GSEA, we used EC-DEPICT, an extension of the GWAS GSEA method DEPICT (Pers et al., 2015). EC-
784 DEPICT has been described elsewhere (Marouli et al., 2017; Turcot et al., 2018). Briefly, the key feature
785 of EC-DEPICT is the use of “reconstituted” gene sets, which are gene sets collected from many different
786 databases (e.g. canonical pathways, protein-protein interaction networks, and mouse phenotypes) that
787 have been extended based on large-scale microarray co-expression data (Fehrmann et al., 2015; Pers et
788 al., 2015).

789 Six groups of variants were analyzed: (1) HbA1c-associated variants only, (2) FI-associated variants only,
790 (3) FG-associated variants only, (4) 2hGlu-associated variants only, (5) all trait-associated variants, and
791 (6) all trait-associated variants except for HbA1c (see **Methods**). For each trait, we clumped the
792 European summary statistics (+/- 500 kb on either side). Then, the most significant nonsynonymous
793 variant for each locus was included in the analysis, with a cut-off of $P < 10^{-5}$. Annotations from the
794 CHARGE consortium were used to assign variants to genes (see **URL**). After GSEA, highly correlated gene
795 sets were grouped by affinity propagation clustering of all 14,462 gene sets (Frey and Dueck, 2007) into
796 “meta-gene sets” using SciKitLearn.clustering.AffinityPropagation version 0.17 (Abraham et al., 2014).
797 For all visualizations, the gene set within a meta-gene set with the best enrichment P -value was used;
798 heat maps were created with the ComplexHeatmap package in R (Gu et al., 2016).

799 **URL:** CHARGE Consortium ExomeChip annotation file (v6):

800 <http://www.chargeconsortium.com/main/exomechip/>

801

802 *Method and choice of data for permutations:* We performed the EC-DEPICT analysis as described
803 elsewhere (Marouli et al., 2017; Turcot et al., 2018). All analyses are based on a group of 14,462
804 “reconstituted” gene sets, which contains a z-score for probability of gene set membership for each
805 gene (for details, see (Fehrmann et al., 2015; Pers et al., 2015)).

806 Briefly, the basic EC-DEPICT method is as follows. We first obtain a list of significant input variants (the
807 most significant nonsynonymous variant per locus) and then map variants to genes based on
808 annotations from the CHARGE consortium (see **URL**). For each gene set, we obtain the gene set

809 membership z-scores for all trait-associated input genes and sum them to generate a test statistic. We
810 then take 2,000 permuted ExomeChip association studies (described in more detail below) and calculate
811 the average permuted test statistic for that gene set, as well as the permuted standard deviation. For
812 each permutation, the number of top genes we take as “input genes” is matched to the actual observed
813 number of input genes. We then calculate (observed test statistic – average permuted test
814 statistic)/(permuted standard deviation) to generate a z-score, which is converted to a p-value via the
815 normal distribution. False discovery rates were calculated by comparing the observed p-values to a
816 permuted *P*-value distribution generated with an additional set of 50 permuted association studies.

817 The permuted ExomeChip association studies are conducted by (1) generating 2,200 sets of normally
818 distributed phenotypes and (2) using these randomly generated phenotypes to conduct 2,200
819 association studies with real ExomeChip data. Using these permutations to adjust the observed test
820 statistics corrects for any inherent structure in the data (e.g. that pathways made up of longer genes
821 may be more likely to come up as significant by chance).

822 For these analyses, we first generated permutations based on ExomeChip data we had used previously
823 for this purpose: 11,899 samples drawn from three cohorts (Malmö Diet and Cancer [MDC], All New
824 Diabetics in Scania [ANDIS], and Scania Diabetes Registry [SDR]). For simplicity, we refer to these cohorts
825 as the “Swedish permutations.”

826 As part of our GSEA pipeline, we remove input trait-associated variants that are not present in the
827 permuted data to ensure that all variants are appropriately modeled. When using the Swedish
828 permutations, this generally results in removing a substantial fraction of the variants, especially of the
829 very rarest variants (due to the smaller sample size of the Swedish data relative to the data being
830 analyzed). We have previously observed that this filtering can actually improve the GSEA signal, possibly
831 due to more heterogeneous biology or a higher false-positive rate in these very rare variants (Turcot et
832 al., 2018). However, in this case, we observed that in performing this filtering, we excluded variants in
833 several known monogenic disease genes, such as *HNF1A* and *SLC2A2*. Therefore, we wished to repeat
834 the analysis with a set of permutations which would allow us to retain these variants. We thus repeated
835 the analysis with a second set of permutations consisting of 152,249 samples from the UK Biobank
836 (referred to as the “UKBB permutations”). The larger sample size in the UKBB permutations means more
837 variants are present and can therefore be included in the analysis.

838 *Concordance of results from two different sets of permuted distributions across phenotypes: For*

839 completeness, we report the results from the use of both sets of permutations. We note that the results
840 are strongly concordant. The larger number of significant gene sets reported based on the UK Biobank
841 permutations is generally a combination of 1) overall improved power (i.e. more variants are included)
842 and 2) the inclusion of variants in key driver genes absent in the Swedish permutations, encompassing
843 both the monogenic genes mentioned above (e.g. *SLC2A2*) and additional genes with clearly relevant
844 biology (e.g. *CTRB2*, *SLC30A8*). The results from both sets of permutations are summarized below. For all
845 analyses, “significance” refers to a false discovery rate of <0.05 .

846 *All-trait analysis:* After filtering, 78 input genes were included for the analysis with the UKBB
847 permutations and 60 for the analysis with the Swedish permutations. (Note that the difference in the
848 number of input genes is due to the presence of a larger number of input variants in the UKBB
849 permutations – see above). We found 234 significant gene sets in 86 meta-gene sets based on the UKBB
850 permutations (**Figure S2**) and 133 gene sets in 51 meta-gene sets based on the Swedish permutations
851 (**Figure S3**). The correlation between the UKBB and Swedish analyses was $r = 0.902$, $P < 10^{-300}$.

852

853 *All-traits-except-HbA1c analysis:* After filtering, 45 input genes were included for the analysis with the
854 UKBB permutations and 33 for the analysis with the Swedish permutations. We found 128 significant
855 gene sets in 53 meta-gene sets based on the UKBB permutations (**Figure S2**) and 45 significant gene sets
856 in 18 meta-gene sets based on the Swedish permutations (**Figure S3**). The correlation between the UKBB
857 and Swedish analyses was $r = 0.882$, $P < 10^{-300}$.

858

859 *HbA1c-only analysis:* After filtering, 41 input genes were included for the analysis with the UKBB
860 permutations and 33 for the analysis with the Swedish permutations. We found 191 significant gene sets
861 in 73 meta-gene sets based on the UKBB permutations (**Figure S2**) and 120 gene sets in 41 meta-gene
862 sets based on the Swedish permutations. (**Figure S3**). The correlation between the UKBB and Swedish
863 analyses was $r = 0.936$, $P < 10^{-300}$.

864 *FG-only analysis:* After filtering, 26 input genes were included for the analysis with the UKBB
865 permutations and 22 for the analysis with the Swedish permutations. We found 106 significant gene sets
866 in 39 meta-gene sets based on the UKBB permutations (**Figure S2**) and 48 significant gene sets in 15
867 meta-gene sets based on the Swedish permutations (**Figure S3**). The correlation between the UKBB and
868 Swedish analyses was $r = 0.939$, $P < 10^{-300}$.

869 *2hGlu-only analysis*: After filtering, 12 input genes were included for the analysis with the UKBB
870 permutations and 7 for the analysis based on the Swedish permutations. We found 56 significant gene
871 sets in 17 meta-gene sets based on the UKBB permutations (**Figure S2**), with no significant gene sets
872 based on the Swedish permutations. The correlation between the UKBB and Swedish analyses was $r =$
873 $0.787, P < 10^{-300}$.

874 *Fl-only analysis*: After filtering, 11 input genes were included for the analysis with the UKBB
875 permutations and 8 for the analysis with the Swedish permutations. There were no significant gene sets
876 from either analysis. The correlation between the UKBB and Swedish analyses was $r = 0.860, P < 10^{-300}$.

877 **Visualization**: As in previous work (Marouli et al., 2017; Turcot et al., 2018), we have included all trait-
878 associated variants in the heat maps, even if they were excluded from the analysis (e.g. because they
879 were absent in the permutations or did not have a nonsynonymous annotation in the CHARGE
880 annotation file). This is because we assume that if the genes harboring those variants have strong
881 predicted membership in significantly trait-associated gene sets, they are still good candidates for
882 prioritization. In fact, this may be even stronger evidence in favor of these genes because they did not
883 contribute to the enrichment analysis and therefore their prioritization is independently derived (and
884 provides even more support to the implicated biology).

885

886 **Tissue enrichment analysis**

887 We analysed identified genes (all 51 effector transcripts) for tissue enrichment using publicly available
888 expression data from the GTEx project, version 7 and publicly-available islet expression data (van de
889 Bunt et al., 2015). We use transcripts per million (TPM) values for gene level analyses. We have excluded
890 genes from non-coding proteins and only used those with unique HGNC IDs ($n = 20,160$). We ranked all
891 genes by median TPM across all samples, and generated 10,000 permutations of each gene set list
892 (golden, silver, and bronze) by selecting a random gene for each entry in the gene set within ± 150 ranks
893 of the transcript for that gene. For each sample in GTEx tissues, the TPM values were converted into
894 ranks for that gene, and sums of ranks within each tissue were computed for each gene. We calculated
895 enrichment p-values for each tissue by taking the total number of instances when the gene list of
896 interest had a lower sum of ranks than the permuted sum of ranks (divided by the total number of
897 permutations). To check that our results were not driven by sample size differences in each of the 45
898 analyzed GTEx tissues and islet tissue, we applied a 'downsampling' strategy. We performed 3 different

899 downsampling analyses with 100, 150 and 175 samples chosen at random from each of the selected
900 GTEx tissues and compared them to the results obtained with the whole dataset. During each
901 downsampling round, we only used those tissues with at least the target number of samples (100, 150
902 or 175) because the random selection was performed under a no-replacement condition. Our results
903 were robust to sample size differences and the trends observed were not driven by differences in
904 sample sizes across tissues.

905

906 **RNA-sequencing of human islets**

907 RNA from human islet samples (n=150) was sequenced on Illumina HiSeq2000 as previously described
908 (van de Bunt et al., 2015). Allele-specific expression was assessed using MAMBA (Pirinen et al., 2015).
909 For the isoform effects, all protein-coding and lincRNA transcripts from GRCh37 (Ensembl release 75)
910 were quantified using Salmon v0.8.1 (Patro et al., 2017). Isoform ratios were calculated by dividing each
911 transcript's expression by the total expression of that gene. For the QTL analysis, all isoforms with
912 expression in < 50% or all samples, with no variance between samples, only 0 or 1 fractions across
913 samples, or those originating from non-autosomal chromosomes were removed. Ratios of the remaining
914 transcripts were rank-transformed to normality. Subsequently, 30 PEER factors to account for potential
915 sources of non-genetic noise were derived from the normalized isoform ratios, and, together with three
916 genotype principal components and a sex covariate, were used in the QTL analysis using FastQTL (Ongen
917 et al., 2016). Finally, the resulting beta-approximated p-values were adjusted for multiple testing across
918 all tested transcripts using the Benjamini-Hochberg procedure.

919

920 **Studies in cellular models**

921

922 **Site directed mutagenesis.** Human *G6PC* (NM_000151.3) and *G6PC2* cDNA (NM_021176.2) within a
923 pCMV6-Entry vector (with a C-terminal Myc-FLAG-tag) was purchased from OriGene (RC215623 and
924 RC211146 respectively). For the study of PTVs, an N-terminal V5 tag sequence (5'-
925 GGTAAGCCTATCCCTAACCTCTCCTCGGTCTCGATTCTACG-3') was cloned into the OriGene vectors. Single
926 nucleotide substitutions were generated in the *G6PC* or *G6PC2* coding sequence using Quikchange II
927 Site-Directed Mutagenesis (Agilent). All mutations were verified by Sanger sequencing and in each case,
928 only the desired nucleotide changes were introduced.

929 **Western blot analyses.** Western blots were performed on total protein lysates collected from human
930 HEK293 and Huh7 cells and rat INS-1 832/13 cells transfected with each wild type or mutant
931 *G6PC/G6PC2* construct using Lipofectamine 2000 (Invitrogen) according to manufacturer's instructions.
932 All cell lines were tested negative for mycoplasma contamination. For the inhibition of cell proteolysis,
933 cells were treated with 10 μ M MG-132 (Calchembio) or 100 μ M chloroquine (Sigma) for 15h. For
934 inhibition of N-linked glycosylation, cells were treated with 1 μ g/ml tunicamycin (Sigma) for 15h. Cells
935 were collected 36-48h after transfection and homogenized in lysis buffer. Cell lysates were separated by
936 4–12% SDS-PAGE (Bio-Rad/Invitrogen). The antibodies used for determining recombinant G6PC/G6PC2
937 expression were: anti-FLAG M2 (Sigma, F1804), anti-V5 (Invitrogen, 46-0705) or anti-myc 4A6 (Millipore,
938 05-724). A β -tubulin antibody (Santa Cruz, sc-9104) was used as a loading control. Secondary antibodies
939 specific to mouse or rabbit IgG were purchased from Thermo Fisher Scientific. Protein bands were
940 detected using the western enhanced chemiluminescence substrate (BioRad).

941 **Immunofluorescence microscopy.** Human HEK293, Huh7 and EndoC- β H1 cells were transfected using
942 FuGene 6 transfection reagent (Promega) according to manufacturer's instructions, in 4-well chamber
943 slides (BD Biosciences). After 48h, cells were fixed with 4% paraformaldehyde in PBS, permeabilized with
944 0.05% Triton X-100 in PBS and blocked with 10% BSA in PBS-Tween 20. Double immunostaining of cells
945 was carried out using anti-FLAG M2 (Sigma, F1804) or anti-V5 (Invitrogen, 46-0705), together with anti-
946 calreticulin (Thermo, PA3-900) or anti-TGN46 (Sigma, T7576) primary antibodies. The secondary
947 antibodies used were anti-mouse Alexa Fluor 488 and anti-rabbit Alexa Fluor 568, both from Life
948 Technologies. DRAQ5 fluorescent probe (Thermo Fisher Scientific) was applied at 20 μ M as a far-red
949 nuclear stain. Finally, slides were mounted with ProLong Gold antifade reagent (Life Technologies) and
950 visualized on a BioRad Radiance 2100 confocal microscope with a 60X 1.0 N.A. objective. Images were
951 acquired with different laser settings that were optimized for each sample and therefore fluorescent
952 intensities are not comparable across samples.

953 **Glucose-6-phosphatase activity of microsomal samples.** For the collection of microsomes, HEK293 cells
954 were transfected with 12.5 μ g of wild type or mutant *G6PC* construct in 10 cm dishes using
955 Lipofectamine 2000 (Invitrogen). For the study of *G6PC2* variant activity, site directed mutagenesis was
956 carried out within the conserved sequence regions on the *G6PC* background. Cells were cultured in
957 Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% fetal bovine serum, 100 U/ml
958 penicillin, and 100 μ g/ml streptomycin. At least two dishes of cells per condition were harvested 48h
959 after transfection and scraped into 0.25M sucrose-5 mM HEPES buffer (SH) followed by several rinses in

960 SH. Cells were mechanically homogenized using a Potter-Elvehjem glass tissue grinder and Teflon pestle,
961 followed by 12 passes through a 27-gauge syringe needle. The homogenate was subjected to
962 centrifugation at 10,000 g for 10 min and the supernatant (post-nuclear fraction) was further
963 centrifuged at 100,000 g for 1h in a TLA 100.4 rotor in an Optima TLX ultracentrifuge (Beckman Coulter).
964 A pellet containing the microsomal fraction was obtained and resuspended in SH. An aliquot of each
965 microsomal sample was lysed in lysis buffer for protein quantification using the Bradford reagent (Bio-
966 Rad) and analysed by western blot (BioRad) to determine the relative levels of recombinant G6PC WT
967 and variant expression. For glucose-6-phosphatase assays, matched amounts of G6PC WT or variant
968 protein (approximately 1-5µg of microsomal protein) were each incubated in a 200 µl reaction mix
969 containing 100 mM MES pH 6.5, and 0 to 20 mM G6P at 37°C for 8 min. The reaction was terminated by
970 addition of 20% trichloroacetic acid and centrifuged at 4,000 rpm for 10 min in a microcentrifuge. The
971 supernatant was mixed in equal parts with a Tausky-Shorr colour reagent (1% ammonium molybdate,
972 5% iron(II) sulphate heptahydrate in 0.5M H₂SO₄) for 7.5 min before measuring absorbance at 660 nm
973 on a spectrophotometer (Molecular Devices Ltd). The amount of phosphates detected was calculated
974 using a KH₂PO₄ standard curve. Results were expressed as mean normalised activity (nmol/mg/min)
975 relative to the activity of wild type at 20 mM G6P for every experiment. Finally, Michaelis-Menten
976 enzyme kinetic analysis and paired t tests of determined kinetic constants were carried out on GraphPad
977 Prism 6.0.

978 **ER stress response reporter assays.** HEK293 cells were co-transfected with *G6PC2* WT or variant
979 constructs and pGL3-Promoter constructs containing ER stress response elements (ERSE-I and ERSE-II) or
980 UPR elements (UPRE-P and UPRE-W) using the FuGene 6 transfection reagent (Promega). A Renilla
981 luciferase gene-containing pRL-CMV was also co-transfected as an internal transfection control. Cells
982 were lysed in passive lysis buffer (Promega) and assayed using the Dual Luciferase Assay System
983 (Promega).

984 **Insulin secretion analysis in EndoC-βH1 cells.** Gene knockdown was carried out on EndoC-βH1 cells
985 using ON-TARGETplus siRNA (Dharmacon, GE Healthcare) and Lipofectamine RNAiMAX (Life
986 Technologies) at a final concentration of 25 nmol/L siRNA. For static incubation experiments, cells were
987 placed in 2.8 mM glucose DMEM (11966, Gibco by Life Technologies) overnight. Cells were starved in 0
988 mM glucose medium for 1h the following day, then stimulated in DMEM containing 1 mM glucose, 6
989 mM glucose, 20 mM glucose, 20 mM glucose with 100 µM tolbutamide (Sigma Aldrich) or 20 mM
990 glucose with 100 µM diazoxide (Sigma Aldrich) at 37°C for 1h. Each condition was carried out in triplicate

991 or quadruplicate wells within each experiment. Viable cell count was measured using the CyQUANT
992 Direct Cell Proliferation Assay kit (C35012, Thermo Scientific). All cell count values were expressed as
993 fluorescent units normalised to mean cell count at 1 mM glucose. Cells were extracted for analysis of
994 insulin content with cold HCl-ethanol (Sigma Aldrich). Insulin levels were measured using the human
995 insulin AlphaLISA detection kit (AL204C, Perkin Elmer).

996

997 **QUANTIFICATION AND STATISTICAL ANALYSIS**

998

999 Western blot bands for protein expression studies were quantified by densitometry analysis using
1000 ImageJ and densitometric data between G6PC/G6PC2 WT and each variant from 3-5 independent
1001 experiments were compared using two-tailed paired Students' t tests. For enzymatic assays, mean
1002 differences in activity between G6Pase WT protein and each variant protein for the substrate G6P were
1003 compared using two-tailed unpaired Students' t tests of the determined kinetic constants V_{max} and K_m .
1004 For the analysis of ER stress luciferase activity data, a two-way analysis of variance (ANOVA) was applied
1005 to compare mean fold difference in reporter activity between G6PC WT and variant. For gene expression
1006 analyses, G6PC2 KO and control cells were analysed using two-tailed unpaired Students' t tests. For the
1007 analysis of insulin secretion data, mean differences between G6PC2 KO cells and control cells for each
1008 condition or time point were compared using two-tailed unpaired Students' t tests. Plotting of graphs
1009 and statistical analyses were carried out on GraphPad Prism 6.0 or 7.0. A P value <0.05 was considered
1010 significant.

1011

1012 **Acknowledgements**

1013

1014 HepG2 cells were the kind gift of the Tomlinson group (OCDEM), INS-1 832/13 cells were the kind gift of
1015 Jochen Lang (Université de Bordeaux), and EndoC- β H1 cells were gifted by Raphael Scharfmann
1016 (INSERM).

1017

<u>Study/Individual</u>	<u>Acknowledgment</u>
--------------------------------	------------------------------

1958 British Birth	Analysis was supported by BHF programme grant (Deloukas) RG/14/5/30893
--------------------	--

cohort

- AGES** This study has been funded by NIH contracts N01-AG-1-2100 and 271201200022C, the NIA Intramural Research Program, Hjartavernd (the Icelandic Heart Association), and the Althingi (the Icelandic Parliament). The study is approved by the Icelandic National Bioethics Committee, VSN: 00-063. The researchers are indebted to the participants for their willingness to participate in the study.
- Andrei I. Tarasov** Andrei I. Tarasov acknowledges the support from Oxford Biomedical Research Centre (OxBRC fellowship)
- Andrew P Morris** Andrew P Morris is a Wellcome Trust Senior Fellow in Basic Biomedical Science.
- Anna L. Gloyn** ALG is a Wellcome Trust Senior Fellow in Basic Biomedical Science. This work was funded by the Wellcome Trust (095101/Z/10/Z and 200837/Z/16/Z).
- ARIC** The Atherosclerosis Risk in Communities (ARIC) study is carried out as a collaborative study supported by the National Heart, Lung, and Blood Institute (NHLBI) contracts (HHSN268201100005C, HHSN268201100006C, HHSN268201100007C, HHSN268201100008C, HHSN268201100009C, HHSN268201100010C, HHSN268201100011C, and HHSN268201100012C). The authors thank the staff and participants of the ARIC study for their important contributions. Funding support for “Building on GWAS for NHLBI-diseases: the U.S. CHARGE consortium” was provided by the NIH through the American Recovery and Reinvestment Act of 2009 (ARRA) (5RC2HL102419). ML was supported by a National Heart, Lung, and Blood Institute T32-HL0072024 Cardiovascular Epidemiology Training Grant.
- ASCOT** This work was supported by Pfizer, New York, NY, USA, for the ASCOT study and the collection of the ASCOT DNA repository; by Servier Research Group, Paris, France; and by Leo Laboratories, Copenhagen, Denmark. We thank all ASCOT trial participants, physicians, nurses, and practices in the participating countries for their important contribution to the study. In particular we thank Clare Muckian and David Toomey for their help in DNA extraction, storage, and handling. This work forms part of the research programme of the NIHR Cardiovascular Biomedical Research Unit at Barts

Athero-Express
Biobank Study

Sander W. Van der Laan is funded through grants from the Interuniversity Cardiology Institute of the Netherlands (ICIN, 09.001), CVON (GENIUS), CVgenes-at-Target (EU FP7/2007-2013 under grant agreement n° HEALTH-F2-2013-601456). Claudia Tersteeg, Krista den Ouden, Mirjam B. Smeets, Sander M. van de Weg, Noortje A.M. van den Dungen, and Loes B. Collé are graciously acknowledged for their work on the DNA extraction. Evelyn Velema, Sara van Laar, Arjan H. Schoneveld, and Petra H. Homoed-van der Kraak are graciously acknowledged for their past and continuing work on the Athero-Express Biobank Study. We would also like to thank all the (former) employees involved in the Athero-Express Biobank Study of the Departments of Surgery of the St. Antonius Hospital Nieuwegein and University Medical Center Utrecht for their continuing work. Jessica van Setten is graciously acknowledged for her help in the quality assurance and quality control of the genotype data. Lastly, we would like to thank all participants of the Athero-Express Biobank Study; without you these kinds of studies would not be possible.

BioMe

The Mount Sinai BioMe Biobank is supported by The Andrea and Charles Bronfman Philanthropies.

Caroline Hayward

Caroline Hayward is supported by an MRC University Unit Programme Grant “QTL in Health and Disease” MC_PC_U127592696.

CHS

This CHS research was supported by NHLBI contracts HHSN268201200036C, HHSN268200800007C, HHSN268201800001C, N01HC55222, N01HC85079, N01HC85080, N01HC85081, N01HC85082, N01HC85083, N01HC85086; and NHLBI grants U01HL080295, R01HL068986, R01HL087652, R01HL105756, R01HL103612, R01HL120393, and U01HL130114 with additional contribution from the National Institute of Neurological Disorders and Stroke (NINDS). Additional support was provided through R01AG023629 from the National Institute on Aging (NIA). A full list of principal CHS investigators and institutions can be found at CHS-NHLBI.org. The provision of genotyping data was supported in part by the National Center for Advancing Translational Sciences, CTSI grant UL1TR001881, and the National Institute of Diabetes and Digestive and Kidney Disease Diabetes Research Center (DRC) grant DK063491 to the Southern California Diabetes Endocrinology Research Center. The content is solely the responsibility of the authors and does not

necessarily represent the official views of the National Institutes of Health.

COPSAC2000

We greatly acknowledge the private and public research funding allocated to COPSAC and listed on www.copsac.com, with special thanks to The Lundbeck Foundation (Grant nr. R16-A1694); Ministry of Health (Grant nr. 903516); Danish Council for Strategic Research (Grant nr.: 0603-00280B); The Danish Council for Independent Research and The Capital Region Research Foundation as core supporters. The funding agencies did not have any influence on study design, data collection and analysis, decision to publish or preparation of the manuscript. No pharmaceutical company was involved in the study. We gratefully express our gratitude to the participants of the COPSAC2000 cohort study for all their support and commitment. We also acknowledge and appreciate the unique efforts of the COPSAC research team.

CROATIA_Korcula

We would like to acknowledge the contributions of the recruitment team in Korcula, the administrative teams in Croatia and Edinburgh and the people of Korcula. Exome array genotyping was performed at the Clinical Research Facility University of Edinburgh, Edinburgh, UK and funded by UK's Medical Research Council

DIABNORD

We are grateful to the study participants who dedicated their time and samples to these studies. We also thank the VHS, the Swedish Diabetes Registry and Umeå Medical Biobank staff for biomedical data and DNA extraction. We also thank M Sterner, G Gremesperger and P Storm for their expert technical assistance with genotyping and genotype data preparation. The current study was funded by Novo Nordisk, the Swedish Research Council, Pålssons Foundation, the Swedish Heart Lung Foundation, and the Skåne Regional Health Authority (all to PWF).

DPS

The DPS has been financially supported by grants from the Academy of Finland (117844 and 40758, 211497, and 118590 (MU)); The EVO funding of the Kuopio University Hospital from Ministry of Health and Social Affairs (5254), Finnish Funding Agency for Technology and Innovation (40058/07), Nordic Centre of Excellence on 'Systems biology in controlled dietary interventions and cohort studies, SYSDIET (070014), The Finnish Diabetes Research Foundation, Yrjö Jahnsson Foundation (56358), Sigrid Juselius Foundation and TEKES grants 70103/06 and 40058/07.

DR's EXTRA The DR's EXTRA Study was supported by the Ministry of Education and Culture of Finland (722 and 627;2004-2011), Academy of Finland (102318; 104943;123885; 211119), Kuopio University Hospital, Finnish Diabetes Association, Finnish Foundations for Cardiovascular Research, Päivikki and Sakari Sohlberg Foundation, by European Commission FP6 Integrated Project (EXGENESIS); LSHM-CT-2004-005272, City of Kuopio and Social Insurance Institution of Finland (4/26/2010).

Edmonton Human islet data The authors thank James Lyon at the Alberta Diabetes Institute IsletCore for his work on human islet isolations. We also thank the Human Organ Procurement and Exchange Program (Edmonton) and the Trillium Gift of Life Network (Toronto) and other organ procurement agencies for their efforts in obtaining human pancreas for research.

EFSOCH This paper presents independent research supported by the National Institute for Health Research (NIHR) Exeter Clinical Research Facility. The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health.

EPIC-Norfolk The EPIC-Norfolk study (DOI 10.22025/2019.10.105.00004) has received funding from the Medical Research Council (MR/N003284/1 and MC-UU_12015/1) and Cancer Research UK (C864/A14136). The genetics work in the EPIC-Norfolk study was funded by the Medical Research Council ((MC_PC_13048)).We are grateful to all the participants who have been part of the project and to the many members of the study teams at the University of Cambridge who have enabled this research

EPIC-Potsdam The study was supported in part by a grant from the German Federal Ministry of Education and Research (BMBF) to the German Center for Diabetes Research (DZD e.V.). The recruitment phase of the EPIC-Potsdam study was supported by the Federal Ministry of Science, Germany (01 EA 9401) and the European Union (SOC 95201408 05 F02). The follow-up of the EPIC-Potsdam study was supported by German Cancer Aid (70-2488-Ha I) and the European Community (SOC 98200769 05 F02). Exome chip genotyping of EPIC-Potsdam samples was carried out under supervision of Per Hoffmann and Stefan Herms at Life & Brain GmbH, Bonn. We are grateful to the Human Study Centre (HSC) of the German Institute of Human Nutrition Potsdam-Rehbrücke, namely the trustee and the data hub for the processing, and the participants for the provision of the data, the biobank for the

processing of the biological samples and the head of the HSC, Manuela Bergmann, for the contribution to the study design and leading the underlying processes of data generation.

EpiHealth

EpiHealth was supported by the Swedish Research Council strategic research network Epidemiology for Health, Uppsala University and Lund University. Genotyping in EpiHealth was supported by Swedish Heart-Lung Foundation (grant no. 20120197 and 20140422), Knut och Alice Wallenberg Foundation (grant no. 2013.0126), and Swedish Research Council (grant no. 2012-1397). Genotyping was performed by the SNP&SEQ Technology Platform in Uppsala. We thank the EpiHealth participants for their dedication and commitment.

ERF STUDY

ERF study is grateful to all study participants and their relatives, general practitioners and neurologists for their contributions and to P. Veraart for her help in genealogy, J. Vergeer for the supervision of the laboratory work and P. Snijders for his help in data collection.

Funding: Erasmus Rucphen Family (ERF) was supported by the Consortium for Systems Biology (NCSB), both within the framework of the Netherlands Genomics Initiative (NGI)/Netherlands Organisation for Scientific Research (NWO). ERF study as a part of EUROSPAN (European Special Populations Research Network) was supported by European Commission FP6 STRP grant number 018947 (LSHG-CT-2006-01947) and also received funding from the European Community's Seventh Framework Programme (FP7/2007-2013)/grant agreement HEALTH-F4-2007-201413 by the European Commission under the programme "Quality of Life and Management of the Living Resources" of 5th Framework Programme (no. QLG2-CT-2002-01254) as well as FP7 project EUROHEADPAIN (nr 602633). High-throughput analysis of the ERF data was supported by joint grant from Netherlands Organisation for Scientific Research and the Russian Foundation for Basic Research (NWO-RFBR 047.017.043). The exome-chip measurements have been funded by the Netherlands Organization for Scientific Research (NWO; project number 184021007) and by the Rainbow Project (RP10; Netherlands Exome Chip Project) of the Biobanking and Biomolecular Research Infrastructure Netherlands (BBMRI-NL; www.bbmri.nl (<http://www.bbmri.nl>)). Ayse Demirkan is supported by a Veni grant (2015) from ZonMw. Ayse Demirkan, Jun Liu and

Cornelia van Duijn have used exchange grants from PRECEDI. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscripts.

- Family Heart Study (FamHS) The Family Heart Study (FamHS) was supported by NIH grants R01-HL-087700 and R01-HL-088215 from NHLBI, and R01-DK-8925601 and R01-DK-075681 from NIDDK.
- Fenland The Fenland Study is funded by the Medical Research Council (MC_U106179471) and Wellcome Trust. We are grateful to all the volunteers for their time and help, and to the General Practitioners and practice staff for assistance with recruitment. We thank the Fenland Study Investigators, Fenland Study Co-ordination team and the Epidemiology Field, Data and Laboratory teams.
- FIA3 Jansson J-H was responsible for the identification of MI cases. The FIA3 study was supported in part by a grant from the Swedish Heart-Lund Foundation (to PWF).
- FIN-D2D 2007 The FIN-D2D 2007 study was supported by funds from the hospital districts of Pirkanmaa; Southern Ostrobothnia; North Ostrobothnia; Central Finland and Northern Savo; the Finnish National Public Health Institute; the Finnish Diabetes Association; the Ministry of Social Affairs and Health in Finland; Finland's Slottery Machine Association; the Academy of Finland [grant number 129293] and Commission of the European Communities, Directorate C-Public Health [grant agreement no. 2004310].
- FINRISK 2007 VS was supported by the Finnish Foundation for Cardiovascular Research. PJ was supported by the Academy of Finland #118065.
- Framingham Heart Study FHS: Genotyping, quality control and calling of the Illumina HumanExome BeadChip in the Framingham Heart Study was supported by funding from the National Heart, Lung and Blood Institute Division of Intramural Research (Daniel Levy and Christopher J. O'Donnell, Principle Investigators). Also supported by National Institute for Diabetes and Digestive and Kidney Diseases (NIDDK) U01 DK078616, NIDDK K24 DK080140 and American Diabetes Association Mentor-Based Postdoctoral Fellowship Award #7-09-MN-32, all to Dr. Meigs, and NIDDK Research Career Award K23 DK65978, a Massachusetts General Hospital Physician Scientist Development Award and a Doris Duke Charitable Foundation Clinical Scientist Development Award to Dr. Florez.

FUSION The FUSION study was supported by DK093757, DK072193, DK062370, and ZIA-HG000024. HAK has received funding from Academy of Finland (support for clinical research careers, grant no 258753).

Generation Scotland We would like to acknowledge the contributions of the families who took part in the Generation Scotland: Scottish Family Health Study, the general practitioners and Scottish School of Primary Care for their help in recruiting them, and the whole Generation Scotland team, which includes academic researchers, IT staff, laboratory technicians, statisticians and research managers. Generation Scotland received core funding from the Chief Scientist Office of the Scottish Government Health Directorate CZD/16/6 and the Scottish Funding Council HR03006. Genotyping of the GS:SFHS samples was carried out by staff at the Genetics Core Laboratory at the Clinical Research Facility, University of Edinburgh, Scotland and was funded by the UK's Medical Research Council.

GENOA Support for GENOA was provided by the National Heart, Lung and Blood Institute (HL054464; HL054481; HL087660; HL119443; HL086694) of the National Institutes of Health. We thank Eric Boerwinkle, PhD and Megan L. Grove, MS, University of Texas Health Science Center, Houston, Texas, USA for their help with genotype calling. We would also like to thank the families that participated in the GENOA study.

GIANT Anne E Justice (AEJ) is funded under NIH 5K99HL130580-02.

GLACIER We are indebted to the study participants who dedicated their time and samples to these studies. We thank J Hutiainen and Å Ågren (Umeå Medical Biobank) for data organization and K Enquist and T Johansson (Västerbottens County Council) for technical assistance with DNA extraction. We also thank M Sterner, G Gremesperger and P Storm for their expert technical assistance with genotyping and genotype data preparation. The current study was funded by Novo Nordisk, the Swedish Research Council, Pålhlssons Foundation, the Swedish Heart Lung Foundation, and the Skåne Regional Health Authority (all to PWF).

GoDARTS GoDARTS study was funded by The Wellcome Trust Study Cohort Wellcome Trust Functional Genomics Grant (2004-2008) (Grant No: 072960/2/03/2) and The Wellcome Trust Scottish Health Informatics Programme (SHIP) (2009-2012). (Grant No: 086113/Z/08/Z).

HABC	The Health, Aging, and Body Composition (HABC) Study is supported by NIA contracts N01AG62101, N01AG62103, and N01AG62106. The genome-wide association study was funded by NIA grant 1R01AG032098-01A1 to Wake Forest University Health Sciences.
HANDLS (Healthy Aging in Neighborhoods of Diversity across the Life Span study)	We would like to thank the Healthy Aging in Neighborhoods of Diversity across the Life Span (HANDLS) study participants, study coordinator, medical staff and field workers. Exome chip genotyping was performed at the Laboratory of Neurogenetics, National Institute on Aging, National Institutes of Health (NIH). The HANDLS study was supported by the Intramural Research Program of the NIH, National Institute on Aging and the National Center on Minority Health and Health Disparities (project # Z01-AG000513 and human subjects protocol number 09-AG-N248). Data analyses for the HANDLS study utilized the computational resources of the NIH HPC Biowulf cluster at the National Institutes of Health, Bethesda, MD. (http://hpc.nih.gov).
Health and Retirement Study (HRS)	HRS is supported by the National Institute on Aging (NIA U01AG009740). The genotyping was funded separately by the National Institute on Aging (RC2 AG036495, RC4 AG039029). Our genotyping was conducted by the NIH Center for Inherited Disease Research (CIDR) at Johns Hopkins University. Genotyping quality control and final preparation of the data were performed by the University of Michigan School of Public Health.
Health2006	The Health2006 was financially supported by grants from the Velux Foundation; The Danish Medical Research Council, Danish Agency for Science, Technology and Innovation; The Aase and Ejner Danielsens Foundation; ALK-Abello A/S, Hørsholm, Denmark, and Research Centre for Prevention and Health, the Capital Region of Denmark.
Health2008	The Health2008 was supported by the Timber Merchant Vilhelm Bang's Foundation, the Danish Heart Foundation (Grant number 07-10-R61-A1754-B838-22392F), and the Health Insurance Foundation (Helsefonden) (Grant number 2012B233).
Heather M. Highland	Heather M. Highland is supported by funding from NHLBI training grant T32 HL007055.

HELIC MANOLIS and HELIC Pomak This work was funded by the Wellcome Trust (098051) and the European Research Council (ERC-2011-StG 280559-SEPI). The MANOLIS cohort is named in honour of Manolis Giannakakis, 1978-2010. We thank the residents of the Mylopotamos villages, and of the Pomak villages, for taking part. The HELIC study has been supported by many individuals who have contributed to sample collection (including A. Athanasiadis, O. Balafouti, C. Batzaki, G. Daskalakis, E. Emmanouil, C. Giannakaki, M. Giannakopoulou, A. Kaparou, V. Kariakli, S. Koinaki, D. Kokori, M. Konidari, H. Koundouraki, D. Koutoukidis, V. Mamakou, E. Mamalaki, E. Mpamiaki, M. Tsoukana, D. Tzakou, K. Vosdogianni, N. Xenaki, E. Zengini), data entry (T. Antonos, D. Papagrighoriou, B. Spiliopoulou), sample logistics (S. Edkins, E. Gray), genotyping (R. Andrews, H. Blackburn, D. Simpkin, S. Whitehead), research administration (A. Kolb-Kokocinski, S. Smee, D. Walker) and informatics (M. Pollard, J. Randall).

Hidetoshi Kitajima Hidetoshi Kitajima was funded by Manpei Suzuki Diabetes Foundation Grant-in-Aid for the young scientists working abroad.

Inês Barroso Inês Barroso is funded by Wellcome (WT206194)

Inter99 The Inter99 was initiated by Torben Jørgensen (PI), Knut Borch-Johnsen (co-PI), Hans Ibsen and Troels F. Thomsen. The steering committee comprises the former two and Charlotta Pisinger. The study was financially supported by research grants from the Danish Research Council, the Danish Centre for Health Technology Assessment, Novo Nordisk Inc., Research Foundation of Copenhagen County, Ministry of Internal Affairs and Health, the Danish Heart Foundation, the Danish Pharmaceutical Association, the Augustinus Foundation, the Ib Henriksen Foundation, the Becket Foundation, and the Danish Diabetes Association.

InterAct Consortium Funding for the InterAct project was provided by the EU FP6 programme (grant number LSHM_CT_2006_037197). We thank all EPIC participants and staff for their contribution to the study. We thank the lab team at the MRC Epidemiology Unit for sample management and Nicola Kerrison for data management.

Jennifer Asimit Medical Research Council Methodology Research Fellowship (MR/K021486/1)

JHS The JHS is supported by contracts HHSN268201300046C, HHSN268201300047C, HHSN268201300048C, HHSN268201300049C, HHSN268201300050C from the National Heart, Lung and Blood Institute and the National Institute on Minority Health and Health Disparities. ExomeChip genotyping was supported by the NHLBI of the National Institutes of Health under award number R01HL107816 to S. Kathiresan. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Joel N. Hirschhorn NIH R01DK075787

KORA The KORA study was initiated and financed by the Helmholtz Zentrum München – German Research Center for Environmental Health, which is funded by the German Federal Ministry of Education and Research (BMBF) and by the State of Bavaria. Furthermore, KORA research was financed by a grant from the BMBF to the German Center for Diabetes Research (DZD) and a grant from the Ministry of Innovation, Science, Research and Technology of the state North Rhine-Westphalia (Düsseldorf, Germany). It was also supported within the Munich Center of Health Sciences (MC-Health), Ludwig-Maximilians-Universität, as part of LMUinnovativ. The authors are grateful to all members of the Helmholtz Zentrum München, the field staff in Augsburg, and the Augsburg registry team who were involved in the planning, organization, and conduct of the KORA studies. In addition, the authors express their appreciation to all study participants.

Leipzig Adults This work was supported by grants from the German Research Council (SFB- 1052 “Obesity mechanisms”; B01; B03), from the German Diabetes Association and from the DHFD (Diabetes Hilfs- und Forschungsfonds Deutschland). IFB AdiposityDiseases is supported by the Federal Ministry of Education and Research (BMBF), Germany, FKZ: 01EO1501 (AD2-060E, AD2-06E99). This work was further supported by the Kompetenznetz Adipositas (Competence network for Obesity) funded by the Federal Ministry of Education and Research (German Obesity Biomaterial Bank; FKZ 01GI1128).

- Leipzig-Childhood-IFB This work was supported by grants from Integrated Research and Treatment Centre (IFB) Adiposity Diseases, from the German Research Foundation for the Clinical Research Group “Atherobesity” KFO 152 (KO3512/1 to AK), and by the European Commission (Beta-JUDO) and by EFRE (LIFE Child Obesity). We are grateful to all the patients and families for contributing to the study. We highly appreciate the support of the Obesity Team and Auxo Team of the Leipzig University Children’s Hospital for management of the patients and to the Pediatric Research Center Lab Team for support with DNA banking.
- LOLIPOP (Exome, OmniEE) The LOLIPOP study is supported by the National Institute for Health Research (NIHR) Comprehensive Biomedical Research Centre Imperial College Healthcare NHS Trust, the British Heart Foundation (SP/04/002), the Medical Research Council (G0601966,G0700931), the Wellcome Trust (084723/Z/08/Z), the NIHR (RP-PG-0407-10371), European Union FP7 (EpiMigrant, 279143) and Action on Hearing Loss (G51). We thank the participants and research staff who made the study possible.
- Lothian Birth Cohort 1921 and Lothian Birth Cohort 1936 We thank the cohort participants and team members who contributed to these studies. Phenotype collection in the Lothian Birth Cohort 1921 was supported by the UK’s Biotechnology and Biological Sciences Research Council (BBSRC), The Royal Society and The Chief Scientist Office of the Scottish Government. Phenotype collection in the Lothian Birth Cohort 1936 was supported by Age UK (The Disconnected Mind project). Genotyping was supported by Centre for Cognitive Ageing and Cognitive Epidemiology (Pilot Fund award), Age UK, and the Royal Society of Edinburgh. The work was undertaken by The University of Edinburgh Centre for Cognitive Ageing and Cognitive Epidemiology, part of the cross council Lifelong Health and Wellbeing Initiative (MR/K026992/1). Funding from the BBSRC and Medical Research Council (MRC) is gratefully acknowledged.
- Mark I McCarthy Mark I McCarthy is a Wellcome Trust Senior Investigator (WT098381); and a National Institute of Health Research Senior Investigator.

MESA MESA and the MESA SHARe project are conducted and supported by the National Heart, Lung, and Blood Institute (NHLBI) in collaboration with MESA investigators. Support for MESA is provided by contracts HHSN268201500003I, N01-HC-95159, N01-HC-95160, N01-HC-95161, N01-HC-95162, N01-HC-95163, N01-HC-95164, N01-HC-95165, N01-HC-95166, N01-HC-95167, N01-HC-95168, N01-HC-95169, UL1-TR-000040, UL1-TR-001079, UL1-TR-001420. The provision of genotyping data was supported in part by the National Center for Advancing Translational Sciences, CTSI grant UL1TR001881, and the National Institute of Diabetes and Digestive and Kidney Disease Diabetes Research Center (DRC) grant DK063491 to the Southern California Diabetes Endocrinology Research Center.

METSIM The METSIM study was supported by the Academy of Finland (contract 124243), the Finnish Heart Foundation, the Finnish Diabetes Foundation, Tekes (contract 1510/31/06), and the Commission of the European Community (HEALTH-F2-2007 201681), and the US National Institutes of Health grants DK093757, DK072193, DK062370, and ZIA- HG000024.

MvdB MvdB is supported by a Novo Nordisk postdoctoral fellowship run in partnership with the University of Oxford

Natasha H J Ng Natasha H J Ng (NHJN) is supported by the National Science Scholarship from the Agency for Science, Technology and Research (A*STAR) in Singapore.

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

Nicola L. Beer NLB was a Naomi Berrie Fellow in Diabetes Research. NLB is now an employee of Novo Nordisk, although all experimental work was carried out under employment at the University of Oxford.

NFBC66 and
NFBC86 NFBC1966 and 1986 received financial support from the Academy of Finland (project grants 104781, 120315, 129269, 1114194, 24300796, Center of Excellence in Complex Disease Genetics and SALVE), University Hospital Oulu, Biocenter, University of Oulu, Finland (75617), NIHM (MH063706, Smalley and Jarvelin), Juselius Foundation, NHLBI grant 5R01HL087679-02 through the STAMPEED program (1RL1MH083268-01), NIH/NIMH (5R01MH63706:02), the European Commission (EURO-BLCS, Framework 5 award QLG1-CT-2000-01643), ENGAGE project and grant agreement HEALTH-F4-2007-201413, EU FP7 EurHEALTHAgeing -277849, the Medical Research Council, UK (G0500539, G0600705, G1002319, PrevMetSyn/SALVE) and the MRC, Centenary Early Career Award. The program is currently being funded by the H2020-633595 DynaHEALTH action and academy of Finland EGEA-project (285547). The DNA extractions, sample quality controls, biobank up-keeping and aliquotting was performed in the National Public Health Institute, Biomedicum Helsinki, Finland and supported financially by the Academy of Finland and Biocentrum Helsinki. We thank the late Professor Paula Rantakallio (launch of NFBCs), and Ms. Outi Tornwall and Ms. Minttu Jussila (DNA biobanking). The authors would like to acknowledge the contribution of the late Academician of Science Leena Peltonen.

OBB The Oxford Biobank is supported by the Oxford Biomedical Research Centre and part of the National NIHR Bioresource.

PIVUS & ULSAM PIVUS and ULSAM are supported by the Swedish Research Council, Swedish Heart-Lung Foundation, Swedish Diabetes Foundation and Uppsala University. The investigators express their deepest gratitude to the study participants. Genotyping and analysis was funded by the Wellcome Trust under awards WT064890, WT090532 and WT098017.

PPP-Botnia The Botnia and The PPP-Botnia studies (L.G., T.T.) have been financially supported by grants from Folkhälsan Research Foundation, the Sigrid Juselius Foundation, The Academy of Finland (grants no. 263401, 267882, 312063 to LG, 312072 to TT), Nordic Center of Excellence in Disease Genetics, EU (EXGENESIS, MOSAIC FP7-600914), Ollqvist Foundation, Swedish Cultural Foundation in Finland, Finnish Diabetes Research Foundation, Foundation for Life and Health in Finland, Signe and Ane Gyllenberg Foundation, Finnish Medical Society, Paavo Nurmi Foundation, Helsinki University Central Hospital Research Foundation, Perklén Foundation, Närpes Health Care Foundation and Ahokas Foundation. The study has also been supported by the Ministry of Education in Finland, Municipal Health Care Center and Hospital in Jakobstad and Health Care Centers in Vasa, Närpes and Korsholm. The research leading to these results has received funding from the European Research Council under the European Union's Seventh Framework Programme (FP7/2007-2013) / ERC grant agreement n° 269045. The skillful assistance of the Botnia Study Group is gratefully acknowledged.

PROMIS Dr. Saleheen has received grants from the National Heart, Lung and Blood Institute, Fogarty International, Pfizer, Regeneron, Eli Lilly, and Genentech.

Robert Sladek RS is the recipient of a Chercheur Boursier award from the Fonds de la Recherche en Santé du Québec and a New Investigator Award from the Canadian Institutes of Health Research; and is supported by operating funds from the Canadian Institutes of Health Research.

Rebecca Fine Rebecca Fine is supported by NHGRI F31 HG009850.

RISC The RISC study was supported by European Union grant QL1-CT-2001-01252 and AstraZeneca. The initial genotyping of the RISC samples was funded by Merck & Co Inc.

Rotterdam study The Rotterdam Study is funded by Erasmus Medical Center and Erasmus University, Rotterdam, Netherlands Organization for the Health Research and Development (ZonMw), the Research Institute for Diseases in the Elderly (RIDE), the Ministry of Education, Culture and Science, the Ministry for Health, Welfare and Sports, the European Commission (DG XII), and the Municipality of Rotterdam. The authors are grateful to the study participants, the staff from the Rotterdam Study and the participating general practitioners and pharmacists. The generation and management of the Illumina exome chip v1.0 array data for the Rotterdam Study (RS-I) was executed by the Human Genotyping Facility of the Genetic Laboratory of the Department of Internal Medicine, Erasmus MC, Rotterdam, The Netherlands. The Exome chip array data set was funded by the Genetic Laboratory of the Department of Internal Medicine, Erasmus MC, from the Netherlands Genomics Initiative (NGI)/Netherlands Organisation for Scientific Research (NWO)-sponsored Netherlands Consortium for Healthy Aging (NCHA; project nr. 050-060-810); the Netherlands Organization for Scientific Research (NWO; project number 184021007) and by the Rainbow Project (RP10; Netherlands Exome Chip Project) of the Biobanking and Biomolecular Research Infrastructure Netherlands (BBMRI-NL; www.bbmri.nl (<http://www.bbmri.nl>)). We thank Ms. Mila Jhamai, Ms. Sarah Higgins, and Mr. Marijn Verkerk for their help in creating the exome chip database, and Carolina Medina-Gomez, Lennart Karssen, and Linda Broer for QC and variant calling. We are grateful to the study participants, the staff from the Rotterdam Study and the participating general practitioners and pharmacists. Dr Franco is employed by ErasmusAGE, a center for aging research across the life course funded by Nestlé Nutrition (Nestec Ltd.) and Metagenics.

SardiNIA The SardiNIA investigators thank all the volunteers who generously participated in this study and made this research possible. The study is supported by National Human Genome Research Institute grants HG005581, HG005552, HG006513, HG007022 and HG007089; by National Heart, Lung, and Blood Institute grant HL117626; by the Intramural Research Program of the US National Institutes of Health, National Institute on Aging, contracts N01-AG-1-2109 and HHSN271201100005C; by Sardinian Autonomous Region (L.R. 7/2009) grant cRP3-

The Singapore Chinese Eye Study (SCES)	The Singapore Chinese Eye Study was funded by the Agency for Science, Technology and Research - Biomedical Research Council (A*STAR BMRC) Grant, Singapore [08/1/35/19/550] and Singapore Ministry of Health's National Medical Research Council [NMRC/CIRG/1417/2015]. The authors gratefully acknowledge use of the services and facilities of the Singapore Eye Research Institute and Singapore National Eye Centre. We also acknowledge the contributions of all participants who volunteered and the personnel responsible for the recruitment and administration of the study.
Sorbs	We thank all those who participated in the study. We would like to thank Knut Krohn (Microarray Core Facility, University of Leipzig, Institute of Pharmacology) for the genotyping support and Joachim Thiery (Institute of Laboratory Medicine, Clinical Chemistry and Molecular Diagnostics, University of Leipzig) for clinical chemistry services. This work was supported by grants from the German Research Council (DFG - SFB 1052 "Obesity mechanisms"; A01, C01, B03 and SPP 1629 TO 718/2-1), from the German Diabetes Association and from the DHFD (Diabetes Hilfs- und Forschungsfonds Deutschland).
Timothy D Spector TwinsUK	Timothy D Spector is holder of an ERC Advanced Principal Investigator award. TwinsUK study was funded by the Wellcome Trust; European Community's Seventh Framework Programme (FP7/2007-2013). The study also receives support from the National Institute for Health Research (NIHR) BioResource Clinical Research Facility and Biomedical Research Centre based at Guy's and St Thomas' NHS Foundation Trust and King's College London.
UKHLS	These data are from Understanding Society: The UK Household Longitudinal Study, which is led by the Institute for Social and Economic Research at the University of Essex and funded by the Economic and Social Research Council. The data were collected by NatCen and the genome wide scan data were analysed by the Wellcome Trust Sanger Institute. The Understanding Society DAC have an application system for genetics data and all use of the data should be approved by them. The application form is at: https://www.understandingsociety.ac.uk/about/health/data .
Vejle Biobank	The Vejle Diabetes Biobank was supported by The Danish Research Council for Independent Research.

WGHS The WGHS is funded by the National Heart, Lung, and Blood Institute (HL043851 and HL080467) and National Cancer Institute (CA047988 and UM1CA182913).
Funding for genotyping on the Exome Array was funded by Amgen.

1018

1019 **Disclosures**

1020

1021 Soren K. Thomsen: SKT is now an employee of Vertex Pharmaceuticals, although all experimental work
1022 was carried out under employment at the University of Oxford.

1023 Martijn van de Bunt: Currently employed by Novo Nordisk.

1024 Audrey Y Chu: Currently employed by Merck.

1025 Dennis O. Mook-Kanamori: Dennis Mook-Kanamori is working as a part-time clinical research consultant
1026 for Metabolon, Inc.

1027 Paul W. Franks: PWF has been a paid consultant for Eli Lilly and Sanofi Aventis and has received research
1028 support from several pharmaceutical companies as part of a European Union Innovative Medicines
1029 Initiative (IMI) project.

1030 Mike A. Nalls: Dr. Mike A. Nalls is supported by a consulting contract between Data Tecnica International
1031 LLC and the National Institute on Aging (NIA), National Institutes of Health (NIH), Bethesda, MD, USA. Dr.
1032 Nalls also consults for Illumina Inc., the Michael J. Fox Foundation, and the University of California
1033 Healthcare.

1034 Mark J. Caulfield: MJC is Chief Scientist for Genomics England, a UK government company.

1035 Joel N. Hirschhorn: JHN is on the scientific advisory board of Camp4 Therapeutics.

1036 Erik Ingelsson: Erik Ingelsson is an advisor and consultant for Precision Wellness, Inc., and advisor for
1037 Cellink.

1038 Inês Barroso: IB and spouse declare stock ownership in GlaxoSmithkline and Incyte Ltd.

1039 **Supplementary Figure Legends**

1040

1041 **Figure S1. GeneMANIA network analysis identifies relevant pathways regulating glycemia.** The

1042 networks represent composite networks for (A) FI and (B) 2hGlu, from the GeneMANIA analysis using

1043 genes with variant associations at $P < 1 \times 10^{-5}$ for each trait as input. Nodes outlined in red correspond to

1044 genes from the input list. Other nodes correspond to related genes based on 50 default databases.

1045 Based on the network, GO terms and Reactome pathways that were significantly enriched are depicted.

1046 To summarize these results, the most significant term of all calculated terms within the same group

1047 (using the Kappa method, see **Methods**) was represented. Each group was assigned a specific color; if a

1048 gene is present in more than one term, it will be displayed in more than one color. Barplots with the

1049 Bonferroni-adjusted $-\log_{10}(p\text{-values})$ of the most significant terms within each group are shown.

1050 Each group was assigned a specific color; if a gene is present in more than one term, it is displayed in

1051 more than one color. Details of the networks are summarized in (C). Related to Figure 2 and Table S7.

1052

1053 **Figure S2. Pathway analysis identifies relevant gene sets regulating glycemia.** EC-DEPICT analysis with

1054 heatmap visualization (UK Biobank permutations) is shown for a. all traits combined; b. HbA1c only; c. all

1055 traits except HbA1c combined; d. FG only; e. 2hGlu only. The columns represent the input genes for the

1056 analysis. We used affinity propagation clustering to define a representative “meta-gene set” for groups

1057 of highly correlated gene sets (see **Methods**); the rows in the heat map represent significant meta-gene

1058 sets (FDR < 0.05). The color of each square indicates DEPICT’s z-score for membership of that gene in

1059 that gene set, where dark red means “very likely a member” and dark blue means “very unlikely a

1060 member”. The gene set annotations indicate whether that meta-gene set was significant at FDR < 0.05 or

1061 not significant (n.s.) for each of the other EC-DEPICT analyses using the UK Biobank permutations (all

1062 traits together, HbA1c only, FG only, 2hGlu only, and all-except-HbA1c). For heatmap intensity and EC-

1063 DEPICT P -values, the meta-gene set values are taken from the most significantly enriched member gene

1064 set. The gene variant annotations are as follows: (1) the European minor allele frequency (MAF) of the

1065 input variant, where rare is MAF < 1%, low-frequency is MAF 1-5%, and common is MAF > 5%, (2)

1066 whether the gene has an Online Mendelian Inheritance in Man (OMIM) annotation as causal for a

1067 diabetes/glycemic-relevant syndrome or blood disorder, (3) the effector transcript classification for that

1068 variant: gold, silver, bronze, or NA (note that only array-wide significant variants were classified, so

1069 suggestively-significant variants are by default classified as “NA”), (4-7) whether each variant was

1070 significant ($P < 2 \times 10^{-7}$), suggestively significant ($P < 10^{-5}$), or not significant in Europeans for each of the
1071 four traits, and (8) whether each variant was classified in the analysis (with UK Biobank permutations) or
1072 excluded by filters (see **Methods**). AWS: array-wide significant. Related to Figure 2 and Table S8.

1073

1074 **Figure S3. Pathway analysis identifies relevant gene sets regulating glycemia.** EC-DEPICT analysis with
1075 heatmap visualization (Swedish permutations) is shown for a. all traits combined; b. HbA1c only; c. all
1076 traits except HbA1c combined; d. FG only. (With these permutations, there was no significance for 2hGlu
1077 only). We used affinity propagation clustering to define a representative “meta-gene set” for groups of
1078 highly-correlated gene sets (see **Methods**); the rows in the heat map represent significant meta-gene
1079 sets (FDR < 0.05). The color of each square indicates DEPICT’s z-score for membership of that gene in
1080 that gene set, where dark red means “very likely a member” and dark blue means “very unlikely a
1081 member”. The gene set annotations indicate whether that meta-gene set was significant at FDR < 0.05 or
1082 not significant (n.s.) for each of the other EC-DEPICT analyses using the Swedish permutations (all traits
1083 together, HbA1c only, FG only, and all-except-HbA1c). For heatmap intensity and EC-DEPICT *P*-values,
1084 the meta-gene set values are taken from the most significantly enriched member gene set. The gene
1085 variant annotations are as follows: (1) the European minor allele frequency (MAF) of the input variant,
1086 where rare is MAF < 1%, low-frequency is MAF 1-5%, and common is MAF > 5%, (2) whether the gene has
1087 an Online Mendelian Inheritance in Man (OMIM) annotation as causal for a diabetes/glycemic-relevant
1088 syndrome or blood disorder, (3) the effector transcript classification for that variant: gold, silver, bronze,
1089 or NA (note that only array-wide significant variants were classified, so suggestively-significant variants
1090 are by default classified as “NA”), (4-7) whether each variant was significant ($P < 2 \times 10^{-7}$), suggestively
1091 significant ($P < 10^{-5}$), or not significant in Europeans for each of the four traits, and (8) whether each
1092 variant was included in the analysis (with Swedish permutations) or excluded by filters (see
1093 **Methods**). AWS: array-wide significant. Related to Figure 2 and Table S8.

1094

1095 **Figure S4. Functional characterisation of G6PC variants.** Related to Figure 4.

1096 (A) Cellular localisation of Q347X was assessed in HEK293 cells and overlaid with a marker for the ER,
1097 calreticulin, (left) or the trans-golgi network, TGN46 (right). White arrows point to positions of the golgi
1098 apparatus. Scale bar indicates 10µm. (B) Glucose-6-phosphatase activity of G6PC-R83C (n=3), with
1099 representative western blot of microsomal protein isolated from HEK293 shown. (C) Glucose-6-

1100 phosphatase activity of G6PC-Q347X (n=2), with representative western blot of microsomal protein
1101 isolated from HEK293 shown. (D) Protein expression levels of G6PC-A204S in microsomal protein
1102 extracted from HEK293 cells was found to be downregulated by 41% compared to WT based on
1103 densitometric analysis (n=4), with representative western blot shown. Data presented as mean \pm SEM
1104 and analysed using paired Students' t test. * p=0.01. Unt: Untransfected; WT: Wild type.

1105

1106 **Figure S5. Functional characterisation of G6PC2 variants and the effect of G6PC2 knockdown on insulin**
1107 **content and secretion in EndoC- β H1 cells.** Related to Figure 5.

1108 (A) Variants prioritised for functional study in the context of the predicted G6PC2 protein structure
1109 (RefSeq NP_066999.1) in the ER membrane. Amino acid residues are coloured as described in the
1110 legend. Variants selected for functional study, in green, are labelled. The N-terminal V5 and C-terminal
1111 Myc-FLAG tags present in the expression constructs are indicated. (B) Quantification of total G6PC2
1112 variant protein expression (both upper and lower bands of representative western blot in Figure 5) in
1113 INS-1 832/13 cells based on western blot densitometric analysis of Myc-tagged G6PC2 constructs
1114 relative to tubulin control (n=5). (C) Expression levels of G6PC2 variant proteins in HEK293 by western
1115 blot densitometric analysis of FLAG-tagged G6PC2 constructs or V5-tagged G6PC2-R283X relative to
1116 tubulin control (n=4). Representative blots are shown for untreated cells and cells treated with
1117 proteasomal inhibitor MG-132 or lysosomal inhibitor chloroquine. (D) Glucose-6-phosphatase activity of
1118 the R281X variant in G6PC (proxy for R283X in G6PC2) in HEK293 (n=2), with representative western blot
1119 of microsomal protein shown. (E) Total insulin secretion and insulin content were assessed at basal and
1120 high glucose conditions (with and without drug treatment) following 96-120h G6PC2 knockdown in
1121 EndoC- β H1. Unpaired two-tailed Students' t tests were used to compare G6PC2 knockdown to control
1122 for each condition, from n=16 across 4 independent experiments. Tol: tolbutamide; Diaz: diazoxide. All
1123 data presented as mean \pm SEM. * p=0.01-0.05; ** p=0.001-0.01; *** p<0.001.

1124

1125 **Figure S6. G6PC2 expression in RNA-Seq data from 150 human islet donor samples.** (A) Allelic balance
1126 was observed for G6PC2 rs146779637 (p.R283X) in two heterozygote human islet samples. (B) The
1127 glucose-raising rs560887-G allele associates significantly (q-value<0.01) with increased expression of the
1128 long G6PC2 isoform (purple) and reduced expression of the short G6PC2 isoform lacking exon 4 (brown).

1129

1130 **Supplementary Table Legends**

1131

1132 **Table S1. Cohort characteristics, genotyping and quality control (QC), glucose, insulin, 2hGlu and**
1133 **HbA1c analyses and covariates.**

1134

1135 **Table S2. Association of identified lead coding variants with T2D and anthropometric traits (height,**
1136 **BMI and WHR) from publicly-available association results.** Alleles E/O: effect allele/other allele; EAF:
1137 effect allele frequency; Neff: Number of samples in the analysis; BETA: effect size; SE: standard error.
1138 Related to Table 1 and Table S3.

1139

1140 **Table S3. Coding variant associations in known glycemc trait loci with conditional results on**
1141 **established signals where available.** Related to Table 1.

1142

1143 **Table S4. Full gene-based results including all variants included in the masks, for both novel and**
1144 **previously-established genes.** Related to Table S9.

1145

1146 **Table S5. HbA1c-associated loci lookup results for blood cell traits.** Related to Table 1.

1147

1148 **Table S6. Annotation and classification of effector transcripts into “gold”, “silver” and “bronze”**
1149 **categories.** Related to Tables 1 and 2 and Figure 1.

1150

1151 **Table S7. Gene Set Enrichment Analysis by GeneMANIA network analysis showing enriched GO terms**
1152 **and Reactome pathways in the network for (A) HbA1c; (B) FG; (C) FI; (D) 2hGlu.** GOID: Gene Ontology
1153 ID; GOTerm: Gene Ontology Term. Gene Set Enrichment (GSE) of networks was performed with ClueGO
1154 using GO terms and REACTOME gene sets. The enrichment results were considered significant when
1155 Bonferroni-adjusted p-value < 0.05 and at least 3% of the genes contained in the tested gene set is
1156 included in the network. Gene sets were also grouped using kappa score into functional groups to
1157 improve visualization of enriched pathways. Related to Figures 2 and S1.

1158

1159 **Table S8. (A-E) EC-DEPICT results (UK Biobank permutations) for (A) all traits combined; (B) all traits**
1160 **except HbA1c combined; (C) HbA1c only; (D) FG only and (E) 2hGlu only. (F-I) EC-DEPICT results**
1161 **(Swedish permutations) for (F) all traits combined; (G) all traits except HbA1c combined; (H) HbA1c**
1162 **only and (I) FG only. Related to Figures 2, S2 and S3.**

1163

1164 **Table S9. Full *G6PC2* gene-based results and conditional analyses for FG and HbA1c. Related to Tables**
1165 **2 and S4.**

1166 References

1167

- 1168 Abraham, A., Pedregosa, F., Eickenberg, M., Gervais, P., Mueller, A., Kossaifi, J., Gramfort, A., Thirion, B.,
1169 and Varoquaux, G. (2014). Machine learning for neuroimaging with scikit-learn. *Front Neuroinform* 8,
1170 14.
- 1171 Almgren, P., Lindqvist, A., Krus, U., Hakaste, L., Ottosson-Laakso, E., Asplund, O., Sonestedt, E., Prasad,
1172 R.B., Laurila, E., Orho-Melander, M., *et al.* (2017). Genetic determinants of circulating GIP and GLP-1
1173 concentrations. *JCI Insight* 2.
- 1174 Astle, W.J., Elding, H., Jiang, T., Allen, D., Ruklisa, D., Mann, A.L., Mead, D., Bouman, H., Riveros-Mckay,
1175 F., Kostadima, M.A., *et al.* (2016). The Allelic Landscape of Human Blood Cell Trait Variation and Links to
1176 Common Complex Disease. *Cell* 167, 1415-1429 e1419.
- 1177 Baerenwald, D.A., Bonnefond, A., Bouatia-Naji, N., Flemming, B.P., Umunakwe, O.C., Oeser, J.K., Pound,
1178 L.D., Conley, N.L., Cauchi, S., Lobbens, S., *et al.* (2013). Multiple functional polymorphisms in the G6PC2
1179 gene contribute to the association with higher fasting plasma glucose levels. *Diabetologia* 56, 1306-
1180 1316.
- 1181 Barrett, J.C., Clayton, D.G., Concannon, P., Akolkar, B., Cooper, J.D., Erlich, H.A., Julier, C., Morahan, G.,
1182 Nerup, J., Nierras, C., *et al.* (2009). Genome-wide association study and meta-analysis find that over 40
1183 loci affect risk of type 1 diabetes. *Nat Genet* 41, 703-707.
- 1184 Battle, A., Brown, C.D., Engelhardt, B.E., and Montgomery, S.B. (2017). Genetic effects on gene
1185 expression across human tissues. *Nature* 550, 204-213.
- 1186 Bindea, G., Mlecnik, B., Hackl, H., Charoentong, P., Tosolini, M., Kirilovsky, A., Fridman, W.H., Pages, F.,
1187 Trajanoski, Z., and Galon, J. (2009). ClueGO: a Cytoscape plug-in to decipher functionally grouped gene
1188 ontology and pathway annotation networks. *Bioinformatics* 25, 1091-1093.
- 1189 Bouatia-Naji, N., Bonnefond, A., Baerenwald, D.A., Marchand, M., Bugliani, M., Marchetti, P., Pattou, F.,
1190 Printz, R.L., Flemming, B.P., Umunakwe, O.C., *et al.* (2010). Genetic and functional assessment of the
1191 role of the rs13431652-A and rs573225-A alleles in the G6PC2 promoter that are strongly associated
1192 with elevated fasting glucose levels. *Diabetes* 59, 2662-2671.
- 1193 Bouatia-Naji, N., Rocheleau, G., Van Lommel, L., Lemaire, K., Schuit, F., Cavalcanti-Proenca, C.,
1194 Marchand, M., Hartikainen, A.L., Sovio, U., De Graeve, F., *et al.* (2008). A polymorphism within the
1195 G6PC2 gene is associated with fasting plasma glucose levels. *Science* 320, 1085-1088.
- 1196 Carrat, G.R., Hu, M., Nguyen-Tu, M.S., Chabosseau, P., Gaulton, K.J., van de Bunt, M., Siddiq, A., Falchi,
1197 M., Thurner, M., Canouil, M., *et al.* (2017). Decreased STARD10 Expression Is Associated with Defective
1198 Insulin Secretion in Humans and Mice. *Am J Hum Genet* 100, 238-256.
- 1199 Chen, W.M., Erdos, M.R., Jackson, A.U., Saxena, R., Sanna, S., Silver, K.D., Timpson, N.J., Hansen, T., Orru,
1200 M., Grazia Piras, M., *et al.* (2008). Variations in the G6PC2/ABCB11 genomic region are associated with
1201 fasting glucose levels. *J Clin Invest* 118, 2620-2628.
- 1202 Chou, J.Y., and Mansfield, B.C. (2008). Mutations in the glucose-6-phosphatase-alpha (G6PC) gene that
1203 cause type Ia glycogen storage disease. *Hum Mutat* 29, 921-930.
- 1204 Cohen, R.M., Franco, R.S., Khera, P.K., Smith, E.P., Lindsell, C.J., Ciralo, P.J., Palascak, M.B., and Joiner,
1205 C.H. (2008). Red cell life span heterogeneity in hematologically normal people is sufficient to alter
1206 HbA1c. *Blood* 112, 4284-4291.
- 1207 Croft, D., Mundo, A.F., Haw, R., Milacic, M., Weiser, J., Wu, G., Caudy, M., Garapati, P., Gillespie, M.,
1208 Kamdar, M.R., *et al.* (2014). The Reactome pathway knowledgebase. *Nucleic Acids Res* 42, D472-477.
- 1209 Dewey, F.E., Murray, M.F., Overton, J.D., Habegger, L., Leader, J.B., Fetterolf, S.N., O'Dushlaine, C., Van
1210 Hout, C.V., Staples, J., Gonzaga-Jauregui, C., *et al.* (2016). Distribution and clinical impact of functional
1211 variants in 50,726 whole-exome sequences from the DiscovEHR study. *Science* 354.

- 1212 Esteghamat, F., Broughton, J.S., Smith, E., Cardone, R., Tyagi, T., Guerra, M., Szabo, A., Ugwu, N., Mani,
1213 M.V., Azari, B., *et al.* (2019). CELA2A mutations predispose to early-onset atherosclerosis and metabolic
1214 syndrome and affect plasma insulin and platelet activation. *Nat Genet* 51, 1233-1243.
- 1215 Fehrmann, R.S., Karjalainen, J.M., Krajewska, M., Westra, H.J., Maloney, D., Simeonov, A., Pers, T.H.,
1216 Hirschhorn, J.N., Jansen, R.C., Schultes, E.A., *et al.* (2015). Gene expression analysis identifies global gene
1217 dosage sensitivity in cancer. *Nat Genet* 47, 115-125.
- 1218 Flannick, J., Thorleifsson, G., Beer, N.L., Jacobs, S.B., Grarup, N., Burt, N.P., Mahajan, A., Fuchsberger, C.,
1219 Atzmon, G., Benediktsson, R., *et al.* (2014). Loss-of-function mutations in SLC30A8 protect against type 2
1220 diabetes. *Nat Genet* 46, 357-363.
- 1221 Foster, J.D., Pederson, B.A., and Nordlie, R.C. (1997). Glucose-6-phosphatase structure, regulation, and
1222 function: an update. *Proc Soc Exp Biol Med* 215, 314-332.
- 1223 Franz, M., Rodriguez, H., Lopes, C., Zuberi, K., Montojo, J., Bader, G.D., and Morris, Q. (2018).
1224 GeneMANIA update 2018. *Nucleic acids research* 46, W60-W64.
- 1225 Frey, B.J., and Dueck, D. (2007). Clustering by passing messages between data points. *Science* 315, 972-
1226 976.
- 1227 Froguel, P., Vaxillaire, M., Sun, F., Velho, G., Zouali, H., Butel, M.O., Lesage, S., Vionnet, N., Clement, K.,
1228 Fougerousse, F., *et al.* (1992). Close linkage of glucokinase locus on chromosome 7p to early-onset non-
1229 insulin-dependent diabetes mellitus. *Nature* 356, 162-164.
- 1230 Gloyn, A.L., Pearson, E.R., Antcliff, J.F., Proks, P., Bruining, G.J., Slingerland, A.S., Howard, N., Srinivasan,
1231 S., Silva, J.M., Molnes, J., *et al.* (2004). Activating mutations in the gene encoding the ATP-sensitive
1232 potassium-channel subunit Kir6.2 and permanent neonatal diabetes. *N Engl J Med* 350, 1838-1849.
- 1233 Goodarzi, M.O., Cui, J., Chen, Y.D., Hsueh, W.A., Guo, X., and Rotter, J.I. (2011). Fasting insulin reflects
1234 heterogeneous physiological processes: role of insulin clearance. *Am J Physiol Endocrinol Metab* 301,
1235 E402-408.
- 1236 Grove, M.L., Yu, B., Cochran, B.J., Haritunians, T., Bis, J.C., Taylor, K.D., Hansen, M., Borecki, I.B., Cupples,
1237 L.A., Fornage, M., *et al.* (2013). Best practices and joint calling of the HumanExome BeadChip: the
1238 CHARGE Consortium. *PLoS One* 8, e68095.
- 1239 Gu, Z., Eils, R., and Schlesner, M. (2016). Complex heatmaps reveal patterns and correlations in
1240 multidimensional genomic data. *Bioinformatics* 32, 2847-2849.
- 1241 Hanson, D., Murray, P.G., Sud, A., Temtamy, S.A., Aglan, M., Superti-Furga, A., Holder, S.E., Urquhart, J.,
1242 Hilton, E., Manson, F.D., *et al.* (2009). The primordial growth disorder 3-M syndrome connects
1243 ubiquitination to the cytoskeletal adaptor OBSL1. *Am J Hum Genet* 84, 801-806.
- 1244 Hara, K., Fujita, H., Johnson, T.A., Yamauchi, T., Yasuda, K., Horikoshi, M., Peng, C., Hu, C., Ma, R.C.,
1245 Imamura, M., *et al.* (2014). Genome-wide association study identifies three novel loci for type 2
1246 diabetes. *Hum Mol Genet* 23, 239-246.
- 1247 Hart, N.J., Aramandla, R., Poffenberger, G., Fayolle, C., Thames, A.H., Bautista, A., Spigelman, A.F.,
1248 Babon, J.A.B., DeNicola, M.E., Dadi, P.K., *et al.* (2018). Cystic fibrosis-related diabetes is caused by islet
1249 loss and inflammation. *JCI Insight* 3.
- 1250 Jia, B., Madsen, L., Petersen, R.K., Techer, N., Kopperud, R., Ma, T., Doskeland, S.O., Ailhaud, G., Wang,
1251 J., Amri, E.Z., *et al.* (2012). Activation of protein kinase A and exchange protein directly activated by
1252 cAMP promotes adipocyte differentiation of human mesenchymal stem cells. *PLoS One* 7, e34114.
- 1253 Kycia, I., Wolford, B.N., Huyghe, J.R., Fuchsberger, C., Vadlamudi, S., Kursawe, R., Welch, R.P., Albanus,
1254 R.D., Uyar, A., Khetan, S., *et al.* (2018). A Common Type 2 Diabetes Risk Variant Potentiates Activity of an
1255 Evolutionarily Conserved Islet Stretch Enhancer and Increases C2CD4A and C2CD4B Expression. *Am J*
1256 *Hum Genet* 102, 620-635.
- 1257 Lei, K.J., Chen, Y.T., Chen, H., Wong, L.J., Liu, J.L., McConkie-Rosell, A., Van Hove, J.L., Ou, H.C., Yeh, N.J.,
1258 Pan, L.Y., *et al.* (1995). Genetic basis of glycogen storage disease type 1a: prevalent mutations at the
1259 glucose-6-phosphatase locus. *Am J Hum Genet* 57, 766-771.

1260 Lesage, S., Drouet, V., Majounie, E., Deramecourt, V., Jacoupy, M., Nicolas, A., Cormier-Dequaire, F.,
1261 Hassoun, S.M., Pujol, C., Ciura, S., *et al.* (2016). Loss of VPS13C Function in Autosomal-Recessive
1262 Parkinsonism Causes Mitochondrial Dysfunction and Increases PINK1/Parkin-Dependent Mitophagy. *Am*
1263 *J Hum Genet* *98*, 500-513.

1264 Liu, S., Liu, Y., Zhang, Q., Wu, J., Liang, J., Yu, S., Wei, G.H., White, K.P., and Wang, X. (2017). Systematic
1265 identification of regulatory variants associated with cancer risk. *Genome Biol* *18*, 194.

1266 Liu, X., Jian, X., and Boerwinkle, E. (2013). dbNSFP v2.0: a database of human non-synonymous SNVs and
1267 their functional predictions and annotations. *Hum Mutat* *34*, E2393-2402.

1268 Mahajan, A., Sim, X., Ng, H.J., Manning, A., Rivas, M.A., Highland, H.M., Locke, A.E., Grarup, N., Im, H.K.,
1269 Cingolani, P., *et al.* (2015). Identification and functional characterization of G6PC2 coding variants
1270 influencing glycaemic traits define an effector transcript at the G6PC2-ABCB11 locus. *PLoS Genet* *11*,
1271 e1004876.

1272 Mahajan, A., Taliun, D., Thurner, M., Robertson, N.R., Torres, J.M., Rayner, N.W., Payne, A.J.,
1273 Steinthorsdottir, V., Scott, R.A., Grarup, N., *et al.* (2018a). Fine-mapping type 2 diabetes loci to single-
1274 variant resolution using high-density imputation and islet-specific epigenome maps. *Nat Genet* *50*, 1505-
1275 1513.

1276 Mahajan, A., Wessel, J., Willems, S.M., Zhao, W., Robertson, N.R., Chu, A.Y., Gan, W., Kitajima, H., Taliun,
1277 D., Rayner, N.W., *et al.* (2018b). Refining the accuracy of validated target identification through coding
1278 variant fine-mapping in type 2 diabetes. *Nat Genet* *50*, 559-571.

1279 Manning, A.K., Hivert, M.F., Scott, R.A., Grimsby, J.L., Bouatia-Naji, N., Chen, H., Rybin, D., Liu, C.T.,
1280 Bielak, L.F., Prokopenko, I., *et al.* (2012). A genome-wide approach accounting for body mass index
1281 identifies genetic variants influencing fasting glycaemic traits and insulin resistance. *Nat Genet* *44*, 659-
1282 669.

1283 Marouli, E., Graff, M., Medina-Gomez, C., Lo, K.S., Wood, A.R., Kjaer, T.R., Fine, R.S., Lu, Y., Schurmann,
1284 C., Highland, H.M., *et al.* (2017). Rare and low-frequency coding variants alter human adult height.
1285 *Nature* *542*, 186-190.

1286 Mehta, Z.B., Fine, N., Pullen, T.J., Cane, M.C., Hu, M., Chabosseau, P., Meur, G., Velayos-Baeza, A.,
1287 Monaco, A.P., Marselli, L., *et al.* (2016). Changes in the expression of the type 2 diabetes-associated
1288 gene VPS13C in the beta-cell are associated with glucose intolerance in humans and mice. *Am J Physiol*
1289 *Endocrinol Metab* *311*, E488-507.

1290 Mithieux, G. (1997). New knowledge regarding glucose-6 phosphatase gene and protein and their roles
1291 in the regulation of glucose metabolism. *Eur J Endocrinol* *136*, 137-145.

1292 Montague, C.T., Farooqi, I.S., Whitehead, J.P., Soos, M.A., Rau, H., Wareham, N.J., Sewter, C.P., Digby,
1293 J.E., Mohammed, S.N., Hurst, J.A., *et al.* (1997). Congenital leptin deficiency is associated with severe
1294 early-onset obesity in humans. *Nature* *387*, 903-908.

1295 Morris, A.P., Voight, B.F., Teslovich, T.M., Ferreira, T., Segre, A.V., Steinthorsdottir, V., Strawbridge, R.J.,
1296 Khan, H., Grallert, H., Mahajan, A., *et al.* (2012). Large-scale association analysis provides insights into
1297 the genetic architecture and pathophysiology of type 2 diabetes. *Nat Genet* *44*, 981-990.

1298 O'Hare, E.A., Yerges-Armstrong, L.M., Perry, J.A., Shuldiner, A.R., and Zaghoul, N.A. (2016). Assignment
1299 of Functional Relevance to Genes at Type 2 Diabetes-Associated Loci Through Investigation of beta-Cell
1300 Mass Deficits. *Mol Endocrinol* *30*, 429-445.

1301 Ongen, H., Buil, A., Brown, A.A., Dermitzakis, E.T., and Delaneau, O. (2016). Fast and efficient QTL
1302 mapper for thousands of molecular phenotypes. *Bioinformatics* *32*, 1479-1485.

1303 Patro, R., Duggal, G., Love, M.I., Irizarry, R.A., and Kingsford, C. (2017). Salmon provides fast and bias-
1304 aware quantification of transcript expression. *Nat Methods* *14*, 417-419.

1305 Pers, T.H., Karjalainen, J.M., Chan, Y., Westra, H.J., Wood, A.R., Yang, J., Lui, J.C., Vedantam, S.,
1306 Gustafsson, S., Esko, T., *et al.* (2015). Biological interpretation of genome-wide association studies using
1307 predicted gene functions. *Nat Commun* *6*, 5890.

1308 Pirinen, M., Lappalainen, T., Zaitlen, N.A., Dermitzakis, E.T., Donnelly, P., McCarthy, M.I., and Rivas, M.A.
1309 (2015). Assessing allele-specific expression across multiple tissues from RNA-seq read data.
1310 *Bioinformatics* 31, 2497-2504.

1311 Rosendahl, J., Kirsten, H., Hegyi, E., Kovacs, P., Weiss, F.U., Laumen, H., Lichtner, P., Ruffert, C., Chen,
1312 J.M., Masson, E., *et al.* (2017). Genome-wide association study identifies inversion in the CTRB1-CTRB2
1313 locus to modify risk for alcoholic and non-alcoholic chronic pancreatitis. *Gut*.

1314 Scarl, R.T., Lawrence, C.M., Gordon, H.M., and Nunemaker, C.S. (2017). STEAP4: its emerging role in
1315 metabolism and homeostasis of cellular iron and copper. *J Endocrinol* 234, R123-R134.

1316 Scott, R.A., Lagou, V., Welch, R.P., Wheeler, E., Montasser, M.E., Luan, J., Magi, R., Strawbridge, R.J.,
1317 Rehnberg, E., Gustafsson, S., *et al.* (2012). Large-scale association analyses identify new loci influencing
1318 glycemic traits and provide insight into the underlying biological pathways. *Nat Genet* 44, 991-1005.

1319 Sever, S., Weinstein, D.A., Wolfsdorf, J.I., Gedik, R., and Schaefer, E.J. (2012). Glycogen storage disease
1320 type 1a: linkage of glucose, glycogen, lactic acid, triglyceride, and uric acid metabolism. *J Clin Lipidol* 6,
1321 596-600.

1322 Shieh, J.J., Terzioglu, M., Hiraiwa, H., Marsh, J., Pan, C.J., Chen, L.Y., and Chou, J.Y. (2002). The molecular
1323 basis of glycogen storage disease type 1a: structure and function analysis of mutations in glucose-6-
1324 phosphatase. *J Biol Chem* 277, 5047-5053.

1325 Sikkeland, J., Sheng, X., Jin, Y., and Saatcioglu, F. (2016). STAMPing at the crossroads of normal
1326 physiology and disease states. *Mol Cell Endocrinol* 425, 26-36.

1327 Soranzo, N., Sanna, S., Wheeler, E., Gieger, C., Radke, D., Dupuis, J., Bouatia-Naji, N., Langenberg, C.,
1328 Prokopenko, I., Stolerman, E., *et al.* (2010). Common variants at 10 genomic loci influence hemoglobin
1329 A(1)(C) levels via glycemic and nonglycemic pathways. *Diabetes* 59, 3229-3239.

1330 Steinthorsdottir, V., Thorleifsson, G., Sulem, P., Helgason, H., Grarup, N., Sigurdsson, A., Helgadottir,
1331 H.T., Johannsdottir, H., Magnusson, O.T., Gudjonsson, S.A., *et al.* (2014). Identification of low-frequency
1332 and rare sequence variants associated with elevated or reduced risk of type 2 diabetes. *Nat Genet* 46,
1333 294-298.

1334 Sveinbjornsson, G., Albrechtsen, A., Zink, F., Gudjonsson, S.A., Oddson, A., Masson, G., Holm, H., Kong,
1335 A., Thorsteinsdottir, U., Sulem, P., *et al.* (2016). Weighting sequence variants based on their annotation
1336 increases power of whole-genome association studies. *Nat Genet* 48, 314-317.

1337 t Hart, L.M., Fritsche, A., Nijpels, G., van Leeuwen, N., Donnelly, L.A., Dekker, J.M., Alsema, M., Fadista,
1338 J., Carlotti, F., Gjesing, A.P., *et al.* (2013). The CTRB1/2 locus affects diabetes susceptibility and
1339 treatment via the incretin pathway. *Diabetes* 62, 3275-3281.

1340 Tewhey, R., Kotliar, D., Park, D.S., Liu, B., Winnicki, S., Reilly, S.K., Andersen, K.G., Mikkelsen, T.S., Lander,
1341 E.S., Schaffner, S.F., *et al.* (2016). Direct Identification of Hundreds of Expression-Modulating Variants
1342 using a Multiplexed Reporter Assay. *Cell* 165, 1519-1529.

1343 Thomsen, S.K., Raimondo, A., Hastoy, B., Sengupta, S., Dai, X.Q., Bautista, A., Censin, J., Payne, A.J.,
1344 Umapathysivam, M.M., Spigelman, A.F., *et al.* (2018). Type 2 diabetes risk alleles in PAM impact insulin
1345 release from human pancreatic beta-cells. *Nat Genet* 50, 1122-1131.

1346 Turcot, V., Lu, Y., Highland, H.M., Schurmann, C., Justice, A.E., Fine, R.S., Bradfield, J.P., Esko, T., Giri, A.,
1347 Graff, M., *et al.* (2018). Protein-altering variants associated with body mass index implicate pathways
1348 that control energy intake and expenditure in obesity. *Nat Genet* 50, 26-41.

1349 Ulirsch, J.C., Nandakumar, S.K., Wang, L., Giani, F.C., Zhang, X., Rogov, P., Melnikov, A., McDonel, P., Do,
1350 R., Mikkelsen, T.S., *et al.* (2016). Systematic Functional Dissection of Common Genetic Variation
1351 Affecting Red Blood Cell Traits. *Cell* 165, 1530-1545.

1352 van de Bunt, M., Manning Fox, J.E., Dai, X., Barrett, A., Grey, C., Li, L., Bennett, A.J., Johnson, P.R.,
1353 Rajotte, R.V., Gaulton, K.J., *et al.* (2015). Transcript Expression Data from Human Islets Links Regulatory
1354 Signals from Genome-Wide Association Studies for Type 2 Diabetes and Glycemic Traits to Their
1355 Downstream Effectors. *PLoS Genet* 11, e1005694.

1356 Wessel, J., Chu, A.Y., Willems, S.M., Wang, S., Yaghoobkar, H., Brody, J.A., Dauriz, M., Hivert, M.F.,
1357 Raghavan, S., Lipovich, L., *et al.* (2015). Low-frequency and rare exome chip variants associate with
1358 fasting glucose and type 2 diabetes susceptibility. *Nat Commun* **6**, 5897.
1359 Wheeler, E., Leong, A., Liu, C.T., Hivert, M.F., Strawbridge, R.J., Podmore, C., Li, M., Yao, J., Sim, X., Hong,
1360 J., *et al.* (2017a). Impact of common genetic determinants of Hemoglobin A1c on type 2 diabetes risk
1361 and diagnosis in ancestrally diverse populations: A transethnic genome-wide meta-analysis. *PLoS Med*
1362 **14**, e1002383.
1363 Wheeler, E., Marenne, G., and Barroso, I. (2017b). Genetic aetiology of glycaemic traits: approaches and
1364 insights. *Hum Mol Genet* **26**, R172-R184.
1365 Williams, A.L., Jacobs, S.B., Moreno-Macias, H., Huerta-Chagoya, A., Churchhouse, C., Marquez-Luna, C.,
1366 Garcia-Ortiz, H., Gomez-Vazquez, M.J., Burt, N.P., Aguilar-Salinas, C.A., *et al.* (2014). Sequence variants
1367 in SLC16A11 are a common risk factor for type 2 diabetes in Mexico. *Nature* **506**, 97-101.
1368 Wolpin, B.M., Rizzato, C., Kraft, P., Kooperberg, C., Petersen, G.M., Wang, Z., Arslan, A.A., Beane-
1369 Freeman, L., Bracci, P.M., Buring, J., *et al.* (2014). Genome-wide association study identifies multiple
1370 susceptibility loci for pancreatic cancer. *Nat Genet* **46**, 994-1000.
1371 Woodmansey, C., McGovern, A.P., McCullough, K.A., Whyte, M.B., Munro, N.M., Correa, A.C., Gatenby,
1372 P.A.C., Jones, S.A., and de Lusignan, S. (2017). Incidence, Demographics, and Clinical Characteristics of
1373 Diabetes of the Exocrine Pancreas (Type 3c): A Retrospective Cohort Study. *Diabetes Care* **40**, 1486-
1374 1493.
1375 Yang, J., Ferreira, T., Morris, A.P., Medland, S.E., Madden, P.A., Heath, A.C., Martin, N.G., Montgomery,
1376 G.W., Weedon, M.N., Loos, R.J., *et al.* (2012). Conditional and joint multiple-SNP analysis of GWAS
1377 summary statistics identifies additional variants influencing complex traits. *Nat Genet* **44**, 369-375, S361-
1378 363.
1379 Yourshaw, M., Taylor, S.P., Rao, A.R., Martin, M.G., and Nelson, S.F. (2015). Rich annotation of DNA
1380 sequencing variants by leveraging the Ensembl Variant Effect Predictor with plugins. *Brief Bioinform* **16**,
1381 255-264.
1382 Zhao, W., Rasheed, A., Tikkanen, E., Lee, J.J., Butterworth, A.S., Howson, J.M.M., Assimes, T.L.,
1383 Chowdhury, R., Orho-Melander, M., Damrauer, S., *et al.* (2017). Identification of new susceptibility loci
1384 for type 2 diabetes and shared etiological pathways with coronary heart disease. *Nat Genet* **49**, 1450-
1385 1457.
1386

Figure 1

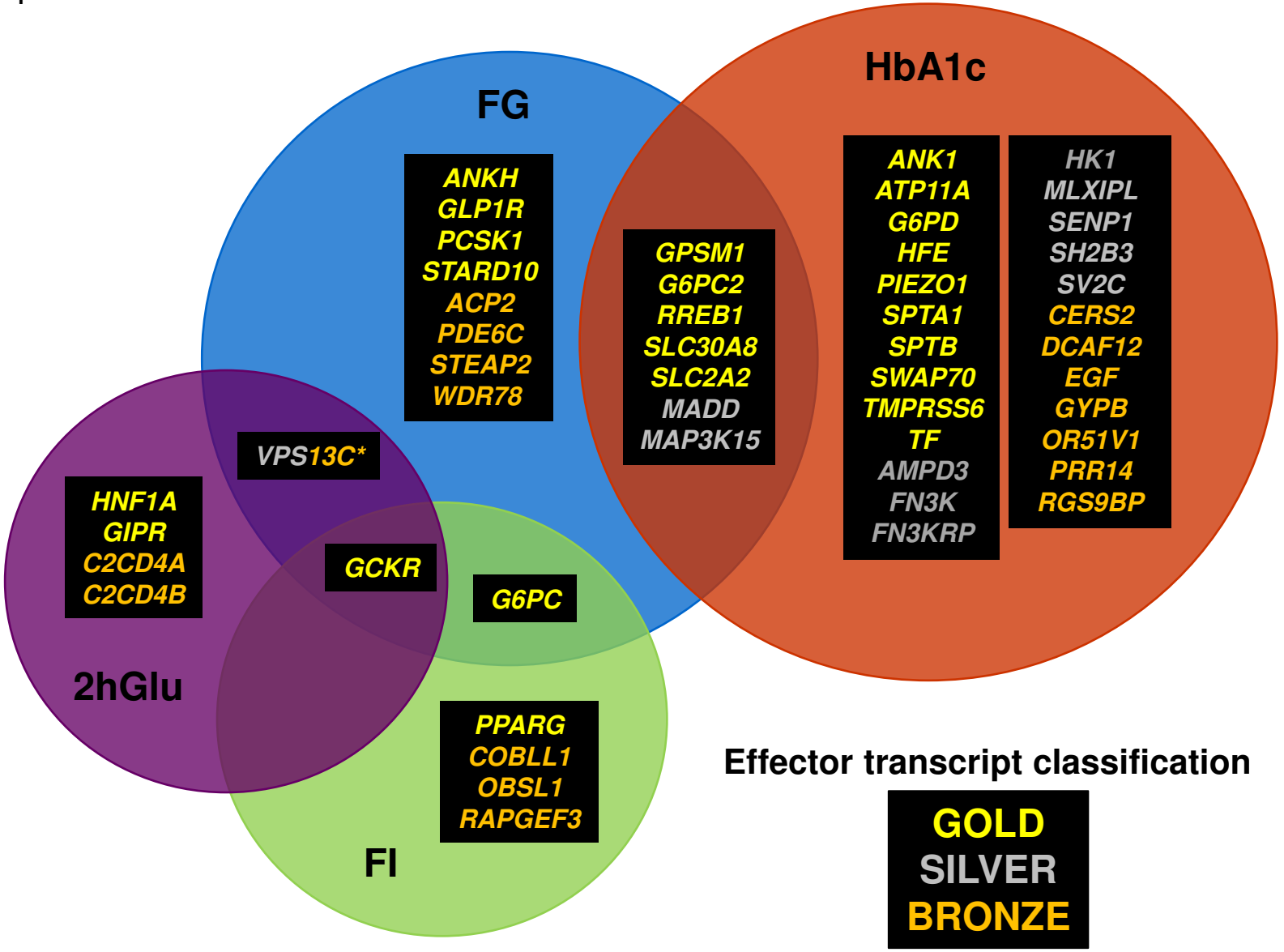


Figure 1. Effector transcript classification into “gold”, “silver” and “bronze” categories based on strength of genetic and biological evidence. A total of 51 effector transcripts from 74 single variant and six gene-based signals were identified, with many of them shared across traits. The classification was undertaken independently by four of the authors and the consensus was used as the final classification for effector transcripts (see **Methods**). *Asterisk indicates “silver” for FG, “bronze” for 2hGlu.

Figure 2

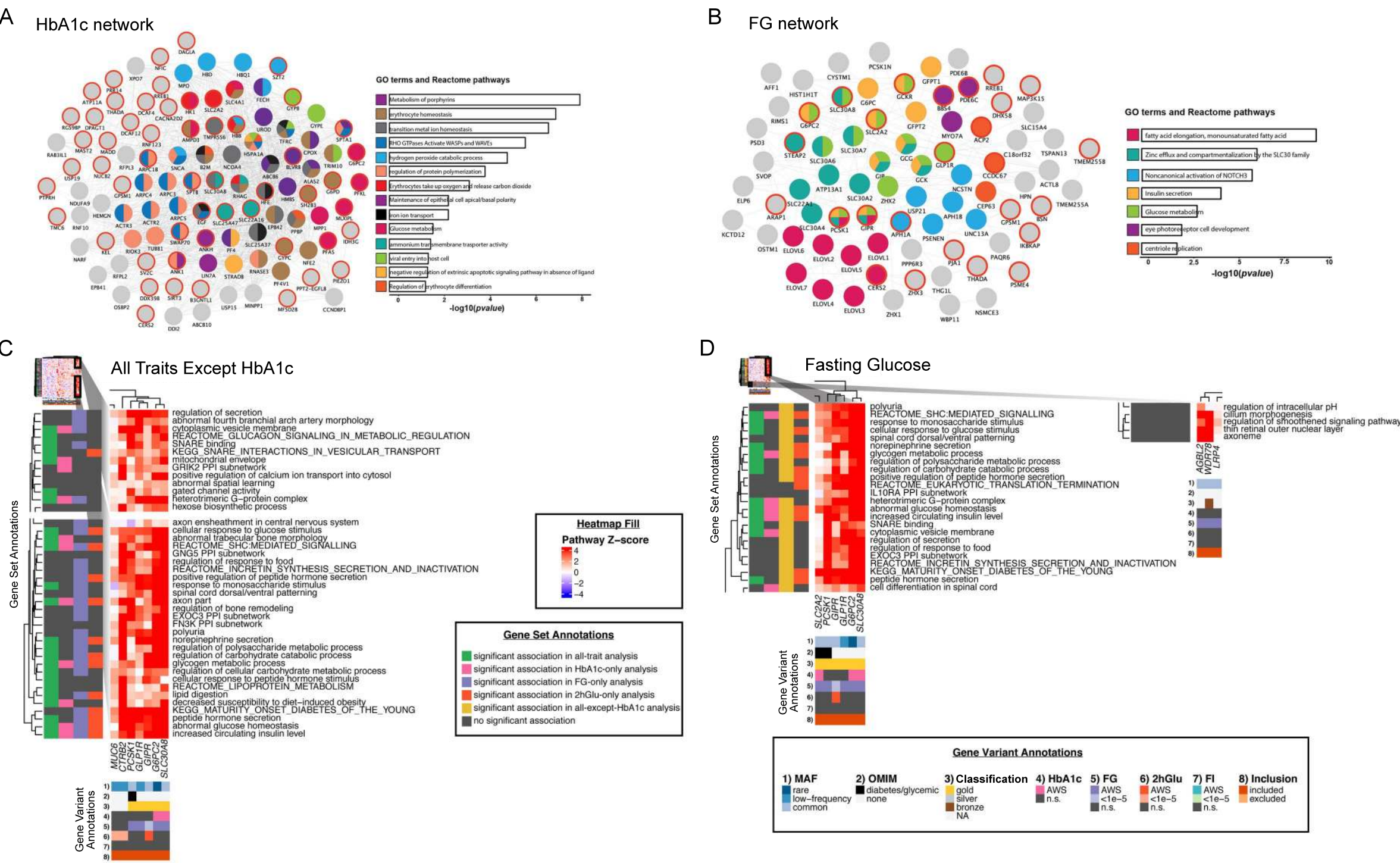


Figure 2. Network and pathway analyses identify relevant gene sets regulating glycemia using two different methods for variant associations with $P < 1 \times 10^{-5}$. (A-B) The networks represent composite networks for (A) HbA1c and (B) FG, from the GeneMANIA analysis using genes with variant associations at $P < 1 \times 10^{-5}$ for each trait as input. Nodes outlined in red correspond to genes from the input list. Other nodes correspond to related genes based on 50 default databases. Based on the network, GO terms and Reactome pathways that were significantly enriched are depicted. To summarize these results, the most significant term of all calculated terms within the same group is represented. Barplots with the Bonferroni-adjusted $-\log_{10}(p\text{-values})$ of the most significant terms within each group are shown. Each group was assigned a specific color; if a gene is present in more than one term, it is displayed in more than one color. (C-D) Heatmaps showing EC-DEPICT results from analysis of (C) all traits except HbA1c and (D) FG. The columns represent the input genes for the analysis. In (C), these are genes with variant associations of $P < 1 \times 10^{-5}$ for FG, FI, and/or 2hGlu, and in (D) these are genes with variant associations of $P < 1 \times 10^{-5}$ for FG. Rows in the heatmap represent significant meta-gene sets (FDR < 0.05). The color of each square indicates DEPICT's z-score for membership of that gene in that gene set, where dark red means "very likely a member" and dark blue means "very unlikely a member." The gene set annotations indicate whether that meta-gene set was significant at FDR < 0.05 or not significant (n.s.) for each of the other EC-DEPICT analyses. For heatmap intensity and EC-DEPICT P-values, the meta-gene set values are taken from the most significantly enriched member gene set. The gene variant annotations are as follows: (1) the European minor allele frequency (MAF) of the input variant, where rare is MAF < 1%, low-frequency is MAF 1-5%, and common is MAF > 5%, (2) whether the gene has an Online Mendelian Inheritance in Man (OMIM) annotation as causal for a diabetes/glycemic-relevant syndrome or blood disorder, (3) the effector transcript classification for that variant: gold, silver, bronze, or NA (note that only array-wide significant variants were classified, so suggestively-significant variants are by default classified as "NA"), (4-7) whether each variant was significant ($P < 2 \times 10^{-7}$), suggestively significant ($P < 1 \times 10^{-5}$), or not significant in Europeans for each of the four traits, and (8) whether each variant was included in the analysis or excluded by filters (see Methods). AWS: array-wide significant. Related to Figures S1 to S3.

Figure 3

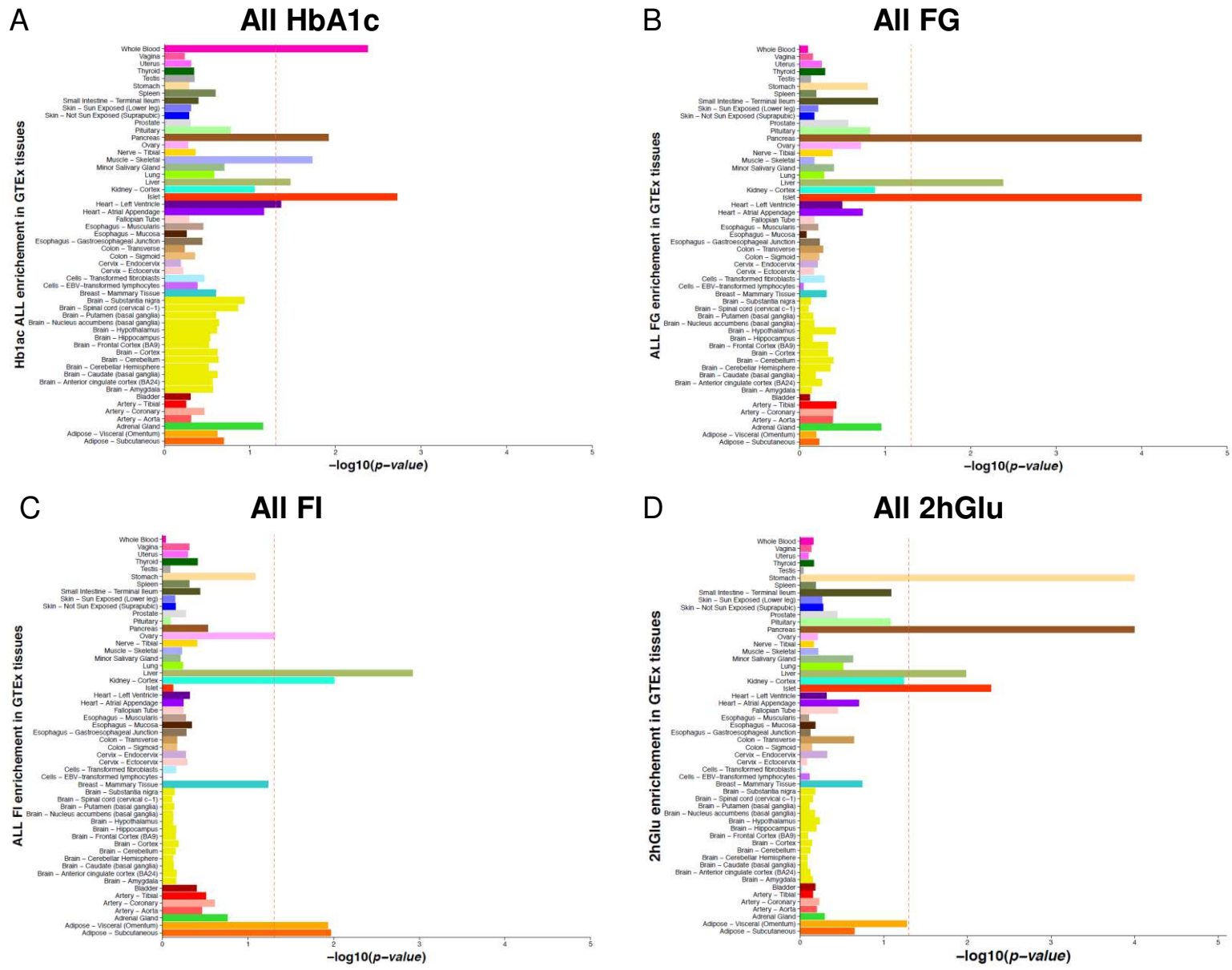


Figure 3. Tissue enrichment analysis reveals the key tissues involved in the regulation of glycemic traits. The figures display expression enrichment of genes from all of the golden, silver, and bronze gene set lists for (A) HbA1c, (B) FG, (C) FI and (D) 2hGlu in GTEx tissue samples plus islet data. Enrichment P-values were assessed empirically for each tissue using a permutation procedure (10,000 iterations), and the red vertical line shows the significance threshold (empirical P<0.05).

Figure 4

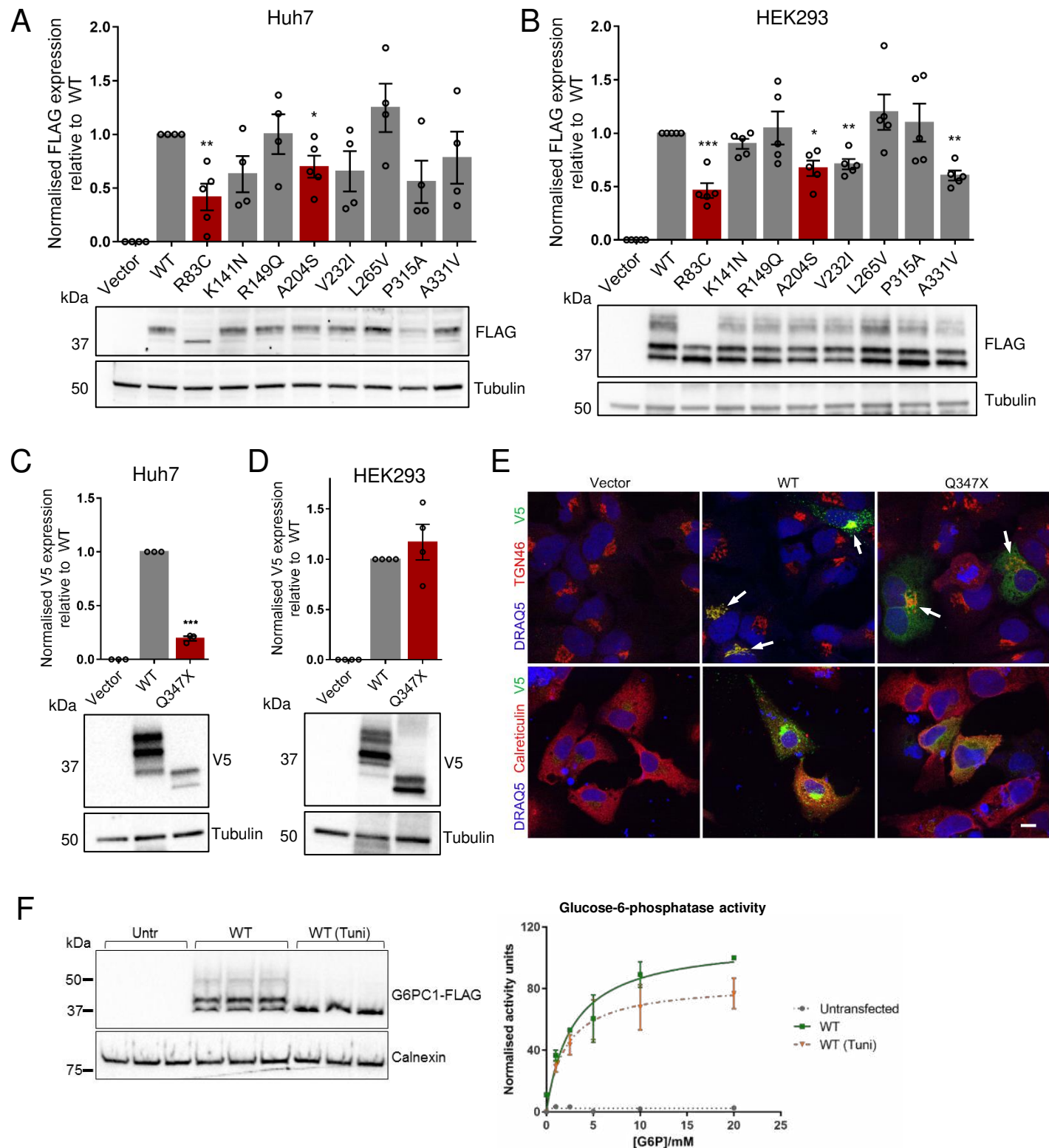


Figure 4. Functional characterisation of G6PC variant proteins. Related to Figure S4.

(A) Protein expression levels of missense G6PC variants were determined in Huh7 cells (n=4-5) and (B) HEK293 cells (n=5) by western blot densitometric analysis of FLAG-tagged G6PC constructs relative to tubulin control, with representative blots shown.

(C) Protein expression levels of PTV Q347X were determined in Huh7 cells (n=3) and (D) HEK293 cells (n=4) by western blot densitometric analysis of V5-tagged G6PC constructs relative to tubulin control, with representative blots shown. Bars in red indicate variants that are statistical drivers of the gene-based signal.

(E) Cellular localisation of V5-tagged G6PC-Q347X was assessed in Huh7 cells and overlaid with markers for the ER (calreticulin) and the trans-golgi network (TGN46). White arrows point to positions of the Golgi apparatus. Scale bar indicates 10 μ m.

(F) Glucose-6-phosphatase activity of unglycosylated WT G6PC protein obtained from tunicamycin-treated (Tuni) HEK293 microsomes (n=2), with representative western blot of microsomal protein shown. All data presented as mean \pm SEM. * p=0.01-0.05; ** p=0.001-0.01; *** p<0.001.

Figure 5

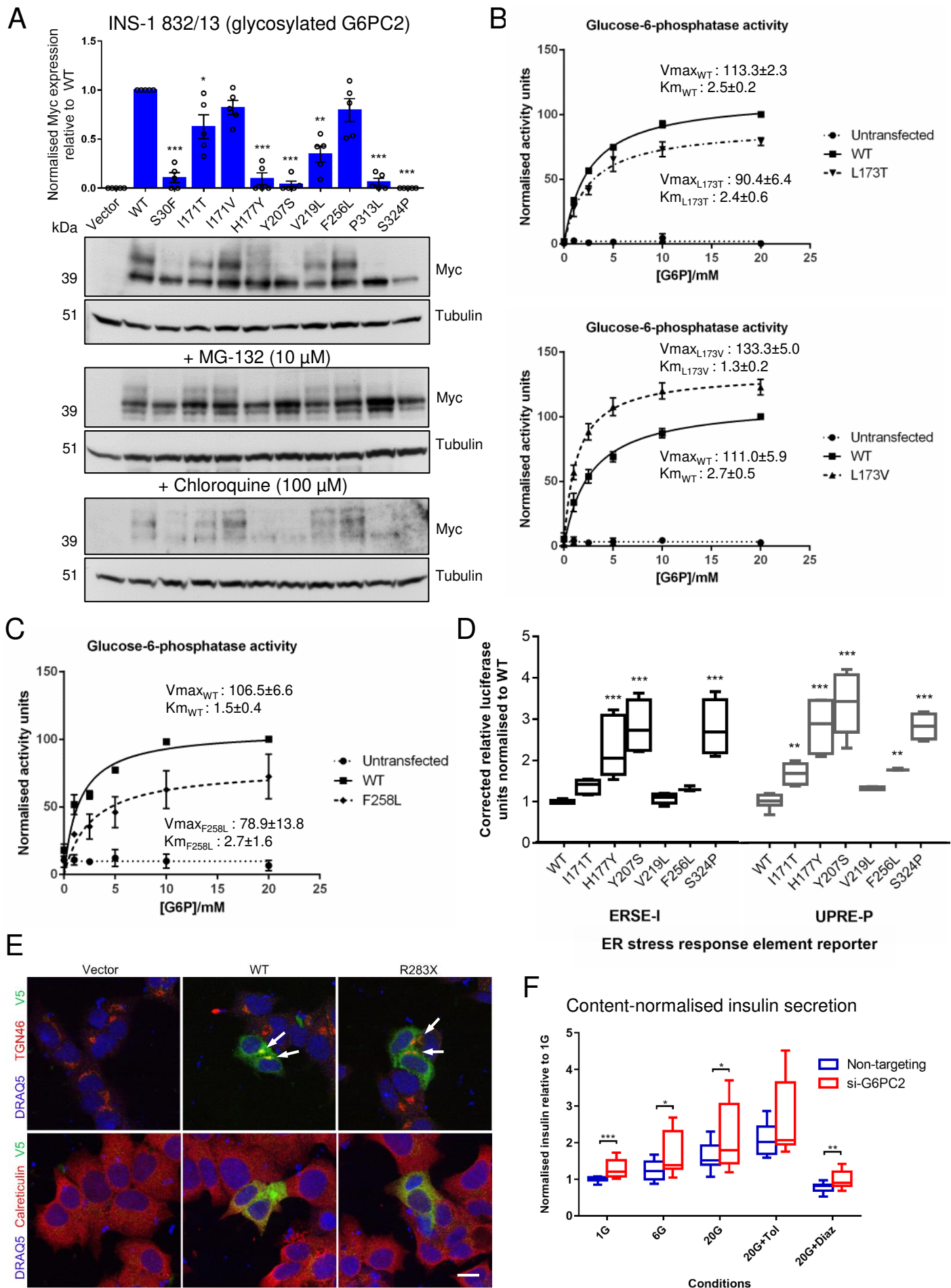


Figure 5. Functional characterisation of G6PC2 variant proteins and the role of G6PC2 in human beta cells. Related to Figure S5.

(A) Expression levels of the glycosylated forms (upper bands only) of G6PC2 variant proteins were determined in INS-1 832/13 cells by western blot densitometric analysis of Myc-tagged G6PC2 constructs relative to tubulin control (n=5). Representative blots are shown for untreated cells together with cells treated with proteasomal inhibitor MG-132 or lysosomal inhibitor chloroquine.

(B) Glucose-6-phosphatase activity of L173T and L173V variants in G6PC (proxy for I171T and I171V in G6PC2 respectively) in HEK293 against increasing glucose-6-phosphate concentrations (n=4), with mean $V_{max} \pm SEM$ and $K_m \pm SEM$ values shown for WT and each variant.

(C) Glucose-6-phosphatase activity of F258L variant in G6PC (proxy for F256L in G6PC2) in HEK293 against increasing glucose-6-phosphate concentrations (n=3), with mean $V_{max} \pm SEM$ and $K_m \pm SEM$ values shown. V_{max} and K_m results were computed based on the Michaelis-Menten kinetic model.

(D) Effect of G6PC2 WT and variant protein expression on luciferase activity driven by ER stress response elements in HEK293 cells. Relative luciferase units corrected for background activity were normalised to WT for each reporter, from n=6 across two independent experiments (except for F256L, n=3 in one experiment) using two-way ANOVA with Fisher's LSD test comparing each variant to WT.

(E) Cellular localisation of R283X in EndoC- β H1 overlaid with markers for the ER (calreticulin) and the trans-golgi network (TGN46). White arrows point to positions of the Golgi apparatus. Scale bar indicates 10 μ m.

(F) Insulin secretion normalised to total content at basal and high glucose conditions (with and without drug treatments) following 96-120h *G6PC2* knockdown in EndoC- β H1. Unpaired two-tailed Students' t tests were used to compare *G6PC2* knockdown to control for each condition, from n=16 across 4 independent experiments. Tol: tolbutamide; Diaz: diazoxide. All data presented as mean $\pm SEM$. * p=0.01-0.05; ** p=0.001-0.01; *** p<0.001.

bioRxiv preprint doi: <https://doi.org/10.1101/2019.10.03.301918>; this version posted October 3, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Track	SNP	Gene	Protein consequence	Alleles	Freq. Effect Allele	Effect (SE)	P	N	Previous glycaemic trait association (if any)	Locus name or previous association
FG	rs1886686	WDR78	p.G12A	G/C	0.739	0.014 (0.002)	2.24×10⁻¹¹	123558	Novel	
HbA1c	rs267738	CERS2	p.E106A	G/T	0.186	-0.01 (0.002)	6.96×10 ⁻¹⁰	144043	HbA1c	CERS2
HbA1c	rs863362	OR10X1	p.W66X	T/C	0.465	0.011 (0.001)	6.76×10 ⁻¹⁵	114945	HbA1c	SPTA1
HbA1c	rs857725	SPTA1	p.K1693Q	G/T	0.262	0.022 (0.001)	1.56×10 ⁻⁵⁰	143956	HbA1c	SPTA1
HbA1c	rs11887523	MFSD2B	p.A60T	A/G	0.007	-0.072 (0.01)	1.44×10 ⁻¹²	122060	HbA1c	ATAD2B
FG	rs1260326²	GCKR	p.L446P	C/T	0.631	0.029 (0.002)	6.36×10⁻⁴⁸	129588	FG, FI, 2hGlu	GCKR
FI	rs1260326²	GCKR	p.L446P	C/T	0.626	0.024 (0.002)	5.55×10⁻³²	104076	FG, FI, 2hGlu	GCKR
2hGlu	rs1260326²	GCKR	p.L446P	C/T	0.618	-0.069 (0.009)	4.48×10⁻¹⁵	57813	FG, FI, 2hGlu	GCKR
FG	rs35720761²	THADA	p.C845Y	T/C	0.108	-0.018 (0.003)	4.35×10⁻⁹	129622	T2D	THADA
HbA1c	rs35720761 ²	THADA	p.C845Y	C/T	0.113	0.014 (0.002)	2.58×10 ⁻¹²	144001	T2D	THADA
FG	rs7578597	THADA	p.T897A	C/T	0.106	-0.019 (0.003)	1.99×10⁻⁸	113162	T2D	THADA
FI	rs7607980²	COBLL1	p.N901D	C/T	0.128	-0.032 (0.003)	1.30×10⁻²⁴	97817	FI	COBLL1
FG	rs2232323	G6PC2	p.Y207S	C/A	0.006	-0.129 (0.012)	1.05×10⁻²⁸	123981	FG, HbA1c	G6PC2
HbA1c	rs2232323	G6PC2	p.Y207S	C/A	0.007	-0.053 (0.007)	3.25×10 ⁻¹³	144038	FG, HbA1c	G6PC2
FG	rs146779637	G6PC2	p.R283X	T/C	0.002	-0.138 (0.02)	1.78×10⁻¹²	127278	FG, HbA1c	G6PC2
HbA1c	rs146779637	G6PC2	p.R283X	T/C	0.002	-0.074 (0.012)	4.58×10 ⁻¹⁰	141728	FG, HbA1c	G6PC2
FI	rs1983210	OBSL1	p.E1365D	G/C	0.729	0.016 (0.003)	8.48×10⁻¹⁰	79767	Novel	
FI	rs3183099	OBSL1	splice region variant	A/G	0.226	-0.013 (0.002)	4.70×10⁻⁸	100713	Novel	
FI	rs1801282²	PPARG	p.P12A	G/C	0.117	-0.031 (0.003)	3.50×10⁻²³	98631	FI	PPARG
HbA1c	rs35726701	RNF123	p.K596E	G/A	0.019	0.025 (0.005)	4.19×10 ⁻⁸	131203	HbA1c	USP4
FG	rs5400	SLC2A2	p.T110I	A/G	0.161	-0.022 (0.003)	2.14×10⁻¹⁷	129591	FG, HbA1c	SLC2A2
HbA1c	rs5400	SLC2A2	p.T110I	A/G	0.153	-0.013 (0.002)	2.27×10 ⁻¹³	144012	FG, HbA1c	SLC2A2
HbA1c¹	rs2237051	EGF	p.M708I	A/G	0.374	-0.007 (0.001)	2.11×10⁻⁷	121204	Novel	
HbA1c	rs7683365	GYPB	p.T48M	A/G	0.312	0.012 (0.002)	1.61×10 ⁻⁸	45191	HbA1c	FREM3
FG	rs146886108²	ANKH	p.R187Q	T/C	0.004	-0.088 (0.014)	5.67×10⁻¹⁰	129647	Novel	
HbA1c	rs31244	SV2C	p.D543N	A/G	0.083	0.012 (0.002)	6.05×10⁻⁸	144000	Novel	
FG	rs6235	PCSK1	p.S690T	G/C	0.264	-0.022 (0.002)	9.22×10⁻²⁴	123560	FG	PCSK1
2hGlu	rs2549782	ERAP2	p.K392N	T/G	0.519	-0.055 (0.009)	6.81×10⁻¹⁰	57836	2hGlu	ERAP2
HbA1c	rs35742417 ³	RREB1	p.S1499Y	A/C	0.173	-0.01 (0.002)	3.76×10 ⁻⁹	143967	FG, T2D	RREB1
FG	rs35742417³	RREB1	p.S1499Y	A/C	0.183	-0.019 (0.002)	1.27×10⁻¹⁶	129577	FG, T2D	RREB1
HbA1c	rs1799945	HFE	p.H63D	G/C	0.129	-0.023 (0.002)	1.20×10 ⁻³⁰	128354	HbA1c	HFE, HIST1H4A
HbA1c	rs1800562	HFE	p.C279Y	A/G	0.051	-0.042 (0.003)	3.30×10 ⁻⁴⁷	138093	HbA1c	HFE, HIST1H4A
FG	rs10305492	GLP1R	p.A316T	A/G	0.014	-0.08 (0.008)	2.37×10⁻²⁵	129601	FG	GLP1R
HbA1c	rs35332062	MLXIPL	p.A358V	A/G	0.117	0.011 (0.002)	6.18×10⁻⁹	144042	Novel	
HbA1c	rs3812316	MLXIPL	p.Q241H	G/C	0.112	0.012 (0.002)	2.15×10⁻⁸	108605	Novel	
FG	rs194524³	STEAP2	p.R456Q	A/G	0.523	0.01 (0.002)	7.65×10⁻⁸	129629	Novel	
HbA1c	rs34664882	ANK1	p.A1503V	A/G	0.026	-0.049 (0.004)	2.43×10 ⁻³⁹	144034	HbA1c	ANK1
FG	rs13266634²	SLC30A8	p.R276W	T/C	0.305	-0.029 (0.002)	1.63×10⁻⁴⁶	129614	FG, HbA1c, T2D	SLC30A8
HbA1c	rs13266634 ²	SLC30A8	p.R276W	T/C	0.300	-0.015 (0.001)	8.50×10 ⁻²⁸	143982	FG, HbA1c, T2D	SLC30A8
HbA1c	rs11557154	DCAF12	p.R113Q	T/C	0.138	-0.009 (0.002)	1.70×10⁻⁷	144045	Novel	
FG	rs17853166	IKBKAP	p.S251G	C/T	0.026	-0.037 (0.006)	4.82×10⁻¹¹	129640	FG	IKBKAP
HbA1c	rs60980157 ²	GPSM1	p.S391L	T/C	0.246	-0.013 (0.002)	6.71×10 ⁻¹⁷	118824	FG, T2D	GPSM1

FG	rs60980157 ²	<i>GPSM1</i>	p.S391L	T/C	0.254	-0.014 (0.002)	2.35×10^{-9}	110915	FG, T2D	<i>GPSM1</i>
HbA1c	rs990220	<i>PDE6C</i>	p.S2701	G/A	0.366	-0.011 (0.002)	1.14×10^{-11}	118580	Novel	
HbA1c	rs61732434	<i>OR51V1</i>	p.S161N	T/C	0.008	-0.052 (0.009)	1.75×10^{-8}	127507	Novel	
HbA1c	rs415895	<i>SWAP70</i>	p.Q447E	G/C	0.641	-0.013 (0.001)	1.15×10^{-21}	138028	Novel	
HbA1c	rs117706710	<i>AMPD3</i>	p.V311L	T/G	0.009	0.037 (0.006)	2.32×10^{-10}	144048	Novel	
FG	rs2167079	<i>ACP2</i>	p.R29Q	T/C	0.340	0.016 (0.002)	7.99×10^{-15}	129580	FG	<i>MADD</i>
HbA1c	rs35233100	<i>MADD</i>	p.R766X	T/C	0.055	-0.015 (0.003)	1.13×10^{-8}	144034	FG	<i>MADD</i>
FG	rs35233100	<i>MADD</i>	p.R766X	T/C	0.054	-0.029 (0.004)	1.46×10^{-12}	126231	FG	<i>MADD</i>
FG	rs56200889 ²	<i>ARAP1</i>	p.Q802E	C/G	0.270	-0.016 (0.002)	1.79×10^{-14}	122674	FG	<i>ARAP1</i>
HbA1c	rs643788	<i>DPAGT1</i>	p.I393V	C/T	0.425	-0.006 (0.001)	1.77×10^{-7}	144009	Novel	
FI ¹	rs145878042	<i>RAPGEF3</i>	p.L300P	G/A	0.011	-0.054 (0.01)	1.15×10^{-7}	91485	Novel	
HbA1c	rs2732481	<i>ZNF641</i>	p.Q363P	G/T	0.315	-0.009 (0.001)	2.07×10^{-11}	142280	HbA1c	<i>SENP1</i>
HbA1c	rs3184504	<i>SH2B3</i>	p.W262R	C/T	0.567	0.007 (0.001)	5.98×10^{-8}	138551	HbA1c	<i>ATXN2</i>
2hGlu	rs1169288 ²	<i>HNF1A</i>	p.I75L	C/A	0.345	0.06 (0.011)	7.90×10^{-9}	44278	T2D	<i>HNF1A</i>
HbA1c	COSM147717	<i>ATP11A</i>	p.M317V	G/A	0.748	0.009 (0.001)	3.77×10^{-12}	144022	HbA1c	<i>ATP11A, TUB GCP3</i>
HbA1c	rs229587	<i>SPTB</i>	p.S439N	T/C	0.357	0.007 (0.001)	2.60×10^{-8}	134780	Novel	
HbA1c	rs35097172	<i>SLC25A47</i>	splice region variant, 5' UTR variant	T/C	0.216	-0.008 (0.002)	5.67×10^{-8}	144028	FG	<i>SLC25A47</i>
2hGlu	rs3784634	<i>VPS13C</i>	p.R974K	T/C	0.540	-0.069 (0.011)	6.40×10^{-10}	37217	2hGlu	<i>VPS13C/ C2CD4A/ C2CD4B</i>
HbA1c ¹	rs3747481	<i>PRR14</i>	p.P359L	T/C	0.261	0.009 (0.002)	3.30×10^{-8}	103338	Novel	
HbA1c	rs201226914	<i>PIEZO1</i>	p.L939M	T/G	0.002	-0.159 (0.015)	4.42×10^{-26}	144024	HbA1c	<i>CDT1, CYBA</i>
2hGlu	rs72839768 ⁴	<i>DVL2</i>	p.T529I	A/G	0.020	0.197 (0.03)	4.10×10^{-11}	57866	T2D	<i>SLC16A13</i>
HbA1c	rs2748427	<i>TMC6</i>	p.W125R	G/A	0.233	0.027 (0.002)	8.56×10^{-70}	132326	HbA1c	<i>TMC6</i>
HbA1c	rs7225887	<i>B3GNTL1</i>	p.A163T	T/C	0.211	-0.015 (0.002)	5.73×10^{-22}	125749	HbA1c	<i>FN3KRP, FN3K</i>
HbA1c	rs35413309	<i>RGS9BP</i>	p.A223V	T/C	0.030	-0.02 (0.004)	1.42×10^{-8}	141598	Novel	
2hGlu	rs1800437 ²	<i>GIPR</i>	p.E318Q	C/G	0.217	0.103 (0.011)	2.59×10^{-23}	56252	2hGlu	<i>GIPR</i>
FG	rs17265513 ³	<i>ZHX3</i>	p.N310S	C/T	0.188	0.016 (0.002)	2.59×10^{-10}	126253	FG	<i>ZHX3</i>
HbA1c	rs855791	<i>TMPRSS6</i>	V727A	G/A	0.577	-0.019 (0.001)	9.46×10^{-51}	143907	HbA1c	<i>TMPRSS6</i>
FG	rs15943	<i>MAP3K15</i>	p.Q1083E	C/G	0.005	-0.084 (0.014)	2.83×10^{-9}	67004	Novel	
FG	rs56381411	<i>MAP3K15</i>	p.G670S	T/C	0.005	-0.085 (0.013)	1.51×10^{-11}	62319	Novel	
HbA1c	rs2229241	<i>RENBP</i>	splice acceptor variant	C/T	0.012	-0.123 (0.007)	1.14×10^{-62}	95622	HbA1c	<i>G6PD</i>
HbA1c	rs1050828	<i>G6PD</i>	p.V68M	T/C	0.007	-0.334 (0.008)	7.41×10^{-322}	112209	HbA1c	<i>G6PD</i>

Table 1. Single-point coding variant associations meeting the significant threshold for coding variants of $P < 2.2 \times 10^{-7}$. This table includes all novel coding variants meeting this threshold, irrespective of whether they fall in completely new loci or in previously-established loci, provided that the association at the established locus was not shown to be due to a non-coding variant (Table S3) or another coding variant at the same locus. Novel loci are highlighted in bold. HbA1c: glycated haemoglobin; FG: fasting glucose; FI: fasting insulin; 2hGlu: 2h glucose; Alleles E/O: effect allele/other allele; Freq. Effect Allele: frequency of effect allele; Effect (SE): effect size (standard error); *P*: p-value; N: number of samples in the analysis; Novel/previous glycemic trait association: Novel corresponds to a new association result; Locus name of previous association – name used for previously-reported locus. ¹Significant in the European-only analysis in our study. ²Genome-wide significant association with T2D since date of analysis (Mahajan et al., 2018b). ³Association with T2D at $P < 1 \times 10^{-4}$ since date of analysis (Mahajan et al., 2018b). ⁴T2D locus identified in Japanese (Hara et al., 2014) and Mexican (Williams et al., 2014) populations only. The date of our exomes analysis is May 2015. Related to Table S3.

Trait	Gene	NSbroad mask			NSstrict mask		
		N var	P_{burden}	P_{SKAT}	N var	P_{burden}	P_{SKAT}
FG	G6PC	9	1.41x10⁻⁶	1.32x10 ⁻⁵	3	1.41x10 ⁻³	7.43x10 ⁻⁴
FI	G6PC	8	1.62x10⁻⁶	8.58x10 ⁻⁶	3	1.85x10 ⁻³	7.80x10 ⁻³
HbA1c	TF	10	2.15x10⁻⁶	5.98x10 ⁻³	3	5.48x10 ⁻²	5.48x10 ⁻²
FG	<i>MAP3K15</i>	18	1.86x10⁻²⁵	1.07x10 ⁻¹⁸	7	1.34x10 ⁻¹⁴	4.01x10 ⁻¹¹
HbA1c	<i>MAP3K15</i>	18	1.27x10⁻⁷	1.53x10 ⁻⁰⁴	7	2.65x10 ⁻⁴	9.46x10 ⁻³
FG	<i>G6PC2</i>	18	4.09x10 ⁻⁶⁷	5.38x10 ⁻⁵⁸	7	7.8x10⁻⁶⁹	3.83x10 ⁻⁵⁶
HbA1c	<i>G6PC2</i>	18	6.18x10 ⁻³⁰	4.65x10 ⁻²⁷	7	1.04x10⁻³¹	1.92x10 ⁻²⁶
FG	<i>SLC30A8</i>	13	5.69x10 ⁻⁴	6.42x10⁻¹¹	7	6.55x10 ⁻¹¹	3.74x10 ⁻¹⁰
HbA1c	<i>SLC30A8</i>	12	7.20x10 ⁻⁸	2.18x10 ⁻⁵	6	5.66x10⁻⁸	3.22x10 ⁻⁶
FG	<i>VPS13C</i>	52	9.66x10 ⁻⁶	3.73x10⁻⁷	26	1.27x10 ⁻⁵	1.44x10 ⁻⁵

Table 2. Gene-based results from broad (NSbroad mask) and strict (NSstrict mask) analyses. Genes in bold are newly discovered from this effort. N var: total number of variants in that gene-based analysis; P_{burden} : p-value from burden test which assumes all variants have the same direction of effect; P_{SKAT} : p-value from SKAT test which allows for different directions of effect between variants. The lowest p-value is highlighted in bold. Related to Table S4.