Article:

(2023) Large-scale exome array summary statistics resources for glycemic traits to aid effector gene prioritization. Wellcome Open Research. 483. ISSN 2398-502X

https://doi.org/10.12688/wellcomeopenres.18754.1

Check for updates

RESEARCH ARTICLE

Large-scale exome array summary statistics resources for glycemic traits to aid effector gene prioritization

[version 1; peer review: 2 approved]

Sara M. Willems[1,2*], Natasha H. J. Ng[3,4*], Juan Fernandez[5], Rebecca S. Fine[6-9], Eleanor Wheeler[1,10], Jennifer Wessel[8,11-13], Hidetoshi Kitajima[5], Gaelle Marenne[10], Xueling Sim[14,15], Hanieh Yaghootkar[16], Shuai Wang[17], Sai Chen[15], Yuning Chen[17], Yii-Der Ida Chen[18], Niels Grarup[19], Ruifang Li-Gao[20], Tibor V. Varga[21], Jennifer L. Asimit[10,22], Shuang Feng[23], Rona J. Strawbridge[24,25], Erica L. Kleinbrink[26,27], Tarunveer S. Ahluwalia[28,29], Ping An[30], Emil V. Appel[19], Dan E. Arking[31], Juha Auvinen[32,33], Lawrence F. Bielak[34], Nathan A. Bihlmeyer[35], Jette Bork-Jensen[19], Jennifer A. Brody[36,37], Archie Campbell[38], Audrey Y. Chu[39], Gail Davies[40,41], Ayse Demirkan[42], James S. Floyd[36,37], Franco Giulianini[39], Xiuqing Guo[18], Stefan Gustafsson[43], Anne U. Jackson[15], Johanna Jakobsdottir[44], Marjo-Riitta Järvelin[32,45,46], Richard A. Jensen[36,37], Stavroula Kanoni[47], Sirkka Keinanen-Kiukaanniemi[48,49], Man Li[50,51], Yingchang Lu[52,53], Jian'an Luan[1], Alisa K. Manning[54,55], Jonathan Marten[56], Karina Meidtner[57,58], Dennis O. Mook-Kanamori[20,59], Taulant Muka[42,60], Giorgio Pistis[61,62], Bram Prins[10], Kenneth M. Rice[36,63], Serena Sanna[61,64], Albert Vernon Smith[44,65], Jennifer A. Smith[34,66], Lorraine Southam[5,10,67], Heather M. Stringham[15], Vinicius Tragante[68], Sander W. van der Laan[69], Helen R. Warren[47,70], Jie Yao[18], Andrianos M. Yiorkas[71,72], Weihua Zhang[73,74], Wei Zhao[34], Mariaelisa Graff[75], Heather M. Highland[75,76], Anne E. Justice[75], Eirini Marouli[47], Carolina Medina-Gomez[42,77], Saima Afaq[45], Wesam A. Alhejily[47,78], Najaf Amin[42], Folkert W. Asselbergs[79,80], Lori L. Bonnycastle[81], Michiel L. Bots[82], Ivan Brandslund[83,84], Ji Chen[10], John Danesh[85], Renée de Mutsert[20], Abbas Dehghan[42,45,86], Tapani Ebeling[87], Paul Elliott[45,88,89], EPIC-Interact Consortium, Aliki-Eleni Farmaki[90,91], Jessica D. Faul[66], Paul W. Franks[7,21,92], Steve Franks[93], Andreas Fritsche[58,94], Anette P. Gjesing[19], Mark O. Goodarzi[95], Vilmundur Gudnason[44,65], Göran Hallmans[96], Tamara B. Harris[44], Karl-Heinz Herzig[97,98], Marie-France Hivert[99,100], Torben Jørgensen[101-103], Marit E. Jørgensen[29,104], Pekka Jousilahti[105], Eero Kajantie[105-108], Maria Karaleftheri[109], Sharon L.R. Kardia[34], Leena Kinnunen[105], Heikki A. Koistinen[105,110,111], Pirjo Komulainen[112], Peter Kovacs[113,114], Johanna Kuusisto[115], Markku Laakso[115], Leslie A. Lange[116], Lenore J. Launer[117], Aaron Leong[118], Jaana Lindström[105], Jocelyn E. Manning Fox[119,120], Satu Männistö[105], Nisa M. Maruthur[51,121,122], Leena Moilanen[123], Antonella Mulas[61,124], Mike A. Nalls[125,126], Matthew Neville[3], James S. Pankow[127], Alison Pattie[41], Eva R.B. Petersen[83], Hannu Puolijoki[128], Asif Rasheed[129], Paul Redmond[41], Frida Renström[21,96], Michael Roden[58,130,131], Danish Saleheen[129,132], Juha Saltevo[133], Kai Savonen[112,134], Sylvain Sebert[46,48], Tea Skaaby[101], Kerrin S. Small[135], Alena Stančáková[115], Jakob Stokholm[28], Konstantin Strauch[136], E-Shyong Tai[14,137,138], Kent D. Taylor[18], Betina H. Thuesen[101], Anke Tönjes[139], Emmanouil Tsafantakis[140], Tiinamaija Tuomi[141-144], Jaakko Tuomilehto[105,145,146], Understanding Society Scientific Group, Matti Uusitupa[147], Marja Vääräsmäki[106,148], Ilonca Vaartjes[82], Magdalena Zoledziewska[61], Goncalo Abecasis[62], Beverley Balkau[149], Hans Bisgaard[28], Alexandra I. Blakemore[71,72], Matthias Blüher[139,150], Heiner Boeing[151], Eric Boerwinkle[152], Klaus Bønnelykke[28], Erwin P. Bottinger[52], Mark J. Caulfield[47,70], John C. Chambers[45,74,153], Daniel I. Chasman[39,154,155], Ching-Yu Cheng[156-158], Francis S. Collins[81], Josef Coresh[51,122], Francesco Cucca[61,124], Gert J. de Borst[159], Ian J. Deary[40,41], George Dedoussis[90], Panos Deloukas[47,160], Hester M. den Ruijter[161], Josée Dupuis[17,162], Michele K. Evans[117], Ele Ferrannini[163], Oscar H. Franco[42,60], Harald Grallert[58,164], Torben Hansen[19,165], Andrew T. Hattersley[166], Caroline Hayward[56], Joel N. Hirschhorn[7,8,167], Arfan Ikram[42], Erik Ingelsson[168-170], Fredrik Karpe[3,171], Kay-Tee Kaw[172], Wieland Kiess[173], Jaspal S. Kooner[74,153,174], Antje Körner[173], Timo Lakka[112,134,175], Claudia Langenberg[1], Lars Lind[176], Cecilia M. Lindgren[5,177], Allan Linneberg[101,178], Leonard Lipovich[27,179], Ching-Ti Liu[17], Jun Liu[42], Yongmei Liu[180], Ruth J.F. Loos[52,181], Patrick E. MacDonald[119,120], Karen L. Mohlke[182], Andrew D. Morris[183], Patricia B. Munroe[47,70], Alison Murray[184], Sandosh Padmanabhan[185], Colin N. A . Palmer[186], Gerard Pasterkamp[161,187], Oluf Pedersen[19], Patricia A. Peyser[34], Ozren Polasek[188], David Porteous[38], Michael A. Province[30], Bruce M. Psaty[36,37,189], Rainer Rauramaa[112], Paul M. Ridker[39,154,190], Olov Rolandsson[191], Patrik Rorsman[3,171], Frits R. Rosendaal[20], Igor Rudan[183], Veikko Salomaa[105], Matthias B. Schulze[57,58], Robert Sladek[192,193], Blair H. Smith[186], Timothy D. Spector[135], John M. Starr[40,194], Michael Stumvoll[139], Cornelia M. van Duijn[42], Mark Walker[195], Nick J. Wareham[1], David R. Weir[66], James G. Wilson[196], Tien Yin Wong[156-158], Eleftheria Zeggini[10,67,197], Alan B. Zonderman[117], Jerome I. Rotter[18], Andrew P. Morris[198], Michael Boehnke[15], Jose C. Florez[55,199,200], Mark I. McCarthy[3,5,171,201], James B. Meigs[118,200], Anubha Mahajan[5,201], Robert A. Scott[1], Anna L. Gloyn[3,5,171,202], Inês Barroso[1,10,203]

[1]MRC Epidemiology Unit, University of Cambridge School of Clinical Medicine, Institute of Metabolic Science, Cambridge Biomedical Campus, Cambridge, CB2 0QQ, UK
[2]General Medicine Center, Saarland University Faculty of Medicine, Homburg, 66421, Germany
[3]Oxford Centre for Diabetes, Endocrinology and Metabolism, University of Oxford, Oxford, OX3 7LE, UK
[4]Stem Cells and Diabetes Laboratory, Institute of Molecular and Cell Biology, Agency for Science, Technology and Research (A*STAR), Singapore, 138673, Singapore
[5]Wellcome Centre for Human Genetics, University of Oxford, Oxford, OX3 7BN, UK
[6]Department of Genetics, Harvard Medical School, Boston, MA, 02115, USA
[7]Division of Endocrinology and Center for Basic and Translational Obesity Research, Boston Children's Hospital, Boston, MA, 02115, USA
[8]Broad Institute of MIT and Harvard, Cambridge, MA, 02142, USA
[9]Current address: Vertex Pharmaceuticals Incorporated, 50 Northern Avenue, Boston, MA, 02210, USA
[10]Department of Human Genetics, Wellcome Sanger Institute, Genome Campus, Hinxton, Cambridge, CB10 1SA, UK
[11]Departments of Epidemiology & Medicine, Schools of Public Health & Medicine, Indiana University, Indianapolis, IN, 46202, USA
[12]Diabetes Translational Research Center, Indiana University School of Medicine, Indianapolis, IN, 46202, USA
[13]General Medicine Division, Massachusetts General Hospital, Boston, MA, USA
[14]Saw Swee Hock School of Public Health, National University Health System, National University of Singapore, Singapore, 117549, Singapore
[15]Department of Biostatistics and Center for Statistical Genetics, University of Michigan, Ann Arbor, MI, 48109, USA
[16]Genetics of Complex Traits, University of Exeter Medical School, University of Exeter, Exeter, EX2 5DW, UK
[17]Department of Biostatistics, Boston University School of Public Health, Boston, MA, USA
[18]The Institute for Translational Genomics and Population Sciences, Department of Pediatrics, The Lundquist Institute for Biomedical Innovation at Harbor-UCLA Medical Center, Torrance, CA, 90502, USA
[19]Novo Nordisk Foundation Center for Basic Metabolic Research, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, 2200, Denmark
[20]Department of Clinical Epidemiology, Leiden University Medical Center, Leiden, 2333 ZA, The Netherlands
[21]Department of Clinical Sciences, Genetic and Molecular Epidemiology Unit, Lund University, Malmö, SE-205 02, Sweden
[22]MRC Biostatistics Unit, University of Cambridge, Cambridge, CB2 0SR, UK
[23]Department of Biostatistics, University of Michigan School of Public Health, Ann Arbor, MI, USA
[24]Mental Health and Wellbeing, School of Health and Wellbeing, College of Medical, Veterinary and Life Sciences, University of Glasgow, Glasgow, G12 8RZ, UK
[25]Cardiovascular Medicine Unit, Department of Medicine Solna, Karolinska Institute, Stockholm, 171 76, Sweden
[26]Quantitative Life Sciences, McGill University, Montreal, Quebec, Canada
[27]Center for Molecular Medicine and Genetics, Wayne State University, Detroit, MI, 48201-1928, USA
[28]COPSAC, Copenhagen Prospective Studies on Asthma in Childhood, Herlev and Gentofte Hospital, University of Copenhagen, Copenhagen, Denmark
[29]Steno Diabetes Center Copenhagen, Gentofte, 2820, Denmark
[30]Department of Genetics, Division of Statistical Genomics, Washington University School of Medicine, St. Louis, Missouri, 63108, USA
[31]McKusick-Nathans Institute, Department of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, MD, USA
[32]Center for Life Course Health Research, University of Oulu, Oulu, 90014, Finland
[33]Unit of Primary Care, Oulu University Hospital, Oulu, Finland
[34]Department of Epidemiology, School of Public Health, University of Michigan, Ann Arbor, MI, 48109, USA
[35]McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, MD, USA
[36]Cardiovascular Health Research Unit, University of Washington, Seattle, WA, 98195, USA
[37]Department of Medicine, University of Washington, Seattle, WA, USA
[38]

Centre for Genomic and Experimental Medicine, Institute of Genetics and Cancer, University of Edinburgh, Edinburgh, EH4 2XU, UK
[39]Division of Preventive Medicine, Brigham and Women's Hospital, Boston, MA, 02215, USA
[40]Centre for Cognitive Ageing and Cognitive Epidemiology, University of Edinburgh, Edinburgh, EH8 9JZ, UK
[41]Department of Psychology, University of Edinburgh, Edinburgh, EH8 9JZ, UK
[42]Department of Epidemiology, Erasmus University Medical Center, Rotterdam, 3015 GE, The Netherlands
[43]Department of Medical Sciences, Molecular Epidemiology and Science for Life Laboratory, Uppsala University, Uppsala, 75237, Sweden
[44]Icelandic Heart Association, Kopavogur, Iceland
[45]Department of Epidemiology and Biostatistics, MRC-PHE Centre for Environment & Health, School of Public Health, Imperial College London, London, W2 1PG, UK
[46]Biocenter Oulu, University of Oulu, Oulu, Finland
[47]William Harvey Research Institute, Barts and The London School of Medicine and Dentistry, Queen Mary University of London, London, UK
[48]Faculty of Medicine, Center for Life Course Health Research, University of Oulu, Oulu, Finland
[49]MRC and Unit of Primary Care, Oulu University Hospital, Oulu, Finland
[50]Division of Nephrology, Internal Medicine, School of Medicine, University of Utah, Salt Lake City, USA
[51]Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, Maryland, USA
[52]The Charles Bronfman Institute for Personalized Medicine, The Icahn School of Medicine at Mount Sinai, New York, NY, 10069, USA
[53]Department of Medicine, Division of Genetic Medicine, Vanderbilt Genetics Institute, Vanderbilt University School of Medicine, Nashville, TN, 37203, USA
[54]Center for Human Genetics Research, Massachusetts General Hospital, Boston, MA, 02114, USA
[55]Department of Medicine, Harvard Medical School, Boston, MA, USA
[56]Medical Research Council Human Genetics Unit, Institute of Genetics and Cancer, University of Edinburgh, Edinburgh, EH4 2XU, UK
[57]Department of Molecular Epidemiology, German Institute of Human Nutrition Potsdam-Rehbruecke (DIfE), Nuthetal, 14558, Germany
[58]German Center for Diabetes Research (DZD), München-Neuherberg, 85764, Germany
[59]Department of Public Health and Primary Care, Leiden University Medical Center, Leiden, 2333 ZA, The Netherlands
[60]Institute of Social and Preventive Medicine, University of Bern, Bern, Switzerland
[61]Italian National Research Council, Institute of Genetics and Biomedic Research, Cittadella Universitaria, Monserrato, 09042, Italy
[62]Center for Statistical Genetics, University of Michigan, Ann Arbor, MI, 48109, USA
[63]Department of Biostatistics, University of Washington, Seattle, WA, USA
[64]University Medical Center Groningen, Department of Genetics, University of Groningen, Groningen, 9700 RB, The Netherlands
[65]Faculty of Medicine, University of Iceland, Reykjavik, Iceland
[66]Survey Research Center, Institute for Social Research, University of Michigan, Ann Arbor, MI, 48104, USA
[67]Institute of Translational Genomics, Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg, Germany
[68]Department of Cardiology, Division Heart & Lungs, University Medical Center Utrecht, Utrecht University, Utrecht, 3584CX, The Netherlands
[69]Central Diagnostics Laboratory, Division Laboratories, Pharmacy, and Biomedical genetics, University Medical Center Utrecht, Utrecht University, Utrecht, The Netherlands
[70]Barts Cardiovascular Research Unit, Barts and The London School of Medicine & Dentistry, Queen Mary University, London, EC1M 6BQ, UK
[71]Section of Investigative Medicine, Department of Medicine, Imperial College London, London, W12 0NN, UK
[72]Department of Life Sciences, Brunel University London, London, UB8 3PH, UK
[73]Department of Epidemiology and Biostatistics, Imperial College London, London, W2 1PG, UK
[74]Ealing Hospital, London North West Healthcare NHS Trust, Middlesex, UB1 3HW, UK
[75]Department of Epidemiology, University of North Carolina at Chapel Hill, Chapel Hill, NC, 27514, USA
[76]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
[77]Department of Internal Medicine, Erasmus University Medical Center, Rotterdam, 3015 GE, The Netherlands
[78]Department of Medicine, King Abdulaziz University, Jeddah, 21589, Saudi Arabia
[79]Amsterdam University Medical Centers, Department of Cardiology, University of Amsterdam, Amsterdam, The Netherlands
[80]Health Data Research UK and Institute of Health Informatics, University College London, London, UK
[81]Center for Precision Health Research, National Human Genome Research Institute, NIH, Bethesda, MD, 20892, USA
[82]

Center for Circulatory Health, University Medical Center Utrecht, Utrecht, 3508GA, The Netherlands
[83]Department of Clinical Biochemistry, Lillebaelt Hospital Vejle, Vejle, 7100, Denmark
[84]Institute of Regional Health Research, University of Southern Denmark, Odense, 5000, Denmark
[85]Department of Public Health and Primary Care, University of Cambridge, Cambridge, CB18RN, UK
[86]UK Dementia Research Institute, Imperial College London, London, UK
[87]Oulu University Hospital, Oulu, 90220, Finland
[88]Imperial College NIHR Biomedical Research Centre, London, UK
[89]Health Data Research UK, Imperial College London, London, UK
[90]Department of Nutrition and Dietetics, School of Health Science and Education, Harokopio University, Athens, 17671, Greece
[91]Department of Population Science and Experimental Medicine, Institute of Cardiovascular Science, University College London, London, UK
[92]Department of Nutrition, Harvard School of Public Health, Boston, MA, USA
[93]Institute of Reproductive and Developmental Biology, Imperial College London, London, W12 0NN, UK
[94]Department of Internal Medicine, Division of Endocrinology, Diabetology, Vascular Medicine, Nephrology, and Clinical Chemistry, University Hospital of Tübingen, Tübingen, Germany
[95]Division of Endocrinology, Diabetes and Metabolism, Cedars-Sinai Medical Center, Los Angeles, CA, 90048, USA
[96]Department of Biobank Research, Umeå University, Umeå, SE-901 87, Sweden
[97]Institute of Biomedicine and Biocenter of Oulu, Faculty of Medicine, Medical Research Center Oulu and Oulu University Hospital, Oulu, Finland
[98]Department of Gastroenterology and Metabolism, Poznan University of Medical Sciences, Poznan, 60-572, Poland
[99]Department of Population Medicine, Harvard Medical School, Harvard Pilgrim Health Care Institute, Boston, MA, USA
[100]Diabetes Unit, Department of Medicine, Massachusetts General Hospital, Boston, MA, USA
[101]Center for Clinical Research and Prevention, Bispebjerg and Frederiksberg Hospital, Frederiksberg, 2000, Denmark
[102]Department of Public Health, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, 2200, Denmark
[103]Faculty of Medicine, University of Aalborg, Aalborg, 9100, Denmark
[104]National Institute of Public Health, Southern Denmark University, Odense, 5000, Denmark
[105]Department of Public Health and Welfare, Finnish Institute for Health and Welfare, Helsinki, FI-00271, Finland
[106]PEDEGO Research Unit, MRC Oulu, Oulu University Hospital and University of Oulu, Oulu, Finland
[107]Department of Clinical and Molecular Medicine, Norwegian University of Science and Technology, Trondheim, Norway
[108]Children's Hospital, Helsinki University Hospital and University of Helsinki, Helsinki, Finland
[109]Echinos Medical Centre, Echinos, Greece
[110]University of Helsinki and Department of Medicine, Helsinki University Hospital, Helsinki, FI-00029, Finland
[111]Minerva Foundation Institute for Medical Research, Biomedicum 2U Helsinki, Helsinki, FI-00290, Finland
[112]Foundation for Research in Health Exercise and Nutrition, Kuopio Research Institute of Exercise Medicine, Kuopio, 70100, Finland
[113]Integrated Research and Treatment (IFB) Center Adiposity Diseases, University of Leipzig, Leipzig, 04103, Germany
[114]Medical Department III – Endocrinology, Nephrology, Rheumatology, University of Leipzig Medical Center, Leipzig, 04103, Germany
[115]Institute of Clinical Medicine, Internal Medicine, University of Eastern Finland, Kuopio, 70210, Finland
[116]Department of Medicine, Division of Bioinformatics and Personalized Medicine, University of Colorado Denver, Denver, CO, USA
[117]Laboratory of Epidemiology and Population Sciences, National Institute on Aging, National Institutes of Health, Baltimore, MD, 21224, USA
[118]Division of General Internal Medicine, Massachusetts General Hospital, Department of Medicine, Harvard Medical School, Boston, MA, USA
[119]Alberta Diabetes Institute IsletCore, University of Alberta, Edmonton, T6G 2E1, Canada
[120]Department of Pharmacology, University of Alberta, Edmonton, T6G 2E1, Canada
[121]Department of Medicine, Division of General Internal Medicine, Johns Hopkins University School of Medicine, Baltimore, MD, USA
[122]Welch Center for Prevention, Epidemiology, and Clinical Research, Johns Hopkins University, Baltimore, MD, USA
[123]Kuopio University Hospital, Kuopio, 70210, Finland
[124]Dipartimento di Scienze Biomediche, Università degli Studi di Sassari, Sassari, 07100, Italy
[125]Laboratory of Neurogenetics, National Institute on Aging, Bethesda, MD, 20892, USA
[126]Data Tecnica International LLC, Glen Echo, MD, 20812, USA
[127]Division of Epidemiology and Community Health, School of Public Health, University of Minnesota, Minneapolis, MN, 55455, USA
[128]South Ostobothnia Central Hospital, Seinajoki, 60220, Finland
[129]Center for Non-Communicable Diseases, Karachi, Pakistan
[130]Institute for Clinical Diabetology, German Diabetes Center, Leibniz Institute for Diabetes Research at Heinrich Heine University Düsseldorf, Düsseldorf, Germany
[131]

Division of Endocrinology and Diabetology, Medical Faculty, University Hospital Düsseldorf, Düsseldorf, Germany
[132]Department of Biostatistics and Epidemiology, University of Pennsylvania, 19104, USA
[133]Central Finland Central Hospital, Jyvaskyla, 40620, Finland
[134]Department of Clinical Physiology and Nuclear Medicine, Kuopio University Hospital, Kuopio, 70029, Finland
[135]Department of Twin Research and Genetic Epidemiology, King's College London, London, SE1 7EH, UK
[136]Institute of Genetic Epidemiology, Helmholtz Center Munich, German Research Center for Environmental Health, German Center for Diabetes Research (DZD e.V.), Neuherberg, Germany
[137]Department of Medicine, Yong Loo Lin School of Medicine, National University of Singapore, Singapore, 119228, Singapore
[138]Duke-NUS Medical School, Singapore, 169857, Singapore
[139]Department of Medicine, University of Leipzig, Leipzig, 04103, Germany
[140]Anogia Medical Centre, Anogia, Greece
[141]Folkhälsan Research Centre, Helsinki, Finland
[142]Department of Endocrinology, Helsinki University Central Hospital, Helsinki, Finland
[143]Institute for Molecular Medicine Finland FIMM, University of Helsinki, Helsinki, Finland
[144]Department of Clinical Sciences, Diabetes and Endocrinology, Lund University Diabetes Centre, Malmö, Sweden
[145]Department of Public Health, University of Helsinki, Helsinki, Finland
[146]Saudi Diabetes Research Group, King Abdulaziz University, Jeddah, 21589, Saudi Arabia
[147]Department of Public Health and Clinical Nutrition, University of Eastern Finland, Kuopio, 70210, Finland
[148]Department of Welfare, Children, Adolescents and Families Unit, National Institute for Health and Welfare, Oulu, Finland
[149]INSERM U1018, Centre de recherche en Épidémiologie et Santé des Populations (CESP), Villejuif, France
[150]Helmholtz Institute for Metabolic, Obesity and Vascular Research (HI-MAG), Helmholtz Zentrum München, University of Leipzig and University Hospital Leipzig, Leipzig, Germany
[151]Department of Epidemiology, German Institute of Human Nutrition Potsdam-Rehbrücke (DIfE), Nuthetal, 14558, Germany
[152]The Human Genetics Center and Institute of Molecular Medicine, University of Texas Health Science Center, Houston, Texas, 77030, USA
[153]Imperial College Healthcare NHS Trust, London, W12 0HS, UK
[154]Harvard School of Medicine, Boston, MA, USA
[155]Division of Genetics, Brigham and Women's Hospital and Harvard Medical School, Boston, MA, USA
[156]Singapore Eye Research Institute, Singapore National Eye Centre, Singapore, 169856, Singapore
[157]Ophthalmology & Visual Sciences Academic Clinical Program (Eye ACP), Duke-NUS Medical School, Singapore, 169857, Singapore
[158]Department of Ophthalmology, Yong Loo Lin School of Medicine, National University of Singapore, Singapore, 119228, Singapore
[159]Department of Vascular Surgery, Division of Surgical Specialties, University Medical Center Utrecht, Utrecht, 3584 CX, The Netherlands
[160]Princess Al-Jawhara Al-Brahim Centre of Excellence in Research of Hereditary Disorders (PACER-HD), King Abdulaziz University, Jeddah, Saudi Arabia
[161]Experimental Cardiology Laboratory, Division Heart and Lungs, University Medical Center Utrecht, Utrecht University, Utrecht, 3584 CX, The Netherlands
[162]Department of Epidemiology, Biostatistics and Occupational Health, McGill University, Montreal, Quebec, Canada
[163]CNR Institute of Clinical Physiology, Department of Clinical & Experimental Medicine, University of Pisa, Pisa, Italy
[164]Institute of Epidemiology II, Research Unit of Molecular Epidemiology, Helmholtz Zentrum München, Munich, Germany
[165]Faculty of Health Sciences, University of Southern Denmark, Odense, 5000, Denmark
[166]University of Exeter Medical School, University of Exeter, Exeter, EX2 5DW, UK
[167]Departments of Pediatrics and Genetics, Harvard Medical School, Boston, MA, 02115, USA
[168]Department of Medicine, Division of Cardiovascular Medicine, Stanford University School of Medicine, Stanford, CA, 94305, USA
[169]Stanford Cardiovascular Institute, Stanford University, Stanford, CA, 94305, USA
[170]Stanford Diabetes Research Center, Stanford University, Stanford, 94305, USA
[171]Oxford NIHR Biomedical Research Centre, Churchill Hospital, Oxford, OX3 7LE, UK
[172]Department of Public Health and Primary Care, Institute of Public Health, University of Cambridge, Cambridge, CB1 8RN, UK
[173]Pediatric Research Center, Department of Women & Child Health, University of Leipzig, Leipzig, Germany
[174]National Heart and Lung Institute, Imperial College London, London, W12 0NN, UK
[175]Institute of Biomedicine, School of Medicine, University of Eastern Finland, Kuopio, 70211, Finland
[176]Department of Medical Sciences, Molecular Epidemiology; EpiHealth, Uppsala University, Uppsala, 75185, Sweden
[177]The Big Data Institute, Li Ka Shing Centre for Health Information and Discovery, University of Oxford, Oxford, OX3 7BN, UK
[178]Department of Clinical Medicine, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, 2200, Denmark
[179]Department of Neurology, Wayne State University School of Medicine, Detroit, MI, USA
[180]Department of Epidemiology & Prevention, Division of Public Health Sciences, Wake Forest University, Winston-Salem, NC, 27157,

USA
[181]The Mindich Child Health and Development Institute, The Icahn School of Medicine at Mount Sinai, New York, NY, 10069, USA
[182]Department of Genetics, University of North Carolina, Chapel Hill, NC, 27599, USA
[183]Usher Institute of Population Health Sciences and Informatics, University of Edinburgh, Edinburgh, EH16 4UX, UK
[184]Aberdeen Biomedical Imaging Centre, University of Aberdeen, Foresterhill Health Campus, Aberdeen, AB25 2ZD, UK
[185]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
[186]Division of Population Health and Genomics, School of Medicine, University of Dundee, Dundee, DD2 4BF, UK
[187]Laboratory of Clinical Chemistry and Hematology, University Medical Center Utrecht, Utrecht, 3584 CX, The Netherlands
[188]Faculty of Medicine, University of Split, Split, Croatia
[189]Departments of Epidemiology, Health Systems and Population Health, University of Washington, Seattle, Seattle, WA, USA
[190]Division of Cardiovascular Medicine, Brigham and Women's Hospital, Boston, MA, 02115, USA
[191]Department of Public Health & Clinical Medicine, Section for Family Medicine, Umeå University, Umeå, SE-901 85, Sweden
[192]Department of Medicine, McGill University, Montreal, Quebec, H4A 3J1, Canada
[193]Department of Human Genetics, McGill University, Montreal, Quebec, H3A 1B1, Canada
[194]Alzheimer Scotland Dementia Research Centre, University of Edinburgh, Edinburgh, EH8 9JZ, UK
[195]Institute of Cellular Medicine, The Medical School, Newcastle University, Newcastle, NE2 4HH, UK
[196]Department of Physiology and Biophysics, University of Mississippi Medical Center, Jackson, MS, USA
[197]Technical University of Munich (TUM) and Klinikum Rechts der Isar, TUM School of Medicine, Munich, Germany
[198]Centre for Genetics and Genomics Versus Arthritis, Centre for Musculoskeletal Research, University of Manchester, Manchester, UK
[199]Diabetes Unit and Center for Genomic Medicine, Massachusetts General Hospital, Boston, MA, USA
[200]Programs in Metabolism and Medical & Population Genetics, Broad Institute, Cambridge, MA, USA
[201]Current address: Genentech, South San Francisco, CA, 94080, USA
[202]Division of Endocrinology, Department of Pediatrics, Stanford School of Medicine, Stanford, CA, USA
[203]Exeter Centre of Excellence in Diabetes (EXCEED), University of Exeter Medical School, Exeter, UK

* Equal contributors

v1 **First published:** 20 Oct 2023, **8**:483
https://doi.org/10.12688/wellcomeopenres.18754.1
Latest published: 20 Oct 2023, **8**:483
https://doi.org/10.12688/wellcomeopenres.18754.1

Abstract

Background

Genome-wide association studies for glycemic traits have identified hundreds of loci associated with these biomarkers of glucose homeostasis. Despite this success, the challenge remains to link variant associations to genes, and underlying biological pathways.

Methods

To identify coding variant associations which may pinpoint effector genes at both novel and previously established genome-wide association loci, we performed meta-analyses of exome-array studies for four glycemic traits: glycated hemoglobin (HbA1c, up to 144,060 participants), fasting glucose (FG, up to 129,665 participants), fasting insulin (FI, up to 104,140) and 2hr glucose post-oral glucose challenge

Open Peer Review

Approval Status ✔ ✔

	1	2
version 1 20 Oct 2023	✔ view	✔ view

1. **Toshimasa Yamauchi**, University of Tokyo Graduate School of Medicine, Tokyo, Japan

2. **Eiji Kutoh**, Gyoda General Hospital, Saitama, Japan
Higashitotsuka Memorial Hospital, Yokohama, Japan

Any reports and responses or comments on the article can be found at the end of the article.

(2hGlu, up to 57,878). In addition, we performed network and pathway analyses.

Results

Single-variant and gene-based association analyses identified coding variant associations at more than 60 genes, which when combined with other datasets may be useful to nominate effector genes. Network and pathway analyses identified pathways related to insulin secretion, zinc transport and fatty acid metabolism. HbA1c associations were strongly enriched in pathways related to blood cell biology.

Conclusions

Our results provided novel glycemic trait associations and highlighted pathways implicated in glycemic regulation. Exome-array summary statistic results are being made available to the scientific community to enable further discoveries.

Keywords
exome chip, glycaemic traits, genetic discovery, effector genes, summary statistics resources

Corresponding author: Inês Barroso (ines.barroso@exeter.ac.uk)

Author roles: Willems SM: Formal Analysis, Methodology, Project Administration, Writing – Original Draft Preparation, Writing – Review & Editing; **Ng NHJ**: Formal Analysis, Methodology, Project Administration, Visualization, Writing – Original Draft Preparation, Writing – Review & Editing; **Fernandez J**: Formal Analysis, Methodology, Visualization, Writing – Original Draft Preparation, Writing – Review & Editing; **Fine RS**: Formal Analysis, Methodology, Visualization, Writing – Original Draft Preparation, Writing – Review & Editing; **Wheeler E**: Formal Analysis, Writing – Original Draft Preparation, Writing – Review & Editing; **Wessel J**: Formal Analysis, Project Administration, Writing – Original Draft Preparation, Writing – Review & Editing; **Kitajima H**: Formal Analysis, Writing – Review & Editing; **Marenne G**: Formal Analysis, Writing – Review & Editing; **Sim X**: Formal Analysis, Writing – Review & Editing; **Yaghootkar H**: Formal Analysis, Writing – Review & Editing; **Wang S**: Formal Analysis, Writing – Review & Editing; **Chen S**: Formal Analysis, Writing – Review & Editing; **Chen Y**: Formal Analysis, Writing – Review & Editing; **Chen YDI**: Formal Analysis, Resources, Writing – Review & Editing; **Grarup N**: Formal Analysis, Writing – Review & Editing; **Li-Gao R**: Formal Analysis, Writing – Review & Editing; **Varga TV**: Formal Analysis, Writing – Review & Editing; **Asimit JL**: Formal Analysis, Writing – Review & Editing; **Feng S**: Formal Analysis, Methodology, Software, Writing – Review & Editing; **Strawbridge RJ**: Formal Analysis, Writing – Review & Editing; **Kleinbrink EL**: Formal Analysis, Resources, Writing – Review & Editing; **Ahluwalia TS**: Formal Analysis, Writing – Review & Editing; **An P**: Formal Analysis, Writing – Review & Editing; **Appel EV**: Formal Analysis, Resources, Writing – Review & Editing; **Arking DE**: Formal Analysis, Writing – Review & Editing; **Auvinen J**: Formal Analysis, Writing – Review & Editing; **Bielak LF**: Formal Analysis, Resources, Writing – Review & Editing; **Bihlmeyer NA**: Formal Analysis, Writing – Review & Editing; **Bork-Jensen J**: Formal Analysis, Writing – Review & Editing; **Brody JA**: Formal Analysis, Resources, Writing – Review & Editing; **Campbell A**: Formal Analysis, Resources, Writing – Review & Editing; **Chu AY**: Formal Analysis, Resources, Writing – Review & Editing; **Davies G**: Formal Analysis, Resources, Writing – Review & Editing; **Demirkan A**: Formal Analysis, Resources, Writing – Review & Editing; **Floyd JS**: Formal Analysis, Writing – Review & Editing; **Giulianini F**: Formal Analysis, Resources, Writing – Review & Editing; **Guo X**: Formal Analysis, Writing – Review & Editing; **Gustafsson S**: Formal Analysis, Writing – Review & Editing; **Jackson AU**: Formal Analysis, Writing – Review & Editing; **Jakobsdottir J**: Formal Analysis, Writing – Review & Editing; **Järvelin MR**: Formal Analysis, Resources, Writing – Review & Editing; **Jensen RA**: Formal Analysis, Writing – Review & Editing; **Kanoni S**: Formal Analysis, Writing – Review & Editing; **Keinanen-Kiukaanniemi S**: Formal Analysis, Resources, Writing – Review & Editing; **Li M**: Formal Analysis, Resources, Writing – Review & Editing; **Lu Y**: Formal Analysis, Resources, Writing – Review & Editing; **Luan J**: Formal Analysis, Writing – Review & Editing; **Manning AK**: Formal Analysis, Writing – Review & Editing; **Marten J**: Formal Analysis, Writing – Review & Editing; **Meidtner K**: Formal Analysis, Writing – Review & Editing; **Mook-Kanamori DO**: Formal Analysis, Writing – Review & Editing; **Muka T**: Formal Analysis, Writing – Review & Editing; **Pistis G**: Formal Analysis, Writing – Review & Editing; **Prins B**: Formal Analysis, Writing – Review & Editing; **Rice KM**: Formal Analysis, Writing – Review & Editing; **Sanna S**: Formal Analysis, Resources, Writing – Review & Editing; **Smith AV**: Formal Analysis, Writing

– Review & Editing; **Smith JA**: Formal Analysis, Resources, Writing – Review & Editing; **Southam L**: Formal Analysis, Resources, Writing – Review & Editing; **Stringham HM**: Formal Analysis, Resources, Writing – Review & Editing; **Tragante V**: Formal Analysis, Writing – Review & Editing; **van der Laan SW**: Formal Analysis, Resources, Writing – Review & Editing; **Warren HR**: Formal Analysis, Writing – Review & Editing; **Yao J**: Formal Analysis, Writing – Review & Editing; **Yiorkas AM**: Formal Analysis, Writing – Review & Editing; **Zhang W**: Formal Analysis, Writing – Review & Editing; **Zhao W**: Formal Analysis, Resources, Writing – Review & Editing; **Graff M**: Formal Analysis, Writing – Review & Editing; **Highland HM**: Formal Analysis, Writing – Review & Editing; **Justice AE**: Formal Analysis, Writing – Review & Editing; **Marouli E**: Formal Analysis, Writing – Review & Editing; **Medina-Gomez C**: Formal Analysis, Writing – Review & Editing; **Afaq S**: Resources, Writing – Review & Editing; **Alhejily WA**: Resources, Writing – Review & Editing; **Amin N**: Resources, Writing – Review & Editing; **Asselbergs FW**: Supervision, Writing – Review & Editing; **Bonnycastle LL**: Resources, Writing – Review & Editing; **Bots ML**: Supervision, Writing – Review & Editing; **Brandslund I**: Resources, Writing – Review & Editing; **Chen J**: Formal Analysis, Writing – Review & Editing; **Danesh J**: Supervision, Writing – Review & Editing; **de Mutsert R**: Resources, Writing – Review & Editing; **Dehghan A**: Resources, Writing – Review & Editing; **Ebeling T**: Resources, Writing – Review & Editing; **Elliott P**: Resources, Writing – Review & Editing; **Farmaki AE**: Resources, Writing – Review & Editing; **Faul JD**: Resources, Writing – Review & Editing; **Franks PW**: Supervision, Writing – Review & Editing; **Franks S**: Resources, Writing – Review & Editing; **Fritsche A**: Resources, Writing – Review & Editing; **Gjesing AP**: Resources, Writing – Review & Editing; **Goodarzi MO**: Resources, Writing – Review & Editing; **Gudnason V**: Resources, Writing – Review & Editing; **Hallmans G**: Resources, Writing – Review & Editing; **Harris TB**: Resources, Writing – Review & Editing; **Herzig KH**: Resources, Writing – Review & Editing; **Hivert MF**: Resources, Writing – Review & Editing; **Jørgensen T**: Resources, Supervision, Writing – Review & Editing; **Jørgensen ME**: Resources, Writing – Review & Editing; **Jousilahti P**: Resources, Writing – Review & Editing; **Kajantie E**: Resources, Writing – Review & Editing; **Karaleftheri M**: Resources, Writing – Review & Editing; **Kardia SLR**: Resources, Writing – Review & Editing; **Kinnunen L**: Resources, Writing – Review & Editing; **Koistinen HA**: Resources, Writing – Review & Editing; **Komulainen P**: Resources, Writing – Review & Editing; **Kovacs P**: Resources, Writing – Review & Editing; **Kuusisto J**: Resources, Writing – Review & Editing; **Laakso M**: Resources, Writing – Review & Editing; **Lange LA**: Resources, Writing – Review & Editing; **Launer LJ**: Resources, Writing – Review & Editing; **Leong A**: Formal Analysis, Writing – Review & Editing; **Lindström J**: Resources, Writing – Review & Editing; **Manning Fox JE**: Resources, Writing – Review & Editing; **Männistö S**: Resources, Writing – Review & Editing; **Maruthur NM**: Resources, Writing – Review & Editing; **Moilanen L**: Resources, Writing – Review & Editing; **Mulas A**: Resources, Writing – Review & Editing; **Nalls MA**: Resources, Writing – Review & Editing; **Neville M**: Resources, Writing – Review & Editing; **Pankow JS**: Resources, Writing – Review & Editing; **Pattie A**: Resources, Writing – Review & Editing; **Petersen ERB**: Resources, Writing – Review & Editing; **Puolijoki H**: Resources, Writing – Review & Editing; **Rasheed A**: Resources, Writing – Review & Editing; **Redmond P**: Resources, Writing – Review & Editing; **Renström F**: Resources, Writing – Review & Editing; **Roden M**: Resources, Writing – Review & Editing; **Saleheen D**: Resources, Writing – Review & Editing; **Saltevo J**: Resources, Writing – Review & Editing; **Savonen K**: Resources, Writing – Review & Editing; **Sebert S**: Resources, Writing – Review & Editing; **Skaaby T**: Resources, Writing – Review & Editing; **Small KS**: Resources, Writing – Review & Editing; **Stančáková A**: Resources, Writing – Review & Editing; **Stokholm J**: Resources, Writing – Review & Editing; **Strauch K**: Resources, Writing – Review & Editing; **Tai ES**: Resources, Writing – Review & Editing; **Taylor KD**: Resources, Writing – Review & Editing; **Thuesen BH**: Resources, Writing – Review & Editing; **Tönjes A**: Resources, Writing – Review & Editing; **Tsafantakis E**: Resources, Writing – Review & Editing; **Tuomi T**: Resources, Writing – Review & Editing; **Tuomilehto J**: Resources, Writing – Review & Editing; **Uusitupa M**: Resources, Writing – Review & Editing; **Vääräsmäki M**: Resources, Writing – Review & Editing; **Vaartjes I**: Resources, Writing – Review & Editing; **Zoledziewska M**: Resources, Writing – Review & Editing; **Abecasis G**: Supervision, Writing – Review & Editing; **Balkau B**: Supervision, Writing – Review & Editing; **Bisgaard H**: Resources, Supervision, Writing – Review & Editing; **Blakemore AI**: Supervision, Writing – Review & Editing; **Blüher M**: Resources, Supervision, Writing – Review & Editing; **Boeing H**: Resources, Supervision, Writing – Review & Editing; **Boerwinkle E**: Resources, Supervision, Writing – Review & Editing; **Bønnelykke K**: Resources, Supervision, Writing – Review & Editing; **Bottinger EP**: Resources, Supervision, Writing – Review & Editing; **Caulfield MJ**: Supervision, Writing – Review & Editing; **Chambers JC**: Resources, Supervision, Writing – Review & Editing; **Chasman DI**: Formal Analysis, Resources, Supervision, Writing – Review & Editing; **Cheng CY**: Resources, Supervision, Writing – Review & Editing; **Collins FS**: Supervision, Writing – Review & Editing; **Coresh J**: Supervision, Writing – Review & Editing; **Cucca F**: Supervision, Writing – Review & Editing; **de Borst GJ**: Resources, Supervision, Writing – Review & Editing; **Deary IJ**: Resources, Supervision, Writing – Review & Editing; **Dedoussis G**: Supervision, Writing – Review & Editing; **Deloukas P**: Supervision, Writing – Review & Editing; **den Ruijter HM**: Supervision, Writing – Review & Editing; **Dupuis J**: Formal Analysis, Supervision, Writing – Review & Editing; **Evans MK**: Resources, Supervision, Writing – Review & Editing; **Ferrannini E**: Resources, Supervision, Writing – Review & Editing; **Franco OH**: Supervision, Writing – Review & Editing; **Grallert H**: Supervision, Writing – Review & Editing; **Hansen T**: Supervision, Writing – Review & Editing; **Hattersley AT**: Supervision, Writing – Review & Editing; **Hayward C**: Formal Analysis, Resources, Supervision, Writing – Review & Editing; **Hirschhorn JN**: Formal Analysis, Supervision, Writing – Review & Editing; **Ikram A**: Supervision, Writing – Review & Editing; **Ingelsson E**: Resources, Supervision, Writing – Review & Editing; **Karpe F**: Resources, Supervision, Writing – Review & Editing; **Kaw KT**: Supervision, Writing – Review & Editing; **Kiess W**: Resources, Supervision, Writing – Review & Editing; **Kooner JS**: Resources, Supervision, Writing – Review & Editing; **Körner A**: Resources, Supervision, Writing – Review & Editing; **Lakka T**: Supervision, Writing – Review & Editing; **Langenberg C**: Supervision, Writing – Review & Editing; **Lind L**: Resources, Supervision, Writing – Review & Editing; **Lindgren CM**: Resources, Supervision, Writing – Review & Editing; **Linneberg A**: Resources, Supervision, Writing – Review & Editing; **Lipovich L**: Formal Analysis, Supervision, Writing – Review & Editing; **Liu CT**: Formal Analysis, Supervision, Writing – Review & Editing; **Liu J**: Formal Analysis, Writing – Review & Editing; **Liu Y**: Resources, Supervision, Writing – Review & Editing; **Loos RJF**: Resources, Supervision, Writing – Review & Editing; **MacDonald PE**: Supervision, Writing – Review & Editing; **Mohlke KL**: Supervision, Writing – Review & Editing; **Morris AD**: Resources, Supervision, Writing – Review & Editing; **Munroe PB**: Supervision, Writing – Review & Editing; **Murray A**: Supervision, Writing – Review & Editing; **Padmanabhan S**: Supervision, Writing – Review & Editing; **Palmer CNA**: Resources, Supervision, Writing – Review & Editing; **Pasterkamp G**: Supervision, Writing – Review & Editing; **Pedersen O**: Supervision, Writing – Review & Editing; **Peyser PA**: Resources, Supervision, Writing – Review & Editing; **Polasek O**: Resources, Supervision, Writing –

Review & Editing; **Porteous D**: Supervision, Writing – Review & Editing; **Province MA**: Supervision, Writing – Review & Editing; **Psaty BM**: Resources, Supervision, Writing – Review & Editing; **Rauramaa R**: Supervision, Writing – Review & Editing; **Ridker PM**: Supervision, Writing – Review & Editing; **Rolandsson O**: Supervision, Writing – Review & Editing; **Rorsman P**: Supervision, Writing – Review & Editing; **Rosendaal FR**: Supervision, Writing – Review & Editing; **Rudan I**: Supervision, Writing – Review & Editing; **Salomaa V**: Resources, Supervision, Writing – Review & Editing; **Schulze MB**: Resources, Supervision, Writing – Review & Editing; **Sladek R**: Supervision, Writing – Review & Editing; **Smith BH**: Supervision, Writing – Review & Editing; **Spector TD**: Resources, Supervision, Writing – Review & Editing; **Starr JM**: Resources, Supervision, Writing – Review & Editing; **Stumvoll M**: Supervision, Writing – Review & Editing; **van Duijn CM**: Supervision, Writing – Review & Editing; **Walker M**: Supervision, Writing – Review & Editing; **Wareham NJ**: Supervision, Writing – Review & Editing; **Weir DR**: Resources, Supervision, Writing – Review & Editing; **Wilson JG**: Resources, Supervision, Writing – Review & Editing; **Wong TY**: Resources, Supervision, Writing – Review & Editing; **Zeggini E**: Supervision, Writing – Review & Editing; **Zonderman AB**: Resources, Supervision, Writing – Review & Editing; **Rotter JI**: Resources, Supervision, Writing – Review & Editing; **Morris AP**: Resources, Supervision, Writing – Review & Editing; **Boehnke M**: Supervision, Writing – Original Draft Preparation, Writing – Review & Editing; **Florez JC**: Supervision, Writing – Original Draft Preparation, Writing – Review & Editing; **McCarthy MI**: Resources, Supervision, Writing – Original Draft Preparation, Writing – Review & Editing; **Meigs JB**: Resources, Supervision, Writing – Original Draft Preparation, Writing – Review & Editing; **Mahajan A**: Formal Analysis, Resources, Writing – Original Draft Preparation, Writing – Review & Editing; **Scott RA**: Conceptualization, Formal Analysis, Methodology, Project Administration, Supervision, Writing – Original Draft Preparation, Writing – Review & Editing; **Gloyn AL**: Conceptualization, Funding Acquisition, Project Administration, Supervision, Writing – Original Draft Preparation, Writing – Review & Editing; **Barroso I**: Conceptualization, Funding Acquisition, Methodology, Project Administration, Supervision, Writing – Original Draft Preparation, Writing – Review & Editing

Competing interests: Rebecca S. Fine: Rebecca S. Fine is currently employed by Vertex Pharmaceuticals Incorporated. Audrey Y Chu: Currently employed by GlaxoSmithkline. Dennis O. Mook-Kanamori: Dennis Mook-Kanamori is working as a part-time clinical research consultant for Metabolon, Inc. Paul W. Franks: PWF has been a paid consultant for Eli Lilly and Sanofi Aventis and has received research support from several pharmaceutical companies as part of a European Union Innovative Medicines Initiative (IMI) project. Mike A. Nalls: Dr. Mike A. Nalls is supported by a consulting contract between Data Tecnica International LLC and the National Institute on Aging (NIA), National Institutes of Health (NIH), Bethesda, MD, USA. Dr. Nalls also consults for Illumina Inc., the Michael J. Fox Foundation, and the University of California Healthcare. Mark J. Caulfield: MJC is Chief Scientist for Genomics England, a UK government company. Joel N. Hirschhorn: JHN is on the scientific advisory board of Camp4 Therapeutics. Erik Ingelsson: Erik Ingelsson is now an employee of GlaxoSmithKline. Anubha Mahajan: Anubha Mahajan is an employee of Genentech since January 2020, and a holder of Roche stock. Mark I McCarthy: The views expressed in this article are those of the author(s) and not necessarily those of the NHS, the NIHR, or the Department of Health. MMcC has served on advisory panels for Pfizer, NovoNordisk and Zoe Global, has received honoraria from Merck, Pfizer, Novo Nordisk and Eli Lilly, and research funding from Abbvie, Astra Zeneca, Boehringer Ingelheim, Eli Lilly, Janssen, Merck, NovoNordisk, Pfizer, Roche, Sanofi Aventis, Servier, and Takeda. As of June 2019, MMcC is an employee of Genentech, and a holder of Roche stock. Inês Barroso: IB and spouse declare stock ownership in GlaxoSmithkline and Incyte Ltd. James B. Meigs: JBM serves as an Academic Associate for Quest Diagnostics R&D Bruce M Psaty serves on the Steering Committee of the Yale Open Data Access Project funded by Johnson & Johnson. Dr. Sander W. van der Laan has received Roche funding for unrelated work. Matthias Blüher received honoraria as a consultant and speaker from Amgen, AstraZeneca, Bayer, Boehringer-Ingelheim, Lilly, Novo Nordisk, Novartis, Pfizer and Sanofi. Vinicius Tragante: VT became an employee of deCODE genetics/Amgen Inc. after the conclusion of this work Dr Franco is employed by ErasmusAGE, a center for aging research across the life course funded by Nestlé Nutrition (Nestec Ltd.) and Metagenics.

Grant information: This work was supported by Wellcome [grant numbers 095101 and 200837 to Anna Gloyn]; Inês Barroso is funded by Wellcome (WT206194) and this work was partly supported by "Expanding excellence in England" award from Research England; The Fenland Study is funded by Wellcome Trust and the Medical Research Council (MC_U106179471); Genotyping of the Generation Scotland GS:SFHS samples was funded by the Wellcome Trust (Wellcome Trust Strategic Award "STratifying Resilience and Depression Longitudinally" (STRADL) Reference 104036/Z/14/Z) and the Medical Research Council UK. 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; 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); For 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 LOLIPOP study is supported by the Wellcome Trust (084723/Z/08/Z), 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 NIHR (RP-PG-0407-10371), European Union FP7 (EpiMigrant, 279143) and Action on Hearing Loss (G51). Mark I McCarthy was funded by Wellcome (grant numbers 090532, 098381, 106130, 203141 and 212259) and NIH U01-DK105535; Genotyping and analysis of PIVUS and ULSAM were funded by the Wellcome Trust under awards WT064890, WT090532 and WT098017. PIVUS and ULSAM are supported by the Swedish Research Council, Swedish Heart-Lung Foundation, Swedish Diabetes Foundation and Uppsala University; TwinsUK is funded by the Wellcome Trust, Medical Research Council, European Union, Chronic Disease Research Foundation (CDRF), and the National Institute for Health Research (NIHR)-funded BioResource, Clinical Research Facility and Biomedical Research Centre based at Guy's and St Thomas' NHS Foundation Trust in partnership with King's College London; 1958 British Birth cohort: analysis was supported by BHF programme grant (Deloukas) RG/14/5/30893; AGES 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 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). 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; Athero-Express Biobank Study: Dr. Sander W. van der Laan is funded through EU H2020 TO_AITION (grant number: 848146). We are thankful for the support of the Netherlands CardioVascular Research Initiative of the Netherlands Heart Foundation (CVON 2011/B019 and CVON 2017-20: Generating the best evidence-based pharmaceutical targets for atherosclerosis [GENIUS I&II]), the ERA-CVD program ‘druggable-MI-targets’ (grant number: 01KL1802), and the Leducq Fondation ‘PlaqOmics’; BioMe: The Mount Sinai BioMe Biobank is supported by The Andrea and Charles Bronfman Philanthropies; Caroline Hayward is supported by an MRC University Unit Programme Grant “QTL in Health and Disease” (U. MC_UU_00007/10); CHS: this CHS research was supported by NHLBI contracts HHSN268201200036C, HHSN268200800007C, HHSN268201800001C, N01HC55222, N01HC85079, N01HC85080, N01HC85081, N01HC85082, N01HC85083, N01HC85086 and 75N92021D00006; 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). 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; 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; CROATIA_Korcula: Exome array genotyping was funded by UK’s Medical Research Council; DIABNORD: the current study was funded by Novo Nordisk, the Swedish Research Council, Påhlssons Foundation, the Swedish Heart Lung Foundation, and the Skåne Regional Health Authority (all to PWF); 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; 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); EFSOCH: this paper presents independent research supported by the National Institute for Health Research (NIHR) Exeter Clinical Research Facility; 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); EPIC-Potsdam: the study was supported in part by a grant from the German Federal Ministry of Education and Research (BMBF) and the State of Brandenburg to the German Center for Diabetes Research (DZD e.V.) (82DZD00302). 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); 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); 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 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; The FIA3 study was supported in part by a grant from the Swedish Heart-Lund Foundation (to PWF); 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; Folkert Asselbergs is supported by UCL Hospitals NIHR Biomedical Research Centre; Framingham Heart Study: 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, UM1

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; 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); Support for GENOA was provided by the National Heart, Lung and Blood Institute (HL054464; HL054481; HL087660; HL119443; HL086694) of the National Institutes of Health; GIANT: Anne E Justice (AEJ) is funded under NIH 5K99HL130580-02; GLACIER: the current study was funded by Novo Nordisk, the Swedish Research Council, Påhlssons Foundation, the Swedish Heart Lung Foundation, and the Skåne Regional Health Authority (all to PWF); 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; 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); 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); 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; 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 is supported by funding from NHLBI training grant T32 HL007055; Hidetoshi Kitajima was funded by Manpei Suzuki Diabetes Foundation Grant-in-Aid for the young scientists working abroad; Inter99: 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; Funding for the InterAct project was provided by the EU FP6 programme (grant number LSHM_CT_2006_037197); Jennifer L. Asimit: Medical Research Council Methodology Research Fellowship (MR/K021486/1); 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; Joel N. Hirschhorn: NIH R01DK075787; 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; 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); Lothian Birth Cohort 1921 and Lothian Birth Cohort 1936: 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; MESA and the MESA SHARe projects 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 75N92020D00001, HHSN268201500003I, N01-HC-95159, 75N92020D00005, N01-HC-95160, 75N92020D00002, N01-HC-95161, 75N92020D00003, N01-HC-95162, 75N92020D00006, N01-HC-95163, 75N92020D00004, N01-HC-95164, 75N92020D00007, N01-HC-95165, N01-HC-95166, N01-HC-95167, N01-HC-95168, N01-HC-95169, UL1-TR-000040, UL1-TR-001079, and UL1-TR-001420, UL1TR001881, DK063491, and R01HL105756. Funding for SHARe genotyping was provided by NHLBI Contract N02-HL-64278; 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; Natasha H J Ng (NHJN) is supported by the National Science Scholarship from the Agency for Science, Technology and Research (A*STAR) in Singapore; 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); 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; The Botnia and The PPP-Botnia studies have been financially supported by grants from Folkhalsan Research Foundation, the Sigrid Juselius Foundation, The Academy of Finland and university of Helsinki (grants no. 263401, 267882, 312063, 312072 and 336826), the European Research Council under the European Union's Seventh Framework Programme (FP7/2007-2013) / ERC grant agreement n° 269045, 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 Heath Care Center and Hospital in Jakobstad and Health Care Centers in Vasa, Närpes and Korsholm; PROMIS: Dr. Saleheen has received grants from the National Heart, Lung and Blood Institute, Fogarty International, Pfizer, Regeneron, Eli Lilly, and Genentech; Robert Sladek 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 is supported by NHGRI F31 HG009850; The RISC study was supported by European Union grant QLG1-CT-2001-01252 and AstraZeneca. The initial genotyping of the RISC samples was funded by Merck & Co Inc.; Rona J Strawbridge is supported by a UKRI Innovation at HDR and University of Glasgow LKAS Fellowship; 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 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)); SardiNIA: 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 cRP3-154; 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]; Sorbs: 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); The UK Household Longitudinal Study is funded by the Economic and Social Research Council; The Vejle Diabetes Biobank was supported by The Danish Research Council for Independent Research; 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.
The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

How to cite this article: Willems SM, Ng NHJ, Fernandez J *et al.* **Large-scale exome array summary statistics resources for glycemic traits to aid effector gene prioritization [version 1; peer review: 2 approved]** Wellcome Open Research 2023, **8**:483 https://doi.org/10.12688/wellcomeopenres.18754.1

First published: 20 Oct 2023, **8**:483 https://doi.org/10.12688/wellcomeopenres.18754.1

Introduction
Genome-wide association studies (GWAS) have identified hundreds of loci associated with glycemic traits and type 2 diabetes (T2D) risk[1–3]. Despite this tremendous success, the challenge remains to link the often lead non-coding variants with effector genes and mechanism of action. To complement these approaches, exome array studies[4,5] and more recently, whole-exome sequencing approaches have focused on coding variant associations[6–9]. These can be helpful to pinpoint potential effector genes for downstream functional studies. Here, we provide exome-array GWAS meta-analysis results for glycated hemoglobin (HbA1c, up to 144,060 participants), fasting glucose (FG, up to 129,665 participants), fasting insulin (FI, up to 104,140) and 2hr glucose post-oral glucose challenge (2hGlu, up to 57,878). Most of the data are from self-reported and genetically clustered European ancestry individuals (85%), with the remaining participants being of African American (6%), South Asian (5%), East Asian (2%) and Hispanic ancestry (2%). We identify single coding variant and gene-based associations to prioritize likely effector genes, and additionally perform pathway analyses to highlight relevant gene sets regulating each glycemic trait. Summary statistics from these analyses are publicly available through our website (www.magicinvestigators.org), as well as through the GWAS catalog (https://www.ebi.ac.uk/gwas/summary-statistics, study accessions GCST90256400 - GCST90256420)[10].

Methods
Study design, cohorts, phenotypes and genotypes
MAGIC (Meta-Analysis of Glucose and Insulin-related traits Consortium) was established to focus on the genetic analysis of glycemic traits in individuals without diabetes. In this MAGIC effort, individuals without diabetes of self-reported and genetically clustered European (85%), African American (6%), South Asian (5%), East Asian (2%) and Hispanic (2%) ancestry from up to 64 cohorts participated. Sample sizes were up to 144,060 for HbA1c, 129,665 for FG, 104,140 for FI and 57,878 for 2hGlu. Participating cohorts and their characteristics are detailed in Supplementary Table S1[11]. Each cohort obtained ethical approval and written informed consent.

Phenotypes
Studied outcomes were FG (mmol/L), Ln-transformed FI (pmol/L), 2hGlu (mmol/L) and HbA1c (% of hemoglobin). Glycemic measurements are described in detail for each contributing cohort in Supplementary Table S1[11]. Individuals with diagnosed or treated diabetes, or those with diabetes based on FG (≥7 mmol/L), 2hGlu (≥11.1 mmol/L) and/or HbA1c (≥6.5%) were excluded from analyses.

Genotyping and QC
The Illumina HumanExome BeadChip is a genotyping array containing variants that have been observed in sequencing data of ~12,000 individuals. Non-synonymous variants seen at least three times across at least two datasets were included on the exome chip. More lenient criteria were used for splice and nonsense variants. Besides the core content of protein-altering variants, the exome chip contains additional variants including common variants identified in GWAS, ancestry informative markers, mitochondrial variants, randomly selected synonymous variants, HLA tag variants and Y chromosome variants. In this study we analyzed association with glycemic traits of 247,470 autosomal and X chromosome variants present on the exome chip. Genotype calling and quality control were performed following protocols developed by the UK Exome Chip or CHARGE consortium[12]. The exact genotyping array, calling algorithm and QC procedure used by each cohort are depicted in Supplementary Table S1[11].

Annotation and functional prediction of variants
Annotation of the exome chip variants was performed using the Ensembl Variant Effect Predictor v78 with plugin dbNSFP v2.9 to add *in silico* functional prediction from Polyphen HumDiv, Polyphen HumVar, LRT, Mutation Taster and SIFT (ensembl66 version)[13,14].

Statistical analyses
Single variant analyses. Individual cohorts ran linear mixed models using the raremetalworker (v 4.13.2) or rvtests (v20140723) software (Supplementary Table S1[11]). For each glycemic outcome, analyses were performed using an additive model for the raw and the inverse normal transformed trait. In the manuscript and in all tables and figures effect estimates and standard errors are for the raw trait, while the p-values are from the inverse normal transformed trait analyses. Analyses were adjusted for age, sex, BMI, study-specific number of PCs and other study-specific covariates (Supplementary Table S1[11]). Raremetal (v4.13.7 or higher) was used to combine results within and across ancestries by fixed-effect meta-analyses. Variants with $P < 10^{-4}$ for deviation from Hardy-Weinberg equilibrium or with call rate <0.99 in individual cohorts were excluded from meta-analyses. In single variant analyses, the threshold for significance was $P < 2.2\times10^{-7}$ for coding variants (stop-gained, stop lost, frameshift, splice donor, splice acceptor, initiator codon, missense, in-frame indel and splice region variants). This *P*-value threshold was based on a Bonferroni correction weighted by the enrichment for complex trait associations among the functional annotation categories[15,16]. We performed so called distance-based clumping; significant association signals located more than 500 kb apart were considered to represent distinct loci. Significantly associated variants located more than 500 kb from any variant already found to be associated in published large-scale glycemic trait and T2D GWAS analyses[1,3,17,18] were considered novel glycemic trait associations. Gene-based and single-variant analyses results presented in the paper are for the meta-analyses of all ancestries combined, unless mentioned otherwise.

Gene-based analyses. Raremetal (v4.13.7 or higher) was used to perform gene-based burden and sequence kernel association (SKAT) tests. For both burden and SKAT tests, two *in silico* masks for inclusion of variants in the test were used: NSstrict and NSbroad. The NSstrict mask includes predicted protein truncating variants (PTVs, splice donor, splice acceptor, stop gained, frameshift, stop lost or initiator codon variant) OR variants that are missense and predicted to be damaging by five prediction algorithms (SIFT, Polyphen HumDiv, Polyphen HumVar, LRT, MutationTaster). The NSbroad mask additionally includes missense variants predicted to be damaging by at least one of the five prediction algorithms AND that have a MAF <1% in each ancestry group. These MAFs were derived from our single variant HbA1c meta-analyses results (N up to 144,060). Gene-based analyses were performed on genes containing at least two variants fulfilling the mask criteria. The *P*-value threshold for significance in gene-based analyses was 2.5×10^{-6} (Bonferroni correction for 20,000 genes).

GeneMANIA network analysis

For network analyses, we used GeneMANIA (v3.5.1), a network approach that searches many large, publicly available biological datasets to find related genes. These include protein-protein, protein-DNA and genetic interactions, pathways, reactions, gene and protein expression data, protein domains and phenotypic screening profiles. GeneMANIA uses a label propagation algorithm for predicting gene function given the composite functional association network (calculated from the databases selected). The weights needed for the label propagation method to work are selected at the beginning of the process. In our case, and according to the defaults, we weighted the network using linear regression, to make genes in the input list interact as much as possible with each other. We analyzed all loci that had at least one non-synonymous variant with $P < 1 \times 10^{-5}$ with any trait, and then mapped the most significant non-synonymous variant at each locus to the gene (input genes). We performed four network analyses: (1) HbA1c-associated variants only, (2) FI-associated variants only, (3) FG-associated variants only, and (4) 2hGlu-associated variants only (Figure 1, Supplementary Figure S1[11]). We selected the 50

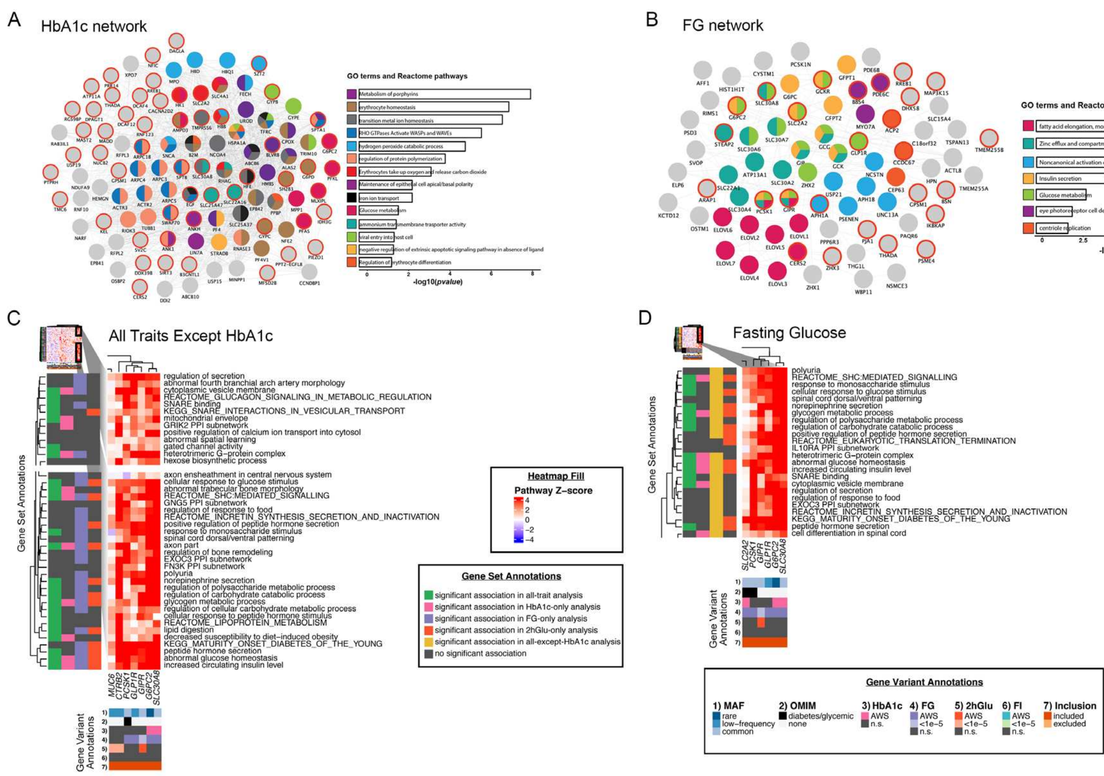

Figure 1. Network and pathway analyses identify relevant gene sets regulating glycemia using two different methods for variant associations with $P < 1 \times 10^{-5}$. (**A**–**B**) The networks represent composite networks for (**A**) HbA1c and (**B**) FG, from the GeneMANIA analysis using genes with 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 -log10(p-values) of the most significant terms within each group are 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) to 6) 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 7) whether each variant was included in the analysis or excluded by filters (see Methods). AWS: array-wide significant.

default databases to create the composite network, and we allowed the method to find at most 50 genes that are related to our query input list. The resultant networks were investigated to find enriched Gene Ontology (GO) terms and Reactome Pathways. Gene Set Enrichment (GSE) of networks and sub-networks were assessed with ClueGO[19] using GO terms and Reactome gene sets[20]. The enrichment results were grouped using a Cohen's Kappa score of 0.4, and terms were considered significant with a Bonferroni-adjusted p-value <0.05, provided that there was an overlap of at least three network genes in the relevant GO gene set when calculating GO enrichment. For the pathway selection (Reactome), we set a threshold that the network genes should represent at least 4% of the pathway. These values were applied given the recommended defaults when running ClueGO[19]. Cohen's Kappa statistic was used to measure the gene-set similarity of GO terms and Reactome pathways and allowed us to group enriched terms into functional groups to improve visualization of enriched pathways. We used all genes with GO annotations and at least one interaction in our network database as the background set.

Gene set enrichment analysis (GSEA)

An extension of the GWAS GSEA method DEPICT[21], EC-DEPICT[22,23], was used for GSEA. The key feature of EC-DEPICT is the use of "reconstituted" gene sets, which are gene sets collected from many different databases (e.g. canonical pathways, protein-protein interaction networks, and mouse phenotypes) that have been extended based on large-scale microarray co-expression data[21,24].

Six groups of variants were analyzed: (1) HbA1c-associated variants only, (2) FI-associated variants only, (3) FG-associated variants only, (4) 2hGlu-associated variants only, (5) all trait-associated variants, and (6) all trait-associated variants except for HbA1c. For each trait, the associated variants based on the European summary statistics were identified and clumped using a +/- 500 kb window. Then, the most significant nonsynonymous variant for each locus was included in the analysis, with a cut-off of $P < 10^{-5}$. Annotations from the CHARGE consortium were used to assign variants to genes (see **URL**). After GSEA, highly correlated gene sets were grouped by affinity propagation clustering of all 14,462 gene sets[25] into "meta-gene sets" using SciKitLearn.clustering. AffinityPropagation version 0.17[26]. For all visualizations, the gene set within a meta-gene set with the best enrichment *P*-value was used; heat maps were created with the ComplexHeatmap package in R[27].

URL: CHARGE Consortium ExomeChip annotation file (v6).

Method and choice of data for permutations: We performed the EC-DEPICT analysis as described elsewhere[22,23]. All analyses are based on a group of 14,462 "reconstituted" gene sets, which contains a z-score for probability of gene set membership for each gene (for details, see[21,24]).

The basic EC-DEPICT method is as follows. We first obtain a list of significant input variants (the most significant nonsynonymous variant per locus) and then map variants to genes based on annotations from the CHARGE consortium (see **URL**). For each gene set, we obtain the gene set membership z-scores for all trait-associated input genes and sum them to generate a test statistic. We then take 2,000 permuted ExomeChip association studies (described in more detail below) and calculate the average permuted test statistic for that gene set, as well as the permuted standard deviation. For each permutation, the number of top genes we take as "input genes" is matched to the actual observed number of input genes. We then calculate (observed test statistic – average permuted test statistic)/(permuted standard deviation) to generate a z-score, which is converted to a p-value via the normal distribution. False discovery rates were calculated by comparing the observed p-values to a permuted *P*-value distribution generated with an additional set of 50 permuted association studies.

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

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

As part of our GSEA pipeline, we remove input trait-associated variants that are not present in the permuted data to ensure that all variants are appropriately modeled. When using the Swedish permutations, this generally results in removing a substantial fraction of the variants, especially of the very rarest variants (due to the smaller sample size of the Swedish data relative to the data being analyzed). We have previously observed that this filtering can actually improve the GSEA signal, possibly due to more heterogeneous biology or a higher false-positive rate in these very rare variants[23]. However, in this case, we observed that in performing this filtering, we excluded variants in several known monogenic disease genes, such as *HNF1A* and *SLC2A2*. Therefore, we wished to repeat the analysis with a set of permutations which would allow us to retain these variants. We thus repeated the analysis with a second set of permutations consisting of 152,249 samples from the UK Biobank (referred to as the "UKBB permutations"). The larger sample size in the UKBB permutations means more variants are present and can therefore be included in the analysis.

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

All-trait analysis: After filtering, 78 input genes were included for the analysis with the UKBB permutations and 60 for the analysis with the Swedish permutations. (Note that the difference in the number of input genes is due to the presence of a larger number of input variants in the UKBB permutations – see above). We found 234 significant gene sets in 86 meta-gene sets based on the UKBB permutations (Supplementary Figure S2[11]) and 133 gene sets in 51 meta-gene sets based on the Swedish permutations (Supplementary Figure S3[11]). The correlation between the UKBB and Swedish analyses was r = 0.902, $P < 10^{-300}$.

All-traits-except-HbA1c analysis: After filtering, 45 input genes were included for the analysis with the UKBB permutations and 33 for the analysis with the Swedish permutations. We found 128 significant gene sets in 53 meta-gene sets based on the UKBB permutations (Supplementary Figure S2[11]) and 45 significant gene sets in 18 meta-gene sets based on the Swedish permutations (Supplementary Figure S3[11]). The correlation between the UKBB and Swedish analyses was r = 0.882, $P < 10^{-300}$.

HbA1c-only analysis: After filtering, 41 input genes were included for the analysis with the UKBB permutations and 33 for the analysis with the Swedish permutations. We found 191 significant gene sets in 73 meta-gene sets based on the UKBB permutations (Supplementary Figure S2[11]) and 120 gene sets in 41 meta-gene sets based on the Swedish permutations. (Supplementary Figure S3[11]). The correlation between the UKBB and Swedish analyses was r = 0.936, $P < 10^{-300}$.

FG-only analysis: After filtering, 26 input genes were included for the analysis with the UKBB permutations and 22 for the analysis with the Swedish permutations. We found 106 significant gene sets in 39 meta-gene sets based on the UKBB permutations (Supplementary Figure S2[11]) and 48 significant gene sets in 15 meta-gene sets based on the Swedish permutations (Supplementary Figure S3[11]). The correlation between the UKBB and Swedish analyses was r = 0.939, $P < 10^{-300}$.

2hGlu-only analysis: After filtering, 12 input genes were included for the analysis with the UKBB permutations and seven for the analysis based on the Swedish permutations. We found 56 significant gene sets in 17 meta-gene sets based on the UKBB permutations (Supplementary Figure S2[11]), with no significant gene sets based on the Swedish permutations. The correlation between the UKBB and Swedish analyses was r = 0.787, $P < 10^{-300}$.

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

Visualization: As in previous work[22,23], we have included all trait-associated variants in the heat maps, even if they were excluded from the analysis (e.g. because they were absent in the permutations or did not have a nonsynonymous annotation in the CHARGE annotation file). This is because we assume that if the genes harboring those variants have strong predicted membership in significantly trait-associated gene sets, they are still good candidates for prioritization. In fact, this may be even stronger evidence in favor of these genes because they did not contribute to the enrichment analysis and therefore their prioritization is independently derived (and provides even more support to the implicated biology).

Results

Study design overview

We performed single-variant and gene-based association analyses with FG, FI, HbA1c, and 2hGlu levels on exome-array coding variants in up to 144,060 individuals without diabetes (to exclude any consequence of diabetes treatments or related interventions on these quantitative traits) of European (85%), African-American (6%), South Asian (5%), East Asian (2%), and Hispanic (2%) ancestry from up to 64 cohorts (Supplementary Table S1[11], Methods). We used a linear mixed model to test single-variant associations in each individual cohort and combined results by fixed-effect meta-analyses within and across ancestries. As body mass index (BMI) is a major risk factor for T2D and is correlated with glycemic traits, all analyses were adjusted for BMI to identify loci influencing glycemia independently from their effects on overall adiposity. We have previously demonstrated that collider bias did not significantly affect results with BMI adjustment[1]. We used distance-based clumping to define distinct loci and considered signals to be novel if they were located more than 500 kb from a variant with an established association with any of the glycemic traits or T2D in large published GWAS (Methods). We considered a coding variant to meet exome-wide significance for association if $P < 2.2 \times 10^{-7}$[15,16] (Table 1, Methods). To increase power to detect rare variant associations, we additionally performed gene-burden and sequence kernel association (SKAT) tests for gene-level analyses to identify genes with significant evidence of association ($P < 2.5 \times 10^{-6}$) (Table 2, Methods). Finally, to identify relevant biological pathways enriched in associations with glycemic traits we conducted pathway and network analyses.

Identification of single-variant associations

Our single variant analyses identified 62 distinct coding variant associations at 58 genes associated with at least one of the glycemic traits at exome-wide significance ($P < 2.2 \times 10^{-7}$) (Table 1). Of these, four variants at three genes represented novel associations. These included a missense (rs1983210,

Table 1. Single-point coding variant associations meeting the significance threshold for coding variants of $P < 2.2 \times 10^{-7}$. This table includes all 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 S2) 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; EAF: effect allele frequency; 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 in this study; Locus name of previous association – name used for previously reported locus. [1]Significant in the European-only analysis in our study. Genes in this table are listed in order of chromosomal position.

Trait	SNP	Gene	Protein Consequence	Alleles E/O	EAF	Effect (SE)	*P*	N	Previous glycemic trait association (if any)	Locus name of 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	rs1260326	*GCKR*	p.L446P	C/T	0.631	0.029 (0.002)	6.36×10^{-48}	129588	FG, FI, 2hGlu	*GCKR*
FI	rs1260326	*GCKR*	p.L446P	C/T	0.626	0.024 (0.002)	5.55×10^{-32}	104076	FG, FI, 2hGlu	*GCKR*
2hGlu	rs1260326	*GCKR*	p.L446P	C/T	0.618	-0.069 (0.009)	4.48×10^{-15}	57813	FG, FI, 2hGlu	*GCKR*
FG	rs35720761	*THADA*	p.C845Y	T/C	0.108	-0.018 (0.003)	4.35×10^{-9}	129622	T2D, FG	*THADA*
HbA1c	rs35720761	*THADA*	p.C845Y	C/T	0.113	0.014 (0.002)	2.58×10^{-12}	144001	T2D, FG	*THADA*
FG	rs7578597	*THADA*	p.T897A	C/T	0.106	-0.019 (0.003)	1.99×10^{-8}	113162	T2D, FG	*THADA*
FI	rs7607980	*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	rs1801282	*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*
HbA1c[1]	rs2237051	*EGF*	p.M708I	A/G	0.374	-0.007 (0.001)	2.11×10^{-7}	121204	HbA1c	*EGF*
HbA1c	rs7683365	*GYPB*	p.T48M	A/G	0.312	0.012 (0.002)	1.61×10^{-8}	45191	HbA1c	*FREM3*
FG	rs146886108	*ANKH*	p.R187Q	T/C	0.004	-0.088 (0.014)	5.67×10^{-10}	129647	T2D	*ANKH*
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	*RREB1*	p.S1499Y	A/C	0.173	-0.01 (0.002)	3.76×10^{-9}	143967	FG, T2D	*RREB1*
FG	rs35742417	*RREB1*	p.S1499Y	A/C	0.183	-0.019 (0.002)	1.27×10^{-16}	129577	FG, T2D	*RREB1*

Trait	SNP	Gene	Protein Consequence	Alleles E/O	EAF	Effect (SE)	*P*	N	Previous glycemic trait association (if any)	Locus name of previous association
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	HbA1c	*MLXIPL*
HbA1c	rs3812316	*MLXIPL*	p.Q241H	G/C	0.112	0.012 (0.002)	2.15×10^{-8}	108605	HbA1c	*MLXIPL*
FG	rs194524	*STEAP2*	p.R456Q	A/G	0.523	0.01 (0.002)	7.65×10^{-8}	129629	FG, T2D, RG	*STEAP2-AS1*
HbA1c	rs34664882	*ANK1*	p.A1503V	A/G	0.026	-0.049 (0.004)	2.43×10^{-39}	144034	HbA1c	*ANK1*
FG	rs13266634	*SLC30A8*	p.R276W	T/C	0.305	-0.029 (0.002)	1.63×10^{-46}	129614	FG, HbA1c, T2D	*SLC30A8*
HbA1c	rs13266634	*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	T2D, HbA1c	*Mahajan 2022 from CMD KP*
FG	rs17853166	*IKBKAP*	p.S251G	C/T	0.026	-0.037 (0.006)	4.82×10^{-11}	129640	FG	*IKBKAP*
HbA1c	rs60980157	*GPSM1*	p.S391L	T/C	0.246	-0.013 (0.002)	6.71×10^{-17}	118824	FG, T2D	*GPSM1*
FG	rs60980157	*GPSM1*	p.S391L	T/C	0.254	-0.014 (0.002)	2.35×10^{-9}	110915	FG, T2D	*GPSM1*
HbA1c	rs906220	*HK1*	p.H7R	G/A	0.916	0.025 (0.003)	2.16×10^{-21}	94970	HbA1c	*HK1*
FG	rs701865	*PDE6C*	p.S270T	A/T	0.366	-0.01 (0.002)	1.14×10^{-7}	118580	FG, RG	*PDE6C*
HbA1c	rs61732434	*OR51V1*	p.S161N	T/C	0.008	-0.052 (0.009)	1.75×10^{-8}	127507	HbA1c	*HBB*
HbA1c	rs415895	*SWAP70*	p.Q447E	G/C	0.641	-0.013 (0.001)	1.15×10^{-21}	138028	HbA1c	*SWAP70*
HbA1c	rs117706710	*AMPD3*	p.V311L	T/G	0.009	0.037 (0.006)	2.32×10^{-10}	144048	HbA1c	*AMPD3*
FG	rs2167079	*ACP2*	p.R29Q	T/C	0.340	0.016 (0.002)	7.99×10^{-15}	129580	FG	*MADD*
HbA1c	rs35233100	*MADD*	p.R766X	T/C	0.055	-0.015 (0.003)	1.13×10^{-8}	144034	FG	*MADD*
FG	rs35233100	*MADD*	p.R766X	T/C	0.054	-0.029 (0.004)	1.46×10^{-12}	126231	FG	*MADD*
FG	rs56200889	*ARAP1*	p.Q802E	C/G	0.270	-0.016 (0.002)	1.79×10^{-14}	122674	FG	*ARAP1*
HbA1c	rs643788	*DPAGT1*	p.I393V	C/T	0.425	-0.006 (0.001)	1.77×10^{-7}	144009	HbA1c	*C2CD2L*
FI[1]	rs145878042	*RAPGEF3*	p.L300P	G/A	0.011	-0.054 (0.01)	1.15×10^{-7}	91485	FI/HbA1c	*HDAC7/ PFKM*
HbA1c	rs2732481	*ZNF641*	p.Q363P	G/T	0.315	-0.009 (0.001)	2.07×10^{-11}	142280	HbA1c	*SENP1*
HbA1c	rs3184504	*SH2B3*	p.W262R	C/T	0.567	0.007 (0.001)	5.98×10^{-8}	138551	HbA1c	*ATXN2*
2hGlu	rs1169288	*HNF1A*	p.I75L	C/A	0.345	0.06 (0.011)	7.90×10^{-9}	44278	T2D, 2hGlu	*HNF1A*
HbA1c	COSM147717	*ATP11A*	p.M317V	G/A	0.748	0.009 (0.001)	3.77×10^{-12}	144022	HbA1c	*ATP11A,TUBGCP3*
HbA1c	rs229587	*SPTB*	p.S439N	T/C	0.357	0.007 (0.001)	2.60×10^{-8}	134780	HbA1c	*SPTB*
HbA1c	rs35097172	*SLC25A47*	splice region variant, 5′ UTR variant	T/C	0.216	-0.008 (0.002)	5.67×10^{-8}	144028	FG	*SLC25A47*

Trait	SNP	Gene	Protein Consequence	Alleles E/O	EAF	Effect (SE)	*P*	N	Previous glycemic trait association (if any)	Locus name of previous association
2hGlu	rs3784634	*VPS13C*	p.R974K	T/C	0.540	-0.069 (0.011)	6.40×10^{-10}	37217	2hGlu	*VPS13C/ C2CD4A/ C2CD4B*
HbA1c[1]	rs3747481	*PRR14*	p.P359L	T/C	0.261	0.009 (0.002)	3.30×10^{-8}	103338	HbA1c	*ITGAD*
HbA1c	rs201226914	*PIEZO1*	p.L939M	T/G	0.002	-0.159 (0.015)	4.42×10^{-26}	144024	HbA1c	*CDT1,CYBA*
2hGlu	rs72839768	*DVL2*	p.T529I	A/G	0.020	0.197 (0.03)	4.10×10^{-11}	57866	T2D, 2hGlu	*SLC16A13*
HbA1c	rs2748427	*TMC6*	p.W125R	G/A	0.233	0.027 (0.002)	8.56×10^{-70}	132326	HbA1c	*TMC6*
HbA1c	rs7225887	*B3GNTL1*	p.A163T	T/C	0.211	-0.015 (0.002)	5.73×10^{-22}	125749	HbA1c	*FN3KRP, FN3K*
HbA1c	rs35413309	*RGS9BP*	p.A223V	T/C	0.030	-0.02 (0.004)	1.42×10^{-8}	141598	HbA1c	*PDCD5*
2hGlu	rs1800437	*GIPR*	p.E318Q	C/G	0.217	0.103 (0.011)	2.59×10^{-23}	56252	2hGlu	*GIPR*
FG	rs17265513	*ZHX3*	p.N310S	C/T	0.188	0.016 (0.002)	2.59×10^{-10}	126253	FG	*ZHX3*
HbA1c	rs855791	*TMPRSS6*	V727A	G/A	0.577	-0.019 (0.001)	9.46×10^{-51}	143907	HbA1c	*TMPRSS6*
FG	rs15943	*MAP3K15*	p.Q1083E	C/G	0.005	-0.084 (0.014)	2.83×10^{-9}	67004	glucose	*PDHA1/MAP3K15*
FG	rs56381411	*MAP3K15*	p.G670S	T/C	0.005	-0.085 (0.013)	1.51×10^{-11}	62319	glucose	*PDHA1/MAP3K15*
HbA1c	rs2229241	*RENBP*	splice acceptor variant	C/T	0.012	-0.123 (0.007)	1.14×10^{-62}	95622	HbA1c	*G6PD*
HbA1c	rs1050828	*G6PD*	p.V68M	T/C	0.007	-0.334 (0.008)	7.41×10^{-322}	112209	HbA1c	*G6PD*

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.

Trait	Gene	NSbroad mask			NSstrict mask		
		N var	P_{burden}	P_{SKAT}	N var	P_{burden}	P_{SKAT}
FG	***G6PC***	9	$\mathbf{1.41\times10^{-6}}$	1.32×10^{-5}	3	1.41×10^{-3}	7.43×10^{-4}
FI	***G6PC***	8	$\mathbf{1.62\times10^{-6}}$	8.58×10^{-6}	3	1.85×10^{-3}	7.80×10^{-3}
HbA1c	***TF***	10	$\mathbf{2.15\times10^{-6}}$	5.98×10^{-3}	3	5.48×10^{-2}	5.48×10^{-2}
FG	*MAP3K15*	18	$\mathbf{1.86\times10^{-25}}$	1.07×10^{-18}	7	1.34×10^{-14}	4.01×10^{-11}
HbA1c	*MAP3K15*	18	$\mathbf{1.27\times10^{-7}}$	1.53×10^{-04}	7	2.65×10^{-4}	9.46×10^{-3}
FG	*G6PC2*	18	4.09×10^{-67}	5.38×10^{-58}	7	$\mathbf{7.8\times10^{-69}}$	3.83×10^{-56}
HbA1c	*G6PC2*	18	6.18×10^{-30}	4.65×10^{-27}	7	$\mathbf{1.04\times10^{-31}}$	1.92×10^{-26}
FG	*SLC30A8*	13	5.69×10^{-4}	$\mathbf{6.42\times10^{-11}}$	7	6.55×10^{-11}	3.74×10^{-10}
HbA1c	*SLC30A8*	12	7.20×10^{-8}	2.18×10^{-5}	6	$\mathbf{5.66\times10^{-8}}$	3.22×10^{-6}
FG	*VPS13C*	52	9.66×10^{-6}	$\mathbf{3.73\times10^{-7}}$	26	1.27×10^{-5}	1.44×10^{-5}

p.E1365D) and a splice region variant (rs3183099) in *OBSL1* associated with FI, another missense variant (rs1886686, p.G12A) in *WDR78* associated with FG, and a missense variant (rs31244, p.D543N) in *SV2C* associated with HbA1c (Table 1). In addition, the missense variant (rs146886108, p.R187Q) in *ANKH* which was previously associated with T2D was associated for the first time with FG.

Identification of gene-based associations

Our gene-based analyses identified six genes associated with glycemic traits, including *G6PC* and *TF* that had not been associated with glycemic traits before (Table 2 and Supplementary Table S2[11]). These findings provide new hypotheses for downstream follow-up studies in the context of glycemic trait biology. *G6PC*, encoding glucose-6-phosphatase, is associated with FG and FI and is a homolog of *G6PC2*. *G6PC2 is* an established effector gene at a GWAS locus which contains multiple coding variants known to influence FG and HbA1c but not FI levels[4,5,28–30]. Loss-of-function variants at *SLC30A8* have been previously associated with reduced risk of T2D[31–33], while *VPS13C* maps to the *VPS13C/C2CD4A/C2CD4B* T2D risk locus. Follow-up studies at this locus have with varying levels of evidence suggested *C2CD4A*, encoding a calcium-dependent nuclear protein, as the causal gene for T2D through its potential role in the pancreatic islets[34–37]. We found evidence of association at *MAP3K15* with reduced levels of FG and HbA1c (Table 2 and Supplementary Table S2[11]), which is consistent with recent reports of the gene's association with reduced levels of HbA1c and glucose, and reduced T2D risk[6,38]. Our analyses also detected *TF* (encoding transferrin) as a novel gene-based association signal associated with HbA1c but not any of the other glycemic traits, consistent with the role of the protein as the main iron carrier in the blood (Table 2 and Supplementary Table S2[11]).

Pathway analyses identify relevant gene sets regulating glycemia

Next, we used our coding variant association results to identify pathways enriched for glycemic trait associations, and to subsequently determine the extent to which different associations within the same trait implicate the same or similar pathways (as indicated by the functional connectivity of the network). To do this we used GeneMANIA network analysis[39], which takes a query list of genes and finds functionally similar genes based on large, publicly available biological datasets, that include protein-protein, protein-DNA and genetic interactions, pathways, protein domains, protein and gene expression data. GeneMANIA taps on updated versions of these databases for its core and network analyses, to identify related genes of known functions based on our input list of genes. To increase power to connect genes in a network, we considered all genes harboring non-synonymous variants that reached $P < 1 \times 10^{-5}$ (Supplementary Table S3[11]) for any of the four glycemic traits in our study and mapped the most significant non-synonymous variant at each locus to the respective gene (totaling 121 associations across all traits) (Methods). A high degree of connectivity was observed within the HbA1c network, with enrichment of processes related to blood cell biology such as porphyrin metabolism, erythrocyte homeostasis and iron transport (Figure 1A and Supplementary Table S4[11]). In comparison, the network generated from FG-associated genes captured several processes known to contribute to glucose regulation and islet function, including insulin secretion, zinc transport and fatty acid metabolism (Figure 1B and Supplementary Table S4[11]). Given that there were fewer genes associated with FI and 2hGlu, we were less powered to draw meaningful insights from the enriched pathways in those traits (Supplementary Figure S1 and Supplementary Table S4[11]).

We also performed gene set enrichment analysis (GSEA) using EC-DEPICT[22,23] (Methods). The primary innovation of EC-DEPICT is the use of 14,462 gene sets extended based on large-scale co-expression data[21,24]. These gene sets take the form of z-scores, where higher z-scores indicate a stronger prediction that a given gene is a member of a gene set. To reduce some of the redundancy in the gene sets (many of which are strongly correlated with one another), we clustered them into 1,396 "meta-gene sets" using affinity propagation clustering[25]. These meta-gene sets are used to simplify visualizations and aid interpretation of results. As before, we considered all loci with variants that reached $P < 1 \times 10^{-5}$ (Supplementary Table S3[11]) for any of the four glycemic traits for defining input genes (Methods). When looking across all traits combined, we found 234 significant gene sets in 86 meta-gene sets with false discovery rate (FDR) of <0.05 (Supplementary Table S5A, Supplementary Figure S2A[11]). As expected, we observed a strong enrichment of insulin- and glucose-related gene sets, as well as hormone secretion and cytoplasmic vesicle gene sets (in keeping with pancreatic beta cell insulin vesicle release). In agreement with the GeneMANIA network analyses, we also noted a particularly strong enrichment for blood-related pathways represented by gene sets such as erythrocyte differentiation and heme metabolic process, which was primarily driven by HbA1c-associated variants. This was likely because HbA1c levels are influenced not only by glycation but also by blood cell turnover rate[1,40,41]. To disentangle blood cell turnover from effects due to glycation, we repeated the analysis excluding variants that were significantly associated with HbA1c only and found 128 significant gene sets in 53 meta-gene sets (FDR <0.05) (Figure 1C, Supplementary Table S5B, Supplementary Figure S2B[11]). Indeed, we noted that majority of the gene sets now implicated pathways relevant to the pancreatic islets and metabolic tissues, such as "abnormal glucose homeostasis", "peptide hormone secretion", "Maturity Onset Diabetes of the Young", and multiple pathways involved in the regulation of glycogen, incretins, and carbohydrate metabolism, that were also seen in the FG only analysis (Figure 1D, Supplementary Table S5D, Supplementary Figure S2D[11]).

We also analyzed each of the four traits separately, to reveal trait-specific enriched gene sets (Supplementary Table S5, Supplementary Figure S2C-E, Supplementary Figure S3C-D[11], Methods). Overall, our network and pathway enrichment analyses provide insight into the biology underlying each glycemic trait and may facilitate the prioritization of specific genes or pathways across multiple different phenotypes.

Discussion

Here we have described large scale meta-analyses results for coding variant and gene-based associations for four glycemic traits, FG, FI, HbA1c and 2Glu, and the downstream pathways and networks that are regulated by the associated genes. Our results identified three genes with novel single-variant associations with glycemic traits *OBSL1* (FI), *WDR78* (FG) and *SVC2* (HbA1c). *OBSL1* encodes a cytoskeletal protein related to obscurin, mutations in which have been shown to lead to an autosomal recessive primordial growth disorder (OMIM: 612921). Loss of OBSL1 leads to downregulation of CUL7, a protein known to interact with IRS-1, downstream of the insulin receptor signaling pathway[42]. *WDR78* encodes a WD repeat-containing protein 78, the same variant rs1886686-C has been previously associated with a decrease in systolic blood pressure[43]. However, none of the *OBSL1* (rs1983210, b = -0.018, p = 1.20 x 10^{-4}, N = 144,114; rs3183099, b = -0.019, p = 1.36 x 10^{-4}, N = 125,397) or *WDR78* (rs1886686, b = -0.017, p = 3.83 x 10^{-5}, N = 164,878) variants we detected here reached exome-wide significance in our recent large multi-ancestry study[1]. This, despite larger sample sizes and good genotype quality (info >0.8 for each of the variants for the majority of cohorts), suggesting caution in the interpretation of these findings, and the need for additional datasets testing these associations. The final variant, p.D543N in *SV2C,* was associated with HbA1c with p = 5.5 x 10^{-5} in the European meta-analysis[1], and with p = 1.37 x 10^{-12} in UK biobank[44]. A second missense variant at this gene, p.T482S, is also strongly associated with HbA1c (p = 1.9 x 10^{-16}) and with red blood cell distribution width in UK biobank (p = 3.3 x 10^{-11})[44], and with mean corpuscular volume (p = 3 x 10^{-11})[45]. Given that variation in red blood cell traits can influence HbA1c levels[1,41], associations between these missense variants suggest *SV2C* as the likely effector gene at this locus. Also, the absence of evidence for association between this gene and other glycemic traits suggests its effect on HbA1c is independent of glycemia.

The novel gene-based association of *G6PC* with FG and FI was notable. Homozygous inactivating alleles in *G6PC*, including both p.R83C and p.Q347X which are contained in our gene-based association (Table S2), are known to give rise to glycogen storage disease type 1a (GSD1a). GSD1a is a rare autosomal recessive metabolic disorder[46,47], but this is the first time that rare coding variants in *G6PC* have been shown to influence FG and FI levels in normoglycemic individuals. The other novel gene-based association was between *TF* and HbA1c. TF encodes transferrin, an iron-binding transport protein that circulates at high levels in blood plasma as an important biological carrier of iron. Dysregulation of iron concentrations due to reduced transferrin levels or function could affect the measurement of HbA1c independently of glycemia[48]. The presence of multiple coding variants within TF associated with red blood cell traits in UK biobank[44] lends additional support to this hypothesis.

Overall, our network and pathway analyses were highly concordant with each other and with other published data identifying processes related to glucose regulation and islet function, including insulin secretion and zinc transport associated with FG loci, and red blood cell biology processes amongst HbA1c associated loci[1]. The FG network revealed linking nodes (that are not among the association signals) with known links to glucose homeostasis and diabetes, such as *GCK* (encoding the beta cell glucose sensor glucokinase), *GCG* (encoding the peptide hormone glucagon secreted by the alpha cells of the pancreas) and *GIP* (encoding the incretin hormone gastric inhibitory polypeptide). Notably, lipid related pathways associated with fasting glucose. One gene within the FG cluster for lipid-related pathways is *CERS2*, which encodes ceramide synthase 2, an enzyme known to be associated with the sphingolipid biosynthetic process (Figure 1B, Supplementary Table S3[11]). Although *CERS2* is only nominally associated with FG and is significantly associated with HbA1c (rs267738: P_{FG} = 3.54 × 10^{-7}; P_{HbA1c} = 6.96 × 10^{-10}), it does not cluster together with any HbA1c-enriched pathway, suggesting that *CERS2* is regulating FG and HbA1c indirectly through its role in lipid metabolism.

Conclusions

In conclusion, our results provided novel glycemic trait associations and highlighted pathways implicated in glycemic regulation. The summary statistics results are being made publicly available through various platforms so they can be harnessed with other data to aid effector gene identification.

Data availability

Underlying data

Open Science Framework (OSF): Underlying data for 'Large-scale exome array summary statistics resources for glycemic traits to aid effector gene prioritization', https://doi.org/10.17605/OSF.IO/K6W3B[11]

This project contains the following underlying data:

- Table S1: Supplementary Table S1 – Cohort characteristics, genotyping and quality control (QC), glucose, insulin, 2hGlu and HbA1c analyses and covariates.
- Table S2: Supplementary Table S2 - Full gene-based results including all variants included in the masks, for both novel and previously-established genes
- Table S3: Supplementary Table S3 - All variants associated with FG, FI, HbA1c and/or 2hGlu in our analyses with P<10-5
- Table S4: Supplementary Table S4 - Gene Set Enrichment Analysis by GeneMANIA network analysis showing enriched GO terms and Reactome pathways in the network for (A) HbA1c; (B) FG; (C) FI; (D) 2hGlu
- Table S5: Supplementary Table S5 - EC-DEPICT results
- Figure S1: Supplementary Figure S1 – GeneMANIA network analysis results
- Figure S2: Supplementary Figure S2 – EC-DEPICT results (UKBB permutations)
- Figure S3: Supplementary Figure S3 - EC-DEPICT results (Swedish permutations)

Accession numbers

GWAS Catalog: meta-analysis summary statistics of 2-hour glucose in African American ancestry. MAGICExome_2hGlu_AFR.tsv.gz, study accession number GCST90256400. https://identifiers.org/gcst:GCST90256400

GWAS Catalog: meta-analysis summary statistics of 2-hour glucose in European ancestry. MAGICExome_2hGlu_EUR.tsv.gz, study accession number GCST90256401. https://identifiers.org/gcst:GCST90256401

GWAS Catalog: multi-ancestry meta-analysis summary statistics of 2 hour glucose. MAGICExome_2hGlu_ALL.tsv.gz, study accession number GCST90256402. https://identifiers.org/gcst:GCST90256402

GWAS Catalog: meta-analysis summary statistics of fasting glucose in African American ancestry. MAGICExome_FG_AFR.tsv.gz, study accession number GCST90256403. https://identifiers.org/gcst:GCST90256403

GWAS Catalog: meta-analysis summary statistics of fasting glucose in East Asian ancestry. MAGICExome_FG_EAS.tsv.gz, study accession number GCST90256404. https://identifiers.org/gcst:GCST90256404

GWAS Catalog: meta-analysis summary statistics of fasting glucose in European ancestry. MAGICExome_FG_EUR.tsv.gz, study accession number GCST90256405. https://identifiers.org/gcst:GCST90256405

GWAS Catalog: meta-analysis summary statistics of fasting glucose in Hispanic ancestry. MAGICExome_FG_HISP.tsv.gz, study accession number GCST90256406. https://identifiers.org/gcst:GCST90256406

GWAS Catalog: meta-analysis summary statistics of fasting glucose in South Asian ancestry. MAGICExome_FG_SAS.tsv.gz, study accession number GCST90256407. https://identifiers.org/gcst:GCST90256407

GWAS Catalog: multi-ancestry meta-analysis summary statistics of fasting glucose. MAGICExome_FG_ALL.tsv.gz, study accession number GCST90256408. https://identifiers.org/gcst:GCST90256408

GWAS Catalog: meta-analysis summary statistics of fasting insulin in African American ancestry. MAGICExome_FI_AFR.tsv.gz, study accession number GCST90256409. https://identifiers.org/gcst:GCST90256409

GWAS Catalog: meta-analysis summary statistics of fasting insulin in East Asian ancestry. MAGICExome_FI_EAS.tsv.gz, study accession number GCST90256410. https://identifiers.org/gcst:GCST90256410

GWAS Catalog: meta-analysis summary statistics of fasting insulin in European ancestry. MAGICExome_FI_EUR.tsv.gz, study accession number GCST90256411. https://identifiers.org/gcst:GCST90256411

GWAS Catalog: meta-analysis summary statistics of fasting insulin in Hispanic ancestry. MAGICExome_FI_HISP.tsv.gz, study accession number GCST90256412. https://identifiers.org/gcst:GCST90256412

GWAS Catalog: meta-analysis summary statistics of fasting insulin in South Asian ancestry. MAGICExome_FI_SAS.tsv.gz, study accession number GCST90256413. https://identifiers.org/gcst:GCST90256413

GWAS Catalog: multi-ancestry meta-analysis summary statistics of fasting insulin. MAGICExome_FI_ALL.tsv.gz, study accession number GCST90256414. https://identifiers.org/gcst:GCST90256414

GWAS Catalog: meta-analysis summary statistics of HbA1c in African American ancestry. MAGICExome_HbA1c_AFR.tsv.gz, study accession number GCST90256415. https://identifiers.org/gcst:GCST90256415

GWAS Catalog: meta-analysis summary statistics of HbA1c in East Asian ancestry. MAGICExome_HbA1c_EAS.tsv.gz, study accession number GCST90256416. https://identifiers.org/gcst:GCST90256416

GWAS Catalog: meta-analysis summary statistics of HbA1c in European ancestry. MAGICExome_HbA1c_EUR.tsv.gz, study accession number GCST90256417. https://identifiers.org/gcst:GCST90256417

GWAS Catalog: meta-analysis summary statistics of HbA1c in Hispanic ancestry. MAGICExome_HbA1c_HISP.tsv.gz, study accession number GCST90256418. https://identifiers.org/gcst:GCST90256418

GWAS Catalog: meta-analysis summary statistics of HbA1c in South Asian ancestry. MAGICExome_HbA1c_SAS.tsv.gz, study accession number GCST90256419. https://identifiers.org/gcst:GCST90256419

GWAS Catalog: multi-ancestry meta-analysis summary statistics of HbA1c. MAGICExome_HbA1c_ALL.tsv.gz, study accession number GCST90256420. https://identifiers.org/gcst:GCST90256420

These data are also available from https://magicinvestigators.org/downloads/

Acknowledgements

Anna L. Gloyn and Inês Barroso contributed equally to the supervision of this work.

We would like to thank Jan-Håkan Jansson, Kurt Lohman, Jung-Jin Lee, Neil Robertson, Hugoline de Haan, Jin Li, Ken Sin Lo, Carola Marzi, Yuan Shi and Salman M. Tajuddin for their contributions to this work. They were part of the collective but could not be listed as authors since they were not involved in the final submission.

The authors would like to thank the Rivas lab for making the Global Biobank Engine resource available.

Study/Individual	Acknowledgment
AGES	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.
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.
ARIC	The authors thank the staff and participants of the ARIC study for their important contributions.
ASCOT	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	Claudia Tersteeg, Krista den Ouden, Mirjam B. Smeets, and Loes B. Collé are graciously acknowledged for their work on the DNA extraction. Astrid E.M.W. Willems, Evelyn Velema, Kristy M. J. Vons, Sara Bregman, Timo R. ten Brinke, Sara van Laar, Louise M. Catanzariti, Joyce E.P. Vrijenhoek, Sander M. van de Weg, Arjan H. Schoneveld, Arnold Koekman, Arjan Boltjes, Petra H. Homoed-van der Kraak, and Aryan Vink 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.
CHS	A full list of principal CHS investigators and institutions can be found at CHS-NHLBI.org. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.
COPSAC2000	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
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 Gremsperger and P Storm for their expert technical assistance with genotyping and genotype data preparation.
EFSOCH	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	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	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	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.
	The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscripts.
Fenland	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.

Study/Individual	Acknowledgment
FIA3	Jansson J-H was responsible for the identification of MI cases.
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. 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.
GENOA	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.
GLACIER	We are indebted to the study participants who dedicated their time and samples to these studies. We 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 Gremsperger and P Storm for their expert technical assistance with genotyping and genotype data preparation.
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). 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)	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.
HELIC MANOLIS and HELIC Pomak	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. Papagrigoriou, 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).
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.
InterAct Consortium	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. More information about EPIC-Ineract can be found here: https://www.mrc-epid.cam.ac.uk/research/studies/interact/ and under cohort reference PMID 21717116
JHS	The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.
KORA	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-Childhood-IFB	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)	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.
Mark I McCarthy	MMcC was a Wellcome Investigator and an NIHR Senior Investigator.
MESA	Genotyping was performed at Affymetrix (Santa Clara, California, USA) and the Broad Institute of Harvard and MIT (Boston, Massachusetts, USA) using the Affymetrix Genome-Wide Human SNP Array 6.0. The authors thank the other investigators, the staff, and the participants of the MESA study for their valuable contributions. A fill list of participating MESA investigators and institutes can be found at http://www.mesa-nhlbi.org.

Study/Individual	Acknowledgment
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.
NFBC66 and NFBC86	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 Academian 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	The investigators express their deepest gratitude to the study participants.
PPP-Botnia	The skillful assistance of the Botnia Study Group is gratefully acknowledged.
Rotterdam study	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. 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.
SardiNIA	The SardiNIA investigators thank all the volunteers who generously participated in this study and made this research possible.
The Singapore Chinese Eye Study (SCES)	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.
Timothy D Spector	Timothy D Spector is holder of an ERC Advanced Principal Investigator award.
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. 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. We would like to thank the following people for their contributions to this work: Michaela Benzeval(1), Jonathan Burton(1), Nicholas Buck(1), Annette Jäckle(1), Meena Kumari(1), Heather Laurie(1), Peter Lynn(1), Stephen Pudney(1), Birgitta Rabe(1), Dieter Wolke(2) (1) Institute for Social and Economic Research (2) University of Warwick

References

1. Chen J, Spracklen CN, Marenne G, *et al.*: **The trans-ancestral genomic architecture of glycemic traits.** *Nat Genet.* 2021; **53**(6): 840–860.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

2. Downie CG, Dimos SF, Bien SA, *et al.*: **Multi-ethnic GWAS and fine-mapping of glycaemic traits identify novel loci in the PAGE Study.** *Diabetologia.* 2022; **65**(3): 477–489.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

3. Mahajan A, Spracklen CN, Zhang W, *et al.*: **Multi-ancestry genetic study of type 2 diabetes highlights the power of diverse populations for discovery and translation.** *Nat Genet.* 2022; **54**(5): 560–572.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

4. Mahajan A, Sim X, Ng HJ, *et al.*: **Identification and functional characterization of *G6PC2* coding variants influencing glycemic traits define an effector transcript at the *G6PC2-ABCB11* locus.** *PLoS Genet.* 2015; **11**(1): e1004876.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

5. Wessel J, Chu AY, Willems SM, *et al.*: **Low-frequency and rare exome chip variants associate with fasting glucose and type 2 diabetes susceptibility.** *Nat Commun.* 2015; **6**: 5897.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

6. Backman JD, Li AH, Marcketta A, *et al.*: **Exome sequencing and analysis of 454,787 UK Biobank participants.** *Nature.* 2021; **599**(7886): 628–634.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

7. Sun BB, Kurki MI, Foley CN, *et al.*: **Genetic associations of protein-coding variants in human disease.** *Nature.* 2022; **603**(7899): 95–102.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

8. Deaton AM, Parker MM, Ward LD, *et al.*: **Gene-level analysis of rare variants in 379,066 whole exome sequences identifies an association of *GIGYF1* loss of function with type 2 diabetes.** *Sci Rep.* 2021; **11**(1): 21565.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

9. Jurgens SJ, Choi SH, Morrill VN, *et al.*: **Analysis of rare genetic variation**

underlying cardiometabolic diseases and traits among 200,000 individuals in the UK Biobank. *Nat Genet.* 2022; **54**(3): 240–250.
PubMed Abstract | Publisher Full Text | Free Full Text

10. Sollis E, Mosaku A, Abid A, *et al.*: **The NHGRI-EBI GWAS Catalog: knowledgebase and deposition resource.** *Nucleic Acids Res.* 2023; **51**(D1): D977–D985.
PubMed Abstract | Publisher Full Text | Free Full Text

11. Willems SM: **Underlying Data for Study ´Large-Scale Exome Array Summary Statistics Resources for Glycemic Traits to Aid Effector Gene Prioritization.** 7 September 2023 edn Web OSF, 2023.

12. Grove ML, Yu B, Cochran BJ, *et al.*: **Best practices and joint calling of the HumanExome BeadChip: the CHARGE Consortium.** *PLoS One.* 2013; **8**(7): e68095.
PubMed Abstract | Publisher Full Text | Free Full Text

13. Yourshaw M, Taylor SP, Rao AR, *et al.*: **Rich annotation of DNA sequencing variants by leveraging the Ensembl Variant Effect Predictor with plugins.** *Brief Bioinform.* 2015; **16**(2): 255–64.
PubMed Abstract | Publisher Full Text | Free Full Text

14. Liu X, Jian X, Boerwinkle E: **dbNSFP v2.0: a database of human non-synonymous SNVs and their functional predictions and annotations.** *Hum Mutat.* 2013; **34**(9): E2393–402.
PubMed Abstract | Publisher Full Text | Free Full Text

15. Sveinbjornsson G, Albrechtsen A, Zink F, *et al.*: **Weighting sequence variants based on their annotation increases power of whole-genome association studies.** *Nat Genet.* 2016; **48**(3): 314–7.
PubMed Abstract | Publisher Full Text

16. Mahajan A, Wessel J, Willems SM, *et al.*: **Refining the accuracy of validated target identification through coding variant fine-mapping in type 2 diabetes.** *Nat Genet.* 2018; **50**(4): 559–571.
PubMed Abstract | Publisher Full Text | Free Full Text

17. Buniello A, MacArthur JAL, Cerezo M, *et al.*: **The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019.** *Nucleic Acids Res.* 2019; **47**(D1): D1005–D1012.
PubMed Abstract | Publisher Full Text | Free Full Text

18. Common Metabolic Diseases Knowledge Portal (cmdkp.org): **Gene pages for all genes in Tables 1 and 2.** (Date accessed 2022 Jun 25 to 2022 Jul 23).
Reference Source

19. Bindea G, Mlecnik B, Hackl H, *et al.*: **ClueGO: a Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks.** *Bioinformatics.* 2009; **25**(8): 1091–3.
PubMed Abstract | Publisher Full Text | Free Full Text

20. Croft D, Mundo AF, Haw R, *et al.*: **The Reactome pathway knowledgebase.** *Nucleic Acids Res.* 2014; **42**(Database issue): D472–7.
PubMed Abstract | Publisher Full Text | Free Full Text

21. Pers TH, Karjalainen JM, Chan Y, *et al.*: **Biological interpretation of genome-wide association studies using predicted gene functions.** *Nat Commun.* 2015; **6**: 5890.
PubMed Abstract | Publisher Full Text | Free Full Text

22. Marouli E, Graff M, Medina-Gomez C, *et al.*: **Rare and low-frequency coding variants alter human adult height.** *Nature.* 2017; **542**(7640): 186–190.
PubMed Abstract | Publisher Full Text | Free Full Text

23. Turcot V, Lu Y, Highland HM, *et al.*: **Protein-altering variants associated with body mass index implicate pathways that control energy intake and expenditure in obesity.** *Nat Genet.* 2018; **50**(1): 26–41.
PubMed Abstract | Publisher Full Text | Free Full Text

24. Fehrmann RSN, Karjalainen JM, Krajewska M, *et al.*: **Gene expression analysis identifies global gene dosage sensitivity in cancer.** *Nat Genet.* 2015; **47**(2): 115–25.
PubMed Abstract | Publisher Full Text

25. Frey BJ, Dueck D: **Clustering by passing messages between data points.** *Science.* 2007; **315**(5814): 972–6.
PubMed Abstract | Publisher Full Text

26. Abraham A, Pedregosa F, Eickenberg M, *et al.*: **Machine learning for neuroimaging with scikit-learn.** *Front Neuroinform.* 2014; **8**: 14.
PubMed Abstract | Publisher Full Text | Free Full Text

27. Gu Z, Eils R, Schlesner M: **Complex heatmaps reveal patterns and correlations in multidimensional genomic data.** *Bioinformatics.* 2016; **32**(18): 2847–9.
PubMed Abstract | Publisher Full Text

28. Soranzo N, Sanna S, Wheeler E, *et al.*: **Common variants at 10 genomic loci influence hemoglobin A_{1c} levels via glycemic and nonglycemic pathways.** *Diabetes.* 2010; **59**(12): 3229–39.
PubMed Abstract | Publisher Full Text | Free Full Text

29. Chen WM, Erdos MR, Jackson AU, *et al.*: **Variations in the *G6PC2/ABCB11* genomic region are associated with fasting glucose levels.** *J Clin Invest.* 2008; **118**(7): 2620–8.
PubMed Abstract | Publisher Full Text | Free Full Text

30. Bouatia-Naji N, Rocheleau G, Van Lommel L, *et al.*: **A polymorphism within the *G6PC2* gene is associated with fasting plasma glucose levels.** *Science.* 2008; **320**(5879): 1085–8.
PubMed Abstract | Publisher Full Text

31. Flannick J, Thorleifsson G, Beer NL, *et al.*: **Loss-of-function mutations in *SLC30A8* protect against type 2 diabetes.** *Nat Genet.* 2014; **46**(4): 357–63.
PubMed Abstract | Publisher Full Text | Free Full Text

32. Dwivedi OP, Lehtovirta M, Hastoy B, *et al.*: **Loss of ZnT8 function protects against diabetes by enhanced insulin secretion.** *Nat Genet.* 2019; **51**(11): 1596–1606.
PubMed Abstract | Publisher Full Text | Free Full Text

33. Kleiner S, Gomez D, Megra B, *et al.*: **Mice harboring the human *SLC30A8* R138X loss-of-function mutation have increased insulin secretory capacity.** *Proc Natl Acad Sci U S A.* 2018; **115**(32): E7642–E7649.
PubMed Abstract | Publisher Full Text | Free Full Text

34. Kuo T, Kraakman MJ, Damle M, *et al.*: **Identification of *C2CD4A* as a human diabetes susceptibility gene with a role in β cell insulin secretion.** *Proc Natl Acad Sci U S A.* 2019; **116**(40): 20033–20042.
PubMed Abstract | Publisher Full Text | Free Full Text

35. Kycia I, Wolford BN, Huyghe JR, *et al.*: **A Common Type 2 Diabetes Risk Variant Potentiates Activity of an Evolutionarily Conserved Islet Stretch Enhancer and Increases *C2CD4A* and *C2CD4B* Expression.** *Am J Hum Genet.* 2018; **102**(4): 620–635.
PubMed Abstract | Publisher Full Text | Free Full Text

36. O'Hare EA, Yerges-Armstrong LM, Perry JA, *et al.*: **Assignment of Functional Relevance to Genes at Type 2 Diabetes-Associated Loci Through Investigation of β-Cell Mass Deficits.** *Mol Endocrinol.* 2016; **30**(4): 429–45.
PubMed Abstract | Publisher Full Text | Free Full Text

37. Mehta ZB, Fine N, Pullen TJ, *et al.*: **Changes in the expression of the type 2 diabetes-associated gene *VPS13C* in the β-cell are associated with glucose intolerance in humans and mice.** *Am J Physiol Endocrinol Metab.* 2016; **311**(2): E488–507.
PubMed Abstract | Publisher Full Text | Free Full Text

38. Sakaue S, Kanai M, Tanigawa Y, *et al.*: **A cross-population atlas of genetic associations for 220 human phenotypes.** *Nat Genet.* 2021; **53**(10): 1415–1424.
PubMed Abstract | Publisher Full Text

39. Franz M, Rodriguez H, Lopes C, *et al.*: **GeneMANIA update 2018.** *Nucleic Acids Res.* 2018; **46**(W1): W60–W64.
PubMed Abstract | Publisher Full Text | Free Full Text

40. Cohen RM, Franco RS, Khera PK, *et al.*: **Red cell life span heterogeneity in hematologically normal people is sufficient to alter HbA1c.** *Blood.* 2008; **112**(10): 4284–91.
PubMed Abstract | Publisher Full Text | Free Full Text

41. Wheeler E, Leong A, Liu CT, *et al.*: **Impact of common genetic determinants of Hemoglobin A1c on type 2 diabetes risk and diagnosis in ancestrally diverse populations: A transethnic genome-wide meta-analysis.** *PLoS Med.* 2017; **14**(9): e1002383.
PubMed Abstract | Publisher Full Text | Free Full Text

42. Hanson D, Murray PG, Sud A, *et al.*: **The primordial growth disorder 3-M syndrome connects ubiquitination to the cytoskeletal adaptor OBSL1.** *Am J Hum Genet.* 2009; **84**(6): 801–6.
PubMed Abstract | Publisher Full Text | Free Full Text

43. Surendran P, Feofanova EV, Lahrouchi N, *et al.*: **Discovery of rare variants associated with blood pressure regulation through meta-analysis of 1.3 million individuals.** *Nat Genet.* 2020; **52**(12): 1314–1332.
PubMed Abstract | Publisher Full Text | Free Full Text

44. Karczewski KJ, Solomonson M, Chao KR, *et al.*: **Systematic single-variant and gene-based association testing of thousands of phenotypes in 394,841 UK Biobank exomes.** *Cell Genom.* 2022; **2**(9): 100168.
PubMed Abstract | Publisher Full Text | Free Full Text

45. Astle WJ, Elding H, Jiang T, *et al.*: **The Allelic Landscape of Human Blood Cell Trait Variation and Links to Common Complex Disease.** *Cell.* 2016; **167**(5): 1415–1429.e19.
PubMed Abstract | Publisher Full Text | Free Full Text

46. Lei KJ, Chen YT, Chen H, *et al.*: **Genetic basis of glycogen storage disease type 1a: prevalent mutations at the glucose-6-phosphatase locus.** *Am J Hum Genet.* 1995; **57**(4): 766–71.
PubMed Abstract | Free Full Text

47. Chou JY, Mansfield BC: **Mutations in the glucose-6-phosphatase-α (G6PC) gene that cause type Ia glycogen storage disease.** *Hum Mutat.* 2008; **29**(7): 921–30.
PubMed Abstract | Publisher Full Text | Free Full Text

48. Sundaram RC, Selvaraj N, Vijayan G, *et al.*: **Increased plasma malondialdehyde and fructosamine in iron deficiency anemia: effect of treatment.** *Biomed Pharmacother.* 2007; **61**(10): 682–5.
PubMed Abstract | Publisher Full Text

Open Peer Review

Current Peer Review Status:

Version 1

Reviewer Report 13 September 2024

https://doi.org/10.21956/wellcomeopenres.20795.r96202

Eiji Kutoh
[1] Gyoda General Hospital, Saitama, Japan
[2] Higashitotsuka Memorial Hospital, Yokohama, Japan

I have completed the review of the research paper titled **"Large-scale exome array summary statistics resources for glycemic traits to aid effector gene prioritization"** by Dr. Willems et al. Below are my comments:

-Overall, the paper is well-written and provides a clear presentation of the background, methods, results, and conclusions. However, there are several points that I would like to address:

-The main conclusions are clear, but their reproducibility might be questionable. The authors could strengthen their findings by validating the results using alternative approaches, preferably not relying solely on in silico methods but incorporating real experiments, such as molecular and cellular biology techniques.

-There are too many authors listed, and it is unclear "who did what." Authorship could be perceived as casual, with individuals included without clear contributions. However, I do not intend to intervene on this matter.

-Are there any potential biases or confounding factors that could have influenced the results? The inclusion and exclusion criteria might introduce selection bias, especially given the narrowly defined population.

-While the study includes participants from various ancestral groups, it is unclear how these individuals were selected or why the study predominantly focuses on certain ethnic groups (85% European). This could introduce bias and raises questions about the robustness of the findings in non-European populations. Will certain associations be more detectable in one group over another? What challenges arise from the smaller representation of other ethnicities? The generalizability of the findings is thus questionable, especially regarding whether ancestry might affect the genetic associations.

-The justification for using exome arrays is unclear. Are there specific advantages of exome arrays compared to whole-exome sequencing? This method is not considered cutting-edge technology.

-The descriptions of phenotypes are somewhat vague. While the metrics for FG, FI, 2hGlu, and HbA1c are provided, there is little detail on how these were measured across different cohorts. Was there any standardization across cohorts?

-The distance-based clumping method for defining loci (500 kb apart) lacks explanation. Why was this particular threshold chosen? Could it exclude significant associations that are closer together?

-What are the clinical interpretations and implications of these results? This aspect seems to be missing. The paper heavily emphasizes in silico data (statistical and computational findings), but more context is needed regarding the physiological and biological significance of the identified gene sets for glycemic traits. For instance, the results mention variants associated with traits but do not thoroughly discuss the clinical relevance or potential functional implications of these variants. How might the novel missense variant rs146886108 in ANKH, for example, influence FG or T2DM risk?

-The exclusion of individuals with diabetes is mentioned, but the rationale could be elaborated upon. Could this exclusion introduce bias? Does it ensure that the identified associations are specific to glycemic traits in non-diabetic individuals?

-The authors report identifying 62 distinct coding variant associations at 58 genes with exome-wide significance. However, there is little detail on the methods used to control for false positives beyond the Bonferroni correction threshold of $P<2.2\times10^{-7}$.

-The association with HbA1c is considered significant, but there is insufficient discussion about potential confounding factors, such as the influence of red blood cell (RBC) traits on HbA1c. I would suggest softening the interpretation of SV2C as an effector gene, given the complex relationship between RBC traits and HbA1c. Experienced clinicians in this field would likely agree that it is not appropriate to base conclusions about glycemic control solely on HbA1c levels.

Is the work clearly and accurately presented and does it cite the current literature?

Yes

Is the study design appropriate and is the work technically sound?

Partly

Are sufficient details of methods and analysis provided to allow replication by others?

Partly

If applicable, is the statistical analysis and its interpretation appropriate?

Partly

Are all the source data underlying the results available to ensure full reproducibility?

Yes

Are the conclusions drawn adequately supported by the results?

Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: diabetology, molecular endocrinology, molecular and cellular biology

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Author Response 27 Sep 2024

Sara Willems

We would like to thank Dr. Kutoh very much for his time and effort! Below our answers to the points raised:

- Our aim here was to generate exome array summary statistics resources to help effector gene prioritization for further (also other than in silico) research. Historically, large-scale meta-analyses results such as those we have shared here have stood the test of time and findings have been widely reproducible, In addition, by making our results publicly available we are ensuring that others can use the data and test their reproducibility.

- Since up to 66 cohorts where included in the meta-analyses presented in our study, many people have substantially contributed to collecting and analysing individual cohort data. More details of the contributions can be found under the tab ´Authors´.

- In this study, we only asked the contributing cohorts to exclude individuals with diagnosed or treated diabetes from the analyses. We did this to exclude any consequence of diabetes treatments or related interventions on the quantitative gycaemic traits that we analysed. To controll for confounding by BMI, all analyses were adjusted for BMI. We have previously demonstrated that collider bias did not significantly affect results with BMI adjustment (Chen J, Spracklen CN, Marenne G, et al.: The trans-ancestral genomic architecture of glycemic traits. Nat Genet. 2021; 53(6): 840–860). Furthermore, gene discovery studies on glycaemic traits using the same inclusion / exclusion criteria and covariates as we did have proven valuable in discovering loci influencing glycaemic traits, a subset of which also influence risk of type 2 diabetes.

- We asked all cohorts with the required data that we knew of at the time of this study to participate. We agree with the reviewer that this study still has over-representation of participants of recent European ancestry. Unfortunately this is a well recognised problem in the broader field of human genetics, and one we tried to mitigate by reaching out to studies that had data from participants on non-European ancestry. Because of different allele frequencies in different ancestries, statistical power for detection of associations can indeed be different in different ancestries. For more on this topic and the value of multiple ancestry analyses, please see our study Chen J, Spracklen CN, Marenne G, *et al.*: The trans-ancestral genomic architecture of glycemic traits. *Nat Genet.* 2021; **53**(6): 840–860.

- At the time of the study, this technology (exome array) was significantly cheaper and easier to implement than whole-exome sequencing, which also made it possible to implement in studies that did not have the resources to undertake whole-exome sequencing. We found it really worthwhile to analyse these data, since it contains a very interesting collection of variants (see our Methods section).

- We asked all cohorts to provide information on their collection method, assay and sample and for insulin additionally the assay sensitivity (see Supplementary Table S1). We asked cohorts to use plasma values for the analyses. If glucose measurement was made in blood, values were adjusted multiplying by 1.13, since plasma values are about 10-15% higher than blood values.

- This (500 kb) is a common threshold in genetic association studies, since variants that are closer together are very likely to be in high LD and thus to represent the same genetic locus. To make sure we didn´t miss distinct variant associations that are closer together at novel loci, we used Raremetal v 4.12.8 to perform analyses conditioning on the most significant variant at the locus and then looked for other significantly associated variants at that locus. These analyses were repeated by including the next most significant and distinct associated variant until no exome- or genome-wide significantly-associated variants were left at the locus. Additionally, gene-based analyses were performed aggregating all variants fullfilling mask criteria (see our Methods section). This was done for all genes with at least 2 variants fulfilling these criteria.

- Our main aim here was to generate exome array summary statistics resources to help effector gene prioritization for further (also other than in silico) research. However, to gain further biological insights, we also used the summary statistics to perform pathway analyses. These identified pathways related to processes like insulin secretion, zinc transport, fatty acid metabolism and, for HbA1c associations, a strong enrichment in pathways related to blood cell biology (for more details on these results, please see our results section). Apart from gaining insight into the biology underlying each glycemic trait, these analyses may further help the prioritization of specific genes or pathways for further research on these important questions raised by the reviewer on clinical interpretations and implications of our results. -The reviewer raises the point ´The exclusion of individuals with diabetes is mentioned, but the rationale could be elaborated upon. Could this exclusion introduce bias? Does it ensure that the identified associations are specific to glycemic traits in non-diabetic individuals?´. Here we refer to the answer regarding biases above, which also includes this point.

- In GWAS analyses (mainly identifying common non-coding variant associations), replication studies have often been performed to additionally control for false positives. To increase power (also to detect potential rarer coding variant associations), we choose to make our discovery cohort as large as possible. In addition, historically, as mentioned above, large-scale meta-analyses results such as those we have shared here have stood the test of time and findings have been widely reproducible. And by making our results publicly available we are ensuring that others can use the data and test their reproducibility.

- We feel we sufficiently acknowledge the influence of red blood cell biology on HbA1c levels and don´t base conclusions about glycemic control solely on HbA1c analyses. For example in the pathway analyses, we describe the strong enrichment for blood-related pathways mainly driven by HbA1c-associated variants and, to disentangle blood cell turnover from effects due to glycation, repeated analyses excluding variants that were significantly associated with HbA1c only. Also regarding SV2C, we describe it´s associations with red blood cell traits and lack of association with other glycemic traits, suggesting its effect on HbA1c is independent of glycemia.

Competing Interests: No competing interests were disclosed.

Reviewer Report 22 January 2024

https://doi.org/10.21956/wellcomeopenres.20795.r69837

Toshimasa Yamauchi
Department of Diabetes and Metabolic Diseases, University of Tokyo Graduate School of Medicine, Tokyo, Japan

In the presented study, the authors have undertaken a comprehensive exome-wide association study (ExWAS) to identify genetic loci linked to glycemic traits. A characteristic aspect of this research is the utilization of ExWAS meta-analysis to examine variants in coding regions, an approach that complements previous studies emphasizing non-coding variants. By analyzing data from a large participant pool, predominantly of European ancestry, the study pinpoints single coding variants and gene-based associations that could act as potential effector genes for glycemic traits such as glycated hemoglobin (HbA1c), fasting glucose (FG), fasting insulin (FI), and 2hr glucose post-oral glucose challenge (2hGlu). Additionally, the study extends to pathway analyses, offering insights into gene sets regulating these traits. The transparency and accessibility of the study are beneficial to the research community, with summary statistics made available on their website and through the GWAS catalog.

The study's methodology, while not novel, adheres to established conventions in the field, ensuring a foundation for their analyses. The discovery of a modest number of new loci and genes associated with glycemic traits, though limited in quantity, is worth reporting. These findings include the identification of four variants in three genes that represent novel associations, underscoring the potential for uncovering new pathways in glycemic regulation. The gene-based analysis further highlights six genes, including G6PC and TF, previously unlinked to glycemic traits.

The findings, while not groundbreaking, are biologically consistent. The study reveals a notable enrichment in blood-related pathways, especially those involving erythrocyte differentiation and heme metabolic processes. This enrichment, predominantly driven by HbA1c-associated variants, underscores the multifaceted influence on HbA1c levels, which are affected by both glycation and blood cell turnover. By excluding variants solely associated with HbA1c, the researchers effectively isolated 128 significant gene sets within 53 meta-gene sets (FDR <0.05). This refinement of analysis illuminated pathways more directly related to pancreatic islet function and metabolic tissues. These pathways, including “abnormal glucose homeostasis”, “peptide hormone secretion”, and “Maturity Onset Diabetes of the Young”, as well as those involved in glycogen regulation, incretin function, and carbohydrate metabolism, align with findings from fasting glucose-only analyses. Such insights could enhance our understanding of the complex genetic and biological mechanisms underlying glycemic control.

Is the work clearly and accurately presented and does it cite the current literature?
Yes

Is the study design appropriate and is the work technically sound?
Yes

Are sufficient details of methods and analysis provided to allow replication by others?
Yes

If applicable, is the statistical analysis and its interpretation appropriate?
Yes

Are all the source data underlying the results available to ensure full reproducibility?
Yes

Are the conclusions drawn adequately supported by the results?
Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Diabetes, Obesity, Genetics

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.