An expanded genome-wide association study of type 2 diabetes in Europeans

Running title: **European T2D genome-wide association study**

Robert A Scott[1*], Laura J Scott[2*], Reedik Mägi[3*], Letizia Marullo[4*], Kyle J Gaulton[5-6*], Marika Kaakinen[7*], Natalia Pervjakova[3], Tune H Pers[8-11], Andrew D Johnson[12], John D Eicher[12], Anne U Jackson[2], Teresa Ferreira[5], Yeji Lee[2], Clement Ma[2], Valgerdur Steinthorsdottir[13], Gudmar Thorleifsson[13], Lu Qi[14-16], Natalie R Van Zuydam[5,17], Anubha Mahajan[5], Han Chen[18-19], Peter Almgren[20], Ben F Voight[21-23], Harald Grallert[24-26], Martina Müller-Nurasyid[27-30], Janina S Ried[27], William N Rayner[5,31-32], Neil Robertson[5,31], Lennart C Karssen[33-34], Elisabeth M van Leeuwen[33], Sara M Willems[1,33], Christian Fuchsberger[2], Phoenix Kwan[2], Tanya M Teslovich[2], Pritam Chanda[35], Man Li [36], Yingchang Lu[37-38], Christian Dina[39], Dorothee Thuillier[40-41], Loic Yengo[40-41], Longda Jiang[7], Thomas Sparso[10], Hans A Kestler[42-43], Himanshu Chheda[44], Lewin Eisele[45], Stefan Gustafsson[46], Mattias Frånberg[47-49], Rona J Strawbridge[47], Rafn Benediktsson[50-51], Astradur B Hreidarsson[51], Augustine Kong[13], Gunnar Sigurðsson[51-52], Nicola D Kerrison[1], Jian'an Luan[1], Liming Liang[14,53], Thomas Meitinger[30,54-55], Michael Roden[56-58], Barbara Thorand[25,57], Tõnu Esko[3,59-60], Evelin Mihailov[3], Caroline Fox[61-62], Ching-Ti Liu[63], Denis Rybin[64], Bo Isomaa[65-66], Valeriya Lyssenko[20], Tiinamaija Tuomi[65,67], David J Couper[68], James S Pankow[69], Niels Grarup[10], Christian T Have[10], Marit E Jørgensen[70], Torben Jørgensen[71-73], Allan Linneberg[71,74-75], Marilyn C Cornelis[76], Rob M van Dam[15,77], David J Hunter[14-15,78-79], Peter Kraft[14,53,79], Qi Sun[15,78], Sarah Edkins[32], Katharine R Owen[31,80], John RB Perry[1], Andrew R Wood[81], Eleftheria Zeggini[32], Juan Tajes-Fernandes[5], Goncalo R Abecasis[2], Lori L Bonnycastle[82], Peter S Chines[82], Heather M Stringham[2], Heikki A Koistinen[83-85], Leena Kinnunen[83-85], Bengt Sennblad[47-48], Thomas W Mühleisen[86-87], Markus M Nöthen[86-87], Sonali Pechlivanis[45], Damiano Baldassarre[88-89], Karl Gertow[47], Steve E Humphries[90], Elena Tremoli[88-89], Norman Klopp[24,91], Julia Meyer[27], Gerald Steinbach[92], Roman Wennauer[93], Johan G Eriksson[65,94-96], Satu Männistö[94], Leena Peltonen[32,44,94,97,149], Emmi Tikkanen[44,98], Guillaume Charpentier[99], Elodie Eury[41], Stéphane Lobbens[41], Bruna Gigante[100], Karin Leander[100], Olga McLeod[47], Erwin P Bottinger[37], Omri Gottesman[37], Douglas Ruderfer[101], Matthias Blüher[102-103], Peter Kovacs[102-103], Anke Tonjes[102-103], Nisa M Maruthur[36,104-105], Chiara Scapoli[4], Raimund Erbel[45], Karl-Heinz Jöckel[45], Susanne Moebus[45], Ulf de Faire[100], Anders Hamsten[47], Michael Stumvoll[102-103], Panagiotis Deloukas[32,106], Peter J Donnelly[5,107], Timothy M Frayling[81], Andrew T Hattersley[108], Samuli Ripatti[32,44,98,109], Veikko Salomaa[83], Nancy L Pedersen[110], Bernhard O Boehm[111-112], Richard N Bergman[113], Francis S Collins[82], Karen L Mohlke[114], Jaakko Tuomilehto[94,115-117], Torben Hansen[10,118], Oluf Pedersen[10], Inês Barroso[32,119], Lars Lannfelt[120], Erik Ingelsson[46,121], Lars Lind[122], Cecilia M Lindgren[5,97], Stephane Cauchi[40], Philippe Froguel[7,40-41], Ruth JF Loos[37-38,123], Beverley Balkau[124-125], Heiner Boeing[126], Paul W Franks[127-128], Aurelio Barricarte Gurrea[129-131], Domenico Palli[132], Yvonne T van der Schouw[133], David Altshuler[97,134-138], Leif C Groop[20,44], Claudia Langenberg[1], Nicholas J Wareham[1], Eric Sijbrands[93], Cornelia M van Duijn[33,139], Jose C Florez[60,135,140], James B Meigs[8,135,141], Eric Boerwinkle[142-143], Christian Gieger[24-25], Konstantin Strauch[27,29], Andres Metspalu[3,144], Andrew D Morris[145], Colin NA Palmer[17,146], Frank B Hu[14-16], Unnur Thorsteinsdottir[13,50], Kari Stefansson[13,50], Josée Dupuis[61,63], Andrew P Morris[3,5,147-148#], Michael Boehnke[2#], Mark I McCarthy[5,31,80#], Inga Prokopenko [5,7,31#] for the DIAbetes Genetics Replication And Meta-analysis (DIAGRAM) Consortium.

Diabetes Publish Ahead of Print, published online May 31, 2017

Affiliations

1)MRC Epidemiology Unit, University of Cambridge, Cambridge, United Kingdom; 2)Department of Biostatistics and Center for Statistical Genetics, University of Michigan, Ann Arbor, MI, USA; 3)Estonian Genome Center, University of Tartu, Tartu, Estonia; 4)Department of Life Sciences and Biotechnology, University of Ferrara, Ferrara, Italy; 5)Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK; 6)Department of Genetics, Stanford University, Stanford, CA, USA; 7)Department of Genomics of Common Disease, Imperial College London, London, UK; 8)Medical and Population Genetics Program, Broad Institute of MIT and Harvard, Cambridge, 02142, USA; 9)Division of Endocrinology and Center for Basic and Translational Obesity Research, Boston Children's Hospital, Boston, 02115, USA; 10)The Novo Nordisk Foundation Center for Basic Metabolic Research, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark; 11)Department of Epidemiology Research, Statens Serum Institut, Copenhagen, Denmark; 12)National Heart, Lung and Blood Institute's The Framingham Heart Study, Population Sciences Branch, Division of Intramural Research, Framingham MA 01702 USA; 13)deCODE Genetics, Amgen inc., Reykjavik, Iceland; 14)Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, MA, USA; 15)Department of Nutrition, Harvard T.H. Chan School of Public Health, Boston, MA, USA; 16)Channing Laboratory, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA, USA; 17)Pharmacogenomics Centre, Biomedical Research Institute, University of Dundee, Ninewells Hospital, Dundee, UK; 18)Human Genetics Center, Department of Epidemiology, Human Genetics and Environmental Sciences, School of Public Health, The University of Texas Health Science Center at Houston, Houston, TX, USA; 19)Center for Precision Health, Schools of Public Health and Biomedical Informatics, The University of Texas Health Science Center at Houston, Houston, TX, USA; 20)Lund University Diabetes Centre, Department of Clincal Science Malmö, University Hospital Scania, Lund University, 20502 Malmö, Sweden ; 21)University of Pennsylvania - Perelman School of Medicine, Department of Pharmacology, Philadelphia PA, 19104, USA; 22)University of Pennsylvania - Perelman School of Medicine, Department of Genetics, Philadelphia PA, 19104, USA; 23)Institute of Translational Medicine and Therapeutics, University of Pennsylvania - Perelman School of Medicine, Philadelphia PA 19104, USA; 24)Research Unit of Molecular Epidemiology, Helmholtz Zentrum Muenchen, German Research Centre for Environmental Health, Neuherberg, Germany; 25)Institute of Epidemiology II, Helmholtz Zentrum Muenchen, German Research Center for Environmental Health, Neuherberg, Germany; 26)German Center for Diabetes Research, Neuherberg, Germany; 27)Institute of Genetic Epidemiology, Helmholtz Zentrum Muenchen, German Research Center for Environmental Health, Neuherberg, Germany; 28)Department of Medicine I, University Hospital Grosshadern, Ludwig-Maximilians-Universität, Munich, Germany; 29)Institute of Medical Informatics, Biometry and Epidemiology, Chair of Genetic Epidemiology, Ludwig-Maximilians-Universität, Munich, Germany; 30)German Center for Cardiovascular Disease (DZHK), Munich Heart Alliance, Munich, Germany; 31)Oxford Centre for Diabetes, Endocrinology and Metabolism, University of Oxford, Oxford, UK; 32)Wellcome Trust Sanger Institute, Hinxton, UK; 33)Department of Epidemiology, Erasmus University Medical Center, Rotterdam, The Netherlands; 34)PolyOmica, 's-Hertogenbosch, The Netherlands; 35)High Throughput Biology Center, Johns Hopkins University School of Medicine, Baltimore, MD, USA; 36)Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, USA; 37)The Charles Bronfman Institute for Personalized Medicine, Icahn School of Medicine at Mount Sinai, New York, NY, USA; 38)The Genetics of Obesity and Related Metabolic Traits Program, The Icahn School of Medicine at Mount Sinai, New York,

NY, USA; 39)l'Institut du thorax, INSERM, CNRS, Université de Nantes, Centre Hospitalier Universitaire (CHU) Nantes, Nantes, France; 40)Lille Institute of Biology, European Genomics Institute of Diabetes, Lille, France; 41)CNRS-UMR-8199, Institute of Biology and Lille 2 University, Pasteur Institute, Lille, France; 42)Leibniz Institute on Aging, Fritz Lipmann Institute, Jena, Germany; 43)Institute of Medical Systems Biology, Ulm University, Germany; 44)Institute for Molecular Medicine Finland (FIMM), University of Helsinki, Helsinki, Finland; 45)Institute for Medical Informatics, Biometry and Epidemiology, University Hospital of Essen, Essen, Germany; 46)Department of Medical Sciences, Molecular Epidemiology and Science for Life Laboratory, Uppsala University, Uppsala, Sweden; 47)Cardiovascular Medicine Unit, Department of Medicine Solna, Karolinska Institutet, Stockholm, Sweden; 48)Science for Life Laboratory, Stockholm, Sweden; 49)Department for Numerical Analysis and Computer Science, Stockholm University, Stockholm, Sweden; 50)Faculty of Medicine, University of Iceland, Reykjavik, Iceland; 51)Landspitali University Hospital, Reykjavik, Iceland; 52)Icelandic Heart Association, Kopavogur, Iceland; 53)Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, MA, USA; 54)Institute of Human Genetics, Helmholtz Zentrum München, Neuherberg, Germany; 55)Institute of Human Genetics, Technische Universität München, Munich, Germany ; 56)Department of Endocrinology and Diabetology, Medical Faculty, Heinrich-Heine University, Düsseldorf, Germany; 57)German Center for Diabetes Research (DZD e.V.), München-Neuherberg, Germany; 58)Institute for Clinical Diabetology, German Diabetes Center, Leibniz Institute for Diabetes Research at Heinrich Heine University Düsseldorf, Düsseldorf, Germany; 59)Division of Genetics and Endocrinology, Children's Hospital, Boston, MA, USA; 60)Program in Medical and Population Genetics, Broad Institute, Cambridge, MA, USA; 61)National Heart, Lung, and Blood Institute's Framingham Heart Study, Framingham, MA, USA; 62)Division of Endocrinology and Metabolism, Brigham and Women's Hospital and Harvard Medical School, Boston, MA, USA; 63)Department of Biostatistics, Boston University School of Public Health, Boston, MA, 02118, USA; 64)Boston University Data Coordinating Center, Boston, MA, USA; 65)Folkhälsan Research Center, FIN-00014 Helsinki, Finland; 66)Department of Social Services and Health Care, 68601 Jakobstad, Finland; 67)Department of Medicine, Helsinki University Hospital, University of Helsinki, 000290 HUS Helsinki , Finland; 68)Collaborative Studies Coordinating Center, Department of Biostatistics, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA; 69)Division of Epidemiology and Community Health, University of Minnesota, Minneapolis, MN, USA; 70)Steno Diabetes Center, Gentofte, Denmark; 71)Research Centre for Prevention and Health, Capital Region of Denmark, Copenhagen, Denmark; 72)Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark; 73)Faculty of Medicine, University of Aalborg, Aalborg, Denmark; 74)Copenhagen University Hospital, Rigshospitalet, Denmark; 75)Department of Clinical Medicine, Faculty of Health and Medical Sciences, University of Copenhagen, Denmark; 76)Department of Preventive Medicine, Northwestern University Feinberg School of Medicine, Chicago, IL, USA; 77)Saw Swee Hock School of Public Health, National University of Singapore, Singapore, Singapore; 78)Channing Division of Network Medicine, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA, USA; 79)Program in Genetic Epidemiology and Statistical Genetics, Harvard T.H. Chan School of Public Health, Boston, MA, USA; 80)Oxford National Institute for Health Research Biomedical Research Centre, Churchill Hospital, Oxford, UK; 81)Genetics of Complex Traits, University of Exeter Medical School, University of Exeter, Exeter, UK; 82)National Human Genome Research Institute, National Institutes of Health, Bethesda, MD, USA; 83)Department of Health, National Institute for Health and Welfare, Helsinki, Finland; 84)University of Helsinki and Helsinki University Central Hospital: Department of Medicine and Abdominal Center: Endocrinology, Helsinki,

Finland; 85)Minerva Foundation Institute for Medical Research, Biomedicum 2U, Helsinki, Finland; 86)Institute of Human Genetics, University of Bonn, Bonn, Germany; 87)Department of Genomics, Life & Brain Center, University of Bonn, Bonn, Germany; 88)Centro Cardiologico Monzino, Istituto di Ricovero e Cura a Carattere Scientifico (IRCCS), Milan, Italy; 89)Dipartimento di Scienze Farmacologiche e Biomolecolari, Università di Milano, Milan, Italy; 90)Cardiovascular Genetics, BHF Laboratories, Institute Cardiovascular Sciences, UCL, London, UK; 91)Hannover Unified Biobank, Hannover Medical School, Hannover, Germany; 92)Department of Clinical Chemistry and Central Laboratory, University of Ulm, Ulm, Germany; 93)Department of Internal Medicine, Erasmus University Medical Center, Rotterdam, The Netherlands; 94)Department of Chronic Disease Prevention, National Institute for Health and Welfare, Helsinki, Finland; 95)Department of General Practice and Primary Health Care, University of Helsinki, Helsinki, Finland; 96)Unit of General Practice, Helsinki University Hospital, Helsinki, Finland; 97)Broad Institute of Harvard and Massachusetts Institute of Technology (MIT), Cambridge, MA 02142, USA; 98)Department of Public Health, Hjelt Institute, University of Helsinki, Helsinki, Finland; 99)Endocrinology-Diabetology Unit, Corbeil-Essonnes Hospital, Corbeil-Essonnes, France; 100)Division of Cardiovascular Epidemiology, Institute of Environmental Medicine, Karolinska Institutet, Stockholm, Sweden; 101)Division of Psychiatric Genomics, Department of Psychiatry, Icahn School of Medicine at Mount Sinai, New York, New York, USA; 102)IFB Adiposity Diseases, University of Leipzig, Leipzig, Germany; 103)Department of Medicine, University of Leipzig, Leipzig, Germany; 104)Department of Medicine, Division of General Internal Medicine, The Johns Hopkins Bloomberg School of Medicine, Baltimore, MD, USA; 105)The Welch Center for Prevention, Epidemiology, and Clinical Research, The Johns Hopkins University, Baltimore, MD, USA; 106)William Harvey Research Institute, Barts and The London School of Medicine and Dentistry, Queen Mary University London, London, UK; 107)Department of Statistics, University of Oxford, Oxford, UK; 108)Institute of Biomedical and Clinical Science, University of Exeter Medical School, Exeter, UK; 109)Public Health Genomics Unit, National Institute for Health and Welfare, Helsinki, Finland; 110)Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden; 111)Division of Endocrinology and Diabetes, Department of Internal Medicine, University Medical Centre Ulm, Ulm, Germany; 112)Lee Kong Chian School of Medicine, Imperial College London and Nanyang Technological University, Singapore, Singapore; 113)Diabetes and Obesity Research Institute, Cedars-Sinai Medical Center, Los Angeles, CA, USA; 114)Department of Genetics, University of North Carolina, Chapel Hill, NC, USA; 115)Dasman Diabetes Institute, Dasman, Kuwait; 116)Centre for Vascular Prevention, Danube-University Krems, Krems, Austria; 117)Diabetes Research Group, King Abdulaziz University, Jeddah, Saudi Arabia; 118)Faculty of Health Sciences, University of Southern Denmark, Odense, Denmark; 119)University of Cambridge Metabolic Research Laboratories and NIHR Cambridge Biomedical Research Centre, Level 4, WT-MRC Institute of Metabolic Science Box 289 Addenbrooke's Hospital Cambridge, Cambridge, UK; 120)Department of Public Health and Caring Sciences, Uppsala University, Uppsala, Sweden; 121)Department of Medicine, Division of Cardiovascular Medicine, Stanford University School of Medicine, Stanford, USA; 122)Department of Medical Sciences, Uppsala University Hospital, Uppsala, Sweden; 123)The Mindich Child Health and Development Institute, Icahn School of Medicine at Mount Sinai, New York, NY, USA; 124)Inserm, CESP, U1018, Villejuif, France; 125)Univ Paris-Sud, UMRS 1018, Villejuif, France; 126)German Institute of Human Nutrition Potsdam-Rehbruecke, Germany; 127)Lund University, Malmö, Sweden; 128)Umeå University, Umeå, Sweden; 129)Navarra Public Health Institute, Pamplona, Spain; 130)Navarra Institute for Health Research (IdiSNA), Pamplona, Spain; 131)CIBER Epidemiology and Public Health (CIBERESP), Madrid, Spain; 132)Cancer Research and

Prevention Institute (ISPO), Florence, Italy; 133)University Medical Center Utrecht, Utrecht, the Netherlands; 134)Center for Human Genetic Research, Massachusetts General Hospital, Boston, MA 02114, USA; 135)Department of Medicine, Harvard Medical School, Boston, MA 02115, USA; 136)Department of Genetics, Harvard Medical School, Boston, MA 02115, USA; 137)Department of Molecular Biology, Harvard Medical School, Boston, Massachusetts 02115, USA; 138)Diabetes Unit, Massachusetts General Hospital, Boston, MA 02144, USA; 139)Netherlands Genomics Initiative, Netherlands Consortium for Healthy Ageing and Center for Medical Systems Biology, Rotterdam, The Netherlands; 140)Diabetes Unit and Center for Human Genetic Research, Massachusetts General Hospital, Boston, MA, USA; 141)General Medicine Division, Massachusetts General Hospital, Boston, MA, USA; 142)Human Genetics Center, University of Texas Health Science Center at Houston, Houston, TX, USA; 143)Human Genome Sequencing Center at Baylor College of Medicine, Houston, TX, USA; 144)Institute of Molecular and Cell Biology, University of Tartu, Tartu, Estonia; 145)Usher Institute of Population Health Sciences and Informatics, University of Edinburgh, Edinburgh, UK; 146)Cardiovascular and Diabetes Medicine, Biomedical Research Institute, University of Dundee, Ninewells Hospital, Dundee, UK; 147)Department of Biostatistics, University of Liverpool, Liverpool, UK; 148)Department of Molecular and Clinical Pharmacology, University of Liverpool, Liverpool, UK; 149) Deceased.

* These authors contributed equally to this research.
# These authors jointly directed this research.

Correspondence should be addressed to:

Dr. Inga Prokopenko
Department of Genomics of Common Disease
School of Public Health, Imperial College London
Burlington Danes Building, Hammersmith Campus,
Du Cane Road, London, W12 0NN, UK
Phone: +4420 759 46501
E-mail: i.prokopenko@imperial.ac.uk

Prof. Mark I. McCarthy
OCDEM, Churchill Hospital, University of Oxford
Old Road, Headington, OX3 7LJ, UK
Email: mark.mccarthy@drl.ox.ac.uk

Prof. Michael Boehnke

Department of Biostatistics and Center for Statistical Genetics
University of Michigan
School of Public Health
1415 Washington Heights
Ann Arbor, MI 48109-2029
Email: boehnke@umich.edu

Abstract word count: 198

Main text word count: 4257

Figures: 3

Tables: 1

References: 51

ABSTRACT

To characterise type 2 diabetes (T2D) associated variation across the allele frequency spectrum, we conducted a meta-analysis of genome-wide association data from 26,676 T2D cases and 132,532 controls of European ancestry after imputation using the 1000 Genomes multi-ethnic reference panel. Promising association signals were followed-up in additional data sets (of 14,545 or 7,397 T2D cases and 38,994 or 71,604 controls). We identified 13 novel T2D-associated loci ($p<5\times10^{-8}$), including variants near the *GLP2R*, *GIP*, and *HLA-DQA1* genes. Our analysis brought the total number of independent T2D associations to 128 distinct signals at 113 loci**.** Despite substantially increased sample size and more complete coverage of low-frequency variation, all novel associations were driven by common SNVs. Credible sets of potentially causal variants were generally larger than those based on imputation with earlier reference panels, consistent with resolution of causal signals to common risk haplotypes. Stratification of T2D-associated loci based on T2D-related quantitative trait associations revealed tissue-specific enrichment of regulatory annotations in pancreatic islet enhancers for loci influencing insulin secretion, and in adipocytes, monocytes and hepatocytes for insulin action-associated loci. These findings highlight the predominant role played by common variants of modest effect and the diversity of biological mechanisms influencing T2D pathophysiology.

MAIN TEXT

Type 2 diabetes (T2D) has rapidly increased in prevalence in recent years and represents a major component of the global disease burden (1). Previous efforts to use genome-wide association studies (GWAS) to characterise the genetic component of T2D risk have largely focused on common variants (minor allele frequency [MAF]>5%). These studies have identified close to 100 loci, almost all of them currently defined by common alleles associated with modest (typically 5-20%) increases in T2D risk (2–6). Direct sequencing of whole genomes or exomes offers the most comprehensive approach for extending discovery efforts to the detection of low-frequency (0.5%<MAF<5%) and rare (MAF<0.5%) risk and protective alleles, some of which might have greater impact on individual predisposition. However, extensive sequencing has, thus far, been limited to relatively small sample sizes (at most, a few thousand cases), restricting power to detect rarer risk alleles, even if they are of large effect (7–9). Whilst evidence of rare variant associations has been detected in some candidate gene studies (10,11), the largest study to date, involving exome sequencing in ~13,000 subjects, found little trace of rare variant association effects (9).

Here, we implement a complementary strategy that makes use of imputation into existing GWAS samples from the DIAbetes Genetics Replication And Meta-analysis (DIAGRAM) Consortium with sequence-based reference panels (12). This strategy allows the detection of common and low-frequency (but not rare) variant associations in extremely large samples (13), and facilitates the fine-mapping of causal variants. We performed a European ancestry meta-analysis of GWAS with 26,676 T2D cases and 132,532 controls, and followed up our findings in additional independent European ancestry studies of 14,545 T2D cases and 38,994 controls genotyped using the Metabochip (4). All contributing studies were imputed against the March 2012 multi-ethnic 1000 Genomes Project (1000G) reference panel of 1,092 whole-genome sequenced individuals (12). Our study provides near-complete evaluation of common

311 variants with much improved coverage of low-frequency variants, and the combined sample
312 size considerably exceeds that of the largest previous T2D GWAS meta-analyses in
313 individuals of European ancestry (4). In addition to genetic discovery, we fine-map novel and
314 established T2D-associated loci to identify regulatory motifs and cell types enriched for
315 potential causal variants, and pathways through which T2D-associated loci increase disease
316 susceptibility.

317 **RESEARCH DESIGN AND METHODS**

318 ***Research participants.*** The DIAGRAM stage 1 meta-analyses is comprised of 26,676 T2D
319 cases and 132,532 controls (effective sample size, N_{eff}=72,143 individuals, defined as
320 $4/[(1/N_{cases})+(1/N_{controls})]$) from 18 studies genotyped using commercial genome-wide single-
321 nucleotide variant (SNV) arrays (**Supplementary Table 1**). The Metabochip stage 2 follow
322 up is comprised of 14,545 T2D cases and 38,994 controls (N_{eff}=38,645) from 16 non-
323 overlapping stage 1 studies (4,14). We performed additional follow-up in 2,796 T2D cases
324 and 4,601 controls from the EPIC-InterAct (15) and 9,747 T2D cases and 61,857 controls
325 from the GERA study (16) (**Supplementary Material**).

326 ***Statistical analyses.*** We imputed autosomal and X chromosome SNVs using the all
327 ancestries 1000G reference panel (1,092 individuals from Africa, Asia, Europe, and the
328 Americas [March, 2012 release]) using miniMAC (17) or IMPUTE2 (18). After imputation,
329 from each study we removed monomorphic variants or those with imputation quality r^2-
330 hat<0.3 (miniMAC) or proper-info<0.4 (IMPUTE2, SNPTEST). Each study performed T2D
331 association analysis using logistic regression, adjusting for age, sex, and principal
332 components for ancestry, under an additive genetic model. We performed inverse-variance
333 weighted fixed-effect meta-analyses of the 18 stage 1 GWAS (**Supplementary Table 1**).
334 Fifteen of the 18 studies repeated analyses also adjusting for body mass index (BMI). SNVs
335 reaching suggestive significance $p<10^{-5}$ in the stage 1 meta-analysis were followed-up. Novel

loci were selected using the threshold for genome-wide significance ($p<5\times10^{-8}$) in the combined stage 1 and stage 2 meta-analysis. For the 23 variants with no proxy ($r^2\geq0.6$) available in Metabochip with 1000G imputation in the fine-mapping regions, the stage 1 result was followed-up in EPIC-InterAct and GERA (N_{eff}=40,637), both imputed to 1000G variant density (**Supplementary Material**).

Approximate conditional analysis with GCTA. We performed approximate conditional analysis in the stage 1 sample using GCTA v1.24 (19,20). We analysed SNVs in the 1Mb-window around each lead variant, conditioning on the lead SNV at each locus (**Supplementary Material**) (21). We considered loci to contain multiple distinct signals if multiple SNVs reached locus-wide significance ($p<10^{-5}$), accounting for the approximate number of variants in each 1Mb window (14).

Fine-mapping analyses using credible set mapping. To identify 99% credible sets of causal variants for each distinct association signal, we performed fine-mapping for loci at which the lead independent SNV reached $p<5\times10^{-4}$ in the stage 1 meta-analysis. We performed credible set mapping using the T2D stage 1 meta-analysis results to obtain the minimal set of SNVs with cumulative posterior probability>0.99 (**Supplementary Material**).

Type 1 diabetes (T1D)/T2D discrimination analysis. Given the overlap between loci previously associated with T1D and the associated T2D loci, we used an inverse variance weighted Mendelian randomisation approach (22) to test whether this was likely to reflect misclassification of T1D cases as individuals with T2D in the current study (**Supplementary Material**).

Expression quantitative trait locus (eQTL) analysis. To look for potential biological overlap of T2D lead variants and eQTL variants, we extracted the lead (most significantly associated) eQTL for each tested gene from existing datasets for a range of tissues (**Supplementary**

Material). We concluded that a lead T2D SNV showed evidence of association with gene expression if it was in high LD (r^2>0.8) with the lead eQTL SNV ($p<5\times10^{-6}$).

Hierarchical clustering of T2D-related metabolic phenotypes. Starting with the T2D associated SNVs, we obtained T2D-related quantitative trait Z-scores from published HapMap-based GWAS meta-analysis for: fasting glucose, fasting insulin adjusted for BMI, homeostasis model assessment for beta-cell function (HOMA-B), homeostasis model assessment for insulin resistance (HOMA-IR) (23); 2-hour glucose adjusted for BMI (24); proinsulin (25); corrected insulin response (CIR) (26); BMI (27); high density lipoprotein cholesterol (HDL-C), low density lipoprotein cholesterol (LDL-C), total cholesterol, and triglycerides (28). When an association result for a SNV was not available, we used the results for the variant in highest LD and only for variants with r^2>0.6. We performed clustering of phenotypic effects using Z-scores for association with T2D risk alleles and standard methods (**Supplementary Material**) (29).

Functional annotation and enrichment analysis. We tested for enrichment of genomic and epigenomic annotations using chromatin states for 93 cell types (after excluding cancer cell lines) from the NIH Epigenome Roadmap project, and binding sites for 165 transcription factors (TF) from ENCODE (30) and Pasquali et al. (31). Using fractional logistic regression, we then tested for the effect of variants with each cell type and TF annotation on the variant posterior probabilities (π_c) using all variants within 1Mb of the lead SNV for each distinct association signal from the fine-mapping analyses (**Supplementary Material**). In each analysis, we considered an annotation significant if it reached a Bonferroni-corrected $p<1.9\times10^{-4}$ (i.e. 0.05/258 annotations).

Pathway analyses with DEPICT. We used the Data-driven Expression Prioritized Integration for Complex Traits (DEPICT) tool (32) to (i) prioritize genes that may represent promising candidates for T2D pathophysiology, and (ii) identify reconstituted gene sets that are

enriched in genes from associated regions and might be related to T2D biological pathways. As input, we used independent SNVs from the stage 1 meta-analysis SNVs with $p<10^{-5}$ and lead variants at established loci (**Supplementary Material**). For the calculation of empirical enrichment *p* values, we used 200 sets of SNVs randomly drawn from entire genome within regions matching by gene density; we performed 20 replications for false discovery rate (FDR) estimation.

RESULTS

Novel loci detected in T2D GWAS and Metabochip-based follow-up. The stage 1 GWAS meta-analysis included 26,676 T2D cases and 132,532 controls and evaluated 12.1M SNVs, of which 11.8M were autosomal and 260k mapped to the X chromosome. Of these, 3.9M variants had MAF between 0.5% and 5%, a near fifteen-fold increase in the number of low-frequency variants tested for association compared to previous array-based T2D GWAS meta-analyses (2,4) (**Supplementary Table 2**). Of the 52 signals showing promising evidence of association ($p<10^{-5}$) in stage 1, 29 could be followed up in the stage 2 Metabochip data. In combined stage 1 and stage 2 data, 13 novel loci were detected at genome-wide significance (**Table 1, Figure 1, Supplementary Figure 1A-D, Supplementary Table 3**).

Lead SNVs at all 13 novel loci were common. Although detected here using 1000G imputed data, all 13 were well captured by variants in the HapMap CEU reference panel (2 directly, 10 via proxies with $r^2>0.8$, and one via proxy with $r^2=0.62$) (**Supplementary Materials**). At all 13, lead variants defined through 1000G and those seen when the SNP density was restricted to HapMap content, had broadly similar evidence of association and were of similar frequency (**Supplementary Figure 2; Supplementary Table 3**). Throughout this manuscript, loci are named for the gene nearest to the lead SNV, unless otherwise specified (**Table 1, Supplementary Materials: Biology box**).

Adjustment for BMI revealed no additional genome-wide significant associations for T2D and, at most known and novel loci, there were only minimal differences in statistical significance and estimated T2D effect size between BMI-adjusted and unadjusted models. The four signals at which we observed a significant effect of BMI adjustment ($p_{heterogeneity}<4.4\times10^{-4}$; based on 0.05/113 variants currently or previously reported to be associated with T2D at genome-wide significance) were *FTO* and *MC4R* (at which the T2D association is known to reflect a primary effect on BMI), and *TCF7L2* and *SLC30A8* (at which T2D associations were strengthened after BMI-adjustment) (**Supplementary Figure 3; Supplementary Table 4**).

Insights into genetic architecture of T2D. In this meta-analysis, we tested 3.9M low-frequency variants ($r^2\geq0.3$ or proper-info$\geq$0.4; minor allele present in $\geq$3 studies) for T2D association, constituting 96.7% of the low-frequency variants ascertained by the 1000G European Panel (March 2012) (**Supplementary Table 2**). For variants with risk-allele frequencies (RAF) of 0.5%, 1%, or 5%, we had 80% power to detect association ($p<5\times10^{-8}$) for allelic ORs of 1.80, 1.48, and 1.16, respectively, after accounting for imputation quality (**Figure 1, Supplementary Table 5**). Despite the increased coverage and sample size, we identified no novel low-frequency variants at genome-wide significance (**Figure 1**). Since we had only been able to test 29 of the 52 promising stage 1 signals on the Metabochip, we investigated whether this failure to detect low-frequency variant associations with T2D could be a consequence of selective variant inclusion on the Metabochip. Amongst the remaining 23 variants, none reached genome-wide significance after aggregating with GWAS data available from EPIC-InterAct. Six of these 23 SNVs had MAF<5%, and for these we performed additional follow-up in the GERA study. However, none reached genome-wide significance in a combined analysis of stage 1, EPIC-InterAct and GERA (a total of 39,219 cases and 198,990 controls) (**Supplementary Table 6**). Therefore, despite substantially

enlarged sample sizes that would have allowed us to detect low-frequency risk alleles with modest effect sizes, the overwhelming majority of variants for which T2D-association can be detected with these sample sizes are themselves common.

To identify loci containing multiple distinct signals, we performed approximate conditional analysis within the established and novel GWAS loci and detected two such novel common variant signals (**Supplementary Table 7**) (19,20). At the *ANKRD55* locus, we identified a previously-unreported distinct ($p_{conditional}<10^{-5}$) association signal led by rs173964 ($p_{conditional}=3.54\times10^{-7}$, MAF=26%) (**Supplementary Table 7, Supplementary Figure 4**). We also observed multiple signals of association at loci with previous reports of such signals (4,14), including *CDKN2A/B* (3 signals in total), *DGKB*, *KCNQ1* (6 signals), *HNF4A*, and *CCND2* (3 signals) (**Supplementary Table 7, Supplementary Figure 4**). At *CCND2,* in addition to the main signal with lead SNV rs4238013, we detected: (i) a novel distinct signal led by a common variant, rs11063018 ($p_{conditional}=2.70\times10^{-7}$, MAF=19%); and (ii) a third distinct signal led by a low-frequency protective allele (rs188827514, MAF=0.6%; $OR_{conditional}=0.60$, $p_{conditional}=1.24\times10^{-6}$) (**Supplementary Figure 5A, Supplementary Table 7**), which represents the same distinct signal as that at rs76895963 ($p_{conditional}=1.0$) reported in the Icelandic population (**Supplementary Figure 5B**) (7). At *HNF4A,* we confirm recent analyses (obtained in partially-overlapping data) (14) that a low-frequency missense variant (rs1800961, p.Thr139Ile, MAF=3.7%) is associated with T2D, and is distinct from the known common variant GWAS signal (which we map here to rs12625671).

We evaluated the trans-ethnic heterogeneity of allelic effects (i.e. discordance in the direction and/or magnitude of estimated odds ratios) at novel loci on the basis of Cochran's Q statistics from the largest T2D trans-ancestry GWAS meta-analysis to date (2). Using reported summary statistics from that study, we observed no significant evidence of heterogeneity of effect size (Bonferroni correction $p_{Cochran's\ Q}<0.05/13=0.0038$) between major ancestral

460 groups at any of the 13 loci (**Supplementary Table 8**). These results are consistent with
461 these loci being driven by common causal variants that are widely distributed across
462 populations.

463 ***1000G variant density for identification of potentially causal genetic variants.*** We used
464 credible set fine-mapping (33) to investigate whether 1000G imputation allowed us to better
465 resolve the specific variants driving 95 distinct T2D association signals at 82 loci
466 (**Supplementary Material**). 99% credible sets included between 1 and 7,636 SNVs; 25
467 included fewer than 20 SNVs, 16 fewer than 10 (**Supplementary Tables 9 and 10**). We
468 compared 1000G-based credible sets with those constructed from HapMap SNVs alone
469 (**Figure 2B**, **Supplementary Table 9**). At all but three of the association signals (two at
470 *KCNQ1* and rs1800961 at *HNF4A*), 1000G imputation resulted in larger credible sets
471 (median increase of 34 variants) spanning wider genomic intervals (median interval size
472 increase of 5kb) (**Figure 2B, Supplementary Table 9**). The 1000G-defined credible sets
473 included >85% of the SNVs in the corresponding HapMap sets (**Supplementary Table 9**).
474 Despite the overall larger credible sets, we asked whether 1000G imputation enabled an
475 increase in the posterior probability afforded to the lead SNVs, but found no evidence to this
476 effect (**Figure 2C**).

477 Within the 50 loci previously associated with T2D in Europeans (4) which had at least
478 modest evidence of association in the current analyses ($p<5x10^{-4}$), we asked whether the lead
479 SNV in 1000G-imputed analysis was of similar frequency to that observed in HapMap
480 analyses. Only at *TP53INP1*, was the most strongly associated 1000G-imputed SNV
481 (rs11786613, OR=1.21, $p=1.6x10^{-6}$, MAF=3.2%) of substantially lower frequency than the
482 lead HapMap-imputed SNV (3) (rs7845219, MAF=47.7%, **Figure 2A**). rs11786613 was
483 neither present in HapMap, nor on the Metabochip (**Supplementary Figure 6**). Reciprocal
484 conditioning of this low-frequency SNV and the previously identified common lead SNV

(rs7845219: OR=1.05, p=5.0x10^{-5}, MAF=47.5%) indicated that the two signals were likely to be distinct but the signal at rs11786613 did not meet our threshold ($p_{conditional}<10^{-5}$) for locus-wide significance (**Supplementary Figure 4**).

Pathophysiological insights from novel T2D associations. Among the 13 novel T2D-associated loci, many (such as those near *HLA-DQA1, NRXN3, GIP, ABO* and *CMIP*) included variants previously implicated in predisposition to other diseases and traits ($r^2>0.6$ with the lead SNV) (**Supplementary Table 3, Supplementary Materials: Biology box**). For example, the novel association at SNV rs1182436 lies ~120Kb upstream of *MNX1,* a gene implicated in pancreatic hypoplasia and neonatal diabetes (34–36).

The lead SNV rs78761021 at the *GLP2R* locus, encoding the receptor for glucagon-like peptide 2, is in strong LD ($r^2=0.87$) with a common missense variant in *GLP2R* (rs17681684, D470N, p=3×10^{-7}). These signals were strongly dependent and mutually extinguished in reciprocal conditional analyses, consistent with the coding variant being causal and implicating *GLP2R* as the putative causal gene (**Supplementary Figure 7**). While previously suggested to regulate energy balance and glucose tolerance (37), GLP2R has primarily been implicated in gastrointestinal function (38,39). In contrast, *GLP1R*, encoding the GLP-1 receptor (the target for a major class of T2D therapies (40)) is more directly implicated in pancreatic islet function and variation at this gene has been associated with glucose levels and T2D risk (41).

We also observed associations with T2D centred on rs9271774 near *HLA-DQA1* (**Table 1**), a region showing a particularly strong association with T1D (42). There is considerable heterogeneity within, and overlap between, the clinical presentations of T1D and T2D, but these can be partially resolved through measurement of islet cell autoantibodies (43). Such measures were not uniformly available across studies contributing to our meta-analysis (**Supplementary Table 1**). We therefore considered whether the adjacency between T1D-

and T2D-risk loci was likely to reflect misclassification of individuals with autoimmune diabetes as cases in the present study.

Three lines of evidence make this unlikely. First, the lead T1D-associated SNV in the HLA region (rs6916742) was only weakly associated with T2D in the present study (p=0.01), and conditioning on this variant had only modest impact on the T2D-association signal at rs9271774 ($p_{unconditional}=3.3x10^{-7}$; $p_{conditional}=9.1x10^{-6}$). Second, of 52 published genome-wide significant T1D-association GWAS signals, 50 were included in the current analysis: only six of these reached even nominal association with T2D (p<0.05; **Supplementary Figure 8**), and at one of these six (*BCAR1*), the T1D risk-allele was *protective* for T2D. Third, in genetic risk score (GRS) analyses, the combined effect of these 50 T1D signals on T2D risk was of only nominal significance (OR =1.02[1.00, 1.03], p=0.026), and significance was eliminated when the 6 overlapping loci were excluded (OR =1.00[0.98, 1.02], p=0.73). In combination, these findings argue against substantial misclassification and indicate that the signal at *HLA-DQA1* is likely to be a genuine T2D signal.

Potential genes and pathways underlying the T2D loci: eQTL and pathway analysis. Cis-eQTLs analyses highlighted four genes as possible effector transcripts: *ABO* (pancreatic islets), *PLEKHA1* (whole blood), *HSD17B12* (adipose, liver, muscle, whole blood) at the respective loci, and *HLA-DRB5* expression (adipose, pancreatic islets, whole blood) at the *HLA-DQA1* locus (**Supplementary Table 11**).

We next asked whether large-scale gene expression data, mouse phenotypes, and protein-protein interaction (PPI) networks could implicate specific gene candidates and gene sets in the aetiology of T2D. Using DEPICT (32), 29 genes were prioritised as driving observed associations (FDR<0.05), including *ACSL1* and *CMIP* among the genes mapping to the novel loci (**Supplementary Table 12**). These analyses also identified 20 enriched reconstituted gene sets (FDR<5%) falling into 4 groups (**Supplementary Figure 9;** complete results,

including gene prioritisation, can be downloaded from https://onedrive.live.com/redir?resid=7848F2AF5103AA1B!1505&authkey=!AIC31supgUwjZVU&ithint=file%2cxlsx). These included pathways related to mammalian target of rapamycin (mTOR), based on co-regulation of the *IDE, TLE1, SPRY2, CMIP*, and *MTMR3* genes (44).

Overlap of associated variants with regulatory annotations. We observed significant enrichment for T2D-associated credible set variants in pancreatic islet active enhancers and/or promoters (log odds [β]=0.74, $p=4.2x10^{-8}$) and FOXA2 binding sites (β=1.40, $p=4.1\times10^{-7}$), as previously reported (**Supplementary Table 13**) (14). We also observed enrichment for T2D-associated variants in coding exons (β=1.56, $p=7.9x10^{-5}$), in EZH2-binding sites across many tissues (β=1.35, $p=5.3x10^{-6}$), and in binding sites for NKX2.2 (β=1.73, $p=4.1x10^{-8}$) and PDX1 (β=1.46, $p=7.4x10^{-6}$) in pancreatic islets (**Supplementary Figure 10**).

Even though credible sets were generally larger, analyses performed on the 1000G imputed results produced stronger evidence of enrichment than equivalent analyses restricted to SNVs present in HapMap. This was most notably the case for variants within coding exons (β=1.56, $p=7.9x10^{-5}$ in 1000G compared to β=0.68, p=0.62 in HapMap), and likely reflects more complete capture of the true causal variants in the more densely imputed credible sets. Single lead SNVs overlapping an enriched annotation accounted for the majority of the total posterior probability ($\pi_c>0.5$) at seven loci. For example, the lead SNV (rs8056814) at *BCAR1* ($\pi_c=0.57$) overlaps an islet enhancer (**Supplementary Figure 11A**), while the newly-identified low-frequency signal at *TP53INP1* overlaps an islet promoter element (rs117866713; $\pi_c=0.53$) (**Figure 2D**) (31).

We applied hierarchical clustering to the results of diabetes-related quantitative trait associations for the set of T2D-associated loci from the present study, identifying three main

clusters of association signals with differing impact on quantitative traits (**Supplementary Table 9**). The first, including *GIPR, C2CDC4A, CDKAL1, GCK, TCF7L2*, *GLIS3*, *THADA*, *IGF2BP2*, and *DGKB* involved loci with a primary impact on insulin secretion and processing (26,29). The second cluster captured loci (including *PPARG*, *KLF14*, and *IRS1*) disrupting insulin action. The third cluster, showing marked associations with BMI and lipid levels, included *NRXN3*, *CMIP, APOE*, and *MC4R*, but not *FTO*, which clustered alone.

In regulatory enhancement analyses, we observed strong tissue-specific enrichment patterns broadly consistent with the phenotypic characteristics of the physiologically-stratified locus subsets. The cluster of loci disrupting insulin secretion showed the most marked enrichment for pancreatic islet regulatory elements (β=0.91, $p=9.5\times10^{-5}$). In contrast, the cluster of loci implicated in insulin action was enriched for annotations from adipocytes (β=1.3, $p=2.7\times10^{-11}$) and monocytes (β=1.4, $p=1.4\times10^{-12}$), and that characterised by associations with BMI and lipids showed preferential enrichment for hepatic annotations (β=1.15, $p=5.8\times10^{-4}$) (**Figure 3A-C**). For example, at the novel T2D-associated *CMIP* locus, previously associated with adiposity and lipid levels (28,45), the lead SNV (rs2925979, π_c=0.91) overlaps an active enhancer element in both liver and adipose tissue, among others (**Supplementary Figure 11B**).

DISCUSSION

In this large-scale study of T2D genetics, in which individual variants were assayed in up to 238,209 subjects, we identify 13 novel T2D-associated loci at genome-wide significance and refine causal variant location for the 13 novel and 69 established T2D loci. We also provide evidence for enrichment in regulatory elements at associated loci in tissues relevant for T2D, and demonstrate tissue-specific enrichment in regulatory annotations when T2D loci were stratified according to inferred physiological mechanism.

584 Together with loci reported in other recent publications (9), we calculate that the present
585 analysis brings the total number of independent T2D associations to 128 distinct signals at
586 113 loci (**Supplementary Table 3**). Lead SNVs at all 13 novel loci were common (MAF >
587 0.15) and of comparable effect size (1.07≤OR≤1.10) to previously-identified common variant
588 associations (2,4). Associations at the novel loci showed homogeneous effects across diverse
589 ethnicities, supporting the evidence for coincident common risk alleles across ancestry groups
590 (2). Moreover, we conclude that misclassification of diabetes subtype is not a major concern
591 for these analyses and that the *HLA-DQA1* signal represents genuine association with T2D,
592 independent of nearby signals that influence T1D.

593 We observed a general increase in the size of credible sets with 1000G imputation compared
594 to HapMap imputation. This is likely due to improved enumeration of potential causal
595 common variants on known risk haplotypes, rather than resolution towards low-frequency
596 variants of larger effect driving common variant associations. These findings are consistent
597 with the inference (arising also from the other analyses reported here) that the T2D-risk
598 signals identified by GWAS are overwhelmingly driven by common causal variants. In such
599 a setting, imputation with denser reference panels, at least in ethnically restricted samples,
600 provides more complete elaboration of the allelic content of common risk haplotypes. Finer
601 resolution of those haplotypes that would provide greater confidence in the location of causal
602 variants will likely require further expansion of trans-ethnic fine-mapping efforts (2). The
603 distinct signals at the established *CCND2* and *TP53INP1* loci point to contributions of low-
604 frequency variant associations of modest effect, but indicate that even larger samples will be
605 required to robustly detect association signals at low frequency. Such new large datasets
606 might be used to expand the follow-up of suggestive signals from our analysis.

607 The discovery of novel genome-wide significant association signals in the current analysis is
608 attributable primarily to increased sample size, rather than improved genomic coverage.

Although we queried a large proportion of the low-frequency variants present in the 1000G European reference haplotypes, and had >80% power to detect genome-wide significant associations with OR>1.8 for the tested low-frequency risk variants, we found no such low-frequency variant associations in either established or novel loci. Whilst low-frequency variant coverage in the present study was not complete, this observation adds to the growing evidence (2,4,9,46) that few low-frequency T2D-risk variants with moderate to strong effect sizes exist in European ancestry samples, and is consistent with a primary role for common variants of modest effect in T2D risk. The present study reinforces the conclusions from a recent study which imputed from whole-genome sequencing data - from 2,657 European T2D cases and controls, rather than 1000G - into a set of GWAS studies partially overlapping with the present meta-analysis. We demonstrated that the failure to detect low frequency associations in that study is not overcome by a substantial increase in sample size (9). It is worth emphasising that we did not, in this study, have sufficient imputation quality to test for T2D associations with rare variants and we cannot evaluate the collective contribution of variants with MAF<0.5% to T2D risk.

The development of T2D involves dysfunction of multiple mechanisms across several distinct tissues (9,29,31,47,48). When coupled with functional data, we saw larger effect estimates for enrichment of coding variants than observed with HapMap SNVs alone, consistent with more complete recovery of the causal variants through imputation using a denser reference panel. The functional annotation analyses also demonstrated that the stratification of T2D-risk loci according to primary physiological mechanism resulted in evidence for consistent and appropriate tissue-specific effects on transcriptional regulation. These analyses exemplify the use of a combination of human physiology and genomic annotation to position T2D GWAS loci with respect to the cardinal mechanistic components of T2D development. Extension of this approach is likely to provide a valuable *in silico*

strategy to aid prioritisation of tissues for mechanistic characterisation of genetic associations. Using the hypothesis-free pathway analysis of T2D associations with DEPICT (32), we highlighted a causal role of mTOR signalling pathway in the aetiology of T2D not observed from individual loci associations. The mTOR pathway has previously been implicated in the link between obesity, insulin resistance, and T2D from cell and animal models (44,49).

The current results emphasize that progressively larger sample sizes, coupled with higher density sequence-based imputation (13), will continue to represent a powerful strategy for genetic discovery in T2D, and in complex diseases and traits more generally. At known T2D-associated loci, identification of the most plausible T2D causal variants will likely require large-scale multi-ethnic analyses, where more diverse haplotypes, reflecting different patterns of LD, in combination with functional (31,50,51) data allow refinement of association signals to smaller numbers of variants (2).

DESCRIPTION OF SUPPLEMENTAL DATA

Supplemental Data include eleven figures and thirteen tables.

AUTHOR CONTRIBUTIONS:

Writing and co-ordination group:

R.A.S., L.J.S., R.M., L.M., K.J.G., M.K., J.D., A.P.M., M.B., M.I.M., I.P.

Central analysis group:

R.A.S., L.J.S., R.M., L.M., C.M., A.P.M., M.B., M.I.M., I.P.

Additional lead analysts:

L.M., K.J.G., M.K., N.P., T.H.P., A.D.J., J.D.E., T.F., Y.Lee, J.R.B.P., L.J., A.U.J.

GWAS cohort-level primary analysts:

R.A.S., L.J.S., R.M., K.J.G., V.S., G.T., L.Q., N.R.V., A.Mahajan, H.Chen, P.A., B.F.V., H.G., M.M., J.S.R., N.W.R., N.R., L.C.K., E.M.L., S.M.W., C.Fuchsberger, P.K., C.M., P.C., M.L., Y.L., C.D., D.T., L.Y., C.Langenberg, A.P.M., I.P.

Metabochip cohort-level primary analysts:

T.S., H.K., H.C., L.E., S.G., T.M.T, M.F., R.J.S.

Cohort sample collection, phenotyping, genotyping or additional analysis:

R.A.S., H.G., R.B., A.B.H., A.K., G.S., N.D.K., J.L., L.L., T.M., M.R., B.T., T.E., E.M., C.F., C.L., D.Rybin, B.I., V.L., T.T., D.J.C., J.S.P., N.G., C.T.H., M.E.J., T.J., A.L., M.C.C., R.M.D., D.J.H., P.Kraft, Q.S., S.E., K.R.O., J.R.B.P., A.R.W., E.Z., J.T.-F., G.R .A., L.L.B., P.S.C., H.M.S., H.A.K., L.K., B.S., T.W.M., M.M.N., S.P., D.B., K.G., S.E.H., E.Tremoli, N.K., J.M., G.Steinbach, R.W., J.G .E., S.M., L.P., E.T., G.C., E.E., S.L., B.G., K.L., O.M., E.P.B., O.G., D.R., M.Blüher, P.Kovacs, A.T., N.M.M., C.S., T.M.F., A.T.H., I.B., B.B., H.B., P.W.F., A.B.G., D.P., Y.T.v.d.S., C.Langenberg, N.J.W., K.Strauch, M.B., M.I.M.

Metabochip cohort principal investigators:

R.E., K.J., S.Moebus, U.d.F., A.H., M.S., P.D., P.J.D., T.M.F., A.T.H., S.R., V.Salomaa, N.L.P., B.O.B., R.N.B., F.S.C., K.L.M., J.T., T.H., O.B.P., I.B., C.Langenberg, N.J.W.

GWAS cohort principal investigators:

L.Lannfelt, E.I., L.Lind, C.M.L., S.C., P.F., R.J.F.L., B.B., H.B., P.W.F., A.B.G., D.P., Y.T.v.d.S., D.A., L.C.G., C.Langenberg, N.J.W., E.S., C.Duijn van, J.C.F., J.B.M., E.B., C.G., K.Strauch, A.M., A.D.M., C.N.A.P., F.B.H., U.T., K.S., J.D., M.B., M.I.M.

References

1. Global Burden of Disease Study 2013 Collaborators. Global, regional, and national incidence, prevalence, and years lived with disability for 301 acute and chronic diseases and injuries in 188 countries, 1990-2013: a systematic analysis for the Global Burden of Disease Study 2013. 2015;386:743–800.
2. DIAbetes Genetics Replication And Meta-analysis (DIAGRAM) Consortium, Asian Genetic Epidemiology Network Type 2 Diabetes (AGEN-T2D) Consortium, South Asian Type 2 Diabetes (SAT2D) Consortium, Mexican American Type 2 Diabetes (MAT2D) Consortium, Type 2 Diabetes Genetic Exploration by Nex-generation sequencing in muylti-Ethnic Samples (T2D-GENES) Consortium, Mahajan A, et al. Genome-wide trans-ancestry meta-analysis provides insight into the genetic architecture of type 2 diabetes susceptibility. Nat Genet. 2014;46:234–44.
3. Voight BF, Scott LJ, Steinthorsdottir V, Morris ADP, Dina C, Welch RP, et al. Twelve type 2 diabetes susceptibility loci identified through large-scale association analysis. Nat Genet. 2010;42:579–89.
4. Morris AP, Voight BF, Teslovich TM, Ferreira T, Segrè AV, Steinthorsdottir V, et al. Large-scale association analysis provides insights into the genetic architecture and pathophysiology of type 2 diabetes. Nat Genet. 2012;44:981–90.
5. Zeggini E, Scott LJ, Saxena R, Voight BF, Marchini JL, Hu T, et al. Meta-analysis of genome-wide association data and large-scale replication identifies additional susceptibility loci for type 2 diabetes. Nat Genet. 2008;40:638–45.
6. Dupuis J, Langenberg C, Prokopenko I, Saxena R, Soranzo N, Jackson AU, et al. New genetic loci implicated in fasting glucose homeostasis and their impact on type 2 diabetes risk. Nat Genet. 2010;42:105–16.
7. Steinthorsdottir V, Thorleifsson G, Sulem P, Helgason H, Grarup N, Sigurdsson A, et al. Identification of low-frequency and rare sequence variants associated with elevated or reduced risk of type 2 diabetes. Nat Genet.2014;46:294–8.
8. Estrada K, Aukrust I, Bjørkhaug L, Burtt NP, Mercader JM, García-Ortiz H, et al. Association of a low-frequency variant in HNF1A with type 2 diabetes in a Latino population. JAMA. 2014;311:2305–14.
9. Fuchsberger C, Flannick J, Teslovich TM, Mahajan A, Agarwala V, Gaulton KJ, et al. The genetic architecture of type 2 diabetes. Nature. 2016;536:41-7.
10. Majithia AR, Flannick J, Shahinian P, Guo M, Bray M-A, Fontanillas P, et al. Rare variants in PPARG with decreased activity in adipocyte differentiation are associated with increased risk of type 2 diabetes. Proc Natl Acad Sci U S A. 2014;111:13127-32.
11. Bonnefond A, Clement N, Fawcett K, Yengo L, Vaillant E, Guillaume J-L, et al. Rare MTNR1B variants impairing melatonin receptor 1B function contribute to type 2 diabetes. Nat Genet. 2012;44:297–301.
12. Abecasis GR, Auton A, Brooks LD, DePristo M a, Durbin RM, Handsaker RE, et al. An integrated map of genetic variation from 1,092 human genomes. Nature. 2012;491:56–65.
13. Yang J, Bakshi A, Zhu Z, Hemani G, Vinkhuyzen AAE, Lee SH, et al. Genetic variance estimation with imputed variants finds negligible missing heritability for human height and body mass index. Nat Genet. 2015;47:1114–20.
14. Gaulton KJ, Ferreira T, Lee Y, Raimondo A, Mägi R, Reschen ME, et al. Genetic fine mapping and genomic annotation defines causal mechanisms at type 2 diabetes susceptibility loci. Nat Genet. 2015;47:1415–25.
15. Langenberg C, Sharp S, Forouhi NG, Franks PW, Schulze MB, Kerrison N, et al. Design and cohort description of the InterAct Project: an examination of the

728 interaction of genetic and lifestyle factors on the incidence of type 2 diabetes in the
729 EPIC Study. Diabetologia. 2011;54:2272–82.
730 16. Cook JP, Morris AP. Multi-ethnic genome-wide association study identifies novel
731 locus for type 2 diabetes susceptibility. Eur J Hum Genet. 2016;24:1175–80.
732 17. Howie B, Fuchsberger C, Stephens M, Marchini J, Abecasis GR. Fast and accurate
733 genotype imputation in genome-wide association studies through pre-phasing. Nat
734 Genet. 2012;44:955–9.
735 18. Howie BN, Donnelly P, Marchini J. A flexible and accurate genotype imputation
736 method for the next generation of genome-wide association studies. PLoS Genet.
737 2009;5:e1000529.
738 19. Yang J, Ferreira T, Morris AP, Medland SE, Madden PAF, Heath AC, et al.
739 Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies
740 additional variants influencing complex traits. Nat Genet. 2012;44:369–75.
741 20. Yang J, Lee SH, Goddard ME, Visscher PM. GCTA: a tool for genome-wide complex
742 trait analysis. Am J Hum Genet. 2011;88:76–82.
743 21. UK10K Consortium, Writing group, Production group, Cohorts group,
744 Neurodevelopmental disorders group, Obesity group, et al. The UK10K project
745 identifies rare variants in health and disease. Nature. 2015;526:82–90.
746 22. Burgess S, Butterworth A, Thompson SG. Mendelian randomization analysis with
747 multiple genetic variants using summarized data. Genet Epidemiol. 2013;37:658–65.
748 23. Manning AK, Hivert M-F, Scott RA, Grimsby JL, Bouatia-Naji N, Chen H, et al. A
749 genome-wide approach accounting for body mass index identifies genetic variants
750 influencing fasting glycemic traits and insulin resistance. Nat Genet. 2012;44:659–69.
751 24. Saxena R, Hivert M-F, Langenberg C, Tanaka T, Pankow JS, Vollenweider P, et al.
752 Genetic variation in GIPR influences the glucose and insulin responses to an oral
753 glucose challenge. Nat Genet. 2010;42:142–8.
754 25. Strawbridge RJ, Dupuis J, Prokopenko I, Barker A, Ahlqvist E, Rybin D, et al.
755 Genome-wide association identifies nine common variants associated with fasting
756 proinsulin levels and provides new insights into the pathophysiology of type 2
757 diabetes. Diabetes. 2011;60:2624–34.
758 26. Prokopenko I, Poon W, Mägi R, Prasad B R, Salehi SA, Almgren P, et al. A central
759 role for GRB10 in regulation of islet function in man. PLoS Genet. 2014;10:e1004235.
760 27. Speliotes EK, Willer CJ, Berndt SI, Monda KL, Thorleifsson G, Jackson AU, et al.
761 Association analyses of 249,796 individuals reveal 18 new loci associated with body
762 mass index. Nat Genet. 2010;42:937–48.
763 28. Willer CJ, Schmidt EM, Sengupta S, Peloso GM, Gustafsson S, Kanoni S, et al.
764 Discovery and refinement of loci associated with lipid levels. Nat Genet.
765 2013;45:1274–83.
766 29. Dimas AS, Lagou V, Barker A, Knowles JW, Mägi R, Hivert MF, et al. Impact of type
767 2 diabetes susceptibility variants on quantitative glycemic traits reveals mechanistic
768 heterogeneity. Diabetes. 2014;63:2158–71.
769 30. Dunham I, Kundaje A, Aldred SF, Collins PJ, Davis CA, Doyle F, et al. An integrated
770 encyclopedia of DNA elements in the human genome. Nature. 2012;489:57–74.
771 31. Pasquali L, Gaulton KJ, Rodríguez-Seguí S a, Mularoni L, Miguel-Escalada I,
772 Akerman I, et al. Pancreatic islet enhancer clusters enriched in type 2 diabetes risk-
773 associated variants. Nat Genet. 2014;46:136–43.
774 32. Pers TH, Karjalainen JM, Chan Y, Westra H, Wood AR, Yang J, et al. Biological
775 interpretation of genome-wide association studies using predicted gene functions. Nat
776 Commun. 2015;6:5890.
777 33. Maller JB, McVean G, Byrnes J, Vukcevic D, Palin K, Su Z, et al. Bayesian
778 refinement of association signals for 14 loci in 3 common diseases. Nat Genet.

2012;44:1294–301.
34. Flanagan SE, De Franco E, Lango Allen H, Zerah M, Abdul-Rasoul MM, Edge JA, et al. Analysis of transcription factors key for mouse pancreatic development establishes NKX2-2 and MNX1 mutations as causes of neonatal diabetes in man. Cell Metab. 2014;19:146–54.
35. Mele M, Ferreira PG, Reverter F, DeLuca DS, Monlong J, Sammeth M, et al. The human transcriptome across tissues and individuals. Science. 2015;348:660–5.
36. Bonnefond A, Vaillant E, Philippe J, Skrobek B, Lobbens S, Yengo L, et al. Transcription factor gene MNX1 is a novel cause of permanent neonatal diabetes in a consanguineous family. Diabetes Metab. 2013;39:276–80.
37. Guan X. The CNS glucagon-like peptide-2 receptor in the control of energy balance and glucose homeostasis. Am J Physiol Regul Integr Comp Physiol. 2014;307:R585-96.
38. Murphy KG, Bloom SR. Gut hormones and the regulation of energy homeostasis. Nature. 2006;444:854–9.
39. GTEx Consortium. Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. Science. 2015;348:648–60.
40. Drucker DJ, Nauck MA. The incretin system: glucagon-like peptide-1 receptor agonists and dipeptidyl peptidase-4 inhibitors in type 2 diabetes. Lancet. 2006;368:1696–705.
41. Wessel J, Chu AY, Willems SM, Wang S, Yaghootkar H, Brody JA, et al. Low-frequency and rare exome chip variants associate with fasting glucose and type 2 diabetes susceptibility. Nat Commun. 2015;6:5897.
42. Bradfield JP, Qu H-Q, Wang K, Zhang H, Sleiman PM, Kim CE, et al. A genome-wide meta-analysis of six type 1 diabetes cohorts identifies multiple associated loci. PLoS Genet. 2011;7:e1002293.
43. NICE guideline. Type 1 diabetes in adults: diagnosis and management [article online]. 2015. Available from: nice.org.uk/guidance/ng17. Accessed 16 March 2017.
44. Zoncu R, Efeyan A, Sabatini DM. mTOR: from growth signal integration to cancer, diabetes and ageing. Nat Rev Mol Cell Biol. 2011;12(1):21–35.
45. Shungin D, Winkler TW, Croteau-Chonka DC, Ferreira T, Locke AE, Mägi R, et al. New genetic loci link adipose and insulin biology to body fat distribution. Nature. 2015;518:187–96.
46. Agarwala V, Flannick J, Sunyaev S, Altshuler D. Evaluating empirical bounds on complex disease genetic architecture. Nat Genet. 2013;45:1418–27.
47. Stumvoll M, Goldstein BJ, Van Haeften TW. Type 2 diabetes: Principles of pathogenesis and therapy. Lancet. 2005;365:1333–46.
48. Parker SCJ, Stitzel ML, Taylor DL, Orozco JM, Erdos MR, Akiyama JA, et al. Chromatin stretch enhancer states drive cell-specific gene regulation and harbor human disease risk variants. Proc Natl Acad Sci U S A. 2013;110:17921–6.
49. Dann SG, Selvaraj A, Thomas G. mTOR Complex1–S6K1 signaling: at the crossroads of obesity, diabetes and cancer. Trends Mol Med. 2007;13:252–9.
50. Claussnitzer M, Dankel SN, Klocke B, Grallert H, Glunk V, Berulava T, et al. Leveraging cross-species transcription factor binding site patterns: from diabetes risk Loci to disease mechanisms. Cell. 2014;156:343–58.
51. Farh KK, Marson A, Zhu J, Kleinewietfeld M, Housley WJ, Beik S, et al. Genetic and epigenetic fine mapping of causal autoimmune disease variants. Nature. 2014;518:337–43.

FIGURE TITLES AND LEGENDS

Figure 1. The effect sizes of the established (blue diamonds, N=69, $p<5\times10^{-4}$, **Supplementary Material**), novel (red diamonds, N=13), and additional distinct (sky blue diamonds, N=13, **Supplementary Table 7**) signals according to their risk allele frequency (**Supplementary Table 3**). The additional distinct signals are based on approximate conditional analyses. The distinct signal at *TP53INP1* led by rs11786613 (**Supplementary Table 7**) is plotted (sky blue diamond). This signal did not reach locus-wide significance, but was selected for follow-up because of its low frequency and absence of LD with previously reported signal at this locus. The power curve shows the estimated effect size for which we had 80% power to detect associations. Established common variants with OR>1.12 are annotated.

Figure 2. A) The number of SNVs included in 99% credible sets when performed on all SNVs compared to when analyses were restricted to those SNVs present in HapMap. B) The cumulative π_c of the top 3 SNVs among all 1000G SNVs and after restriction to HapMap SNVs is shown. While the low frequency SNV at *TP53INP1* (rs11786613) did not reach the threshold for a distinct signal in approximate conditional analyses, we fine-mapped both this variant and the previous common signal separately after reciprocal conditioning, which suggested they were independent. C) The minor allele frequency of the lead SNV identified in current analyses compared to that identified among SNVs present in HapMap. D) The association of the low frequency variant rs11786613 (blue) and that of the previous lead variant at this locus, rs7845219 (purple). The low frequency variant overlaps regulatory annotations active in pancreatic islets, among other tissues, and the sequence surrounding the A allele of this variant has a *in silico* recognition motif for a FOXA1:AR (androgen receptor) protein complex.

Figure 3. Type 2 diabetes loci stratified by patterns of quantitative trait (e.g. glycaemic, insulin, lipid, and anthropometric) effects show distinct cell-type annotation patterns. We hierarchically clustered loci based on endophenotype data and identified groups of T2D loci associated with measures of A) insulin secretion, B) insulin resistance, and C) BMI/lipids. We then tested the effect of variants in cell-type enhancer and promoter chromatin states on the posterior probabilities of credible sets for each group. We identified most significant effects among pancreatic islet chromatin for insulin secretion loci, CD14+ monocyte and adipose chromatin for insulin resistance loci, and liver chromatin for BMI/lipid loci.

GUARANTOR'S STATEMENT

Dr. Inga Prokopenko is the guarantor of this work and, as such, had full access to all the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

COMPETING FINANCIAL INTERESTS STATEMENT

Inês Barroso and spouse own stock in GlaxoSmithKline and Incyte.

Jose C Florez has received consulting honoraria from Pfizer and PanGenX.

Valgerdur Steinthorsdottir, Gudmar Thorleifsson, Augustine Kong, Unnur Thorsteinsdottir, and Kari Stefansson are employed by deCODE 4 Genetics/Amgen inc.

Erik Ingelsson is a scientific advisor for Precision Wellness, Cellink and Olink Proteomics for work unrelated to the present project.

Mark I McCarthy sits on Advisory Panels for Pfizer and NovoNordisk, has received honoraria from Pfizer NovoNordisk and EliLilly, and is also a recipient of research funding from Pfizer, NovoNordisk, EliLilly, Takeda, Sanofi-Aventis, Merck, Boehringer-Ingelheim, Astra Zeneca, Janssen, Roche, Servier and Abbvie.

Robert A Scott, Laura J Scott, Reedik Mägi, Letizia Marullo, Kyle J Gaulton, Marika Kaakinen, Natalia Pervjakova, Tune H Pers, Andrew D Johnson, John D Eicher, Anne U Jackson, Teresa Ferreira, Yeji Lee, Clement Ma, Lu Qi, Natalie R Van Zuydam, Anubha Mahajan, Han Chen, Peter Almgren, Ben F Voight, Harald Grallert, Martina Müller-Nurasyid, Janina S Ried, N William Rayner, Neil Robertson, Lennart C Karssen, Elisabeth M van Leeuwen, Sara M Willems, Christian Fuchsberger, Phoenix Kwan, Tanya M Teslovich, Pritam Chanda, Man Li , Yingchang Lu, Christian Dina, Dorothee Thuillier, Loic Yengo, Longda Jiang, Thomas Sparso, Hans A Kestler, Himanshu Chheda, Lewin Eisele, Stefan Gustafsson, Mattias Frånberg, Rona J Strawbridge, Rafn Benediktsson, Astradur B Hreidarsson, Gunnar Sigurðsson, Nicola D Kerrison, Jian'an Luan, Liming Liang, Thomas Meitinger, Michael Roden, Barbara Thorand, Tõnu Esko, Evelin Mihailov, Caroline Fox, Ching-Ti Liu, Denis Rybin, Bo Isomaa, Valeriya Lyssenko, Tiinamaija Tuomi, David J Couper, James S Pankow, Niels Grarup, Christian T Have, Marit E Jørgensen, Torben Jørgensen, Allan Linneberg, Marilyn C Cornelis, Rob M van Dam, David J Hunter, Peter Kraft, Qi Sun, Sarah Edkins, Katharine R Owen, John RB Perry, Andrew R Wood, Eleftheria Zeggini, Juan Tajes-Fernandes, Goncalo R Abecasis, Lori L Bonnycastle, Peter S Chines, Heather M Stringham, Heikki A Koistinen, Leena Kinnunen, Bengt Sennblad, Thomas W Mühleisen, Markus M Nöthen, Sonali Pechlivanis, Damiano Baldassarre, Karl Gertow, Steve E Humphries, Elena Tremoli, Norman Klopp, Julia Meyer, Gerald Steinbach, Roman Wennauer, Johan G Eriksson, Satu Männistö, Leena Peltonen, Emmi Tikkanen, Guillaume Charpentier, Elodie Eury, Stéphane Lobbens, Bruna Gigante, Karin Leander, Olga McLeod,

Erwin P Bottinger, Omri Gottesman, Douglas Ruderfer, Matthias Blüher, Peter Kovacs, Anke Tonjes, Nisa M Maruthur, Chiara Scapoli, Raimund Erbel, Karl-Heinz Jöckel, Susanne Moebus, Ulf de Faire, Anders Hamsten, Michael Stumvoll, Panagiotis Deloukas, Peter J Donnelly, Timothy M Frayling, Andrew T Hattersley, Samuli Ripatti, Veikko Salomaa, Nancy L Pedersen, Bernhard O Boehm, Richard N Bergman, Francis S Collins, Karen L Mohlke, Jaakko Tuomilehto, Torben Hansen, Oluf Pedersen, Lars Lannfelt, Lars Lind, Cecilia M Lindgren, Stephane Cauchi, Philippe Froguel, Ruth JF Loos, Beverley Balkau, Heiner Boeing, Paul W Franks, Aurelio Barricarte Gurrea, Domenico Palli, Yvonne T van der Schouw, David Altshuler, Leif C Groop, Claudia Langenberg, Nicholas J Wareham, Eric Sijbrands, Cornelia M van Duijn, James B Meigs, Eric Boerwinkle, Christian Gieger, Konstantin Strauch, Andres Metspalu, Andrew D Morris, Colin NA Palmer, Frank B Hu, Josée Dupuis, Andrew P Morris, Michael Boehnke, and Inga Prokopenko declare to have no competing financial interest.

Table 1. Novel loci associated with T2D from the combination of 1000G-imputed GWAS meta-analysis (stage 1) and Metabochip follow-up (stage 2).

	Stage 1						Stage 2							Stage1+Stage2	
Locus name*	**Chr:Position**	**SNV†**	**EA/ NEA**	**EAF**	**OR (CI 95%)**	***P*-value**	**Chr:Position**	**SNV‡**	**r^2 with lead SNV**	**EA/ NEA**	**EAF**	**OR (95% CI)**	***P*-value**	**OR (95% CI)ᵋ**	***P*-value**
ACSL1	4:185708807	rs60780116	T/C	0.84	1.09 (1.06-1.13)	$7.38x10^{-8}$	4:185714289	rs1996546	0.62	G/T	0.86	1.08 (1.03-1.13)	$5.60x10^{-4}$	1.09 (1.06-1.12)	$1.98x10^{-10}$
HLA-DQA1	6:32594309	rs9271774	C/A	0.74	1.10 (1.06-1.14)	$3.30x10^{-7}$	6:32594328	rs9271775	0.91	T/C	0.80	1.08 (1.03-1.13)	$7.59x10^{-4}$	1.09 (1.06-1.12)	$1.11x10^{-9}$
SLC35D3	6:137287702	rs6918311	A/G	0.53	1.07 (1.04-1.10)	$6.67x10^{-7}$	6:137299152	rs4407733	0.92	A/G	0.52	1.05 (1.02-1.08)	$1.63x10^{-3}$	1.06 (1.04-1.08)	$6.78x10^{-9}$
MNX1	7:157027753	rs1182436	C/T	0.80	1.08 (1.05-1.12)	$8.30x10^{-7}$	7:157031407	rs1182397	0.92	G/T	0.85	1.06 (1.02-1.11)	$4.38x10^{-3}$	1.08 (1.05-1.10)	$1.71x10^{-8}$
ABO	9:136155000	rs635634	T/C	0.18	1.08 (1.05-1.12)	$3.59x10^{-7}$	9:136154867	rs495828	0.83	T/G	0.20	1.06 (1.01-1.10)	$1.23x10^{-2}$	1.08 (1.05-1.10)	$2.30x10^{-8}$
PLEKHA1	10:124186714	rs2292626	C/T	0.50	1.09 (1.06-1.11)	$1.75x10^{-12}$	10:124167512	rs2421016	0.99	C/T	0.50	1.05 (1.02-1.08)	$2.30x10^{-3}$	1.07 (1.05-1.09)	$1.51x10^{-13}$
HSD17B12	11:43877934	rs1061810	A/C	0.28	1.08 (1.05-1.11)	$5.29x10^{-9}$	11:43876435	rs3736505	0.92	G/A	0.30	1.05 (1.01-1.08)	$4.82x10^{-3}$	1.07 (1.05-1.09)	$3.95x10^{-10}$
MAP3K11	11:65364385	rs111669836	A/T	0.25	1.07 (1.04-1.10)	$7.43x10^{-7}$	11:65365171	rs11227234	1.00	T/G	0.24	1.05 (1.01-1.08)	$8.77x10^{-3}$	1.06 (1.04-1.09)	$4.12x10^{-8}$
NRXN3	14:79945162	rs10146997	G/A	0.21	1.07 (1.04-1.10)	$4.59x10^{-6}$	14:79939993	rs17109256	0.98	A/G	0.21	1.07 (1.03-1.11)	$1.27x10^{-4}$	1.07 (1.05-1.09)	$2.27x10^{-9}$
CMIP	16:81534790	rs2925979	T/C	0.30	1.08 (1.05-1.10)	$2.72x10^{-8}$	16:81534790	rs2925979	1.00	T/C	0.31	1.05 (1.02-1.08)	$3.06x10^{-3}$	1.07 (1.04-1.09)	$2.27x10^{-9}$
ZZEF1	17:4014384	rs7224685	T/G	0.30	1.07 (1.04-1.10)	$2.00x10^{-7}$	17:3985864	rs8068804	0.95	A/G	0.31	1.07 (1.03-1.11)	$4.11x10^{-4}$	1.07 (1.05-1.09)	$3.23x10^{-10}$
GLP2R	17:9780387	rs78761021	G/A	0.34	1.07 (1.05-1.10)	$5.49x10^{-8}$	17:9791375	rs17676067	0.87	C/T	0.31	1.03 (1.00-1.07)	$3.54x10^{-2}$	1.06 (1.04-1.08)	$3.04x10^{-8}$
GIP	17:46967038	rs79349575	A/T	0.51	1.07 (1.04-1.09)	$2.61x10^{-7}$	17:47005193	rs15563	0.78	G/A	0.54	1.04 (1.01-1.07)	$2.09x10^{-2}$	1.06 (1.03-1.08)	$4.43x10^{-8}$

*The nearest gene is listed; this does not imply this is the biologically relevant gene; †Lead SNV types: all map outside transcripts except rs429358 (missense variant) and rs1061810 (3'UTR); ‡Stage 2: proxy SNV (r^2>0.6 with stage 1 lead SNV) was used when no stage 1 SNV was available. ᵋThe meta-analysis OR is aligned to the Stage 1 SNV risk allele. Abbreviations: Chr – chromosome, CI – confidence interval, EA - effect allele, EAF – effect allele frequency, OR – odds ratio, NEA – non-effect allele.

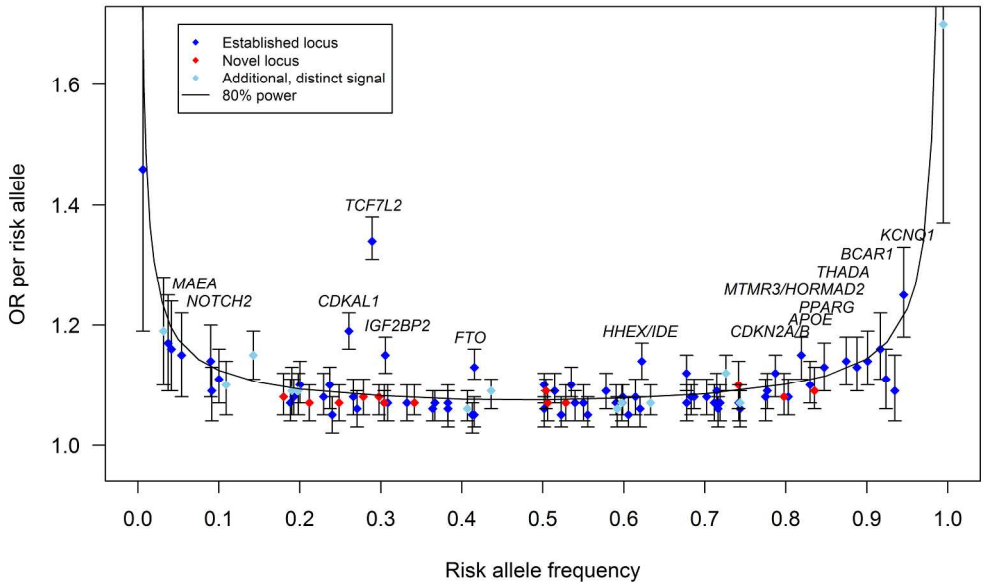


Figure 1. The effect sizes of the established (blue diamonds, N=69, p<5×10-4, Supplementary Methods), novel (red diamonds, N=13), and additional distinct (sky blue diamonds, N=13, Supplementary Table 7) signals according to their risk allele frequency (Supplementary Table 3). The additional distinct signals are based on approximate conditional analyses. The distinct signal at TP53INP1 led by rs11786613 (Supplementary Table 7) is plotted (sky blue diamond). This signal did not reach locus-wide significance, but was selected for follow-up because of its low frequency and absence of LD with previously reported signal at this locus. The power curve shows the estimated effect size for which we had 80% power to detect associations. Established common variants with OR>1.12 are annotated.

119x71mm (600 x 600 DPI)

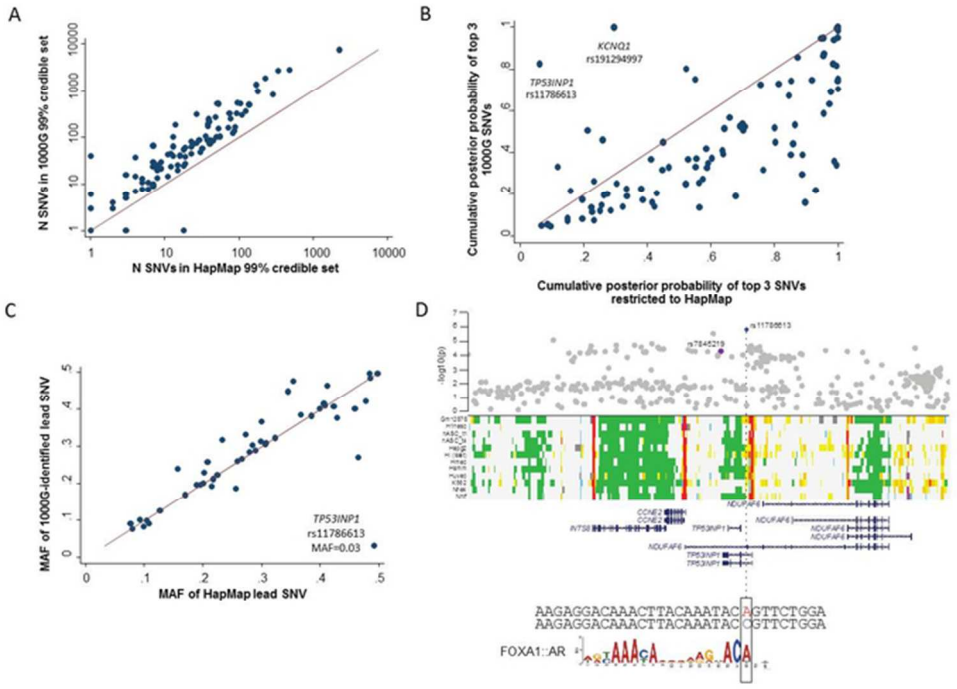


Figure 2. A) The number of SNVs included in 99% credible sets when performed on all SNVs compared to when analyses were restricted to those SNVs present in HapMap. B) The cumulative πc of the top 3 SNVs among all 1000G SNVs and after restriction to HapMap SNVs is shown. While the low frequency SNV at TP53INP1 (rs11786613) did not reach the threshold for a distinct signal in approximate conditional analyses, we fine-mapped both this variant and the previous common signal separately after reciprocal conditioning, which suggested they were independent. C) The minor allele frequency of the lead SNV identified in current analyses compared to that identified among SNVs present in HapMap. D) The association of the low frequency variant rs11786613 (blue) and that of the previous lead variant at this locus, rs7845219 (purple). The low frequency variant overlaps regulatory annotations active in pancreatic islets, among other tissues, and the sequence surrounding the A allele of this variant has a in silico recognition motif for a FOXA1:AR (androgen receptor) protein complex.

29x21mm (600 x 600 DPI)

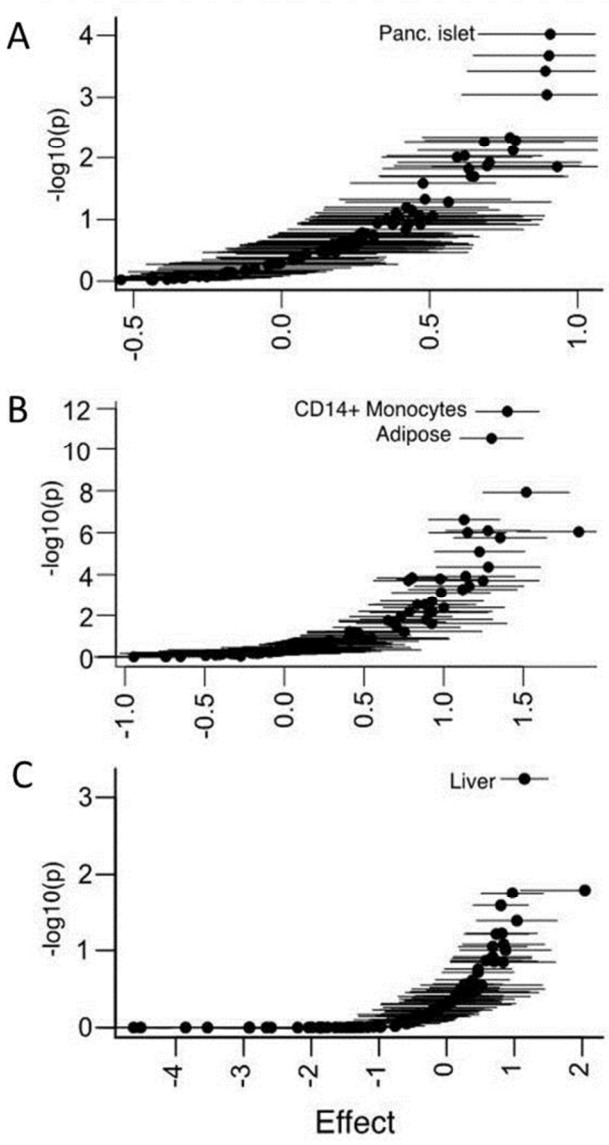


Figure 3. Type 2 diabetes loci stratified by patterns of quantitative trait (e.g. glycaemic, insulin, lipid, and anthropometric) effects show distinct patterns of tissue-specific epigenomic annotation. We hierarchically clustered loci based on endophenotype data and identified groups of T2D loci associated with measures of A) insulin secretion, B) insulin resistance, and C) BMI/lipids. We then looked for enrichment of credible set posterior probabilities for variants mapping to tissue-specific chromatin state annotations. We identified the most significant effects among pancreatic islet annotations for insulin secretion loci, CD14+ monocyte and adipose annotations for insulin resistance loci, and hepatic annotations for BMI/lipid loci.

34x65mm (600 x 600 DPI)

An expanded genome-wide association study of type 2 diabetes in Europeans

Running title: **European T2D genome-wide association study**

DIAbetes Genetics Replication And Meta-analysis (DIAGRAM) Consortium

COMPETING FINANCIAL INTERESTS STATEMENT

Inês Barroso and spouse own stock in GlaxoSmithKline and Incyte.

Jose C Florez has received consulting honoraria from Pfizer and PanGenX.

Valgerdur Steinthorsdottir, Gudmar Thorleifsson, Augustine Kong, Unnur Thorsteinsdottir, and Kari Stefansson are employed by deCODE Genetics/Amgen inc.

Erik Ingelsson is a scientific advisor for Precision Wellness, Cellink and Olink Proteomics for work unrelated to the present project.

Mark I McCarthy sits on Advisory Panels for Pfizer and NovoNordisk, has received honoraria from Pfizer NovoNordisk and EliLilly, and is also a recipient of research funding from Pfizer, NovoNordisk, EliLilly, Takeda, Sanofi-Aventis, Merck, Boehringer-Ingelheim, Astra Zeneca, Janssen, Roche, Servier and Abbvie.

Robert A Scott, Laura J Scott, Reedik Mägi, Letizia Marullo, Kyle J Gaulton, Marika Kaakinen, Natalia Pervjakova, Tune H Pers, Andrew D Johnson, John D Eicher, Anne U Jackson, Teresa Ferreira, Yeji Lee, Clement Ma, Lu Qi, Natalie R Van Zuydam, Anubha Mahajan, Han Chen, Peter Almgren, Ben F Voight, Harald Grallert, Martina Müller-Nurasyid, Janina S Ried, N William Rayner, Neil Robertson, Lennart C Karssen, Elisabeth M van Leeuwen, Sara M Willems, Christian Fuchsberger, Phoenix Kwan, Tanya M Teslovich, Pritam Chanda, Man Li , Yingchang Lu, Christian Dina, Dorothee Thuillier, Loic Yengo, Longda Jiang, Thomas Sparso, Hans Kestler, Himanshu Chheda, Lewin Eisele, Stefan Gustafsson, Mattias Frånberg, Rona J Strawbridge, Rafn Benediktsson, Astradur B Hreidarsson, Gunnar Sigurðsson, Nicola D Kerrison, Jian'an Luan, Liming Liang, Thomas Meitinger, Michael Roden, Barbara Thorand, Tõnu Esko, Evelin Mihailov, Caroline Fox, Ching-Ti Liu, Denis Rybin, Bo Isomaa, Valeriya Lyssenko, Tiinamaija Tuomi, David J Couper, James S Pankow, Niels Grarup, Christian T Have, Marit E Jørgensen, Torben Jørgensen, Allan Linneberg, Marilyn C Cornelis, Rob M van Dam, David J Hunter, Peter Kraft, Qi Sun, Sarah Edkins, Katharine R Owen, John RB Perry, Andrew R Wood, Eleftheria Zeggini, Juan Tajes-Fernandes, Goncalo R Abecasis, Lori L Bonnycastle, Peter S Chines, Heather M Stringham, Heikki A Koistinen, Leena Kinnunen, Bengt Sennblad, Thomas W Mühleisen, Markus M Nöthen, Sonali Pechlivanis, Damiano Baldassarre, Karl Gertow, Steve E Humphries, Elena Tremoli, Norman Klopp, Julia Meyer, Gerald Steinbach, Roman Wennauer, Johan G Eriksson, Satu Männistö, Leena Peltonen, Emmi Tikkanen, Guillaume Charpentier, Elodie Eury, Stéphane Lobbens, Bruna Gigante, Karin Leander, Olga McLeod, Erwin P Bottinger, Omri Gottesman, Douglas Ruderfer, Matthias Blüher, Peter Kovacs, Anke Tonjes, Nisa M Maruthur, Chiara Scapoli, Raimund Erbel, Karl-Heinz Jöckel, Susanne Moebus, Ulf de Faire, Anders Hamsten, Michael Stumvoll, Panagiotis Deloukas, Peter J Donnelly, Timothy M Frayling, Andrew T Hattersley, Samuli Ripatti, Veikko Salomaa, Nancy L Pedersen, Bernhard O Boehm, Richard N Bergman, Francis S Collins, Karen L Mohlke, Jaakko Tuomilehto, Torben Hansen, Oluf Pedersen, Lars Lannfelt, Lars Lind, Cecilia M Lindgren, Stephane Cauchi, Philippe Froguel, Ruth JF Loos, Beverley Balkau, Heiner Boeing, Paul W Franks, Aurelio Barricarte Gurrea, Domenico Palli, Yvonne T van der Schouw, David Altshuler, Leif C

Groop, Claudia Langenberg, Nicholas J Wareham, Eric Sijbrands, Cornelia M van Duijn, James B Meigs, Eric Boerwinkle, Christian Gieger, Konstantin Strauch, Andres Metspalu, Andrew D Morris, Colin NA Palmer, Frank B Hu, Josée Dupuis, Andrew P Morris, Michael Boehnke, and Inga Prokopenko declare to have no competing financial interest.

ACKNOWLEDGEMENTS

ARIC: The Atherosclerosis Risk in Communities Study is carried out as a collaborative study supported by National Heart, Lung, and Blood Institute contracts (HHSN268201100005C, HHSN268201100006C, HHSN268201100007C, HHSN268201100008C, HHSN268201100009C, HHSN268201100010C, HHSN268201100011C, and HHSN268201100012C), R01HL087641, R01HL59367 and R01HL086694; National Human Genome Research Institute contract U01HG004402; and National Institutes of Health contract HHSN268200625226C. Infrastructure was partly supported by Grant Number UL1RR025005, a component of the National Institutes of Health and NIH Roadmap for Medical Research. We wish to acknowledge the many contributions of Dr. Linda Kao, who helped direct the diabetes genetics working group in the ARIC Study until her passing in 2014. We thank the staff and participants of the ARIC study for their important contributions.

BioMe: This work is funded by The Mount Sinai IPM Biobank Program is supported by The Andrea and Charles Bronfman Philanthropies.

D2D2007: The FIN-D2D study has been financially supported by the hospital districts of Pirkanmaa, South Ostrobothnia, and Central Finland, the Finnish National Public Health Institute (National Institute for Health and Welfare), the Finnish Diabetes Association, the Ministry of Social Affairs and Health in Finland, the Academy of Finland (grant number 129293), the European Commission (Directorate C-Public Health grant agreement number 2004310), and Finland's Slottery Machine Association.

DANISH: The study was funded by the Lundbeck Foundation and produced by the Lundbeck Foundation Centre for Applied Medical Genomics in Personalised Disease Prediction, Prevention and Care (LuCamp, www.lucamp.org), and Danish Council for Independent Research. The Novo Nordisk Foundation Center for Basic Metabolic Research is an independent Research Center at the University of Copenhagen, partially funded by an unrestricted donation from the Novo Nordisk Foundation (www.metabol.ku.dk).

DGI: This work was supported by a grant from Novartis. The Botnia study was supported by grants from the Signe and Ane Gyllenberg Foundation, Swedish Cultural Foundation in Finland, Finnish Diabetes Research Society, the Sigrid Juselius Foundation, Folkhälsan Research Foundation, Foundation for Life and Health in Finland, Jakobstad Hospital, Medical Society of Finland, Närpes Research Foundation and the Vasa and Närpes Health centers, the European Community's Seventh Framework Programme (FP7/2007-2013), the European Network for Genetic and Genomic Epidemiology (ENGAGE), the Collarative European Effort to Develop Diabetes Diagnostics (CEED/2008-2012), and the Swedish Research Council, including a Linné grant (No.31475113580).

DGDG: This work was funded by Genome Canada, Génome Quebec, and the Canada Foundation for Innovation. Cohort recruitment was supported by the Association Française des Diabetiques, INSERM, CNAMTS, Centre Hospitalier Universitaire Poitiers, La Fondation de France and the Endocrinology-Diabetology Department of the Corbeil-Essonnes Hospital. C. Petit, J-P. Riveline and S. Franc were instrumental in recruitment and S. Brunet, F. Bacot, R. Frechette, V. Catudal, M. Deweirder, F. Allegaert, P. Laflamme, P. Lepage, W. Astle, M. Leboeuf and S. Leroux provided technical assistance. K. Shazand and N. Foisset provided organizational guidance. We thank all individuals who participated as cases or controls in this study.

deCODE: The study was funded by deCODE Genetics/Amgen inc. and partly supported by ENGAGE HEALTH-F4-2007-201413. We thank the Icelandic study participants and the staff of deCODE Genetics core facilities and recruitment center for their contributions to this work.

DILGOM: The DILGOM study was supported by the Academy of Finland (grant number 118065). V.Salomaa was supported by the Academy of Finland (grant number 139635) and the Finnish Foundation for Cardiovascular Research. S.Mannisto was supported by the Academy of Finland (grant numbers 136895 and 263836). S.R. was supported by the Academy of Finland Center of Excellence in Complex Disease Genetics (grant numbers 213506 and 129680), the Academy of Finland (grant number 251217), the Finnish Foundation for Cardiovascular Research, and the Sigrid Juselius Foundation.

DRsEXTRA: The DR's EXTRA Study was supported by the Ministry of Education and Culture of Finland (627;2004-2011), the Academy of Finland (grant numbers 102318 and 123885), Kuopio University Hospital , the Finnish Diabetes Association, the Finnish Heart Association, the Päivikki and Sakari Sohlberg Foundation, and by grants from European Commission FP6 Integrated Project (EXGENESIS, LSHM-CT-2004-005272), the City of Kuopio, and the Social Insurance Institution of Finland (4/26/2010).

EGCUT: EU grant through the European Regional Development Fund (Project No. 2014-2020.4.01.15-0012), PerMedI (TerVE EstRC), EU H2020 grants 692145, 676550, 654248, and Estonian Research Council, Grant IUT20-60.

EMIL-Ulm: The EMIL Study received support by the State of Baden-Württemberg, Germany, the City of Leutkirch, Germany, and the German Research Council to B.O.B. (GRK 1041). The Ulm Diabetes Study Group received support by the German Research Foundation (DFG-GRK 1041) and the State of Baden-Wuerttemberg Centre of Excellence Metabolic Disorders to B.O.B.

EPIC-InterAct: This work was funded by the EU FP6 programme (grant number LSHM_CT_2006_037197). We thank all EPIC participants and staff for their contribution to the EPIC-InterAct study. We thank the lab team at the MRC Epidemiology Unit for sample management. I.B. was supported by grant WT098051.

FHS: This research was conducted in part using data and resources from the Framingham Heart Study of the National Heart Lung and Blood Institute of the National Institutes of Health and Boston University School of Medicine. The analyses reflect intellectual input and resource development from the Framingham Heart Study investigators participating in the SNP Health Association Resource (SHARe) project. This work was partially supported by the National Heart, Lung and Blood Institute's Framingham Heart Study (contract number N01-HC-25195) and its contract with Affymetrix, Inc for genotyping services (contract number N02-HL-6-4278). A portion of this research utilized the Linux Cluster for Genetic Analysis (LinGA-II) funded by the Robert Dawson Evans Endowment of the Department of Medicine at Boston University School of Medicine and Boston Medical Center. The work is also supported by National Institute for Diabetes and Digestive and Kidney Diseases (NIDDK) R01 DK078616 to J.B.M., J.D. and J.C.F., NIDDK K24 DK080140 to J.B.M., NIDDK U01 DK085526 to H.C., J.D. and J.B.M., and a Massachusetts General Hospital Research Scholars Award to J.C.F..

FUSION: This work was funded by NIH grants U01 DK062370 , R01-HG000376, R01-DK072193, and NIH intramural project number ZIA HG000024. Genome-wide genotyping was conducted by

the Johns Hopkins University Genetic Resources Core. Facility SNP Center at the Center for Inherited Disease Research (CIDR), with support from CIDR NIH contract number N01-HG-65403.

GERA: Data came from a grant, the Resource for Genetic Epidemiology Research in Adult Health and Aging (RC2 AG033067; Schaefer and Risch, PIs) awarded to the Kaiser Permanente Research Program on Genes, Environment, and Health (RPGEH) and the UCSF Institute for Human Genetics. The RPGEH was supported by grants from the Robert Wood Johnson Foundation, the Wayne and Gladys Valley Foundation, the Ellison Medical Foundation, Kaiser Permanente Northern California, and the Kaiser Permanente National and Northern California Community Benefit Programs.

GoDARTS: This study was funded by the Wellcome Trust (084727/Z/08/Z, 085475/Z/08/Z, 085475/B/08/Z) and as part of the EU IMI-SUMMIT program. We acknowledge the support of the Health Informatics Centre, University of Dundee for managing and supplying the anonymised data and NHS Tayside, the original data owner. We are grateful to all the participants who took part in the Go-DARTS study, to the general practitioners, to the Scottish School of Primary Care for their help in recruiting the participants, and to the whole team, which includes interviewers, computer and laboratory technicians, clerical workers, research scientists, volunteers, managers, receptionists, and nurses.

HEINZ NIXDORF RECALL (HNR): We thank the Heinz Nixdorf Foundation [Chairman: M. Nixdorf; Past Chairman: G. Schmidt (deceased)], the German Ministry of Education and Science (BMBF) for the generous support of this study. An additional research grant was received from Imatron Inc., South San Francisco, CA, which produced the EBCT scanners, and GE-Imatron, South San Francisco, CA, after the acquisition of Imatron Inc. We acknowledge the support of the Sarstedt AG & Co. (Nümbrecht, Germany) concerning laboratory equipment. We received support of the Ministry of Innovation, Science and Research, Nordrhine Westfalia for the genotyping of the Heinz Nixdorf Recall study participants. Technical support for the imputation of the Heinz Nixdorf Recall Study data on the Supercomputer Cray XT6m was provided by the Center for Information and Media Services, University of Duisburg-Essen. We are indebted to all the study participants and to the dedicated personnel of both the study center of the Heinz Nixdorf Recall study and the EBT-scanner facilities D. Grönemeyer, Bochum, and R. Seibel, Mülheim, as well as to the investigative group, in particular to U. Roggenbuck, U. Slomiany, E. M. Beck, A. Öffner, S. Münkel, M. Bauer, S. Schrader, R. Peter, and H. Hirche.

HPFS: This work was funded by the NIH grants P30 DK46200, DK58845, U01HG004399, and UM1CA167552.

IMPROVE and SCARFSHEEP: The IMPROVE study was supported by the European Commission (LSHM-CT-2007-037273), the Swedish Heart-Lung Foundation, the Swedish Research Council (8691), the Knut and Alice Wallenberg Foundation, the Foundation for Strategic Research, the Torsten and Ragnar Söderberg Foundation, the Strategic Cardiovascular Programme of Karolinska Institutet, and the Stockholm County Council (560183). The SCARFSHEEP study was supported by the Swedish Heart-Lung Foundation, the Swedish Research Council, the Strategic Cardiovascular Programme of Karolinska Institutet, the Strategic Support for Epidemiological Research at Karolinska Institutet, and the Stockholm County Council. B.S. acknowledges funding from the Magnus Bergvall Foundation and the Foundation for Old Servants. M.F. acknowledges funding from the Swedish e-science Research Center (SeRC). R.J.S. is supported by the Swedish Heart-Lung Foundation, the Tore Nilsson Foundation, the Thuring Foundation, and the Foundation for Old Servants. S.E.H. is funded by the British Heart Foundation (PG08/008).

KORAgen: The KORA research platform (KORA, Cooperative Research in the Region of Augsburg) 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. The KORA research was supported within the Munich Center of Health Sciences (MC Health), Ludwig-Maximilians-Universität, as part of LMUinnovativ. Part of this project was supported by the German Center for Diabetes Research (DZD).

METSIM: The METSIM study was funded by the Academy of Finland (grant numbers 77299 and 124243).

NHS: This work was funded by the NIH grants P30 DK46200, DK58845, U01HG004399, and UM1CA186107.

PPP-MALMO-BOTNIA (PMB): The PPP-Botnia study has been financially supported by grants from the Sigrid Juselius Foundation, the Folkhälsan Research Foundation, the Ministry of Education in Finland, the Nordic Center of Excellence in Disease Genetics, the European Commission (EXGENESIS), the Signe and Ane Gyllenberg Foundation, the Swedish Cultural Foundation in Finland, the Finnish Diabetes Research Foundation, the Foundation for Life and Health in Finland, the Finnish Medical Society, the Paavo Nurmi Foundation, the Helsinki University Central Hospital Research Foundation, the Perklén Foundation, the Ollqvist Foundation, and the Närpes Health Care Foundation. The study has also been supported by the Municipal Heath Care Center and Hospital in Jakobstad and Health Care Centers in Vasa, Närpes and Korsholm. Studies from Malmö were supported by grants from the Swedish Research Council (SFO EXODIAB 2009-1039, LUDC 349-2008-6589, 521-2010-3490, 521-2010-3490, 521-2010-3490, 521-2007-4037, 521-2008-2974, ANDIS 825-2010-5983), the Knut and Alice Wallenberg Foundation (KAW 2009.0243), the Torsten and Ragnar Söderbergs Stiftelser (MT33/09), the IngaBritt and Arne Lundberg's Research Foundation (grant number 359), and the Heart-Lung Foundation.

PIVUS and ULSAM: This work was funded by the Swedish Research Council, Swedish Heart-Lung Foundation, Knut och Alice Wallenberg Foundation, and Swedish Diabetes Foundation. Genome-wide genotyping was funded by the Wellcome Trust and performed by the SNP&SEQ Technology Platform in Uppsala (www.genotyping.se). We thank Tomas Axelsson, Ann-Christine Wiman, and Caisa Pöntinen for their assistance with genotyping. The SNP Technology Platform is supported by Uppsala University, Uppsala University Hospital, and the Swedish Research Council for Infrastructures.

Rotterdam Study: This work 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. This study is funded by the Research Institute for Diseases in the Elderly (014-93-015; RIDE2), the Netherlands Genomics Initiative (NGI)/Netherlands Organisation for Scientific Research (NWO) project nr. 050-060-810. The generation and management of GWAS genotype data for the Rotterdam Study is supported by the Netherlands Organisation of Scientific Research NWO Investments (nr. 175.010.2005.011, 911-03-012). We thank Pascal Arp, Mila Jhamai, Marijn Verkerk, Lizbeth Herrera and Marjolein Peters for their help in creating the GWAS database. The authors thank the study participants, the staff from the Rotterdam Study and the participating general practitioners and pharmacists.

SWEDISH TWIN REGISTRY (STR): This work was supported by grants from the US National Institutes of Health (AG028555, AG08724, AG04563, AG10175, AG08861), the Swedish Research Council, the Swedish Heart-Lung Foundation, the Swedish Foundation for Strategic Research, the Royal Swedish Academy of Science, and ENGAGE (within the European Union FP7 HEALTH-F4-2007-201413). Genotyping was performed by the SNP&SEQ Technology Platform in Uppsala (www.genotyping.se). We thank Tomas Axelsson, Ann-Christine Wiman, and Caisa Pöntinen for their excellent assistance with genotyping. The SNP Technology Platform is supported by Uppsala University, Uppsala University Hospital, and the Swedish Research Council for Infrastructures.

WARREN 2/58BC and WELLCOME TRUST CASE CONTROL CONSORTIUM (WTCCC): Collection of the UK type 2 diabetes cases was supported by Diabetes UK, BDA Research, and the UK Medical Research Council (Biomedical Collections Strategic Grant G0000649). The UK Type 2 Diabetes Genetics Consortium collection was supported by the Wellcome Trust (Biomedical Collections Grant GR072960). Metabochip genotyping was supported by the Wellcome Trust (Strategic Awards 076113, 083948, and 090367, and core support for the Wellcome Trust Centre for Human Genetics 090532), and analysis by the European Commission (ENGAGE HEALTH-F4-2007-201413), MRC (Project Grant G0601261), NIDDK (DK073490, DK085545 and DK098032), and Wellcome Trust (083270 and 098381). WTCCC is funded by Wellcome 076113 and 085475.

Institutional support for study design and analysis: This work was funded by MRC (G0601261), NIDDK (RC2-DK088389, U01-DK105535, U01-DK085545, U01-DK105535), FP7 (ENGAGE HEALTH-F4-2007-201413) and the Wellcome Trust (090532, 098381, 106130, and 090367)

Individual funding for study design and analysis: J.T.-F. is a Marie-Curie Fellow (PIEF-GA-2012-329156). M.K. is supported by the European Commission under the Marie Curie Intra-European Fellowship (project MARVEL, PIEF-GA-2013-626461). C.Langenberg, R.A.S. and N.J.W. are funded by the Medical Research Council (MC_UU_12015/1). L.M. is partially supported by 2010-2011 PRIN funds of the University of Ferrara – Holder: Prof. Guido Barbujani – and in part sponsored by the European Foundation for the Study of Diabetes (EFSD) Albert Renold Travel Fellowships for Young Scientists, and by the fund promoting internationalisation efforts of the University of Ferrara – Holder: Prof. Chiara Scapoli. A.P.M. is a Wellcome Trust Senior Fellow in Basic Biomedical Science (grant number WT098017). M.I.M. is a Wellcome Trust Senior Investigator. J.R.B.P is supported by the Wellcome Trust (WT092447MA). T.H.P. is supported by The Danish Council for Independent Research Medical Sciences (FSS) The Lundbeck Foundation and The Alfred Benzon Foundation. I.P. was in part funded by the Elsie Widdowson Fellowship, the Wellcome Trust Seed Award in Science (205915/Z/17/Z) and the European Union's Horizon 2020 research and innovation programme (DYNAhealth, project number 633595). B.F.V. is supported by the NIH/NIDDK (R01DK101478) and the American Heart Association (13SDG14330006). E. Z. is supported by the Wellcome Trust (098051). S.E.H. is funded by British Heart Foundation PG08/008 and UCL BRC. V.Salomaa was supported by the Academy of Finland (grant # 139635) and by the Finnish Foundation for Cardiovascular Research.

An expanded genome-wide association study of type 2 diabetes in Europeans

DIAbetes Genetics Replication And Meta-analysis (DIAGRAM) Consortium

SUPPLEMENTARY INFORMATION

Contents

SUPPLEMENTARY MATERIAL

Research participants

The DIAGRAM stage 1 analyses comprised a total of 26,676 T2D cases and 132,532 control participants from 18 GWAS. The Metabochip stage 2 follow up comprised 16 studies (D2D2007, DANISH, DIAGEN, DILGOM, DRsEXTRA, EMIL-Ulm, FUSION2, NHR, IMPROVE, InterACT-CMC, Leipzig, METSIM, HUNT/TROMSO, SCARFSHEEP, STR, Warren2/58BC) with Metabochip data (1), in which the participants did not overlap those included in stage 1. Stage 1 study sizes ranged between 80 and 7,249 T2D cases and from 455 to 83,049 controls. The study characteristics are described in detail in **Supplementary Table 1**. The Metabochip follow-up study sizes ranged from 101 and 3,553 T2D cases and from 586 to 6,603 controls. Details of Metabochip replication cohorts have been described in detail previously (1,2). For SNVs not captured on Metabochip directly or by proxy, we performed follow-up in 2,796 individuals with T2D and 4,601 controls from the EPIC-InterAct study (3). In addition, we used 9,747 T2D cases and 61,857 controls from the GERA study (4) to follow-up six low frequency variants not captured on Metabochip. All study participants were of European ancestry and were from the United States and Europe. All studies were approved by local research ethic committees, and all participants gave written informed consent.

Overview of Study Design and Analysis Strategy

We performed inverse-variance weighted fixed-effect meta-analyses of 18 stage 1 GWAS (**Supplementary Table 1**). Following imputation to the 1000G multi-ethnic reference panel, each study performed T2D association analysis using logistic regression, adjusting for age, sex, and study-specific covariates, under an additive genetic model. Fifteen of the 18 studies repeated analyses also adjusting for body mass index (BMI). A total of 40 loci reached genome-wide significance ($p=5x10^{-8}$) in the stage 1 meta-analysis, of which four mapped >500kb from previously-known T2D-associated loci, and were therefore considered likely to represent novel signals. At a lesser level of significance ($p<10^{-5}$), we identified 48 additional putative novel signals. In stage 1, we identified fifty-two regions in which the most strongly associated SNP had a $p<10^{-5}$, was greater than 500kb distant from the nearest known T2D associated variant and was in r^2 <.02 with all known T2D associated variants. Of the combined set of 52 putative novel signals, 46 featured a lead SNV with MAF >5%. From each of these 52 regions, we selected the most strongly-associated variant for follow-up in stage 2. As the stage 1 meta-analysis had exhausted most European-ancestry studies with available GWAS data, stage 2 was primarily based on 16 independent European-ancestry studies (2) genotyped on the Metabochip custom array (5). Of the 52 putative lead variants from stage 1, 29 variants or their LD proxies ($r^2\geq0.6$) were present in MetaboChip. Specifically, four SNVs were themselves present on the Metabochip, 20 were represented by a proxy ($r^2>0.8$) and an additional 5 by a proxy in lower linkage disequilibrium (LD) ($0.8>r^2>0.6$) (**Table 1, Supplementary Table 6, Supplementary Figure 1A-C**). Novel loci were defined using the threshold for genome-wide significance in the combined stage 1 and stage 2 meta-analysis or in stage 1 alone, when no suitable proxy was available. The remaining 23 variants were followed-up in EPIC-InterAct study. We neither observed any additional signals attaining genome-wide significance threshold, nor detected any nominally significant effects in this follow-up stage alone. Six low-frequency variants were followed-up additionally in the GERA study (**Supplementary Table 6**).

Genotyping, imputation and quality control

Genotyping of individual stage 1 studies was carried out using commercial genome-wide single-nucleotide variant (SNV) arrays as detailed in **Supplementary Table 1**. We excluded samples and SNPs as described in **Supplementary Table 1.** We imputed autosomal and X chromosome SNVs using the all ancestries 1000 Genomes Project (1000G) reference panel (1,092 individuals from Africa, Asia, Europe, and the Americas, (March, 2012 release)) using miniMAC (6) or IMPUTE2 (7). EPIC-InterAct was genotyped on the Illumina HumanCoreExome chip and imputed using the 1000G reference panel (March, 2012 release). The imputation parameters are given in **Supplementary Table 1**. Insertion/deletion variants were not analysed due to the lower quality of their calls in the 1000G reference panel release used as compared to later panel releases. After imputation, from each study we removed monomorphic imputed variants or those with study-

specific imputation quality r^2-hat<0.3 (miniMAC) or proper-info<0.4 (IMPUTE2, SNPTEST). Metabochip studies were imputed using with the same 1000G panel (1,2) as used in Stage 1.

To compare the variant imputation quality and distribution of minor allele frequency (MAF) for variants imputed using the 1000G March 2012 reference panel to those imputed using the HapMap2 reference panel European individuals, we also imputed into the WTCCC sample using HapMap2 reference panel European individuals. We independently binned the SNVs from the two imputation panels by allele frequency and computed the per-bin SNP number and the average proper_info score.

Statistical analyses

In stage 1, in each study we performed logistic regression association analysis of T2D with genotype dosage using an additive genetic model including as covariates age, sex and principal components derived from the genetic data to account for population stratification. We further applied genomic control (GC) correction to study-level association summary statistics to correct for residual population structure not accounted for by principal components adjustment. We combined the association results using inverse variance-weighted fixed effect meta-analysis using both GWAMA (8) and METAL (9) , and observed identical results. The stage 1 meta-analysis had 11.7M autosomal and 260k chromosome X SNVs that 1) had a total minor allele count >5 and 2) were present in ≥3 studies. The lambda (GC) value was 1.08, while inflation estimates from LDscore regression (10) showed no evidence of population stratification suggesting lambda (GC)=1. We performed inverse variance weighted fixed-effects meta-analysis of the 16 stage 2 Metabochip studies (lambda GC correction applied based on QT-interval variant set (1)) and the 18 stage 1 studies using GWAMA (8) and METAL (9) software. Heterogeneity was assessed using the I^2 index from the complete study-level meta-analysis. We combined stage 1 and stage 2 results by inverse variance-weighted fixed-effect meta-analysis.

We performed a secondary T2D association analysis by modelling body mass index (BMI) as covariate in 15 studies (not including DGDG, GoDARTS and WTCCC). The total sample size for this analysis was 21,440 T2D cases and 97,052 controls, (N_{eff}=70,242). The lambda (GC) was 1.05. Genetic effect sizes (beta coefficients) estimated from models with and without BMI adjustments were compared using a matched analysis within the same subset of 15 studies: $\frac{(\beta_{noBMI}-\beta_{BMI})}{\sqrt{SE(\beta_{noBMI})^2+SE(\beta_{BMI})^2-2\rho\times SE(\beta_{noBMI})\times SE(\beta_{BMI})}}$, where β_{BMI} and β_{noBMI} are the estimated genetic effect from models with and without BMI adjustment, $SE(\beta)$ is the estimated standard error of the estimates, and ρ is the estimated correlation between β_{BMI} and β_{noBMI} obtained from all genetic variants (ρ=0.90).

Comparison between HapMap and 1000G reference variant sets

We made LocusZoom(11) regional plots of the Stage 1 meta-analysis results indexed by lead SNV for the 13 novel loci, and estimated LD using the EUR 1000G March 2012 variant set (**Supplementary Figure 2**). We also made regional plots indexed by the lead 1000G SNV, but otherwise only including SNVs present in the previous HapMap2-imputed analyses(1,12).

Power calculations

We performed power calculations[10] over a range of odds ratios (ORs), using the corresponding genotype relative risk (GRR) in the power calculation, to (i) determine the effect size that would yield 80% power based on a grid search and (ii) to provide power estimates for pre-specified ORs, for specified risk allele frequency (RAF). The RAF is defined as the frequency of the allele that increases T2D risk in the stage 1 meta-analysis. We determined power as a function of the GRR, RAF, alpha=5×10^{-8}, and the average weighted effective case sample size, assuming a 1:1 ratio of cases and controls. For each variant, we defined weighted effective case sample size as the product of the variant-specific effective case sample size and the average variant-specific imputation quality (based on r^2 hat or info measures available from each included study). To calculate the average weighted effective case sample size, for each RAF we selected the 10,000 stage 1 meta-analysis variants with RAF closest to the target RAF (taking equal proportions of variants above and below the RAF), and took the average of the 10,000 weighted effective case sample sizes.

Approximate conditional analysis with GCTA

To identify if multiple statistically independent signals were present in known and novel T2D associated regions, we performed approximate conditional analysis in the stage 1 sample using GCTA (v1.24) (13). Among 70 established T2D-associated and 13 novel loci ($p<5\times10^{-4}$), we analysed SNVs in the 1Mb-window around each lead variant, conditioning on the lead SNV at each locus. We ran the GCTA analysis using three separate genotype reference panels for estimation of LD between variants (14): UK10K project (N=3,621), Genetics of Diabetes Audit and Research in Tayside Scotland (GoDARTS (15)) study (3,298 T2D cases and 3,708 controls) and Prospective Investigation of the Vasculature in Uppsala Seniors (PIVUS (16)) study (n=949). We considered loci as containing distinct signals (in the initial and further rounds of analysis) if a SNV reached locus-wide significance after accounting for region-specific multiple testing ($p<10^{-5}$) in all three reference panels. Where we observed distinct signals, we then conditioned on the original lead SNV, and the newly observed distinct SNV(s) to detect further signals, until no additional signal was identified at $p<10^{-5}$. We identified six regions with more than one independent signal (18 distinct signals). In each region with multiple signals, for each independent variant we conditioned on all other independent variants in the region and used these results were used for finemapping (below). At *KCNQ1*, we performed conditioning using GCTA model selection which better handles the large number of independent signals (using the UK10K reference panel).

Finemapping analyses using credible set mapping

The goal of finemapping was to identify sets of 99% credible causal variants for the lead independent variants at known and novel loci. We used credible set fine-mapping (17) within 95 distinct signals (at 82 loci) with T2D-association signals $p<5x10^{-4}$ in the present stage 1 to investigate whether 1000G-imputation allowed us to better resolve the specific variants driving these associations (**Supplementary Tables 3 and 9**). We included in the credible set analysis all signals where the lead independent SNV reached $p<5x10^{-4}$ in the stage 1 meta-analysis, as SNVs with weak association, mostly those identified in non-European GWASs, generally yield very large credible SNP sets. In regions with multiple independent variants, we used the signal remaining following approximate conditional analysis on all other independent variants in the region (see above). To define the locus boundaries, for each lead SNV we identified the outermost variants from the set of variants in $r^2 \geq .2$ with the lead SNV and added an additional flanking region of .02 cM to each side. To perform credible set mapping, the T2D stage 1 meta-analysis results were converted to Bayes' factors (BF) for each variant within the variant/locus boundary (17). The posterior probability that SNV_j was causal was defined by:

$$\varphi_j = \frac{BF_j}{\sum_k BF_k}$$

where, BF_j denotes the BF for the jth SNV, and the denominator is the sum of all included BFs. A 99% credible set of variants was created by ranking the posterior probabilities from highest to lowest and summing them until the cumulative posterior probability exceeded 0.99. To estimate the credible set sizes we would have observed with HapMap imputation-based meta-analysis results, we recomputed the posterior probabilities after first restricting to variants observed in previous HapMap-imputed analyses.

T1D/T2D discrimination analysis

Given the overlap between loci previously associated with T1D and the newly associated T2D loci, we used an inverse variance weighted Mendelian randomisation approach (18) to test whether this was likely to reflect misclassification of T1D cases as individuals with T2D in the current study. Briefly, using 50 SNVs associated with T1D at genome-wide significance (19), we tested the association of genetic predisposition to T1D with T2D in the present analysis. If some proportion of T2D cases in the current study actually are T1D, we would expect that the T1D risk variants to consistently predict T2D risk. We performed analysis with and without the lead SNVs showing associations with both T1D and T2D ($p<0.05$ for T2D).

Expression quantitative trait loci (eQTL) analysis

Lead SNVs at all 13 novel loci mapped to non-coding sequence, leaving uncertain the identities of the effector transcripts through which the T2D-risk effects are mediated. To highlight potential effectors, we first considered RNA expression data, focusing on data from pancreatic islets, adipose, muscle, liver, and whole blood, and seeking coincidence (r^2>0.8) between the lead T2D-associated SNVs and drivers of regional cis-eQTLs ($p<5x10^{-6}$) (**Supplementary Table 10**). To look for potential biological overlap of T2D lead variants and eQTL variants, we extracted the lead (most significantly associated) eQTL for each tested gene from existing datasets for pancreatic islets (20), skeletal muscle (21,22), adipose tissue(22–26), liver (22,24,27–30) and whole blood (which has the largest sample size of available eQTL studies) (22,23,26,31–47) . Additional eQTL data was integrated from online sources including ScanDB (http://www.scandb.org/newinterface/about.html), the Broad Institute GTEx Portal (http://www.gtexportal.org/home/), and the Pritchard Lab (eqtl.uchicago.edu). Additional liver eQTL data was downloaded from ScanDB and cis-eQTLs were limited to those with $p<10^{-6}$.We considered that a lead T2D SNV showed potential evidence of influencing gene expression if it was in high LD (r^2>0.8) with the lead eQTL SNP, and if the lead eQTL SNP had $p< 5 \times 10^{-6}$

Hierarchical clustering of T2D-related metabolic phenotypes

Starting with the T2D associated SNV variants in the finemapping set, we identified sets of variants with similar patterns of T2D related quantitative trait association. For the T2D associated SNVs, we obtained T2D-related quantitative trait z scores from published HapMap-based GWAS meta-analysis for: fasting glucose (FG (48)), fasting insulin adjusted for BMI (FIadjBMI (48)), homeostasis model assessment for beta-cell function (HOMA-B (48)), homeostasis model assessment for insulin resistance (HOMA-IR (48)), 2-h glucose adjusted for BMI (2hGluadjBMI (49)), proinsulin (PR (50)), corrected insulin response (CIR (51)), body mass index (52), high density lipoprotein (HDL-C), low density lipoprotein (LDL-C), total cholesterol (TC), triglycerides (TG), all from the Global Lipids Genetics Consortium (53). When the result for a SNV was not available, we used the results from the variant in highest r^2 (r^2>0.6). We coded the z-scores such that a positive sign indicated that the trait value was higher for the T2D risk allele, a negative sign that the trait value was lower for the T2D risk allele. We performed complete linkage hierarchical clustering and used the Euclidian distance dissimilarity measure L^2=15% as a threshold to define the loci clusters. We tested the validity of groups through multi-scale bootstrap resampling with 50,000 bootstrap replicates, as described previously(54). All distances, clustering analyses and statistical calculations were done using *stats*, *gplots*, *pvclust*, *fpc* and *vegan* packages in the R programming language (R Core Team (2013) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/).

Functional annotation and enrichment analysis

We tested for enrichment of genomic and epigenomic annotations obtained from two sources. First, we obtained chromatin states for 93 cell types (after excluding cancer cell lines) from the NIH Epigenome Roadmap project. For each cell type, we collapsed active enhancer (EnhA) and promoter (TssA) states into one annotation for that cell type. Secondly, we obtained binding sites for 165 transcription factors (TF) from ENCODE (55) and Pasquali et al. (56). We first sought to extend these analyses to the denser variant coverage and expanded number of GWAS signals in the present meta-analysis (**Supplementary Table 9**). Across credible sets for the 95 distinct signals with $p<5x10^{-4}$ in the present stage 1 European analysis (**Supplementary Tables 3 and 9**), we used a fractional logistic regression model to compare a binary indicator of variants overlapping a total of 261 functional annotations to the posterior probabilities for association derived from the fine-mapping analysis (π_c) (**Supplementary Table 12**). For each TF, we collapsed all binding sites into one annotation. We then tested for the effect of variants with each cell type and TF annotation on the variant posterior probabilities (π_c) using all variants in the 95 credible regions (ie 100% credible sets). We used a generalized linear model where the dependent variable is π_c value for each variant and the predictor variable is a binary indicator of overlap of the variant and the annotation, *a* (1 if yes, 0 if no). We included several additional binary indicators for generic gene-based annotations in the

model for each annotation - 3UTR (*u*), 5UTR (*v*), coding exon (*c*), and within 1kb upstream of GENCODE Tss (*t*) - as well as a categorical variable for locus membership (*l*).

$$\log\left(\frac{\pi_c}{1-\pi_c}\right) = \beta_0 + \beta_1 a + \beta_2 u + \beta_3 v + \beta_4 c + \beta_5 t + \beta_6 l \ , \qquad \pi_c \sim Binomial$$

For each annotation, we obtained the estimated effect size and standard error from this model. We then re-calculated the standard error using the sandwich variance estimator (R package sandwich). We calculated a z-score by dividing the effect size by the re-estimated standard error, and calculated a two-sided p-value from the z-score. We also applied this model to the three subsets of loci visually identified from the hierarchical clustering as having similar T2D-related trait association patterns. In each analysis, we considered an annotation significant if it reached a Bonferroni-corrected p-value threshold of $2x10^{-4}$ (.05/256 annotations).

Pathway analyses with DEPICT

We used the Data-driven Expression Prioritized Integration for Complex Traits (DEPICT) tool (57) to i) prioritize genes that may represent promising candidates for T2D pathophysiology, and (ii) identify reconstituted gene sets that are enriched in genes from associated regions and might be related to T2D biological pathways. As input we used independent SNVs (LD-pruning parameters: r^2<0.05 in the 1000 Genomes project phase 1 reference panel including 268 unrelated individuals from CEU, GBR and TSI populations; release date 2011-05-21; physical distance threshold=500kb) selected from the set including stage1 meta-analysis SNVs with $p<10^{-5}$ and lead variants at established loci. We then used the DEPICT method (57) to construct associated regions by mapping genes to independently associated SNVs, if they overlapped or resided within LD window (r^2 >.0.5) with the independently associated SNV. Variants within the major histocompatibility complex region (chromosome 6, base pairs 25,000,000 through 35,000,000) were excluded. This gave 206 independent regions covering 328 genes for the analysis with DEPICT. For the calculation of empirical enrichment *p* values, we used 200 sets of SNVs randomly drawn from entire genome within regions matching by gene density; we performed 20 replications for FDR estimation. For each significantly enriched reconstituted gene set, we plotted the five genes that most strongly mapped to the given gene sets and resided within an associated T2D locus. The mapping strength between a gene and a reconstituted gene set was denoted by a Z-score shown in parenthesis after the gene identifier in **Supplementary Table 10**. After the gene set enrichment analysis, we omitted reconstituted gene sets for which genes in the original gene set were not nominally enriched (Wilcoxon rank-sum test). By design, genes in the original gene set are expected to be enriched in the reconstituted gene set; lack of enrichment complicates interpretation of the reconstituted gene set because the label of the reconstituted gene set will be inaccurate. Using this procedure the "Megacephaly" reconstituted gene set was removed from the results. To visualize the 20 reconstituted gene sets with $p<10^{-5}$ in Cytoscape (58) (**Supplementary Figure 10**), we estimated their overlap by computing the pairwise Pearson correlation coefficient r between each pair of gene sets followed by discretization into one of three bins; $0.3\leq\rho<0.5$ as low overlap, $0.5\leq\rho<0.7$ as medium overlap, and $\rho\geq0.7$ as high overlap.

Supplementary material and methods references

1. Morris AP, Voight BF, Teslovich TM, Ferreira T, Segrè A V, Steinthorsdottir V, et al. Large-scale association analysis provides insights into the genetic architecture and pathophysiology of type 2 diabetes. Nat Genet. 2012;44(9):981–90.

2. Gaulton KJ, Ferreira T, Lee Y, Raimondo A, Mägi R, Reschen ME, et al. Genetic fine mapping and genomic annotation defines causal mechanisms at type 2 diabetes susceptibility loci. Nat Genet. 2015;47(12):1415–25.

3. Langenberg C, Sharp S, Forouhi NG, Franks PW, Schulze MB, Kerrison N, et al. Design and cohort description of the InterAct Project: an examination of the interaction of genetic and lifestyle factors on the incidence of type 2 diabetes in the EPIC Study. Diabetologia. 2011;54(9):2272–82.

4. Cook JP, Morris AP. Multi-ethnic genome-wide association study identifies novel locus for type 2 diabetes susceptibility. Eur J Hum Genet. England; 2016;24(8):1175–80.

5. Voight BF, Kang HM, Ding J, Palmer CD, Sidore C, Chines PS, et al. The metabochip, a custom genotyping array for genetic studies of metabolic, cardiovascular, and anthropometric traits. PLoS Genet. 2012;8(8):e1002793.

6. Howie B, Fuchsberger C, Stephens M, Marchini J, Abecasis GR. Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. Nat Genet. 2012;44(8):955–9.

7. Howie BN, Donnelly P, Marchini J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. PLoS Genet. 2009;5(6):e1000529.

8. Mägi R, Morris AP. GWAMA: software for genome-wide association meta-analysis. BMC Bioinformatics. 2010;11:288.

9. Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. Bioinformatics. 2010;26(17):2190–1.

10. Bulik-Sullivan BK, Loh P-R, Finucane HK, Ripke S, Yang J, Consortium SWG of the PG, et al. LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. Nat Genet. 2015; 47(3):326–41.

11. Pruim RJ, Welch RP, Sanna S, Teslovich TM, Chines PS, Gliedt TP, et al. LocusZoom: Regional visualization of genome-wide association scan results. Bioinformatics. 2010;26(18):2336–7.

12. Voight BF, Scott LJ, Steinthorsdottir V, Morris ADP, Dina C, Welch RP, et al. Twelve type 2 diabetes susceptibility loci identified through large-scale association analysis. Nat Genet. 2010;42(7):579–89.

13. Yang J, Lee SH, Goddard ME, Visscher PM. GCTA: a tool for genome-wide complex trait analysis. Am J Hum Genet. 2011;88(1):76–82.

14. UK10K Consortium, Writing group, Production group, Cohorts group, Neurodevelopmental disorders group, Obesity group, et al. The UK10K project identifies rare variants in health and disease. Nature. 2015;526(7571):82–90.

15. Morris AD, Boyle DI, MacAlpine R, Emslie-Smith A, Jung RT, Newton RW, et al. The diabetes audit and research in Tayside Scotland (DARTS) study: electronic record linkage to create a diabetes register. DARTS/MEMO Collaboration. BMJ. 1997;315(7107):524–8.

16. Lind L, Fors N, Hall J, Marttala K, Stenborg A. A comparison of three different methods to evaluate endothelium-dependent vasodilation in the elderly: the Prospective Investigation of the Vasculature in Uppsala Seniors (PIVUS) study. Arterioscler Thromb Vasc Biol. 2005;25(11):2368–75.

17. Maller JB, McVean G, Byrnes J, Vukcevic D, Palin K, Su Z, et al. Bayesian refinement of association signals for 14 loci in 3 common diseases. Nat Genet. 2012 Dec;44(12):1294–301.

18. Burgess S, Butterworth A, Thompson SG. Mendelian randomization analysis with multiple genetic variants using summarized data. Genet Epidemiol. 2013;37(7):658–65.

19. Burren OS, Adlem EC, Achuthan P, Christensen M, Coulson RMR, Todd J a. T1DBase: Update 2011, organization and presentation of large-scale data sets for type 1 diabetes research. Nucleic Acids Res. 2011;39(SUPPL. 1):997–1001.

20. Fadista J, Vikman P, Laakso EO, Mollet IG, Esguerra J Lou, Taneera J, et al. Global genomic and transcriptomic analysis of human pancreatic islets reveals novel genes influencing glucose metabolism. Proc Natl Acad Sci U S A. 2014;111(38):13924–9.

21. Keildson S, Fadista J, Ladenvall C, Hedman AK, Elgzyri T, Small KS, et al. Expression of phosphofructokinase in skeletal muscle is influenced by genetic variation and associated with insulin

sensitivity. Diabetes. 2014;63(3):1154–65.

22. Foroughi Asl H, Talukdar H a., Kindt ASD, Jain RK, Ermel R, Ruusalepp A, et al. Expression Quantitative Trait Loci Acting Across Multiple Tissues Are Enriched in Inherited Risk for Coronary Artery Disease. Circ Cardiovasc Genet. 2015;8(2):305–15.

23. Emilsson V, Thorleifsson G, Zhang B, Leonardson AS, Zink F, Zhu J, et al. Genetics of gene expression and its effect on disease. Nature. 2008; 452(7186):423-8.

24. Greenawalt DM, Dobrin R, Chudin E, Hatoum IJ, Suver C, Beaulaurier J, et al. A survey of the genetics of stomach, liver, and adipose gene expression from a morbidly obese cohort. Genome Res. 2011;21(7):1008–16.

25. Grundberg E, Small KS, Hedman ÅK, Nica AC, Buil A, Keildson S, et al. Mapping cis- and trans-regulatory effects across multiple tissues in twins. Nat Genet. 2012;44(10):1084–9.

26. GTEx Consortium. The Genotype-Tissue Expression (GTEx) project. Nat Genet. 2013 Jun;45(6):580–5.

27. Schadt EE, Molony C, Chudin E, Hao K, Yang X, Lum PY, et al. Mapping the genetic architecture of gene expression in human liver. PLoS Biol. 2008;6(5):1020–32.

28. Schröder A, Klein K, Winter S, Schwab M, Bonin M, Zell a, et al. Genomics of ADME gene expression: mapping expression quantitative trait loci relevant for absorption, distribution, metabolism and excretion of drugs in human liver. Pharmacogenomics J. 2011; 13(1):12–20.

29. Wang X, Tang H, Teng M, Li Z, Li J, Fan J, et al. Mapping of hepatic expression quantitative trait loci (eQTLs) in a Han Chinese population. J Med Genet. 2014;51(5):319–26.

30. Innocenti F, Cooper GM, Stanaway IB, Gamazon ER, Smith JD, Mirkov S, et al. Identification, replication, and functional fine-mapping of expression quantitative trait loci in primary human liver tissue. PLoS Genet. 2011;7(5):e1002078.

31. Westra H-J, Peters MJ, Esko T, Yaghootkar H, Schurmann C, Kettunen J, et al. Systematic identification of trans eQTLs as putative drivers of known disease associations. Nat Genet. 2013;45(10):1238–43.

32. Fehrmann RSN, Jansen RC, Veldink JH, Westra HJ, Arends D, Bonder MJ, et al. Trans-eqtls reveal that independent genetic variants associated with a complex phenotype converge on intermediate genes, with a major role for the HLA. PLoS Genet. 2011;7(8):e1002197.

33. Mehta D, Heim K, Herder C, Carstensen M, Eckstein G, Schurmann C, et al. Impact of common regulatory single-nucleotide variants on gene expression profiles in whole blood. Eur J Hum Genet. 2013;21(1):48–54.

34. Zhernakova D V., de Klerk E, Westra HJ, Mastrokolias A, Amini S, Ariyurek Y, et al. DeepSAGE Reveals Genetic Variants Associated with Alternative Polyadenylation and Expression of Coding and Non-coding Transcripts. PLoS Genet. 2013;9(6):e1003594.

35. Sasayama D, Hori H, Nakamura S, Miyata R, Teraishi T, Hattori K, et al. Identification of Single Nucleotide Polymorphisms Regulating Peripheral Blood mRNA Expression with Genome-Wide Significance: An eQTL Study in the Japanese Population. PLoS One. 2013;8(1):e54967.

36. Landmark-Høyvik H, Dumeaux V, Nebdal D, Lund E, Tost J, Kamatani Y, et al. Genome-wide association study in breast cancer survivors reveals SNPs associated with gene expression of genes belonging to MHC class I and II. Genomics. 2013;102(4):278–87.

37. van Eijk KR, de Jong S, Boks MPM, Langeveld T, Colas F, Veldink JH, et al. Genetic analysis of DNA methylation and gene expression levels in whole blood of healthy human subjects. BMC Genomics. 2012;13(1):636.

38. Battle A, Mostafavi S, Zhu X, Potash JB, Weissman MM, McCormick C, et al. Characterizing the genetic basis of transcriptome diversity through RNA-sequencing of 922 individuals. Genome Res. 2014;24(1):14–24.

39. Benton MC, Lea R a., Macartney-Coxson D, Carless M a., G??ring HH, Bellis C, et al. Mapping eQTLs in the Norfolk Island genetic isolate identifies candidate genes for CVD risk traits. Am J Hum Genet. 2013;93(6):1087–99.

40. Narahara M, Higasa K, Nakamura S, Tabara Y, Kawaguchi T, Ishii M, et al. Large-scale East-Asian eQTL mapping reveals novel candidate genes for LD mapping and the genomic landscape of transcriptional effects of sequence variants. PLoS One. 2014;9(6):e100924.

41. Quinlan J, Idaghdour Y, Goulet J-P, Gbeha E, de Malliard T, Bruat V, et al. Genomic architecture of sickle cell disease in West African children. Front Genet. 2014;5:26.

42. Wright F a, Sullivan PF, Brooks AI, Zou F, Sun W, Xia K, et al. Heritability and genomics of gene expression in peripheral blood. Nat Genet. Nature Publishing Group; 2014;46(5):430–7.

43. Schramm K, Marzi C, Schurmann C, Carstensen M, Reinmaa E, Biffar R, et al. Mapping the genetic architecture of gene regulation in whole blood. PLoS One. 2014;9(4):e93844.

44. Lock E, Soldano K, Garrett M, Cope H, Markunas C, Fuchs H, et al. Joint eQTL assessment of whole blood and dura mater tissue from individuals with Chiari type I malformation. BMC Genomics. 2015;16(1):11.

45. Powell JE, Henders AK, McRae AF, Caracella A, Smith S, Wright MJ, et al. The Brisbane systems genetics study: Genetical genomics meets complex trait genetics. PLoS One. 2012;7(4):e35430.

46. Pierce BL, Tong L, Chen LS, Rahaman R, Argos M, Jasmine F, et al. Mediation Analysis Demonstrates That Trans-eQTLs Are Often Explained by Cis-Mediation: A Genome-Wide Analysis among 1,800 South Asians. PLoS Genet. 2014;10(12):e1004818.

47. Chen W, Brehm JM, Lin J, Wang T, Forno E, Acosta-Pérez E, et al. Expression quantitative trait loci (eQTL) mapping in Puerto Rican children. PLoS One. 2015;10(3):e0122464.

48. Manning AK, Hivert M-F, Scott RA, Grimsby JL, Bouatia-Naji N, Chen H, et al. A genome-wide approach accounting for body mass index identifies genetic variants influencing fasting glycemic traits and insulin resistance. Nat Genet. 2012;44(6):659–69.

49. Saxena R, Hivert M-F, Langenberg C, Tanaka T, Pankow JS, Vollenweider P, et al. Genetic variation in GIPR influences the glucose and insulin responses to an oral glucose challenge. Nat Genet. 2010;42(2):142–8.

50. Strawbridge RJ, Dupuis J, Prokopenko I, Barker A, Ahlqvist E, Rybin D, et al. Genome-wide association identifies nine common variants associated with fasting proinsulin levels and provides new insights into the pathophysiology of type 2 diabetes. Diabetes. 2011;60(10):2624–34.

51. Prokopenko I, Poon W, Mägi R, Prasad B R, Salehi SA, Almgren P, et al. A central role for GRB10 in regulation of islet function in man. PLoS Genet. 2014;10(4):e1004235.

52. Speliotes EK, Willer CJ, Berndt SI, Monda KL, Thorleifsson G, Jackson AU, et al. Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index. Nat Genet. 2010;42(11):937–48.

53. Willer CJ, Schmidt EM, Sengupta S, Peloso GM, Gustafsson S, Kanoni S, et al. Discovery and refinement of loci associated with lipid levels. Nat Genet. 2013;45(11):1274–83.

54. Dimas AS, Lagou V, Barker A, Knowles JW, Mägi R, Hivert MF, et al. Impact of type 2 diabetes susceptibility variants on quantitative glycemic traits reveals mechanistic heterogeneity. Diabetes. 2014;63(6):2158–71.

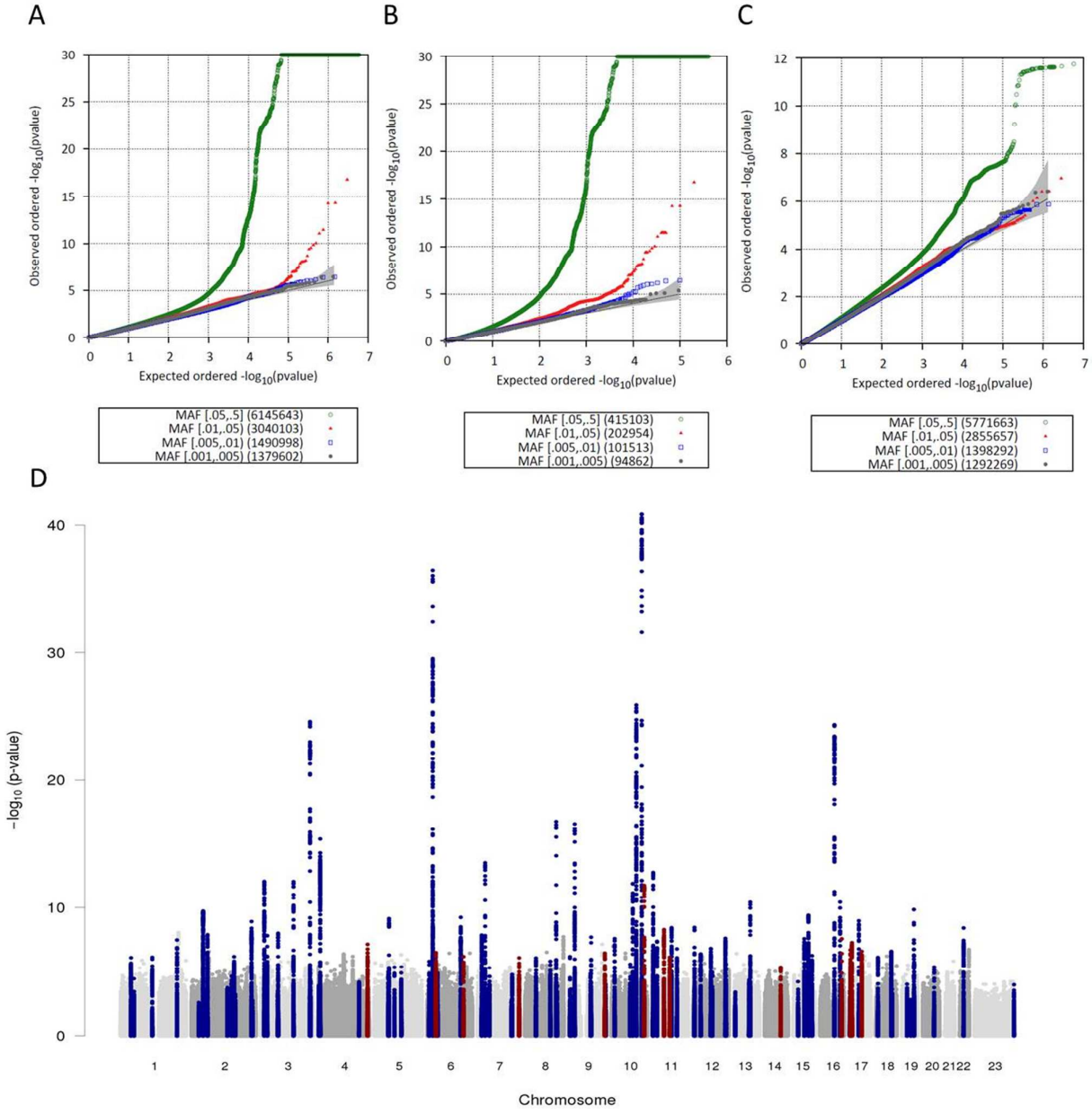
55. Dunham I, Kundaje A, Aldred SF, Collins PJ, Davis C a, Doyle F, et al. An integrated encyclopedia of DNA elements in the human genome. Nature. 2012;489(7414):57–74.

56. Pasquali L, Gaulton KJ, Rodríguez-Seguí S a, Mularoni L, Miguel-Escalada I, Akerman I, et al. Pancreatic islet enhancer clusters enriched in type 2 diabetes risk-associated variants. Nat Genet. 2014;46(2):136–43.

57. Pers TH, Karjalainen JM, Chan Y, Westra H, Wood AR, Yang J, et al. Biological interpretation of genome-wide association studies using predicted gene functions. Nat Commun. 2015;6:5890.

58. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. 2003;13(11):2498–504.

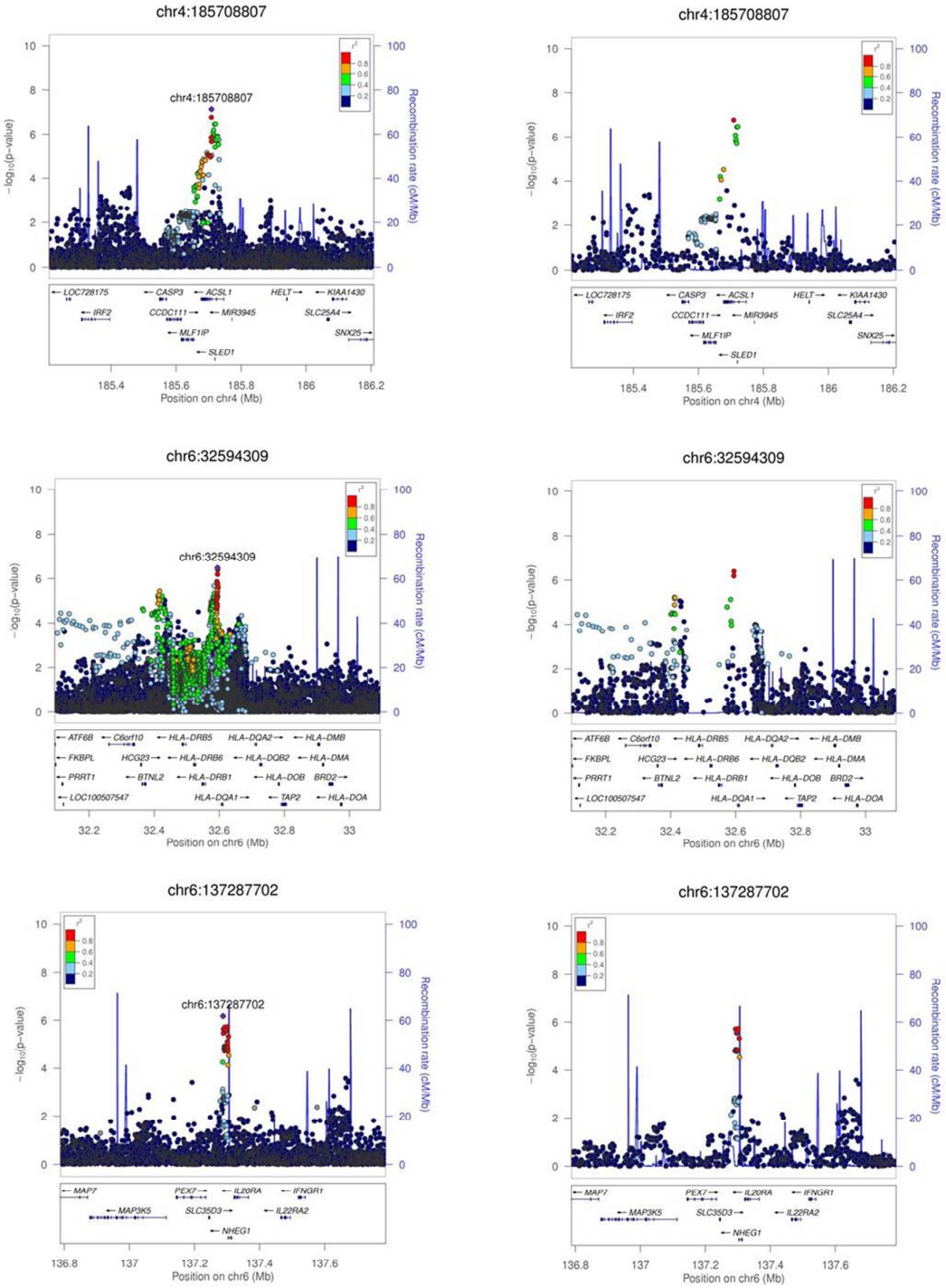
SUPPLEMENTARY FIGURES

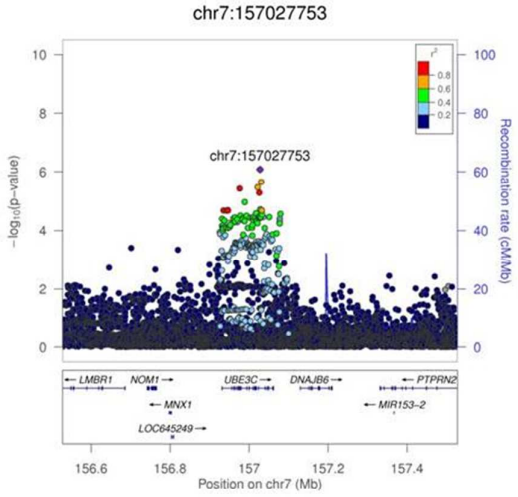


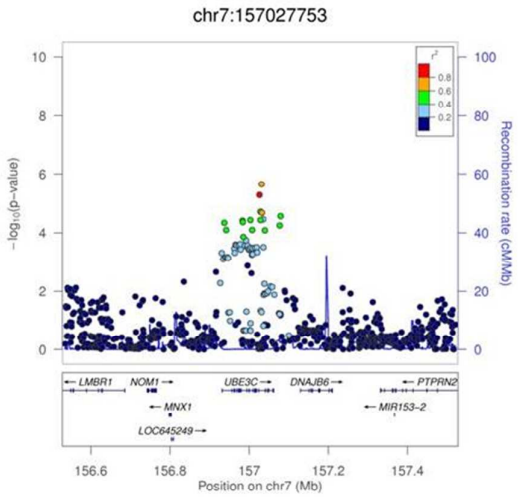
Supplementary Figure 1. QQ- and Manhattan plots of the discovery association meta-analysis results. A) QQ-plot of all the signals. B) QQ-plot of previously established signals. C) QQ-plot of novel signals. D) Manhattan plot. Signals of association reaching genome-wide significance for the first time in the present study ($p<5x10^{-8}$) are colored in red; blue dots represent previously established loci (**Supplementary Table 3**). The Y-axis was trimmed at $-\log_{10}$(p-value)=40 for easier visualisation; the *TCF7L2* association signal ($p=1.35\times10^{-81}$) falls far beyond this range (**Supplementary Table 3**).

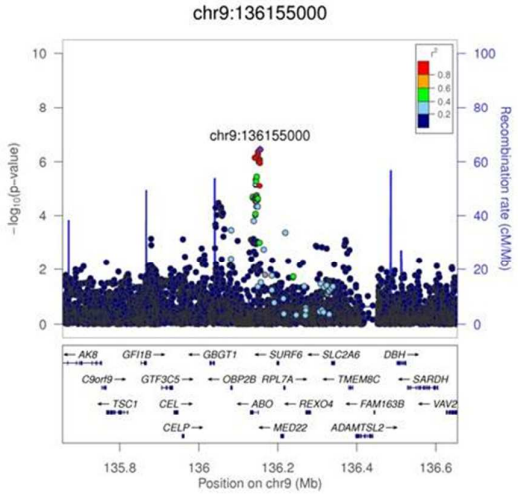
All 1000G SNVs

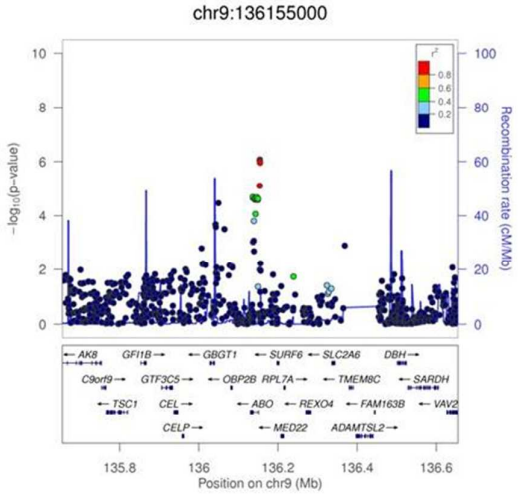
Restricted to HapMap SNVs

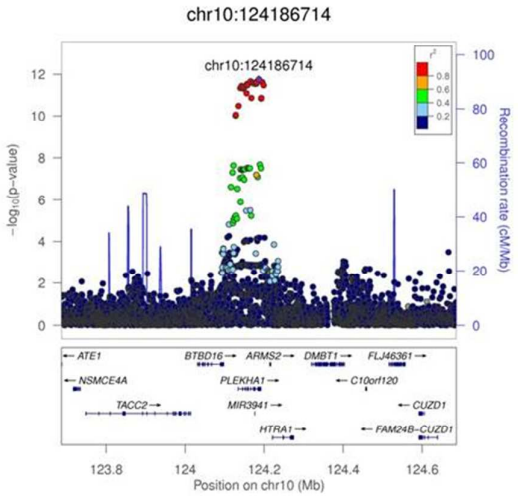


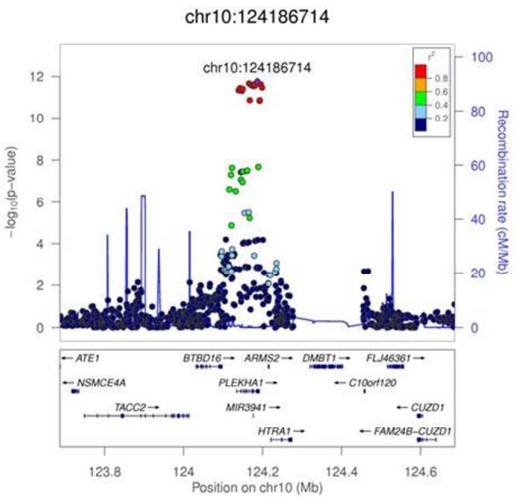

chr7:157027753
chr7:157027753
−log10(p-value)
Recombination rate (cM/Mb)
LMBR1
NOM1
UBE3C
DNAJB6
PTPRN2
MNX1
MIR153-2
LOC645249
156.6
156.8
157
157.2
157.4
Position on chr7 (Mb)

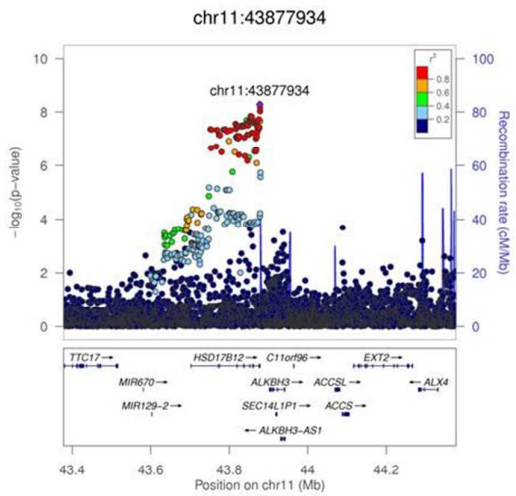

chr7:157027753
−log10(p-value)
Recombination rate (cM/Mb)
LMBR1
NOM1
UBE3C
DNAJB6
PTPRN2
MNX1
MIR153-2
LOC645249
156.6
156.8
157
157.2
157.4
Position on chr7 (Mb)

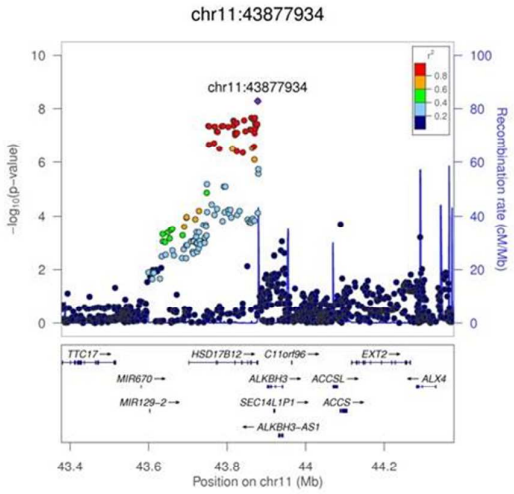

chr9:136155000
chr9:136155000
−log10(p-value)
Recombination rate (cM/Mb)
AK8
GFI1B
GBGT1
SURF6
SLC2A6
DBH
C9orf9
GTF3C5
OBP2B
RPL7A
TMEM8C
SARDH
TSC1
CEL
ABO
REXO4
FAM163B
VAV2
CELP
MED22
ADAMTSL2
135.8
136
136.2
136.4
136.6
Position on chr9 (Mb)

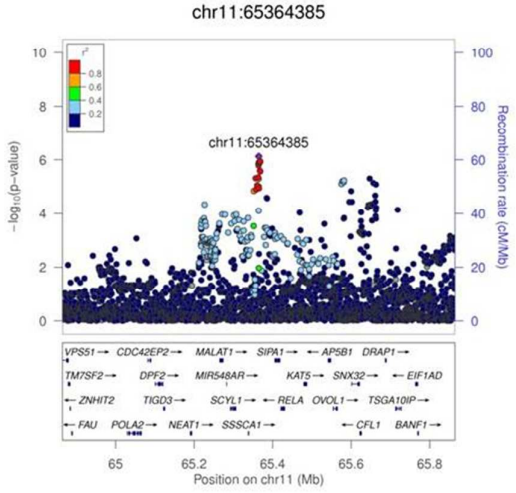

chr9:136155000
−log10(p-value)
Recombination rate (cM/Mb)
AK8
GFI1B
GBGT1
SURF6
SLC2A6
DBH
C9orf9
GTF3C5
OBP2B
RPL7A
TMEM8C
SARDH
TSC1
CEL
ABO
REXO4
FAM163B
VAV2
CELP
MED22
ADAMTSL2
135.8
136
136.2
136.4
136.6
Position on chr9 (Mb)

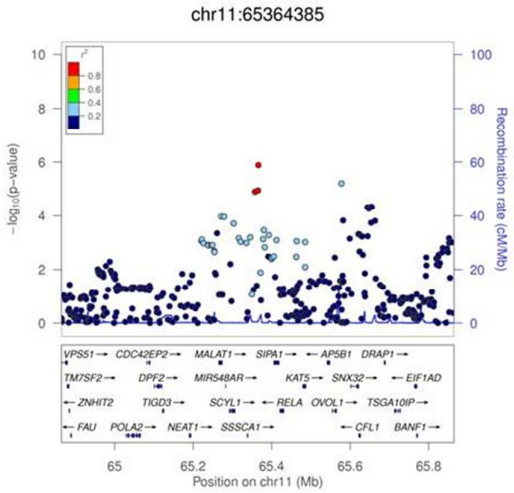

chr10:124186714
chr10:124186714
−log10(p-value)
Recombination rate (cM/Mb)
ATE1
BTBD16
ARMS2
DMBT1
FLJ46361
NSMCE4A
PLEKHA1
C10orf120
TACC2
MIR3941
CUZD1
HTRA1
FAM24B-CUZD1
123.8
124
124.2
124.4
124.6
Position on chr10 (Mb)

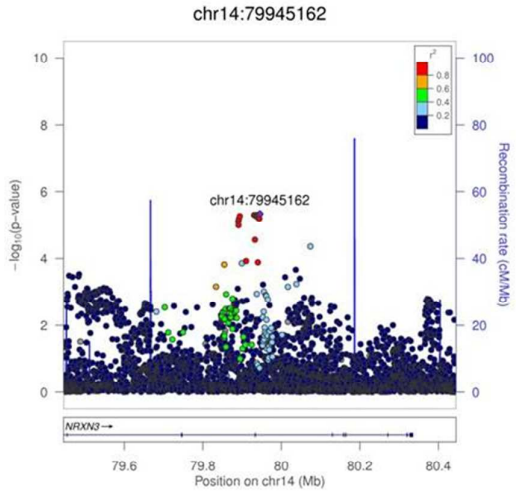

chr10:124186714
−log10(p-value)
Recombination rate (cM/Mb)
ATE1
BTBD16
ARMS2
DMBT1
FLJ46361
NSMCE4A
PLEKHA1
C10orf120
TACC2
MIR3941
CUZD1
HTRA1
FAM24B-CUZD1
123.8
124
124.2
124.4
124.6
Position on chr10 (Mb)

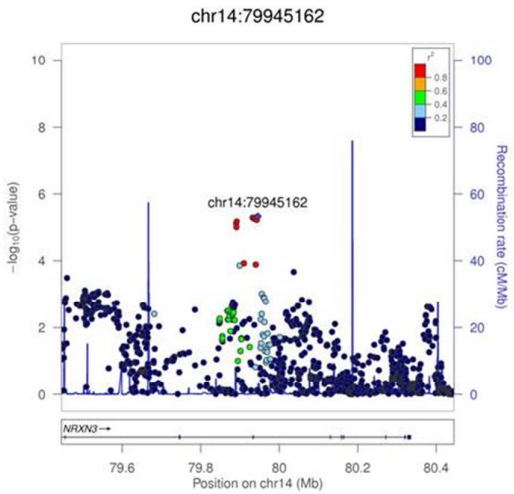

chr11:43877934
chr11:43877934
Position on chr11 (Mb)

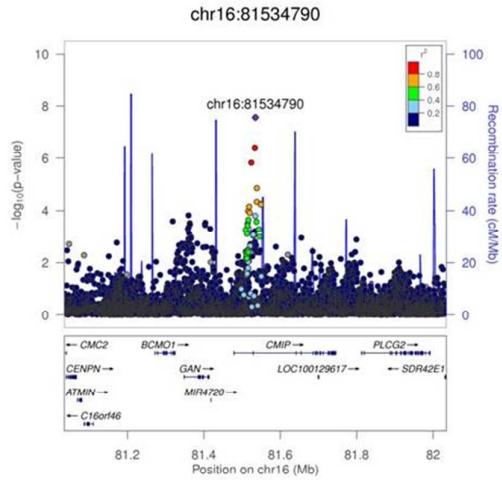

chr11:43877934
chr11:43877934
Position on chr11 (Mb)

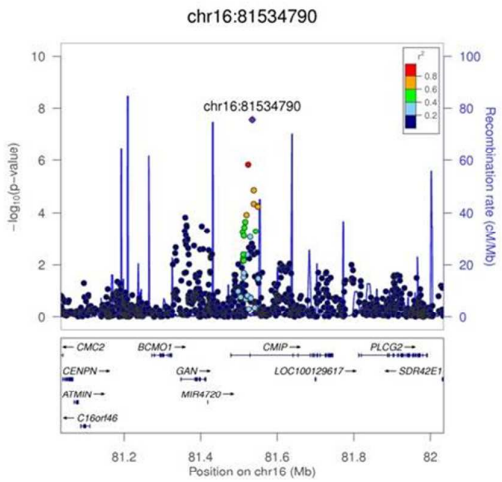

chr11:65364385
chr11:65364385
Position on chr11 (Mb)

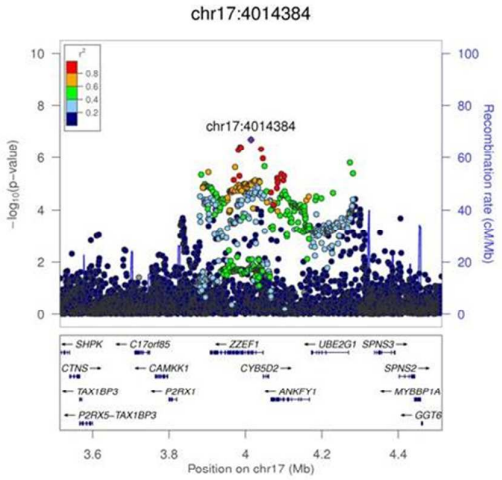

chr11:65364385
chr11:65364385
Position on chr11 (Mb)

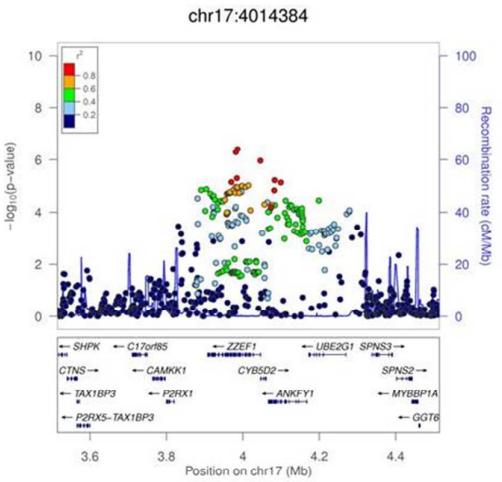

chr14:79945162
chr14:79945162
Position on chr14 (Mb)

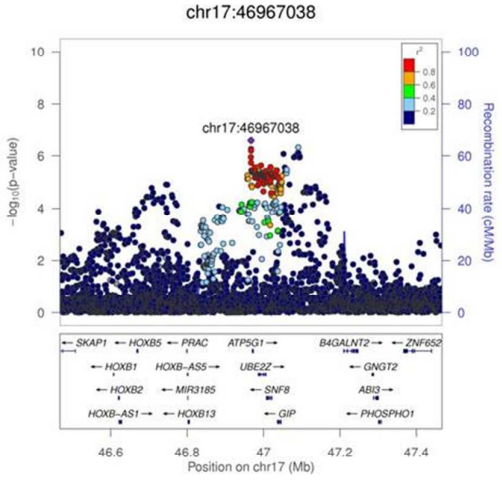

chr14:79945162
chr14:79945162
Position on chr14 (Mb)

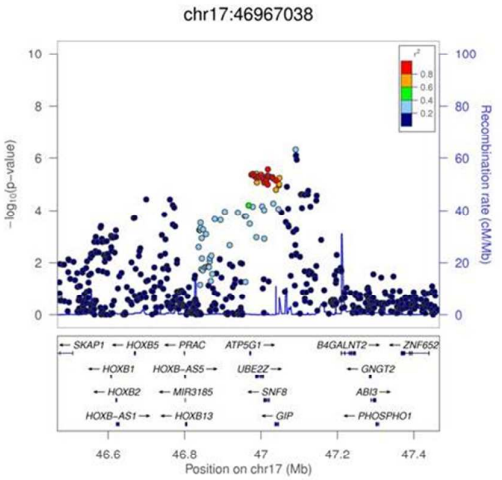

chr16:81534790
chr16:81534790
−log10(p−value)
Recombination rate (cM/Mb)
CMC2
BCMO1
CMIP
PLCG2
CENPN
GAN
LOC100129617
SDR42E1
ATMIN
MIR4720
C16orf46
Position on chr16 (Mb)

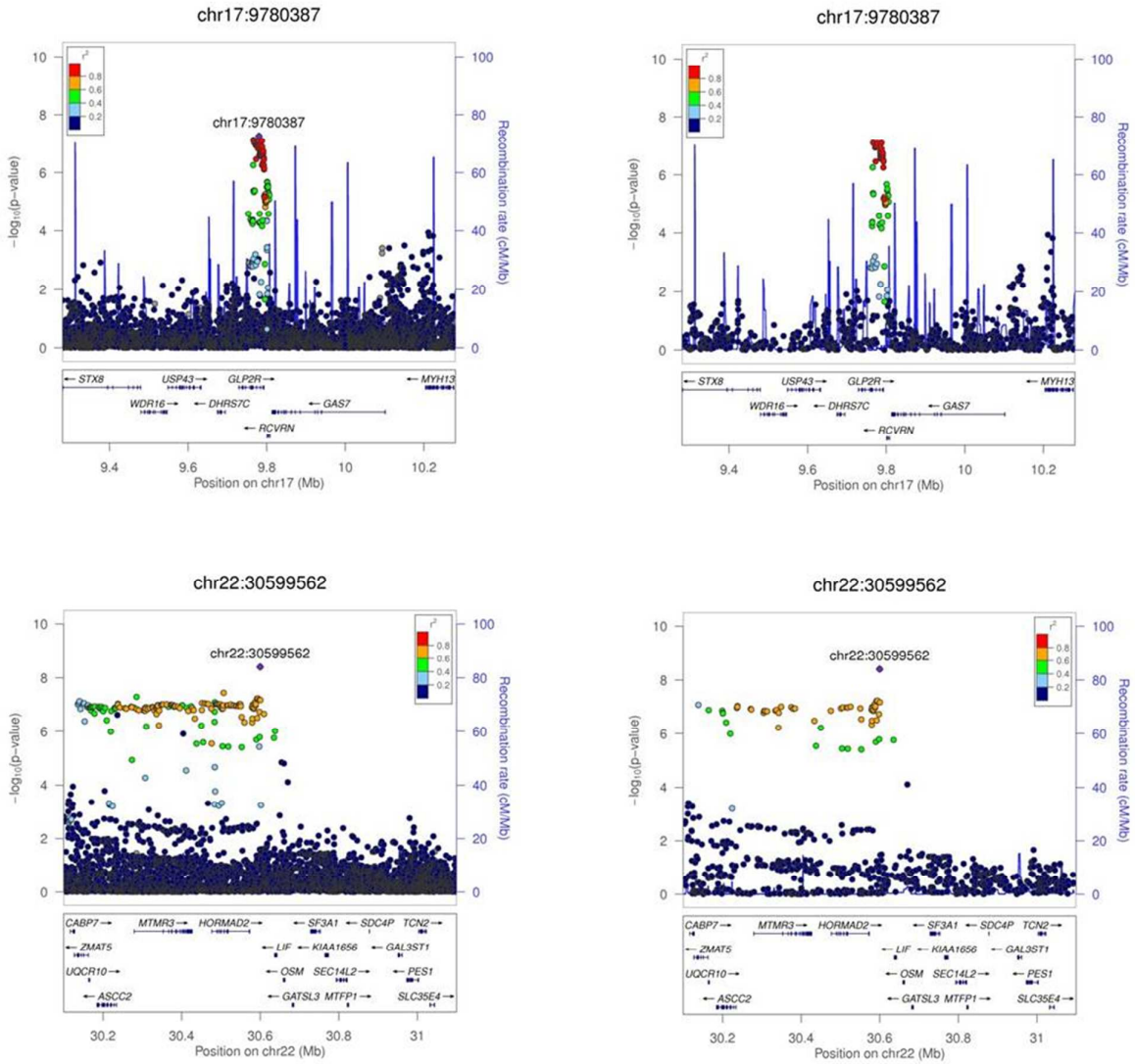

chr16:81534790
chr16:81534790
−log10(p−value)
Recombination rate (cM/Mb)
CMC2
BCMO1
CMIP
PLCG2
CENPN
GAN
LOC100129617
SDR42E1
ATMIN
MIR4720
C16orf46
Position on chr16 (Mb)


chr17:4014384
chr17:4014384
−log10(p−value)
Recombination rate (cM/Mb)
SHPK
C17orf85
ZZEF1
UBE2G1
SPNS3
CTNS
CAMKK1
CYB5D2
SPNS2
TAX1BP3
P2RX1
ANKFY1
MYBBP1A
P2RX5−TAX1BP3
GGT6
Position on chr17 (Mb)

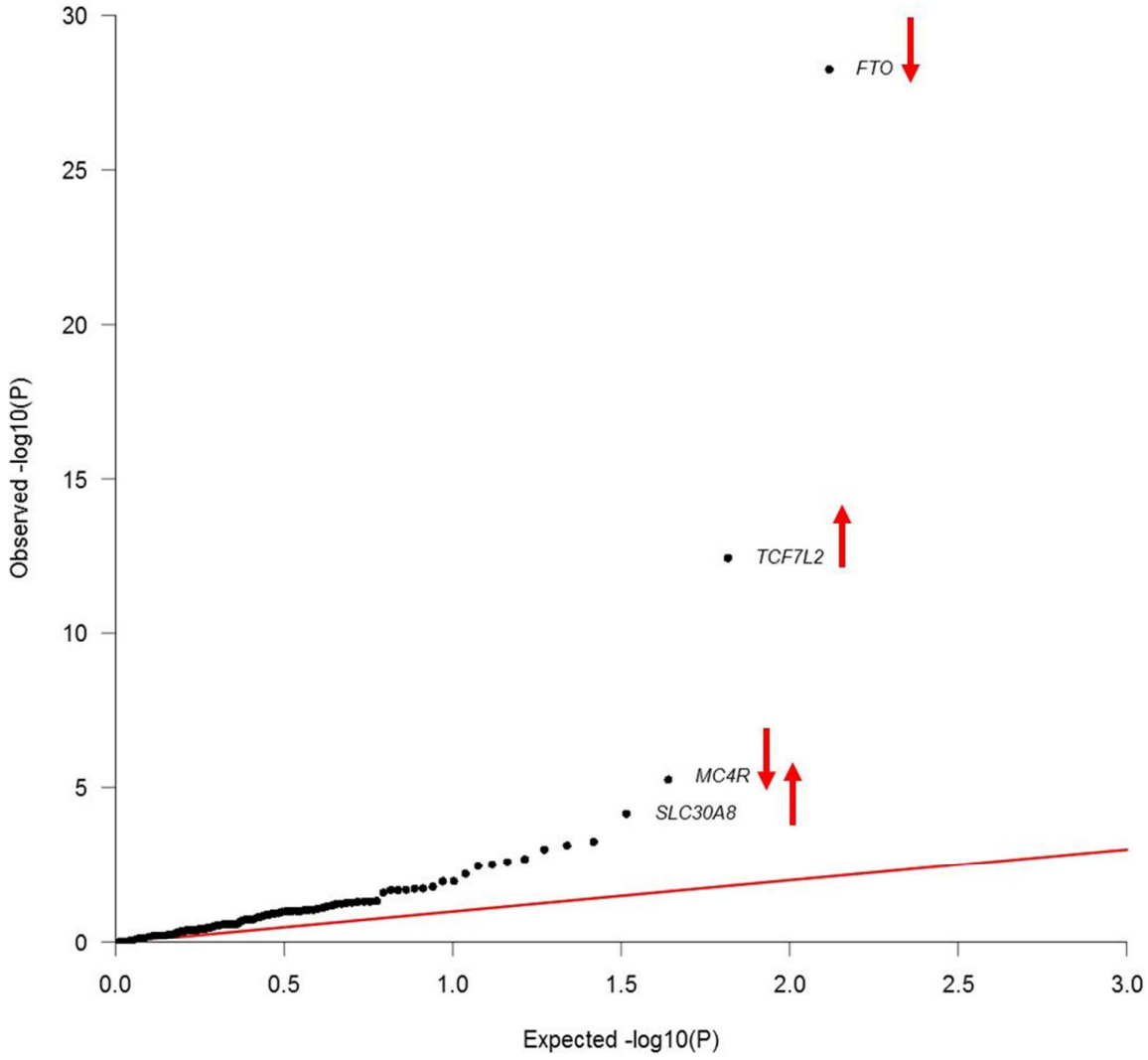

chr17:4014384
chr17:4014384
−log10(p−value)
Recombination rate (cM/Mb)
SHPK
C17orf85
ZZEF1
UBE2G1
SPNS3
CTNS
CAMKK1
CYB5D2
SPNS2
TAX1BP3
P2RX1
ANKFY1
MYBBP1A
P2RX5−TAX1BP3
GGT6
Position on chr17 (Mb)


chr17:46967038
chr17:46967038
−log10(p−value)
Recombination rate (cM/Mb)
SKAP1
HOXB5
PRAC
ATP5G1
B4GALNT2
ZNF652
HOXB1
HOXB−AS5
UBE2Z
GNGT2
HOXB2
MIR3185
SNF8
ABI3
HOXB−AS1
HOXB13
GIP
PHOSPHO1
Position on chr17 (Mb)


chr17:46967038
chr17:46967038
−log10(p−value)
Recombination rate (cM/Mb)
SKAP1
HOXB5
PRAC
ATP5G1
B4GALNT2
ZNF652
HOXB1
HOXB−AS5
UBE2Z
GNGT2
HOXB2
MIR3185
SNF8
ABI3
HOXB−AS1
HOXB13
GIP
PHOSPHO1
Position on chr17 (Mb)



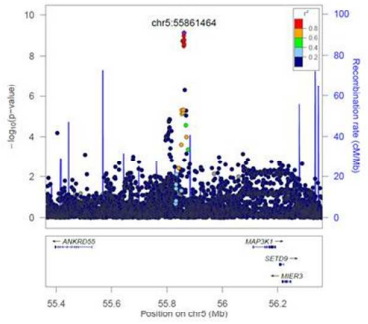
Supplementary Figure 2. Regional plots for the thirteen novel T2D loci. In the left panel, the plot is based using all 1000 Genomes March 2012 multi-ethnic SNV set, whereas in the right panel the plot is restricted to SNVs present in HapMap CEU reference set.



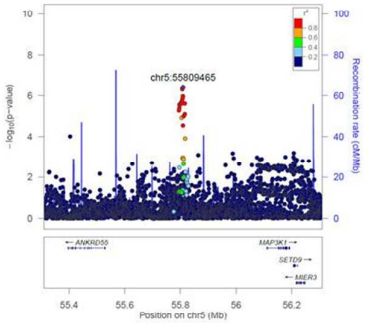
Supplementary Figure 3. QQ-plot of the expected vs. observed P-values for heterogeneity between BMI-adjusted and unadjusted association analysis models for established and novel T2D loci. The *FTO, TCF7L2, MC4R* and *SLC30A8* loci show large differences between models ($p_{heterogeneity}$=5.70x10^{-29}, 3.51x10^{-13}, 5.54x10^{-6} and 6.94x10^{-5}, respectively).

1) *ANKRD55*

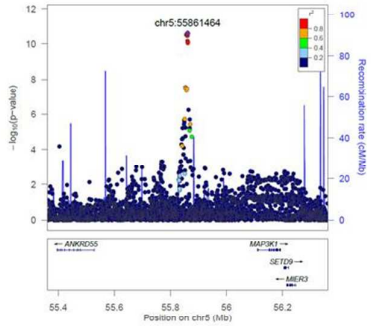
Lead SNV, unconditional



Distinct signal, conditional on lead SNV

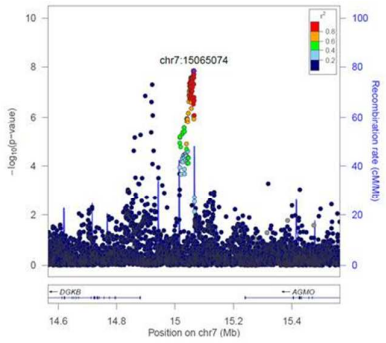


Lead SNV, conditional on distinct signal

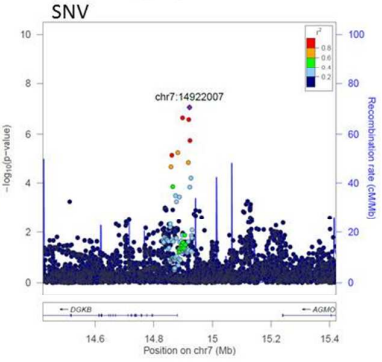


2) *DGKB*

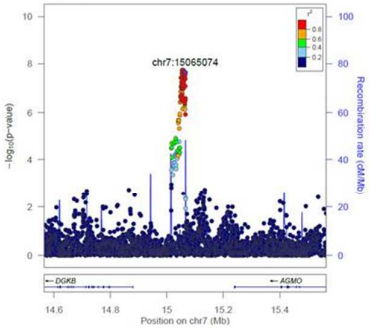
Lead SNV, unconditional



Distinct signal, conditional on lead SNV

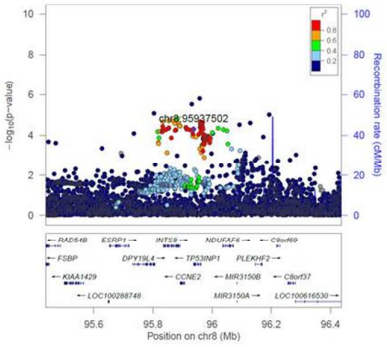


Lead SNV, conditional on distinct signal

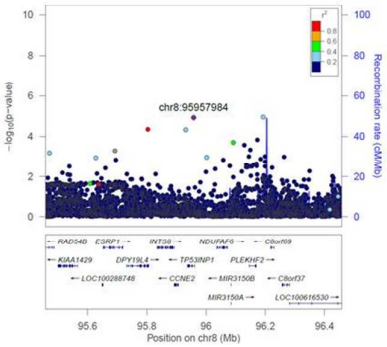


3) *TP53INP1*

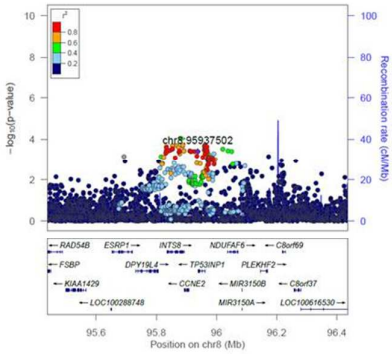
Lead SNV, unconditional



Distinct signal, conditional on lead SNV

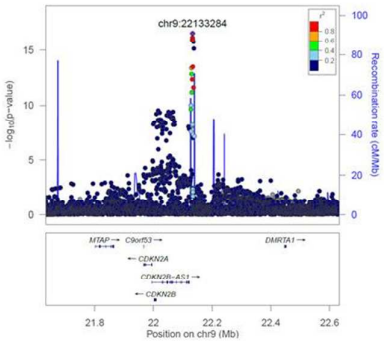


Lead SNV, conditional on distinct signal

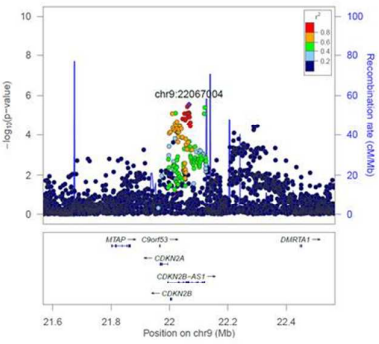


4) *CDKN2A/B*

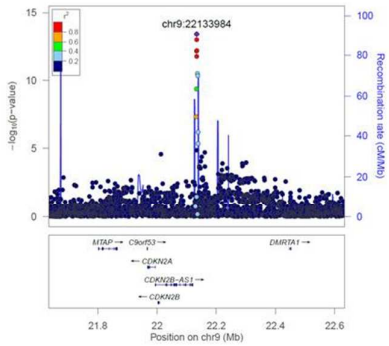
Lead SNV, unconditional



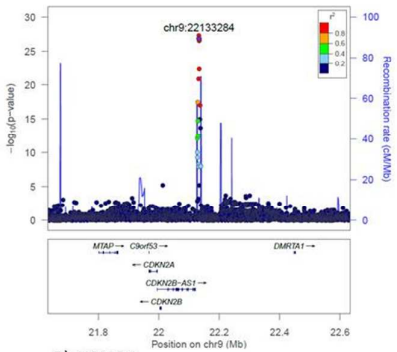
First distinct signal, conditional on lead SNV and other distinct signals



Second distinct signal, conditional on lead SNV and other distinct signals

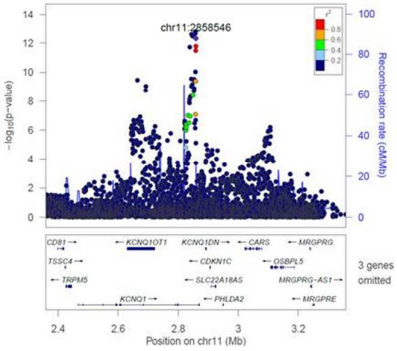


Lead SNV, conditional on distinct signals

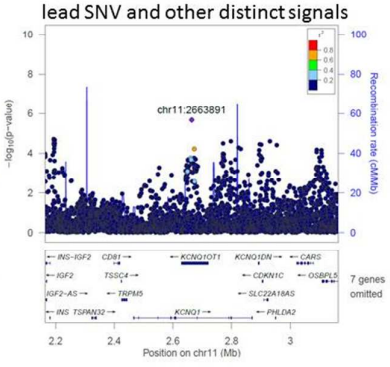


5) *KCNQ1*

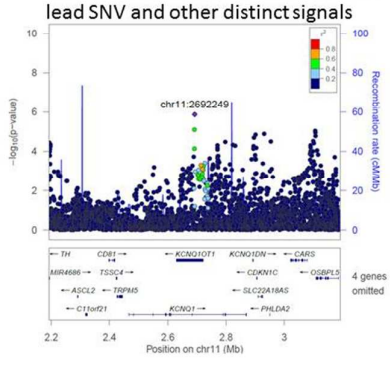
Lead SNV, unconditional



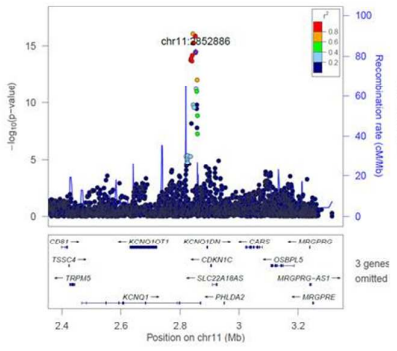
First distinct signal, conditional on lead SNV and other distinct signals



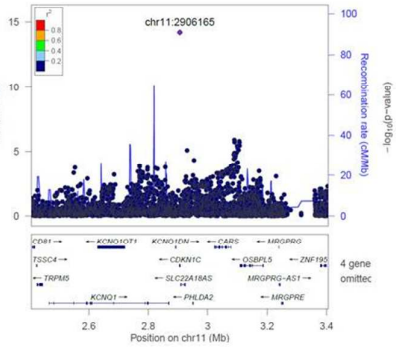
Second distinct signal, conditional on lead SNV and other distinct signals



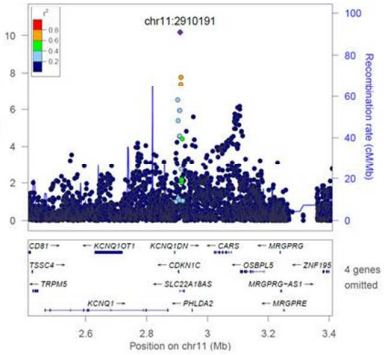
Third distinct signal, conditional on lead SNV and other distinct signals



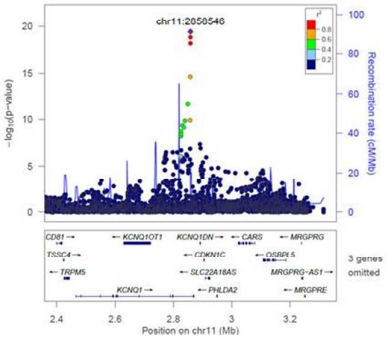
Fourth distinct signal, conditional on lead SNV and other distinct signals

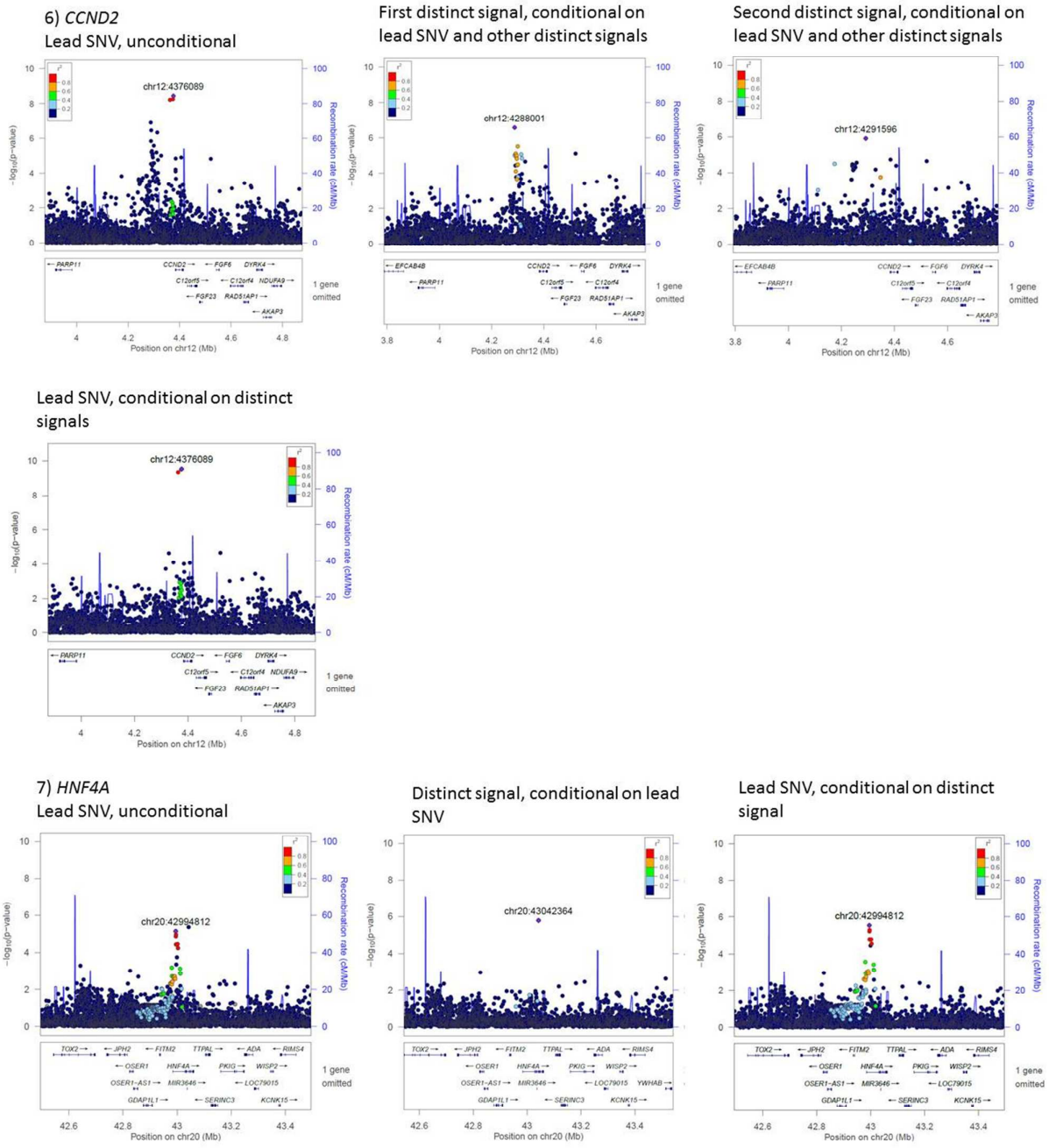


Fifth distinct signal, conditional on lead SNV and other distinct signals

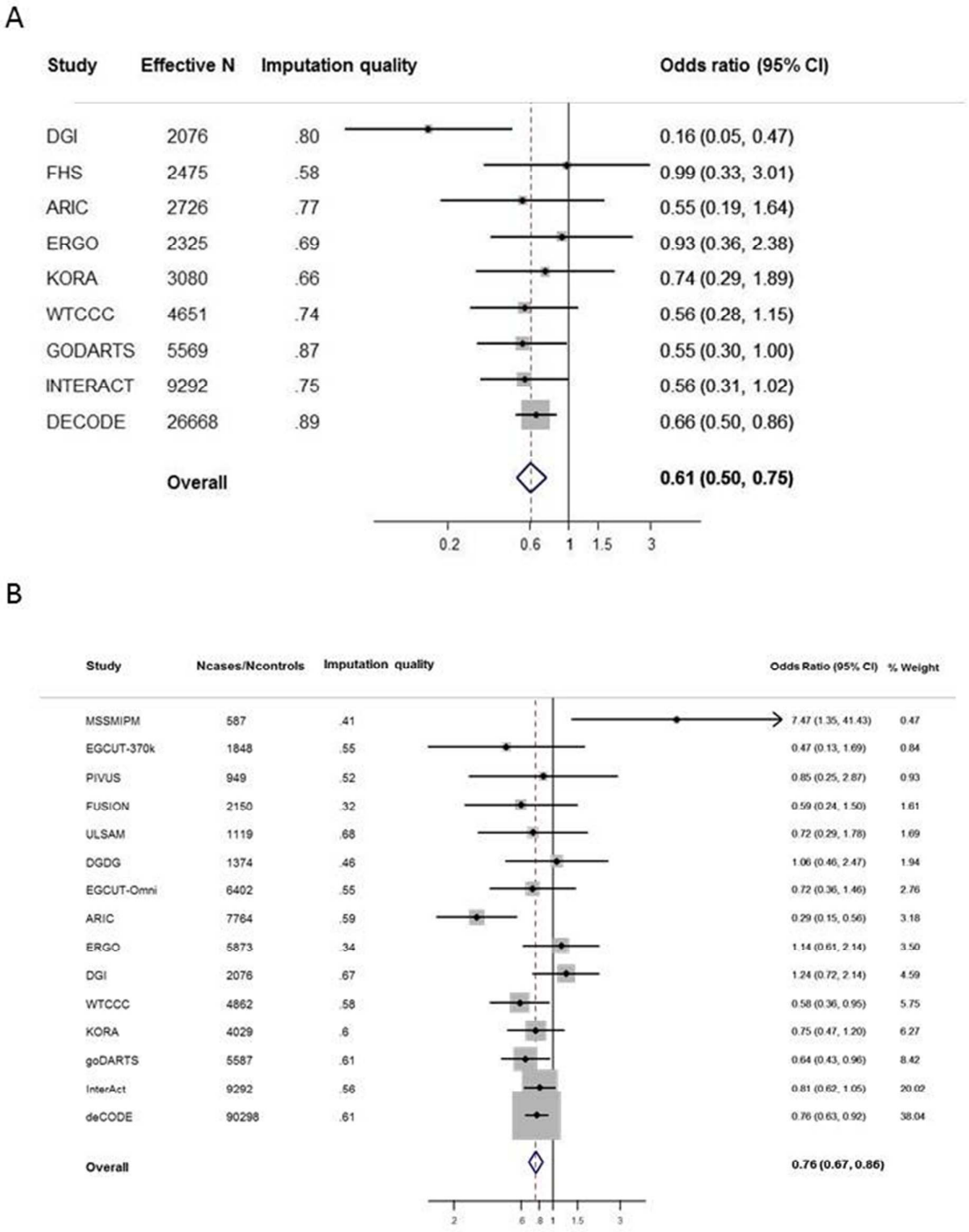


Lead SNV, conditional on distinct signals

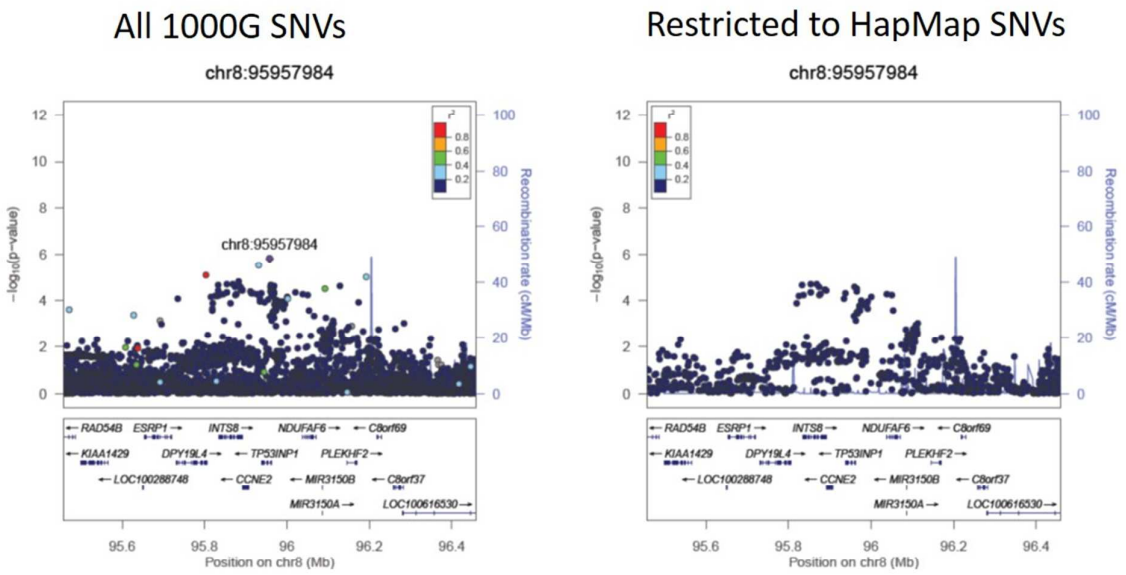




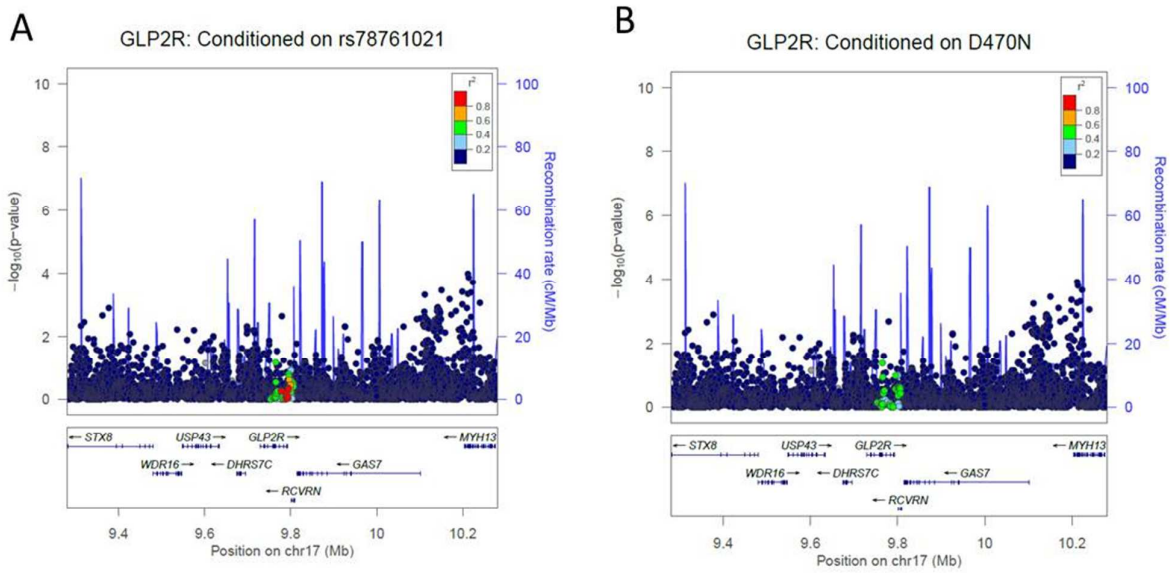
Supplementary Figure 4. Regional plots for T2D loci showing additional distinct signals ($p<10^{-5}$) in the approximate conditional analysis. First, unconditional analysis results are shown, followed by results conditioned on the lead SNV and other distinct signals. In the last plot for each locus the results for lead SNV conditional on the distinct signal(s) are shown.



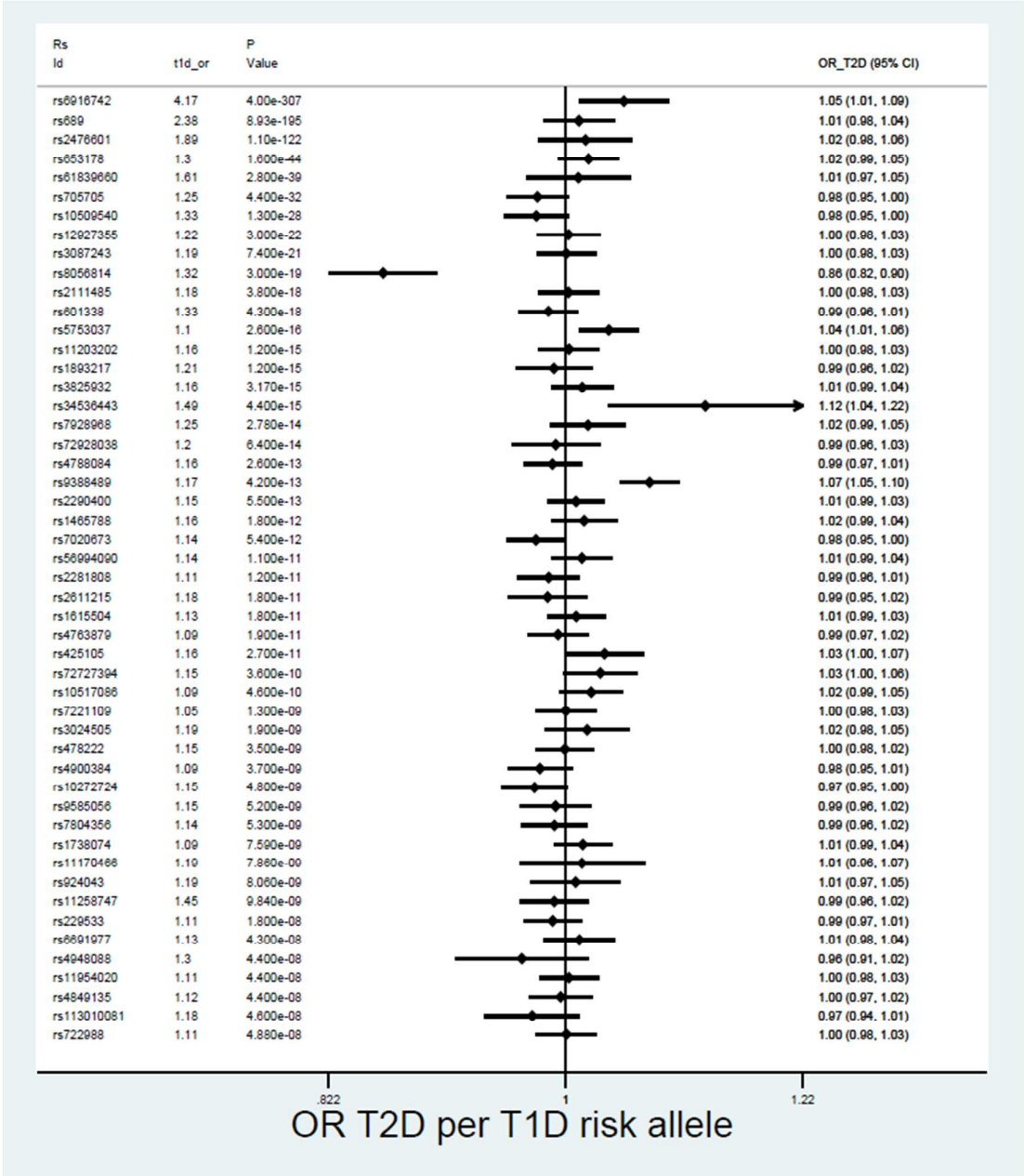
Supplementary Figure 5. Forest plots of the A) putative low frequency distinct signal (rs188827514) and B) previously established (Steinthorsdottir et al.) **low-frequency variant (rs76895963) at *CCND2* for their associations with T2D.** Odds ratios (OR) with their 95% confidence intervals (CI) are shown from unconditioned models.



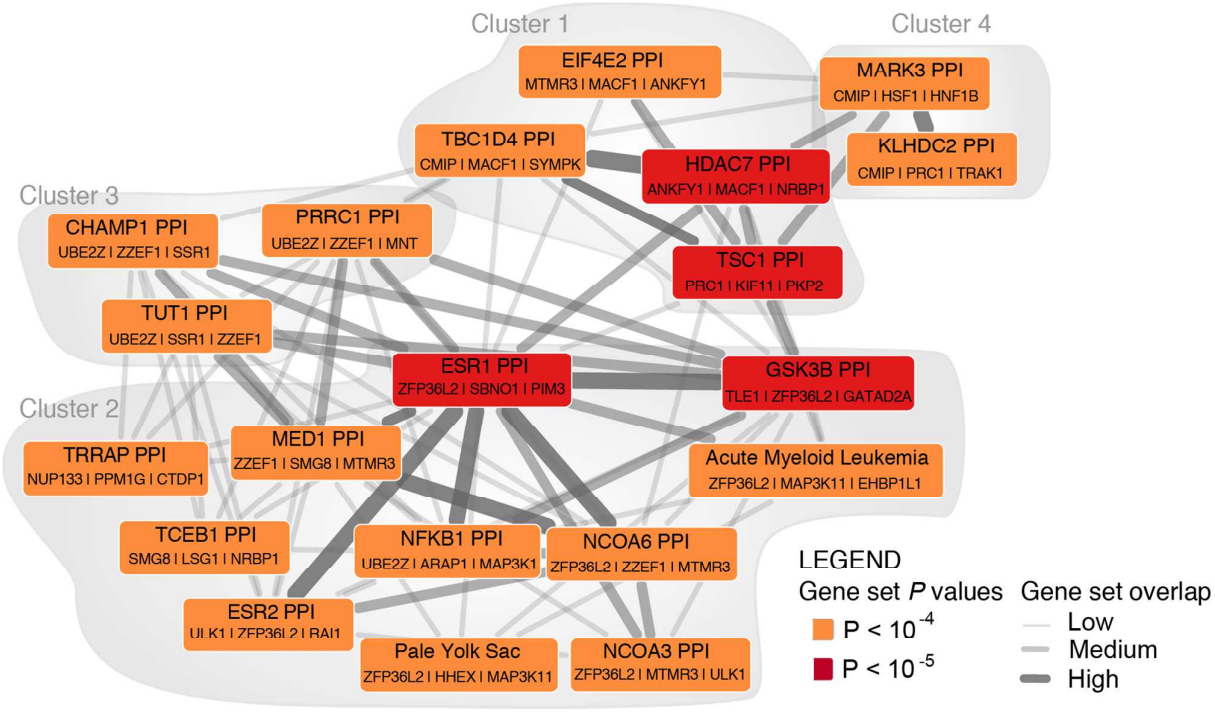
Supplementary Figure6. Regional architecture of *TP53INP1* locus. In the right panel the figure is plotted using all 1000 Genomes SNVs and highlights the new lead SNV (rs11786613) independent from the previous lead variant, signal visible in the left panel the plot is restricted to SNVs present in HapMap.



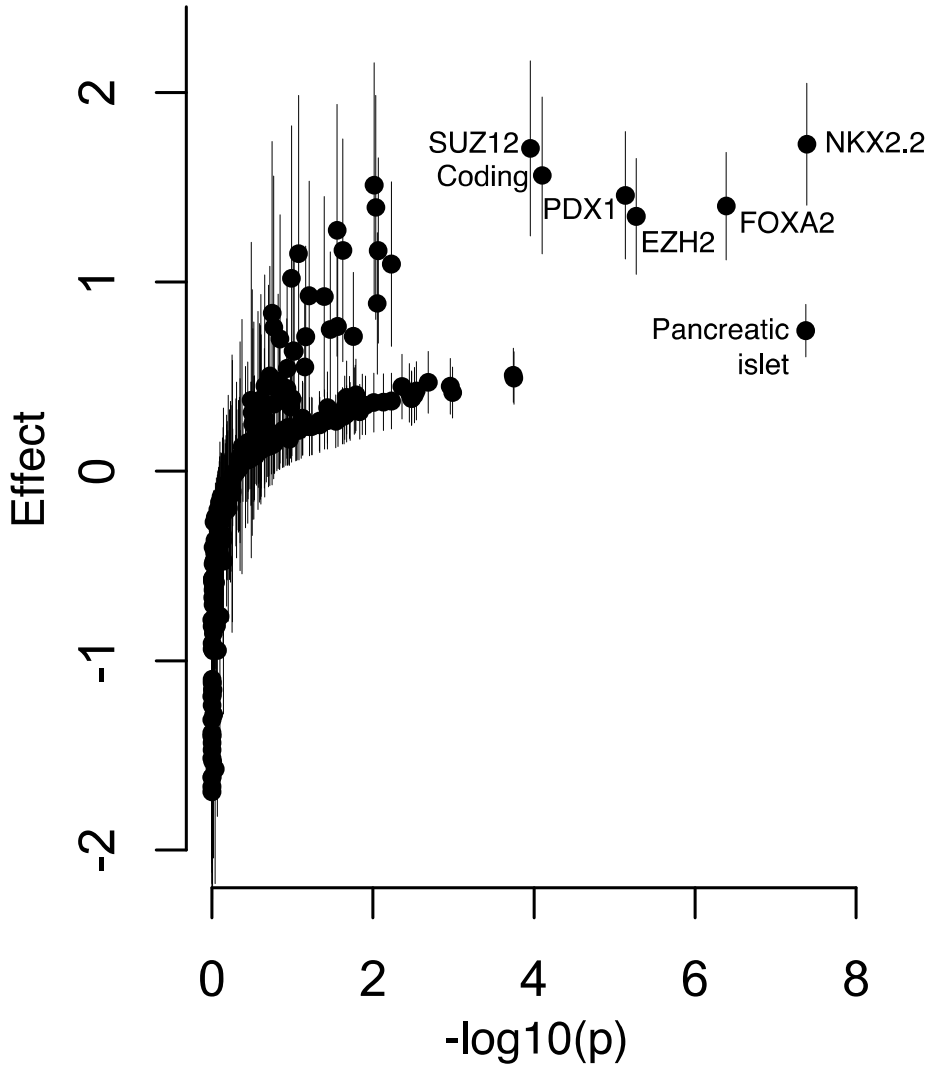
Supplementary Figure 7. Association of variation in *GLP2R* with T2D after approximate conditional analyses on either A) the lead SNV (rs78761021), or B) D470N.



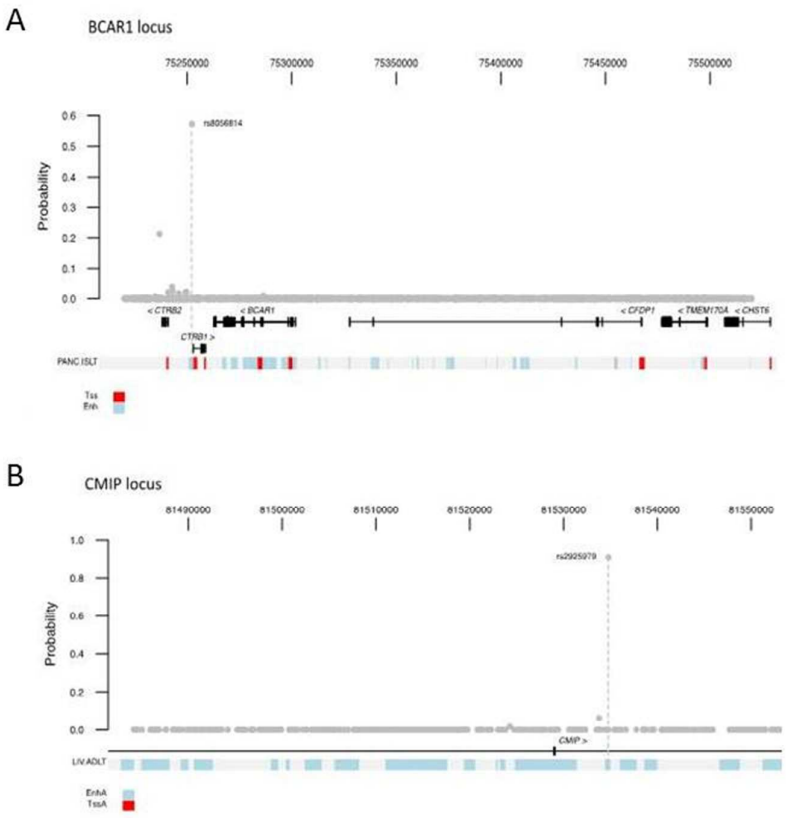
Supplementary Figure 8. Effects on T2D of 50 established T1D variants. All effects are aligned to T1D risk-raising allele. Loci are sorted from top to bottom by the magnitude of association with T1D.



Supplementary Figure 9. Significantly enriched reconstituted gene sets by DEPICT. We report 20 significantly enriched reconstituted gene sets (FDR<0.05, **Supplementary Table 11**). Reconstituted gene sets are represented by nodes and their overlap by edges. Reconstituted gene sets are colour-coded based on their degree of enrichment in genes at the associated T2D loci (darker means more significant). DEPICT identified 21 significantly enriched reconstituted gene sets; one gene set was omitted due to a potential mismatch between the reconstituted gene set identifier and the reconstituted gene set (see Methods). For each gene set, the three genes exhibiting the highest likelihood within the given gene set and being within associated T2D loci are shown. Pairwise overlap between reconstituted gene sets were estimated by computing the Pearson correlation coefficient r between two reconstituted gene sets followed by discretization into one of three bins; $0.3 \leq r < 0.5$ denotes low overlap, $0.5 \leq r < 0.7$ denotes medium overlap, and $r \geq 0.7$ denotes high overlap. Edges representing overlap corresponding to $r < 0.3$ are not shown.



Supplementary Figure 10. Type 2 diabetes credible sets are enriched for genomic annotations. We calculated the posterior probability of causality for all variants at 95 established T2D loci. We then tested the effect of variants annotated with protein-coding genes, cell type chromatin state, and transcription factor binding on the posterior probabilities across all loci. We identified significant effects among coding exons and pancreatic islet chromatin, and for binding sites of the FOXA2, NKX2.2, PDX1, and EZH2 transcription factors.



Supplementary Figure 11. **Genomic annotation at credible sets of novel loci**. A) The T2D signal at the *BCAR1* locus contains a variant rs8056814 with a 57% probability of being causal for the signal. This variant overlaps an enhancer active in pancreatic islets proximal to the *CTRB1* gene. B) The novel T2D signal at the *CMIP* locus is also associated with BMI and lipid phenotypes. The variant rs2925979 has a 91% probability of being causal for the *CMIP* signal and overlaps an enhancer active in liver, which is the most enriched cell type in the BMI/lipid physiology group.

BIOLOGY BOX

ACSL1: chr4:185708807 (rs60780116) is an intronic variant in acyl-CoA synthetase long chain family member 1 coding gene (***ACSL1***), an isozyme that converts free long-chain fatty acids into fatty acyl-CoA esters, playing a key role in lipid biosynthesis and fatty acid degradation. *ACSL1* is highly expressed in adipose, liver, skeletal muscle tissue and in whole blood, but expressed at lower levels in pancreas(1). Recent reports have implicated *ACSL1* in regulating systemic glucose homeostasis(2), potentially via an effect on metabolic flexibility and capacity to switch between fatty acid and glucose metabolism. Variants in *ACSL1* have previously been associated with Kawasaki disease(3) (r^2=0.12).

HLA-DQA1: Variation in the HLA region has been strongly associated with T1D(4) (r^2=0.08) and other autoimmune diseases, including multiple sclerosis(5) (r^2=0.47) and inflammatory bowel disease(6) (r^2=0.13). Associations with total cholesterol and LDL cholesterol have also been reported(7) (r^2=0.06). The lead SNV for T2D association in the HLA region (chr6:32594309; rs9271774) lies ~2kb upstream of ***HLA-DQA1***. It is in high LD (r^2=0.82) with a SNV strongly associated with expression of *HLA-DRB5* in pancreatic islets(8). Analyses (see main text) suggest that the T2D association is not the result of misclassification of individuals with T1D as T2D cases in the present study.

SLC35D3: Index variant chr6:137287702 (rs6918311) is located ~20kb downstream of the RNA gene ***NHEG1*** (neuroblastoma highly expressed 1), which has no well characterized function. Also proximal to the lead SNV are: (1) ***SLC35D3,*** which is a member of the solute carrier family 35 and a regulator of the biosynthesis of platelet-dense granules with possible role in carbohydrate transport; (2) ***PEX7,*** (peroxisomal biogenesis factor 7) encoding for the cytosolic receptor for the set of peroxisomal matrix enzymes, which is involved in cell metabolism and is associated with peroxisome biogenesis disorders and implicated in autism; and (3) ***IL20RA***, which encodes for a subunit of the receptor for interleukin 20, and is a cytokine suggested to be involved in epidermal function.

MNX1*:* chr7:157027753 (rs1182436) is an intronic variant in ***UBE3C***, which encodes for a ubiquitin protein ligase. The lead SNV in the locus lies ~100kb upstream of ***MNX1***, which is highly expressed in pancreas(1) containing coding mutations recently implicated in neonatal diabetes(9).

ABO*:* chr9:136155000 (rs635634) variant lies ~5kb upstream of ***ABO*** gene, which determines blood group by modifying the oligosaccharides on cell surface glycoproteins. Variation in or near *ABO* has been associated with a very wide range of phenotypes, including glycaemic(10), lipid traits (7) (r^2=1), coronary artery disease(11) and stroke(12) (r^2=0.83). The lead variant at this locus is in low LD (r^2<0.05) with blood group-defining markers(13).

PLEKHA1*:* chr10:124186714 (rs2292626) is an intronic variant in ***PLEKHA1*** (pleckstrin homology domain containing, family A member 1). The encoded protein localises to the plasma membrane where it specifically binds phosphatidylinositol 3,4-bisphosphate. This protein may be involved in the formation of signalling complexes in the plasma membrane. Variants in modest LD (rs10490924; r^2=0.27) have been associated with age-related macular degeneration(14).

HSD17B12: chr11:43877934 is a 3'UTR variant of ***HSD17B12*** encoding the enzyme 17-beta hydroxysteroid dehydrogenase-12. *HSD17B12* encodes 17beta-hydroxysteroid dehydrogenase, involved in fatty acid metabolism(15) and estrogen sex steroid hormone formation. *HSD17B12* has been identified as central to adipocyte differentiation(16), and a correlated variant (rs2176598; r^2=0.68) was recently associated with BMI(17). However, rs1061810 remained associated with T2D after adjustment for BMI, and we found only a nominal difference in the association of rs1061810 with T2D in meta-analyses with or without adjustment for BMI (**Supplementary Table 4**), potentially indicating a role for *HSD17B12* in risk of

diabetes independently of associations with adiposity. Other associations from this locus have been reported with forced vital capacity(18) (r^2=0.59) and neuroblastoma(19) (r^2=0.24).

MAP3K11: chr11:65364385 (rs111669836) is located next to ***KCNK7*** (potassium channel, subfamily K, member 7) gene, a member of the superfamily of potassium channel proteins. ***MAP3K11*** encodes the Mitogen-activated protein kinase 11, part of the serine/threonine kinase family. MAP3K11 has been implicated in regulation of pancreatic beta-cell death(20). Variation at this locus has previously been associated with e.g. height(21) (r^2=0.02) and lipid levels(7) (r^2=0.08).

NRXN3*:* chr14:79945162 (rs10146997) is an established variant associated with waist circumference(22), BMI(23) and obesity(24). It is an intronic variant in the ***NRXN3*** (Homo sapiens neurexin 3) gene, which is part of a family of central nervous adhesion molecules It is expressed in the same sub-cortical regions where reward training neuronal pathways are expressed.

CMIP*:* chr16:81534790 (rs2925979). This gene encodes a c-Maf inducing protein that plays a role in the T-cell signalling pathway. C-mip down-regulates NF-κB activity and promotes apoptosis in podocytes(25) in cases of idiopathic nephrotic syndrome (INS). Associations with WHR(26), adiponectin(27) and HDL cholesterol(7) levels have been reported for this same variant.

ZZEF1: chr17:4014384 (rs7224685) is an intronic variant in the ***ZZEF1*** (zinc finger, ZZ-type with EF-hand domain 1) gene related to calcium ion binding. This locus was previously implicated in functional impairment in major depressive disorder, bipolar disorder and schizophrenia(28).

GLP2R: chr17:9780387 (rs78761021) is an intronic variant in the glucagon-like peptide 2 receptor (***GLP2R***) gene belonging to a G protein-coupled receptor superfamily. It is closely related to the glucagon receptor (GCGR) and GLP1R. Glucagon-like peptide-2 (GLP2) is a 33-amino acid proglucagon-derived peptide produced by intestinal enteroendocrine cells.

GIP: the nearest gene to the detected signal (chr17:46967038, rs12941263) in this region is ***ATP5G1***, coding for a subunit of mitochondrial ATP synthase and involved in "energy production", in lipid transports and in cellular metabolism. Another gene within locus, ***GIP*** encodes an incretin hormone that belongs to the glucagon superfamily and is gastric inhibitory polypeptide. GIP is a potent stimulator of insulin secretion from pancreatic beta-cells following food ingestion and nutrient absorption via its G protein-coupled receptor activation of adenylyl cyclase and other signal transduction pathways(29). Variants (rs46522, rs318095) in high LD (r^2=0.97) with our identified SNV at *GIP* have been associated with susceptibility to coronary heart disease(11) and height(30). Variation in the receptor for *GIP* (*GIPR*) have previously been associated with glycemic traits and T2D(31,32).

BIOLOGY BOX REFERENCES:

1. Mele M, Ferreira PG, Reverter F, DeLuca DS, Monlong J, Sammeth M, et al. The human transcriptome across tissues and individuals. Science. 2015;348(6235):660–5.
2. Li LO, Grevengoed TJ, Paul DS, Ilkayeva O, Koves TR, Pascual F, et al. Compartmentalized Acyl-CoA Metabolism in Skeletal Muscle Regulates Systemic Glucose Homeostasis. Diabetes. 2015;64(1):23–35.
3. Onouchi Y, Ozaki K, Burns JC, Shimizu C, Terai M, Hamada H, et al. A genome-wide association study identifies three new risk loci for Kawasaki disease. Nat Genet. 2012;44(5):517–21.
4. Barrett JC, Clayton DG, Concannon P, Akolkar B, Cooper JD, Erlich H a, et al. Genome-wide association study and meta-analysis find that over 40 loci affect risk of type 1 diabetes. Nat Genet. 2009;41(6):703–7.
5. Patsopoulos NA, Esposito F, Reischl J, Lehr S, Bauer D, Heubach J, et al. Genome-wide meta-analysis identifies novel multiple sclerosis susceptibility loci. Ann Neurol. 2011;70(6):897–912.
6. Kugathasan S, Baldassano RN, Bradfield JP, Sleiman PMA, Imielinski M, Guthery SL, et al. Loci on 20q13 and 21q22 are associated with pediatric-onset inflammatory bowel disease. Nat Genet. 2008;40(10):1211–5.
7. Willer CJ, Schmidt EM, Sengupta S, Peloso GM, Gustafsson S, Kanoni S, et al. Discovery and

refinement of loci associated with lipid levels. Nat Genet. 2013;45(11):1274–83.
8. Fadista J, Vikman P, Laakso EO, Mollet IG, Esguerra J Lou, Taneera J, et al. Global genomic and transcriptomic analysis of human pancreatic islets reveals novel genes influencing glucose metabolism. Proc Natl Acad Sci U S A. 2014;111(38):13924–9.
9. Flanagan SE, De Franco E, Lango Allen H, Zerah M, Abdul-Rasoul MM, Edge JA, et al. Analysis of transcription factors key for mouse pancreatic development establishes NKX2-2 and MNX1 mutations as causes of neonatal diabetes in man. Cell Metab. 2014;19(1):146–54.
10. Wessel J, Chu AY, Willems SM, Wang S, Yaghootkar H, Brody JA, et al. Low-frequency and rare exome chip variants associate with fasting glucose and type 2 diabetes susceptibility. Nat Commun. 2015;6:5897.
11. Schunkert H, Konig IR, Kathiresan S, Reilly MP, Assimes TL, Holm H, et al. Large-scale association analysis identifies 13 new susceptibility loci for coronary artery disease. Nat Genet. 2011;43(4):333–8.
12. Dichgans M, Malik R, König IR, Rosand J, Clarke R, Gretarsdottir S, et al. Shared genetic susceptibility to ischemic stroke and coronary artery disease : A genome-wide analysis of common variants. Stroke. 2014;45(1):24–36.
13. Olsson ML, Chester MA. Polymorphism and recombination events at the ABO locus: A major challenge for genomic ABO blood grouping strategies. Transfus Med. 2001;11(4):295–313.
14. Fritsche LG, Chen W, Schu M, Yaspan BL, Yu Y, Thorleifsson G, et al. Seven new loci associated with age-related macular degeneration. Nat Genet. 2013;45(4):433–9, 439-2.
15. Saloniemi T, Jokela H, Strauss L, Pakarinen P, Poutanen M. The diversity of sex steroid action: Novel functions of hydroxysteroid (17β) dehydrogenases as revealed by genetically modified mouse models. J Endocrinol. 2012;212(1):27–40.
16. Söhle J, Machuy N, Smailbegovic E, Holtzmann U, Grönniger E, Wenck H, et al. Identification of new genes involved in human adipogenesis and fat storage. PLoS One. 2012;7(2):e31193.
17. Locke A, B K, SI B, AE J, TH P, FR D, et al. Genetic studies of body mass index yield new insights for obesity biology. Nature. 2015;518:197–206.
18. Loth DW, Artigas MS, Gharib S a, Wain L V, Franceschini N, Koch B, et al. Genome-wide association analysis identifies six new loci associated with forced vital capacity. Nat Genet. 2014;46(7):669–77.
19. Diskin SJ, Capasso M, Schnepp RW, Cole KA, Attiyeh EF, Hou C, et al. Common variation at 6q16 within HACE1 and LIN28B influences susceptibility to neuroblastoma. Nat Genet. 2012;44(10):1126–30.
20. Humphrey RK, Yu SMA, Bellary A, Gonuguntla S, Yebra M, Jhala US. Lysine 63-linked ubiquitination modulates mixed lineage kinase-3 interaction with JIP1 scaffold protein in cytokine-induced pancreatic b cell death. J Biol Chem. 2013;288(4):2428–40.
21. Lango Allen H, Estrada K, Lettre G, Berndt SI, Weedon MN, Rivadeneira F, et al. Hundreds of variants clustered in genomic loci and biological pathways affect human height. Nature. 2010;467(7317):832–8.
22. Heard-Costa NL, Zillikens MC, Monda KL, Johansson A, Harris TB, Fu M, et al. NRXN3 is a novel locus for waist circumference: a genome-wide association study from the CHARGE Consortium. PLoS Genet. 2009;5(6):e1000539.
23. Speliotes EK, Willer CJ, Berndt SI, Monda KL, Thorleifsson G, Jackson AU, et al. Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index. Nat Genet. 2010;42(11):937–48.
24. Berndt SI, Gustafsson S, Mägi R, Ganna A, Wheeler E, Feitosa MF, et al. Genome-wide meta-analysis identifies 11 new loci for anthropometric traits and provides insights into genetic architecture. Nat Genet. 2013;45(5):501–12.
25. Ory V, Fan Q, Hamdaoui N, Zhang SY, Desvaux D, Audard V, et al. C-mip down-regulates NF-??B activity and promotes apoptosis in podocytes. Am J Pathol. 2012;180(6):2284–92.
26. Shungin D, Winkler TW, Croteau-Chonka DC, Ferreira T, Locke AE, Mägi R, et al. New genetic loci link adipose and insulin biology to body fat distribution. Nature. 2015;518(7538):187–96.
27. Wu Y, Gao H, Li H, Tabara Y, Nakatochi M, Chiu YF, et al. A meta-analysis of genome-wide association studies for adiponectin levels in East Asians identifies a novel locus near WDR11-FGFR2. Hum Mol Genet. 2014;23(4):1108–19.

28. McGrath LM, Cornelis MC, Lee PH, Robinson EB, Duncan LE, Barnett JH, et al. Genetic predictors of risk and resilience in psychiatric disorders: A cross-disorder genome-wide association study of functional impairment in major depressive disorder, bipolar disorder, and schizophrenia. Am J Med Genet Part B Neuropsychiatr Genet. 2013;162(8):779–88.
29. Takeda J, Seino Y, Tanaka K, Fukumoto H, Kayano T, Takahashi H, et al. Sequence of an intestinal cDNA encoding human gastric inhibitory polypeptide precursor. Proc Natl Acad Sci U S A. 1987;84(20):7005–8.
30. Wood AR, Esko T, Yang J, Vedantam S, Pers TH, Gustafsson S, et al. Defining the role of common variation in the genomic and biological architecture of adult human height. Nat Genet. 2014;46(11):1173–86.
31. Morris AP, Voight BF, Teslovich TM, Ferreira T, Segrè A V, Steinthorsdottir V, et al. Large-scale association analysis provides insights into the genetic architecture and pathophysiology of type 2 diabetes. Nat Genet. 2012;44(9):981–90.
32. Scott RA, Lagou V, Welch RP, Wheeler E, Montasser ME, Luan J, et al. Large-scale association analyses identify new loci influencing glycemic traits and provide insight into the underlying biological pathways. Nat Genet. 2012;44(9):991–1005.

Supplementary Table 6. Novel signals with suggestive association in Stage 1 (P<10-5) but with no r

Stage1							Stage2*
Chr	Position	SNV	EA/NEA	EAF	OR (95% CI)	P-value	Chr
5	3048750	rs16870903	T/C	0.0021	2.98 (1.9-4.68)	2.10E-06	
8	64660127	rs187357831	T/G	0.009	1.46 (1.25-1.72)	3.10E-06	
17	73841419	rs3893328	A/G	0.0097	0.59 (0.48-0.74)	2.26E-06	
8	17730962	rs145953760	A/G	0.0169	1.27 (1.14-1.41)	8.56E-06	
13	86575869	rs7329157	T/C	0.0283	1.18 (1.1-1.27)	9.72E-06	13
4	129526996	rs4975241	C/G	0.0607	1.14 (1.08-1.2)	1.32E-06	
18	77548685	rs28620500	A/G	0.071	0.85 (0.79-0.91)	3.40E-06	
4	83563582	rs4693043	A/G	0.144	1.08 (1.05-1.12)	3.16E-06	
6	65590847	rs7774169	A/G	0.1927	0.93 (0.9-0.96)	4.85E-06	6
7	30728452	rs917195	T/C	0.2349	0.93 (0.9-0.96)	1.91E-06	
12	21752108	rs10841855	T/G	0.2496	0.93 (0.9-0.96)	1.54E-06	
5	101620174	rs2548724	T/C	0.2554	1.07 (1.04-1.1)	4.77E-07	
17	48632401	rs898453	A/G	0.274	0.94 (0.91-0.96)	2.05E-06	17
3	170727351	rs1879442	A/G	0.2767	0.94 (0.92-0.97)	4.76E-06	3
17	27613677	rs12452857	A/G	0.2882	1.06 (1.04-1.09)	5.60E-06	17
1	219771721	rs4846569	T/C	0.2943	0.93 (0.9-0.95)	8.83E-09	1
17	17649172	rs11655029	T/C	0.3223	1.06 (1.03-1.09)	6.08E-06	17
15	54776716	rs11858061	A/G	0.3752	1.06 (1.04-1.09)	1.70E-06	15
8	145536056	rs62530366	G/A	0.38	1.08 (1.05-1.11)	1.90E-08	
12	133683261	rs905226	T/C	0.4508	0.95 (0.92-0.97)	8.80E-06	
9	126123009	rs2491353	T/C	0.4528	0.94 (0.92-0.97)	1.99E-06	9
4	95109078	rs1509946	T/G	0.4776	0.94 (0.92-0.96)	4.16E-07	4
22	50435480	rs5771069	A/G	0.4837	0.94 (0.91-0.96)	1.85E-07	22
18	40772286	rs816750	C/G	0.5046	1.06 (1.03-1.08)	2.55E-06	
20	45757655	rs4809627	T/C	0.5223	1.06 (1.03-1.08)	4.85E-06	
7	13894939	rs7801928	T/C	0.5413	1.06 (1.04-1.09)	1.29E-06	7
2	65642097	rs6731993	A/T	0.6107	1.07 (1.04-1.09)	2.60E-07	2
5	112823768	rs1057827	T/C	0.651	1.06 (1.04-1.09)	2.99E-06	5
3	73633701	rs9847947	C/G	0.7371	1.07 (1.04-1.1)	1.69E-06	
3	31176875	rs1625526	A/G	0.7496	1.07 (1.04-1.1)	1.11E-06	
3	114913508	rs6438234	A/G	0.763	0.93 (0.91-0.96)	1.79E-06	3
12	77398721	rs17815608	A/T	0.8276	1.08 (1.05-1.12)	6.20E-06	
6	148963919	rs150268806	T/C	0.8292	0.93 (0.9-0.96)	3.93E-06	
8	82343438	rs182719694	A/G	0.8546	1.1 (1.06-1.14)	3.07E-07	
7	121954105	rs62476011	T/C	0.8628	0.92 (0.89-0.95)	4.28E-06	7
1	88416590	rs6691335	T/C	0.9016	0.9 (0.87-0.94)	2.34E-06	
8	105560821	rs13268287	A/G	0.929	1.13 (1.07-1.19)	7.37E-06	
5	142172314	rs80020232	T/G	0.9819	0.58 (0.46-0.73)	6.41E-06	
19	22530857	rs191030109	T/C	0.9984	0.38 (0.25-0.57)	3.17E-06	

* - Stage 2 SNPs available on Metabochip are reported by their position and rsID. Other 22 variants v

rs187357831 variant rs185032206 (r^2=0.75), for rs3893328 variant rs75830455 (r2=0.53), for rs8002

rs13268508 (r2=0.85).

replication (P>5x10-8) in Stage 2 or Independent InterAct/Interact+GERA study analysis.

Position	SNV	r2 with lea	EA/NEA	EAF	OR (95% CI)	P-value	Stage1+Stage2 OR (95% CI)
			T/C	0.006	1.03 (0.86-1.23)	0.74	1.19 (1.01-1.40)
			T/G	0.009	0.88 (0.72-1.08)	0.23	1.20 (1.06-1.36)
			A/G	0.0097	0.95 (0.83-1.08)	0.40	0.84 (0.75-0.94)
			A/G	0.015	1.02 (0.91-1.14)	0.79	1.15 (1.06-1.24)
86575869	rs7329157	1	T/C	0.031	0.93 (0.84-1.02)	0.12	1.08 (1.02-1.15)
			C/G	0.057	1.11(0.95-1.28)	0.19	1.14 (1.08-1.19)
			A/G	0.060	1.17(0.95-1.46)	0.15	0.88 (0.82-0.94)
			A/G	0.156	0.96(0.87-1.05)	0.39	1.07 (1.03-1.10)
65533066	rs10498828	0.94	T/C	0.180	0.97 (0.94-1.01)	0.16	0.95 (0.93-0.97)
			T/C	0.215	0.95 (0.87-1.05)	0.33	0.93 (0.90-0.96)
			T/G	0.237	0.91 (0.83-1.01)	0.07	0.93 (0.90-0.96)
			T/C	0.232	1.07 (0.99-1.17)	0.10	1.07 (1.04-1.10)
48636534	rs989128	0.60	A/G	0.359	0.98 (0.95-1.01)	0.11	0.95 (0.93-0.97)
170724883	rs8192675	0.97	C/T	0.295	0.95 (0.92-0.99)	0.01	0.95 (0.93-0.97)
27647630	rs797973	0.84	G/T	0.267	1.03 (1.00-1.07)	0.04	1.05 (1.03-1.07)
219771721	rs4846569	1.00	T/C	0.284	0.99 (0.94-1.04)	0.61	0.94 (0.92-0.96)
17654319	rs11656775	0.95	A/G	0.332	1.04 (1.01-1.08)	0.03	1.05 (1.03-1.08)
54756628	rs4776231	0.91	A/C	0.382	1.02 (0.99-1.05)	0.25	1.04 (1.02-1.06)
			G/A	0.362	1.04 (0.99-1.04)	0.32	1.05 (1.03-1.07)
			T/C	0.421	0.95 (0.89-1.03)	0.20	0.95 (0.93-0.97)
126112812	rs10760280	0.66	T/C	0.571	0.99 (0.97-1.02)	0.70	0.96 (0.95-0.98)
95012684	rs1904096	0.82	C/A	0.516	1.00 (0.94-1.07)	0.98	0.95 (0.93-0.97)
50440296	rs137848	0.97	T/C	0.487	0.97 (0.94-1.00)	0.02	0.95 (0.93-0.97)
			C/G	0.535	1.07 (0.99-1.15)	0.08	1.06 (1.04-1.08)
			T/C	0.546	0.99 (0.92-1.07)	0.89	1.05 (1.03-1.08)
13894276	rs1019029	0.66	G/A	0.479	1.02 (0.99-1.05)	0.20	1.05 (1.03-1.07)
65627406	rs2661796	0.60	T/C	0.576	1.00 (0.97-1.03)	0.84	1.04 (1.02-1.06)
112809728	rs367943	1.00	C/T	0.660	1.03 (1.00-1.07)	0.03	1.05 (1.03-1.07)
			C/G	0.741	0.98 (0.90-1.07)	0.65	1.06 (1.03-1.09)
			A/G	0.742	1.01 (0.93-1.10)	0.80	1.06 (1.04-1.09)
114913508	rs6438234	1.00	A/G	0.747	0.97 (0.94-1.01)	0.12	0.95 (0.93-0.97)
			A/T	0.839	1.04 (0.93-1.15)	0.49	1.08 (1.04-1.11)
			T/C	0.829	1.08 (0.98-1.18)	0.14	0.94 (0.92-0.97)
			A/G	0.855	0.94 (0.85-1.05)	0.30	1.08 (1.05-1.12)
122017812	rs1859351	0.83	C/T	0.843	0.98 (0.94-1.02)	0.30	0.95 (0.92-0.97)
			T/C	0.896	1.03 (0.90-1.18)	0.67	0.91 (0.88-0.94)
			A/G	0.920	1.14 (0.98-1.32)	0.08	1.13 (1.08-1.19)
			T/G	0.982	1.012(0.85-1.21)	0.89	0.83 (0.72-0.95)
			T/C	0.998	0.99 (0.84-1.17)	0.95	0.87 (0.75-1.01)

were either directly available in the InterAct and GERA GWAS, or proxies were used in GERA as follows: fo

.0232 variant rs71587235 (r2=1.0), for rs191030109 variant rs146989164 (r2=0.60), for rs62530366 varia

P-value
4.06E-02
3.75E-03
1.82E-03
5.50E-04
1.05E-02
4.70E-07
1.25E-04
3.28E-05
9.02E-06
7.28E-06
1.99E-06
5.67E-07
4.27E-06
1.05E-07
1.60E-06
1.24E-07
6.50E-07
1.00E-05
8.98E-06
6.02E-05
8.45E-05
2.17E-06
6.00E-08
3.06E-07
5.43E-06
5.56E-06
1.33E-04
6.07E-07
1.47E-05
5.16E-06
3.41E-06
3.04E-06
2.04E-04
5.83E-06
3.72E-05
5.39E-07
1.38E-06
9.33E-03
7.12E-02

or

int