SUPPLEMENTARY NOTE, FIGURES and TABLES

Twelve type 2 diabetes susceptibility loci identified through large-scale association analysis

Benjamin F Voight, Laura J Scott, Valgerdur Steinthorsdottir, Andrew P Morris, Christian Dina, Ryan P Welch, Eleftheria Zeggini, Cornelia Huth, Yurii S Aulchenko, Gudmar Thorleifsson, Laura J McCulloch, Teresa Ferreira, Harald Grallert, Najaf Amin, Guanming Wu, Cristen J Willer, Soumya Raychaudhuri, Steve A McCarroll, Claudia Langenberg, Oliver M Hofmann, Josée Dupuis, Lu Qi, Ayellet V Segrè, Mandy van Hoek, Pau Navarro, Kristin Ardlie, Beverley Balkau, Rafn Benediktsson, Amanda J Bennett, Roza Blagieva, Eric Boerwinkle, Lori L Bonnycastle, Kristina Bengtsson Boström, Bert Bravenboer, Suzannah Bumpstead, Noël P Burtt, Guillaume Charpentier, Peter S Chines, Marilyn Cornelis, David J Couper, Gabe Crawford, Alex SF Doney, Katherine S Elliott, Amanda L Elliott, Michael R Erdos, Caroline S Fox, Christopher S Franklin, Martha Ganser, Christian Gieger, Niels Grarup, Todd Green, Simon Griffin, Christopher J Groves, Candace Guiducci, Samy Hadjadj, Neelam Hassanali, Christian Herder, Bo Isomaa, Anne U Jackson, Paul RV Johnson, Torben Jørgensen, Wen HL Kao, Norman Klopp, Augustine Kong, Peter Kraft, Johanna Kuusisto, Torsten Lauritzen, Man Li, Aloysius Lieverse, Cecilia M Lindgren, Valeriya Lyssenko, Michel Marre, Thomas Meitinger, Kristian Midthjell, Mario A Morken, Narisu Narisu, Peter Nilsson, Katharine R Owen, Felicity Payne, John RB Perry, Ann-Kristin Petersen, Carl Platou, Christine Proença, Inga Prokopenko, Wolfgang Rathmann, N William Rayner, Neil R Robertson, Ghislain Rocheleau, Michael Roden, Michael J Sampson, Richa Saxena, Beverley M Shields, Peter Shrader, Gunnar Sigurdsson, Thomas Sparsø, Klaus Strassburger, Heather M Stringham, Qi Sun, Amy J Swift, Barbara Thorand, Jean Tichet, Tiinamaija Tuomi, Rob M van Dam, Timon W van Haeften, Thijs van Herpt, Jana V van Vliet-Ostaptchouk, G Bragi Walters, Michael N Weedon, Cisca Wijmenga, Jacqueline Witteman, The MAGIC investigators, The GIANT consortium, Richard N Bergman, Stephane Cauchi, Francis S Collins, Anna L Gloyn, Ulf Gyllensten, Torben Hansen, Winston A Hide, Graham A Hitman, Albert Hofman, David J Hunter, Kristian Hveem, Markku Laakso, Karen L Mohlke, Andrew D Morris, Colin NA Palmer, Peter P Pramstaller, Igor Rudan, Eric Sijbrands, Lincoln D Stein, Jaakko Tuomilehto, Andre Uitterlinden, Mark Walker, Nicholas J Wareham, Richard M Watanabe, Goncalo R Abecasis, Bernhard O Boehm, Harry Campbell, Mark J Daly, Andrew T Hattersley, Frank B Hu, James B Meigs, James S Pankow, Oluf Pedersen, H.-Erich Wichmann, Inês Barroso, Jose C Florez, Timothy M Frayling, Leif Groop, Rob Sladek, Unnur Thorsteinsdottir, James F Wilson, Thomas Illig, Philippe Froguel, Cornelia M van Duijn, Kari Stefansson, David Altshuler, Michael Boehnke, Mark I McCarthy

Contents

Sample descriptions

Stage 1 and 2 Sample Descriptions: The Stage 1 autosomal and X chromosome meta-analysis sample comprises 8,130 cases and 38,987 controls. The Stage 2 autosomal meta-analysis sample comprises up to 34,412 cases and 59,925 controls, with *in silico* data from 2,832 cases and 15,843 controls and *de novo* genotyping (or in one instance, combined *de novo* and *in silico* data) for a further 31,580 cases and 44,082 controls. The Stage 2 X chromosome meta-analysis sample comprises 8,535 cases and 12,326 controls with de novo genotyping. The sample counts are provided in **Supplementary Table 1,** sample characteristics in **Supplementary Table 2** and details of the SNP genotyping platform, genotype and sample quality control and T2D association analysis approach in **Table SN1**.

Stage 1 Sample Descriptions

Stage 1 results for both autosomes and X markers were based on meta-analysis from eight GWAS studies.

DGDG (Diabetes Gene Discovery Group): Details of the samples and the GWA scan performed on them have been described elsewhere[S1]. Diabetic subjects were recruited at the UMR8090 CNRS unit in Lille (N=266) or at the Endocrinology-Diabetology Department of the Corbeil-Essonnes Hospital (N=413). Controls were drawn from participants in the Epidemiological Study on the Insulin Resistance syndrome (DESIR) program and were mainly resident in west-central France. In all, we included 679 T2D cases and 697 normoglycemic controls in this study: all were of French Caucasian origin. Cases were defined on the basis of: (i) T2D diagnosed according to 1997 American Diabetes Association (ADA)[S2,S3] criteria; (ii) a family history of diabetes in first degree relatives; and (iii) BMI <30 kg/m^2. Cases with a diagnosis of diabetes below 45 years of age were screened for known MODY mutations. Cases from the Corbeil-Essonnes Hospital were also systematically tested for fasting C-peptide levels: if fasting C-peptide levels were below 0.4 mg/l, subjects were also tested for anti-GAD antibodies: those with anti-GAD antibodies >10U/ml were excluded. A subset (76/266) of the cases from UMR8090 CNRS were tested for anti-islet-antibodies (positive if >5.5% islet beta-cells stained positive) and/or anti-insulin antibodies and positive cases were excluded. Controls were defined as having: (i) age at examination >45yr; (ii) normal fasting glucose according to 1997 ADA criteria[S2] (FG<5.7 mmol/l); and (iii) BMI <27 kg/m^2.

deCODE: Details of the samples and the GWA scan performed on them have been described elsewhere[S4,S5]. Cases were recruited based on lists of diabetic individuals diagnosed either through a long-term epidemiology study performed at the Icelandic Heart Association or at one of two major hospitals in Reykjavík. Additional cases were recruited through screening for T2D amongst participants in deCODE genetics-funded family-based studies on cardiovascular and metabolic traits. These Icelandic samples consist of 1,465 T2D cases and 23,194 population-based controls who were not known to have diabetes. All participants answered a questionnaire, including questions on medication and age at diagnosis. For those individuals with known diabetes, T2D was defined on the basis of treatment with diet and/or oral glucose-lowering agents. Individuals currently treated with insulin were classified as having T2D if they were also using or had previously used oral glucose-lowering agents. For previously-undiagnosed individuals, the diagnosis of T2D was based on the criteria set by the ADA[S2,S3]. Anti-glutamic acid decarboxylase (anti-GAD) antibodies were measured in 75% of the patients and, of those, 90% tested negative (<32 ng/ml). However, in this study, to maintain consistency with previous reports, subjects with positive anti-GAD antibodies were not excluded. Controls were defined as individuals with no known T2D or

diabetes medication use.

DGI (Diabetes Genetics Initiative): Details of the samples and the GWA scan performed on them have been described elsewhere[S6,S7]. For this analysis, however, we excluded the discordant-sibship component included in some earlier DGI studies, so that we could perform an OR based meta-analysis. The Finnish samples (656 cases, 720 controls) were predominantly ascertained from the Botnia region of Finland and the Swedish samples (366 cases, 355 controls) were recruited in Southern Sweden and Skara. Thus, the combined data set includes 1022 cases and 1075 NGT controls. Cases from both countries were defined as meeting WHO (1999) criteria[S8] with FG ≥ 7.0 mmol/l or a 2-hour glucose ≥ 11.1 mmol/l during an oral glucose tolerance test (OGTT). Cases had an age of diagnosis >35 years and no detectable glutamic acid decarboxylase antibodies (defined as anti-GAD antibody levels <32 IU/ml in the Finnish samples or <1.3 anti-GAD relative units in Sweden). Controls were defined as normal glucose tolerant (NGT) and having no first degree relatives with T2D. Cases and controls were selected to be matched on age (within 5 years of onset), sex, BMI, and geographic similarity.

EUROSPAN (European Special Population Network): Details of the samples and the previous GWA scan have been described elsewhere[S9,S10,S11,S12]. Cases and controls were ascertained from four genetically isolated European populations recruited from: South Tyrol, Italy (MICROS: Study of Micro-isolates in South Tyrol); Vis island, Croatia; Rucphen, Netherlands (ERF: the Erasmus Rucphen Family study); and Orkney Islands, Scotland (ORCADES: Orkney Complex Disease Study). The Scottish, South Tyrolean and Croatian populations are small populations founded by a limited number of individuals and/or have undergone a population bottleneck, followed by long isolation and no immigration[S13]. The Dutch population is a recently isolated population located in the southwest of the Netherlands[S14]. Participants ranged in age from 18-100 years and have decreased genetic diversity, increased linkage disequilibrium, and higher degrees of relatedness compared to non-isolated populations (though these factors were taken into account in the analyses). A total of 268 T2D cases and 3,710 normoglycemic controls were included in the present study. For all samples, T2D was defined as FPG >= 7.0mmol/l and/or use of anti-diabetic medication. No islet autoantibodies were measured. Controls were normoglycemic and did not have any history of using medication for T2D.

FUSION (Finland-United States Investigation of NIDDM Genetics): Details of the samples and the GWA scan performed on them have been described elsewhere[S15]. The FUSION study collection includes Finnish T2D sibpairs (where the T2D proband had a reported age-of-diagnosis between 35 and 60), their family members and spouses and also contains (as additional controls) individuals from Vantaa, Finland, who were born in 1925 and were normal glucose tolerant (NGT) at ages 65 and 70. Only one T2D case per family was included (n=789); controls were selected from the spouses (n=304) and from the older individuals from Vantaa (n=219). A further 372 cases and 651 controls (also of Finnish origin) were selected from Finrisk2002, a population-based risk factor survey[S16]. In total, therefore, the FUSION sample comprises 1,161 T2D cases and 1,174 NGT controls. T2D cases were defined by WHO 1999 criteria[S8] of fasting plasma glucose ≥7.0 mmol/l or 2-h plasma glucose ≥11.1 mmol/l, by report of diabetes medication use, or based on medical record review. FUSION cases were excluded if they had: (i) known or probable type 1 diabetes amongst first degree relatives; (ii) insulin treatment initiated within 10 years of disease diagnosis, detectable levels of anti-GAD antibodies and fasting C-peptide ≤0.30 nmol/l; or (iii) insulin treatment initiated within 4 years of diagnosis and fasting C-peptide ≤0.30 nmol/l. Controls had NGT as defined by WHO 1999 criteria[S8]. Cases and controls were approximately frequency matched by 5-year age category, sex, and birth province.

KORAgen (Cooperative Health Research in the Region of Augsburg, Southern Germany): Details of the samples and the GWA scan performed on them have been described elsewhere[S17,S18,S19]. KORA is a regional research platform for population-based studies, subsequent follow-up studies and family studies, established in 1996 to expand the WHO (World Health Organization) MONICA (Monitoring of Trends and Determinants in Cardiovascular Disease) project in Augsburg. For DIAGRAM+ Stage 1, we included 433 T2D cases and 1,438 non-diabetic controls, all of German ethnicity. Participants were ascertained from the KORA surveys S3 (1994/1995), F3 (follow-up of S3, 2004/2005) and S4 (1999-2001). Cases were identified by self report of T2D in a personal interview which was validated by a questionnaire mailed to the treating physician and/or by medical chart review. No islet autoantibodies were measured. Controls were non-diabetic as defined by self-report.

Rotterdam Study: Details of the samples and the GWA scan performed on them have been described elsewhere[S20]. The Rotterdam Study is an ongoing prospective, population-based, cohort study among inhabitants of a Rotterdam suburb to investigate chronic diseases in persons aged 55 years and above [S20]. Baseline examinations took place from 1990-1993 and in 1999. In addition to continual surveillance for major disease outcomes, follow-up examinations were performed at ~3 year intervals. These samples used in DIAGRAM+ Stage 1 comprise 1,178 T2D cases and 4,761 non-diabetic population controls, all of Dutch ethnicity. At baseline, prevalent cases of diabetes were diagnosed by a non-fasting or postload glucose level (after oral glucose tolerance testing) ≥11.1 mmol/l and/or treatment with oral antidiabetic medication or insulin. During follow-up, incident diabetes was diagnosed as fasting plasma glucose levels ≥7.0 mmol/l, non-fasting plasma glucose levels ≥11.0 mmol/l and/or treatment with oral antidiabetic medication or insulin. All diagnoses of T2D were confirmed by a general practitioner. Patients registered in general practitioner records as having type 1 diabetes were excluded from the present analyses (n = 15). Islet cell specific antibodies were not measured. At baseline, controls from the 1990-1993 cohort were defined on the basis of NGT or nonfasting glucose level <11.0 mmol/l, and from the 1999 cohort on the basis of FPG <7.0 mmol/l. During the follow-up, controls from either cohort had to maintain FPG <7.0 mmol/l and have no physician diagnosis of diabetes.

WTCCC/UKT2D (Wellcome Trust Case Control Consortium/United Kingdom Type 2 Diabetes Genetics consortium) Details of the samples and the GWA scan performed on them have been described elsewhere[S7,S21,S22]. The T2D cases were selected from UK Caucasian subjects who are part of the Diabetes UK Warren 2 repository. Approximately 30% were explicitly recruited as part of multiplex sibships[S23] and ~25% were offspring in parent-offspring "trios" or "duos" (that is families comprising only one parent complemented by additional sibs)[S24]. The remainder were recruited as isolated cases but these cases were (compared to population-based cases) of relatively early onset and had a high proportion of T2D parents and/or siblings[S25]. Controls were derived from two sources in approximately equal numbers. First, the 1958 Birth Cohort (also known as the National Child Development Study) includes all births in England, Wales and Scotland, during one week in 1958. In a biomedical examination at 44-45 years[S26] (http://www.b58cgene.sgul.ac.uk/followup.php), blood was drawn and EBV transformed cell lines prepared. Second, the WTCCC in collaboration with the UK Blood Services set up a UK national repository of anonymised samples of DNA and viable mononuclear cells from consenting blood donors, age range 18-69 years. The samples included in the DIAGRAM+ Stage 1 are those from the WTCCC study and comprise 1,924 T2D cases and 2,938 population controls. Cases were individuals who reported all four grandparents having exclusively British and/or Irish origin, by both self-reported ethnicity and place of birth. T2D was defined as current prescribed treatment with sulphonylureas, biguanides, other oral agents and/or insulin or,

in the case of individuals treated with diet alone, historical or contemporary laboratory evidence of hyperglycemia (as defined by WHO [S8]). Individuals with other known forms of diabetes, e.g. maturity-onset diabetes of the young and mitochondrial diabetes, were excluded. Other inclusion criteria were: absence of first-degree relatives with type 1 diabetes; an interval of ≥1 year between diagnosis and institution of regular insulin therapy; and negative testing for antibodies to glutamic acid decarboxylase (anti-GAD). An anti-GAD titer >10 U (corresponding to ~8 SD above the mean of 88 normal control subjects) in duplicate samples was considered positive. All cases were diagnosed between age 25 and 75. Controls from both sources were selected without reference to T2D status. 1958 Birth Cohort controls were of self-reported white ethnicity and were representative of gender and each geographical region. The UK blood donor controls were selected based on sex and geographical region to reproduce the distribution of the samples of the 1958 Birth Cohort.

Stage 2 Sample Descriptions

We obtained association summary results for up to 23 of the most strongly associated autosomal SNPs from 19 independent samples (not including Breda). As the Stage 1 X chromosome signal only emerged some time after the autosomal signals, and since X chromosome imputation had not been performed for ARIC, FHS and NHS, replication genotyping for rs5945326 has been confined to Breda, DARTS, a subset of DGDG (Stage2) and Ulm.

***In silico* Stage 2 data:** We obtained association summary results for the 23 most strongly associated autosomal SNPs from three GWA studies (not included in the 8-study GWAS meta-analysis).

ARIC (Atherosclerosis Risk in Communities Study): Details of the samples and the GWA scan performed on them have been described elsewhere[S27,S28]. The Atherosclerosis Risk in Communities (ARIC) study is a prospective study to investigate the etiology of atherosclerosis and another outcomes in each of four US communities (Forsyth County, North Carolina; Jackson, Mississippi; suburbs of Minneapolis, Minnesota; Washington County, Maryland). We included data from European ancestry samples only – in total, 691 T2D cases and 6,425 non-T2D controls. Cases and controls were ascertained at the baseline examination (1987-89). Diabetes was defined as self-reported physician diagnosis of diabetes, self-reported use of diabetes medications in the last two weeks, fasting glucose ≥ 126 mg/dL or casual glucose ≥ 200 mg/dL. No islet antibodies were measured. Admixture of T1D cases within the diabetes group is estimated at 2% based on self-reported age of diagnosis and insulin use at baseline. Controls had fasting glucose <110 mg/dl.

FHS (Framingham Heart Study): Details of the samples and the GWA scan performed on them have been described elsewhere[S29,S30,S31]. These data were collected over three epochs, beginning in 1948, in a study designed to allow prospective study of cardiovascular and metabolic outcomes over several generations. The FHS sample for which GWA data could be included in DIAGRAM+ Stage 2 replication comprised 674 T2D cases and 7,664 non-diabetic controls from all three generations of Framingham participants. Almost all are of European ancestry. Diabetes was defined as: (i) for cohort "Gen 1"[S29], casual glucose ≥200 mg/dl at any of the study examinations (1-22) or taking diabetes medication (oral or insulin) at any examination ; (ii) for the offspring (Gen 2) cohort[S30], fasting plasma glucose ≥126 mg/dl at any examination (1-7) or diabetes treatment at any visit (in the offspring, >99% of diabetes is type 2 diabetes); and (iii) for the "Gen 3" cohort[S31]: fasting plasma glucose ≥126 mg/dl at exam 1 or diabetes treatment at exam 1. No islet autoantibodies were measured in any individual. Controls had no diabetes (fasting glucose <126 mg/dl and no T2D medication) at their last study visits.

NHS (Nurses Health Study): Details of the NHS have been described previously[S32]. NHS is a longitudinal study of diet, lifestyle, and biological factors that influence the development of chronic diseases in nurses from the United States. For DIAGRAM+ Stage 2, we included European ancestry samples for a total of 1,467 T2D cases and 1,754 non-diabetic controls. Cases and controls were selected using a "nested" case-control study design. Before 1998, T2D was defined using criteria consistent with those proposed by the National Diabetes Data Group (NDDG)[S33]. During the 1998 and 2000 cycles, T2D was defined using the ADA criteria[S2,S3]. Individuals reporting diagnosis of type 1 diabetes were excluded. No islet autoantibodies were measured. Controls did not have diabetes at baseline and remained without diabetes through the 2006 follow-up.

***De novo* Stage 2 genotyping or combined *de novo* genotyping and GWAS**: Additional autosomal SNP genotyping was performed in 16 case-control studies. In 15 samples all of this additional genotyping was de novo; but in the last (Diagene/Rotterdam), the analysis was based on a combination of de novo typing and GWAS data not otherwise included in Stage 1. Genotyping on the X chromosome SNP was performed in 4 studies (Breda, DARTS, a subset of DGDG (Stage2) and Ulm).

ADDITION/Ely (Anglo-Danish-Dutch study of Intensive Treatment In PeOple with screeN detected diabetes in primary care/MRC Ely Study): Detailed characteristics of these study populations have been described elsewhere[S7,S34,S35]. The Cambridge arm of ADDITION study identified previously undiagnosed T2D cases via a population-based stepwise screening strategy among 40 to 69 year olds (generating 800 cases). The MRC Ely study is a population-based cohort of white European men and women aged 35 to 79 years (and provided 92 cases and 1612 controls). Therefore, for purposes of Stage 2, we had access to 892 T2D cases and 1612 non-diabetic controls. All participants were of British, Irish and/or other white ethnic origin. T2D was defined based on WHO criteria [S8]. No islet autoantibodies were measured. Controls had no diagnosed diabetes based on WHO criteria and were from a similar sampling frame as the cases[S35].

Breda (Breda Study): Detailed characteristics of the study sample have been described elsewhere[S36,S37]. Cases were recruited in 1998–1999 in collaboration with the Diabetes Service Breda and 80 general practitioners from the city of Breda in The Netherlands. The control sample is made up of healthy blood bank donors with unknown T2D status. We had access to 545 T2D cases and 880 controls of unknown T2D status. T2D was diagnosed according to WHO criteria[S8,S38] and cases had at least one additional family member with T2D. Cases were excluded if they had received insulin within a year of diagnosis, had a history of ketoacidosis, displayed undetectable C-peptide concentrations, a positive family history of T1D or a family history indicative of monogenic diabetes. No islet autoantibodies were measured. As controls were blood donors, no information on T2D status was available. All cases and controls were from The Netherlands and of European descent, and at least three of their four grandparents were also born in The Netherlands. These samples were used only for the X chromosome SNP genotyping.

CCCS (Cambridgeshire Case-Control Study): Detailed characteristics of these study samples have been described elsewhere[S7,S39]. Cases aged 45-76 years were randomly selected from 41 general practitioner diabetes registers in Cambridgeshire, UK. Controls were recruited at random from the same population sampling frame. For DIAGRAM+ Stage 2 replication, we had access to 541 T2D cases and 527 non-diabetic controls of white European UK origin. T2D was defined using diagnosis of diabetes at age >30 years and lack of insulin use in the first year after diagnosis [S39]. Anti-GAD antibodies were measured in a subset of the cases (n=394) to examine the proportion of late onset autoimmune diabetes: 12% of the cases had some evidence of islet autoimmunity.

Diabetes was excluded in controls by medical record search and by a glycated haemoglobin measurement of less than 6%. Controls were individually matched to cases for age, sex and GP practice.

DARTS (The Diabetes Audit and Research in Tayside Scotland): Sample characteristics have been reported elsewhere[S7,S22]. For the DIAGRAM+ Stage 2 analysis, we included 5,113 T2D cases and 6,615 normoglycemic population-based controls, all of European UK origin. The numbers included in this study are larger than those previously reported for this sample[S7,S22] due to ongoing recruitment to the same criteria. Cases had T2D diagnosed between the ages of 35-70 years (inclusive). The diagnosis of diabetes was based on either current prescribed treatment with diabetes-specific medication or, in the case of individuals treated with diet alone, laboratory evidence of diabetes as defined by the WHO[S8,S38]. Cases were excluded if they had an established (clinical and/or molecular) diagnosis of monogenic diabetes (e.g. maturity-onset diabetes of the young, mitochondrial diabetes) or if they had been treated with regular insulin therapy within 1 year of diagnosis. No islet autoantibodies were measured. Controls were defined as having no diagnosis of diabetes at the time of recruitment (or subsequently), fasting glucose ≤7.0 mmol/l, HbA1c ≤6.4% and age < 80 years.

DGDG (Diabetes Gene Discovery Group) (Stage 2): Sample characteristics have been described previously[S1,S40]. Both cases and control groups were recruited through hospital attendance, media campaigns and epidemiological studies. Diabetic individuals were recruited at the UMR8090 CNRS unit in Lille (n=452), from incident cases in the D.E.S.I.R cohort (n=262), at the Endocrinology-Diabetology Department of the Corbeil-Essonnes Hospital (n=1,124) and at the Endocrinology-Diabetology Department of the Poitiers University Hospital (n=1,362). Controls were ascertained as part of the D.E.S.I.R. cohort (n=1,977), from the "Fleurbaix-Laventie Ville Santé" cohorts[S41] (n=212), or at the UMR8090 CNRS unit in Lille (n=936). In total therefore, 3,200 T2D cases and 3,125 normoglycemic controls of French Caucasian samples were available. T2D recruitment was based on 1997 ADA criteria[S2] and BMI <35 kg/m². As described for the Stage 1 DGDG sample, cases from the Corbeil-Essonnes Hospital (n=891) were tested for C-peptide, while a subset of cases from UMR8090 (n=121) were tested for anti-islets auto-antibodies and/or anti-insulin antibodies. The remaining cases were not tested for auto antibodies markers. Controls were defined on the basis of normoglycemia according to 1997 ADA criteria[S2] (FG <5.7 mmol/L), age at exam >40 years, and BMI <35 kg/m².

DiaGene/Rotterdam Study: Sample characteristics and recruitment of the Rotterdam Study and DiaGene study have been described in detail elsewhere[S20]. Rotterdam Study participants for Stage 2 were recruited to the Rotterdam Study between 2006-2008[S20] and were not included in the Rotterdam Stage 1 GWAS sample. The DiaGene study is an ongoing collection of population-based T2D cases and controls from the city of Eindhoven, the Netherlands. The Stage 2 sample includes a total of 1,699 cases (Rotterdam Study, n=239; DiaGene, n=1,460) and 1,656 non-diabetic controls from the Rotterdam Study. Cases were diagnosed according to WHO and ADA guidelines[S2,S3,S8,S38] and/or based on treatment with oral anti-diabetic medication or insulin. No islet autoantibodies were measured. Controls are aged ≥55years and did not have diabetes and remained without diabetes throughout follow-up. The Rotterdam Study sample had GWAS data available (but this arrived too late for inclusion in the Stage 1 meta-analysis) and had been imputed as described above for the Stage 1 samples. The DiaGene sample was specifically genotyped for the SNPs of interest.

FUSION2 (Finland-United States Investigation of NIDDM Genetics, Stage 2): Detailed characteristics of the study sample have been described elsewhere[S15]. The cases and controls were selected from the Dehko2D, Health 2000, Finrisk 1987, Finrisk 2002 and Savitaipale Diabetes studies, with additional cases provided by the Action

LADA study. The Dehko2D, Health 2000, Finrisk and Savitaipale Diabetes studies are population-based surveys; Action LADA is a study of latent autoimmune diabetes in adults (LADA) in recently-diagnosed diabetes patients (though recruitment here was restricted to those who were GAD-antibody negative). In all, we had access to 1,211 T2D cases and 1,266 NGT controls, all of Finnish origin. T2D was defined according to WHO 1999 criteria[S8]. No islet autoantibodies were measured in the four other studies. Controls were NGT and were approximately frequency-matched to the T2D cases by five-year age category, sex, and birth province.

GCI-Poland (Genomics Collaborative, Inc., Polish sample): Detailed characteristics of this study sample have been described elsewhere[S6]. These samples originated from a repository of individuals enrolled in studies of multifactorial disease, as well as samples from healthy individuals enrolled as control subjects. We included data from 790 T2D cases and 803 normoglycemic controls, of Polish origin. All samples came from a single location in Poland and had two parents and four grandparents with Polish ethnicity. T2D patients were defined according to current WHO (1999) criteria [S8]. No islet autoantibodies were measured. Control subjects were healthy individuals with FPG <7 mmol/l and with no known history of chronic disease. Cases and controls were matched by age, sex and grand-parental country of origin.

GCI-US (Genomics Collaborative, Inc., US sample): Detailed characteristics of this study sample have been described elsewhere[S6]. These samples originated from a repository of individuals enrolled in studies of multifactorial disease, as well as samples from healthy individuals enrolled as control subjects. Here, we included data from 1,190 T2D cases and 1,168 normoglycemic controls, all of US European origin. All were non-Hispanic and had two parents and four grandparents with self-reported US European ethnicity. T2D patients were defined according to current WHO (1999) criteria[S8]. No islet autoantibodies were measured. Control subjects were healthy individuals with FPG <7 mmol/l and with no known history of chronic disease. Cases and controls were matched by age, sex and grand-parental country of origin.

HUNT (Nord-Trøndelag Health Study): Details of the study sample have been described elsewhere[S7]. Cases and controls were from the county of Nord-Trøndelag in central Norway and selected from the HUNT2 study. We included 1,213 population-based T2D cases and 3,221 non-diabetic controls, all of Norwegian origin. Cases were initially identified by self-report of treatment for diabetes (diet only, oral anti-diabetic medication, or insulin treatment started at least 12 months after the onset of diabetes). After additional phenotyping, T2D cases were further defined by anti-GAD antibody <0.08 units and at least one of the following: C-peptide levels >150 pmol/l or age of diagnosis >35years. The controls were individuals who gave no report of a personal history of diabetes.

KORAgen (Cooperative Health Research in the Region of Augsburg, Southern Germany): Details of the KORA Surveys and the family study have been described elsewhere[S17,S18,S19]. In DIAGRAM+ Stage 2, we included 1,016 T2D cases and 1,498 non-diabetic controls, all of German origin. Of this total, 447 cases (independent of those included in the Stage 1 KORAgen GWAS sample) were drawn from four KORA surveys (S1-S4, 1984-2001)[S17] and 569 cases were drawn from the KORA T2DM Family Study (T2DMFAM, 2001/2002)[S19]. Age-and sex-matched controls were from KORA survey S4. Case and control definitions were the same as those described earlier for the Stage 1 KORAgen sample.

Malmo (Malmo Diabetes Registry/Malmo Diet and Cancer Study): Detailed characteristics of this study sample have been described elsewhere[S6]. Cases were selected from the Malmo Diabetes Registry. Controls were chosen from the Malmo Diet and Cancer study. In all, for DIAGRAM+ Stage 2, we included 2,449 cases and 3,363

normoglycemic controls. T2D was defined according to WHO (1999)[S8]. In addition, all cases had Scandinavian origin, age of diagnosis of T2D >35 years, C-peptide ≥0.3 nmol/l and were anti-GAD antibody negative (as described for the Stage 1 DGI study). Controls were Swedes defined as having fasting blood glucose <5.5 mmol/l and HbA1c <6.0%.

METSIM (METabolic Syndrome In Men): Detailed sample characteristics have been reported elsewhere[S42]**.** The METSIM study sample is randomly selected from the population register of the town of Kuopio in Eastern Finland. For the present study, we included 940 T2D cases and 4,152 NGT controls, all of Finnish origin. T2D was defined using WHO 1999 criteria[S8]. Type 1 diabetic subjects were excluded on the basis of clinical history according to WHO criteria[S8] including ketoacidosis at diagnosis and insulin treatment started at diagnosis. Measurement of GAD auto-antibodies in a subset of the cohort suggests <5% have LADA. Controls had NGT (NGT).

NDCCS (EPIC and Norfolk Diabetes Case Control Study): Detailed characteristics of the samples have been described elsewhere[S7]. The Norfolk Diabetes Case-Control Study is an ongoing study of men and women with T2D in Norfolk. All T2D patients identified through general practice diabetes registers in Norfolk and local hospital diabetes clinic and retinal screening programme patient registers are invited to participate. The sample included in DIAGRAM Stage 2 includes 6,056 T2D cases and 6,428 controls reporting British, Irish and/ or other white ethnic origin. T2D was defined as clinically diagnosed diabetes, with no record of insulin use during the first year of diagnosis, cystic fibrosis, chronic pancreatitis or long term steroid use. No islet autoantibodies were measured. Controls were randomly selected from EPIC-Norfolk participants and were free of known diabetes at baseline or during follow-up. Diabetes was excluded in controls based on self-report (self reported history of diabetes, doctor diagnosed diabetes, anti-diabetic drug use), linkage to primary care registers, secondary care registers, hospital admissions and mortality data.

Steno (Steno Diabetes Center Study): Detailed sample characteristics have been reported elsewhere[S7,S34,S43,S44]. The cases are made up of: (a) T2D individuals from the population-based Inter99 sample of middle-aged people sampled at Research Centre for Prevention and Health[S43]; (b) T2D patients sampled through the out-patient clinic at Steno Diabetes Center; and (c) T2D cases from the Danish ADDITION study sampled through the Department of General Practice at University of Aarhus[S34,S44]. Control subjects were derived from: (a) Inter99, and (b) from a population-based group of middle-aged subjects recruited from Steno Diabetes Center, whose phenotype characterisation was similar to that of Inter99. As a result, a total of 3,672 T2D cases and 5,046 NGT controls were available for this study, all of Danish origin. T2D was diagnosed in accordance with WHO criteria[S8,S38]. Diabetic individuals with the following known conditions/clinical features were excluded from the category of type 2 diabetes: chronic pancreatitis; hemochromatosis; severe insulin resistance; maturity-onset diabetes of the young; maternally inherited diabetes and deafness; family history of first-degree relatives with type 1 diabetes; insulin requirement within the first year after diabetes diagnosis; fasting serum C-peptide level ≤150 pmol/l at the time of recruitment; anti-GAD antibody positivity (level >0 U/ml). Controls were all of NGT.

Ulm (Center of Excellence-Baden-Wuerttemberg Type 2 Diabetes Cohort): Detailed sample characteristics have been reported elsewhere[S45]. All samples were ascertained through the Centre of Excellence for Metabolic Disorders Division at Ulm University. For this study, we had access to 944 T2D cases and 1,949 NGT controls, all of German origin. T2D was defined as FPG >125 mg/dl, or 2 hour glucose >200 mg/dl after oral glucose tolerance test (OGTT), or, alternatively, use of oral anti-diabetic agents and/or insulin. In addition, all individuals were tested for serum autoantibodies including anti-GAD, anti-insulinoma-associated protein 2 (IA2) and anti-islet cell

(ICA)[S46,S47]. Positive IA2 or anti-GAD antibody levels were defined as levels ≥99th% percentile based on 2000 healthy adults. ICA levels ≥20 JDF units were considered to be positive. Controls were NGT (in 1100, confirmed by HbA1c <6%) and had no evidence of islet autoimmunity.

W2C/UKBS (Warren 2 Consortium/United Kingdom Blood Services): Sample characteristics have been reported elsewhere[S7,S22]. The cases were individuals from the wider Warren 2 consortium collection who had not been included in the WTCCC GWA (as described under Stage 1). The controls were samples from the UK Blood Services Controls (UKBS) (see Stage 1) also not included in the WTCCC GWAS study. A total of 654 T2D cases and 1,653 UKBS population controls of UK origin were included. Inclusion and exclusion criteria are as described for Stage 1 WTCCC, but these 654 cases were not recruited as part of sibships, duos or trios.

All samples used had been collected with appropriate informed consent consistent with their use in the present study.

Study	Genotyping						Imputation			Association Analysis						Sample QC	
	Platform	MAF	Call Rate	p for HWE	Autosomal SNPs Meeting QC Criteria	Chr X SNPs Meeting QC Criteria	Imputation Software[a]	MAF	Imputation Quality	Total Autosomal SNPs in Meta-Analysis	Total X chr SNPs in Meta-Analysis	Autosomal genotyped λ_{GC}	Autosomal imputed λ_{GC}	Analysis Software[b]	Covariates	Call Rate	Sample duplication/relatedness checks
Stage 1																	
DGDG	Illumina Human Hap 300 Bead Array	>0.01	>0.95	$>10^{-4}$	297,972	8,863	IMPUTE	>0.01	proper-info > 0.5	2,051,387	57,968	1.1	1.098	SNPTEST	none	>0.95	Yes, removed
deCODE	Illumina Human Hap 300/300-duo+ or CNV370-duo Bead Arrays	>0.01	>0.96	$>10^{-6}$	281,406	8,456	IMPUTE	>0.01	proper-info > 0.5	2,338,113	58,783	1.308[b]	1.305[b]	SNPTEST	none	>0.98	Yes, duplicates removed, related individuals accounted for by use of genomic control correction
DGI	Affymetrix GeneChip Human Mapping 500k Array Set	>0.01	>0.95	$>10^{-6}$	380,748	7,908	MACH1	>0.01	proper-info > 0.5	2,230,032	55,071	1.05	1.06	PLINK Genotyped/ SNPTEST Imputed	Age, Sex, BMI, Study Center	>0.95	Yes, removed
EUROSPAN	Illumina Human Hap 300 or HapMap 370 CNV Bead Arrays	>0.01	>0.98	$>10^{-6}$	309,642	8,983	MACH1	>0.01	$r^2 > 0.5$	2,359,525	64,982[c]	0.974	0.977	GenABEL Genotyped/ SNPTEST Imputed	Age, Sex	>0.98	Yes, removed
FUSION	Illumina Human Hap 300 Bead Array	>0.01	>0.90	$>10^{-6}$	306,239	8,979	MACH1	>0.01	$r^2 > 0.3$	2,413,085	60,218	1.03	1.04	In-house Genotyped/ mach2dat Imputed	Age, Sex, Birth Province	>0.975	Yes, removed
KORAgen	Affymetrix GeneChip Human Mapping 500k Array Set	>0.01	>0.95	$>10^{-6}$	356,183	8,176	IMPUTE	>0.01	proper-info > 0.5	2,325,232	54,062	1.04	1.04	SNPTEST	Age, Sex	>0.93	Yes, checked
Rotterdam	Illumina Human Hap 550 Bead Array	>0.01	>0.98	$>10^{-6}$	500,241	12,601	MACH1	>0.01	$r^2 > 0.5$	2,439,672	60,287	1.006	1.01	GenABEL Genotyped/ SNPTEST Imputed	None	>0.975	Yes, removed
WTCCC/UKT2D	Affymetrix GeneChip Human Mapping 500k Array Set	>0.01	>0.95 for MAF>0.05 >0.99 for MAF<0.05	$>10^{-3}$	393,142	8,204	IMPUTE	>0.01	proper-info > 0.5	2,308,535	56,641	1.06	1.08	PLINK Genotyped/ SNPTEST Imputed	none	>0.97	Yes, removed
Stage 2																	
In silico replication																	
ARIC	Affymetrix Genome-Wide Human SNP Array 6.0	>0.01	>0.90	$>10^{-4}$	4		MACH1	>0.01	$r^2 > 0.3$	20		n/a	n/a	ProbABEL	Age, Sex, Study Center	>0.95	Yes, removed
FHS	Affymetrix GeneChip Human Mapping 500k Array Set + MIPS 50K	>0.01	>0.95	$>10^{-4}$	2		MACH1	>0.01	$r^2 > 0.5$	23		1.14	1.02	generalized estimating equation (GEE) in R	age, sex, cohort	>0.95	Yes, duplicates removed, relatedness accounted for in the analysis

NHS	Affymetrix Genome-Wide Human SNP Array 6.0	>0.02	>0.98	$>10^{-4}$	1		MACH1	>0.02	$r^2 > 0.3$	23		1.02	1.02	PLINK Genotyped/ ProABEL Imputed	age, BMI	>0.98	Yes, removed
de novo genotyping or combined in silico/de novo genotyping																	
ADDITION/Ely	Sequenom and Custom TaqMan SNP Assays	>0.01	>0.90	$>10^{-4}$	19		-	-	-	19	0	-	-	Stata	Age, Sex, BMI	-	-
Breda	Taqman SNP Assays	>0.01	>0.90	$>10^{-4}$		1				0	1			SPSS	none		
CCCS	Sequenom and Custom TaqMan SNP Assays	>0.01	>0.90	$>10^{-4}$	20					20	0	-	-	Stata	Age, Sex, BMI	-	-
DARTS	Taqman SNP Assays	>0.01	>0.90	>10-4	20	1	-	-	-	20	1	-	-	Stata	none	-	-
DGDG	Sequenom iPLEX Assay	>0.01	>0.90	$>10^{-4}$	22	1	-	-	-	22	1	-	-	PLINK	none	-	-
DiaGene[d]	Taqman SNP Assays	>0.01	>0.90	$>10^{-4}$	23		-	-	-	23	0	-	-	SPSS		-	-
Rotterdam[d]	Illumina Infinium II HumanHap550, v3 and Taqman SNP Assays (3 SNPs)	>0.01	>0.98	$>10^{-6}$	3		MACH1	>0.01	$r^2 > 0.9$	20	0	n/a	n/a	SPSS	Age, Sex, BMI	>.975	Yes, removed
FUSION2	Sequenom iPLEX Assays	>0.01	>0.90	$>10^{-4}$	17		-	-	-	17	0	-	-	In-house	Age, Sex, Birth province	-	-
GCI-Poland	Sequenom iPLEX Assays	>0.01	>0.90	$>10^{-4}$	19		-	-	-	19	0	-	-	PLINK	Age, Sex, BMI	-	-
GCI-US	Sequenom iPLEX Assays	>0.01	>0.90	$>10^{-4}$	19		-	-	-	19	0	-	-	PLINK	Age, Sex, BMI	-	-
HUNT	Sequenom iPLEX Assays	>0.01	>0.90	$>10^{-4}$	18		-	-	-	18	0	-	-	In-house	Sex	-	-
KORAgen	Sequenom iPLEX Assays	>0.01	>0.90	$>10^{-4}$	14		-	-	-	14	0	-	-	SAS	Age, Sex	-	-
Malmo	Sequenom iPLEX Assays	>0.01	>0.90	$>10^{-4}$	19		-	-	-	19	0	-	-	PLINK	Age, Sex, BMI	-	-
METSIM	Sequenom iPLEX Assays	>0.01	>0.90	$>10^{-4}$	17		-	-	-	17	0	-	-	In-house	Age	-	-
NDCCS	Sequenom and Custom TaqMan SNP Assays	>0.01	>0.85	$>10^{-4}$	20		-	-	-	20	0	-	-	Stata	Age, Sex, BMI	-	-
Steno	KASPar SNP genotyping	>0.01	>0.95	$>10^{-4}$	23		-	-	-	23	0	-	-	Rgui/Fishers Exact Test	none		
ULM	Sequenom iPLEX Assays	>0.01	>0.90	$>10^{-4}$	21	1	-	-	-	21	1	-	-	PLINK	none	-	-
W2C/UKBS	Taqman SNP Assays	>0.01	>0.90	>10-4	20		-	-	-	20	0	-	-	Stata	none	-	-

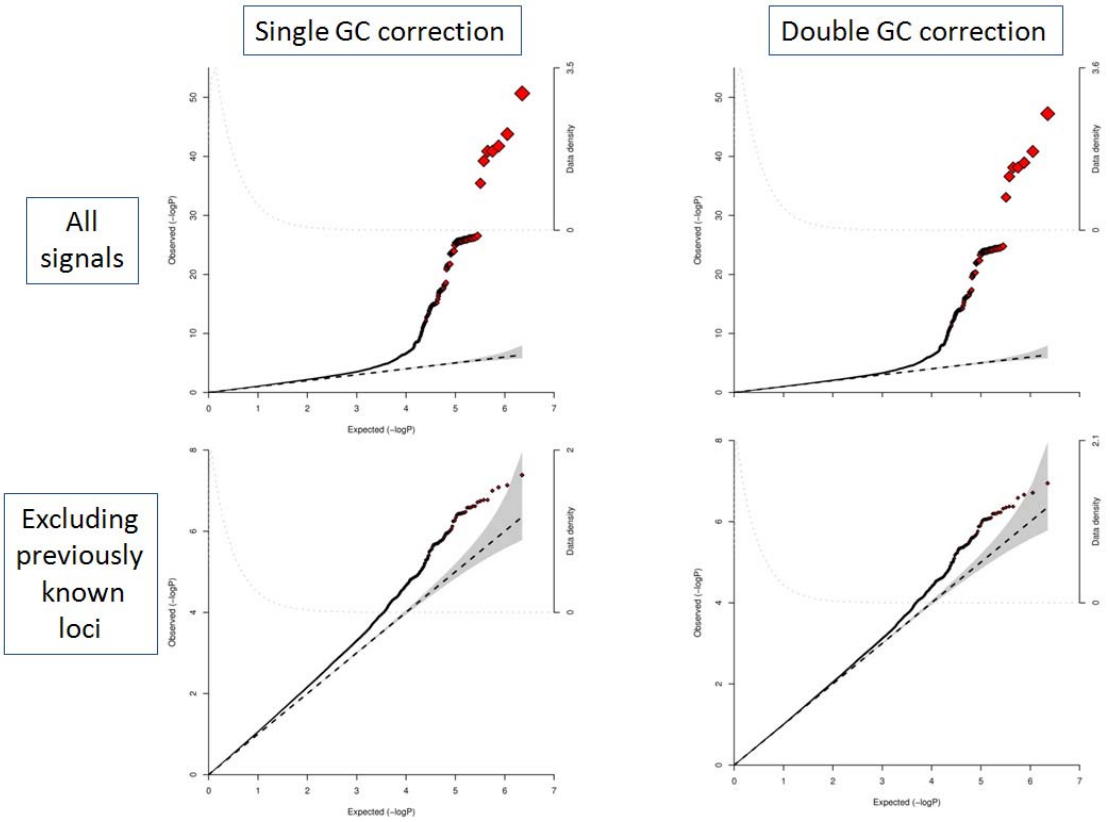
Supplementary Note Table SN1: Stage 1 and Stage 2 Genotyping, Imputation and Analysis

[a]Autosomal SNP imputation was performed using IMPUTE[S48] or MACH1 (Li, Y., Willer, C.J., Ding, J., Scheet, P., Abecasis, G.R., personal communication). For all groups, X chromosome imputation was performed using IMPUTE [S48].

[b]Unless otherwise stated, logistic regression analysis was performed using the specified software or the genetic analysis program SNPTEST[S48], PLINK [S49], GenABEL [S50], or ProbABEL [S50].

[c]Imputed only in ERF component of EUROSPAN. [d]DiaGene and Rotterdam (Stage 2) analyzed together.

Meta-analysis



Supplementary Note Figure SN1: Quantile-quantile plots for the Stage 1 meta-analysis (autosomal chromosomes only) showing all data (top plots) and after excluding 17 previously reported signals (bottom panels). Those excluded were *TCF7L2, PPARG, KCNJ11, CDKAL1, CDKN2A, IGF2BP2, FTO, HHEX, SLC30A8, JAZF1, THADA, CDC123, TSPAN8, NOTCH2, ADAMTS9, HNF1B* and *WFS1*. In the left-hand panels, all individual Stage 1 samples were subjected to genome-control correction; in the right panels a second round of post-meta-analysis genomic control correction has been applied.

Evaluation of potential effects of admixture of cases with autoimmune diabetes (either type 1 diabetes or latent autoimmune diabetes of adulthood [LADA]).

As described above, the case collections included in the present study varied in the degree to which they were able to exclude all individuals with diabetes due to causes other than type 2 diabetes. A particular issue arises with autoimmune diabetes arising in later life (often termed LADA) which often results in a clinical diagnosis of type 2 diabetes unless specific measures are taken to obtain a definitive diagnosis. LADA represents a late-onset, more indolent form of type 1 diabetes and a number of studies have shown that variants that have an impact on type 1 diabetes susceptibility are also associated with LADA[S51-54].

To establish the extent of any overlap, and therefore the potential that some of the signals we uncovered are the result of case misclassification, we examined the extent to which variants known to influence susceptibility to T1D and LADA were detected in our stage 1 analysis (**Supplementary Note Table SN2).** The list of loci associated with type 1 diabetes is long, and most of the strongest T1D-association signals (*HLA, INS, PTPN22, CD25*) have also been shown to be associated with LADA[S51-54]. As the effect sizes of these variants for T1D (and LADA) are considerably greater than those of the T2D variants uncovered by DIAGRAM+, presence/absence of these T1D/LADA signals in our T2D meta-analysis should provide a sensitive test of misclassification.

Locus	SNP	Risk allele	Rationale	P-value	OR in DIAGRAM+ Stage 1 analysis	Effective sample size
HLA	rs1270942	G	Tags rs1048709 and strong assoc with T1D	0.28	1.04 (0.97-1.11)	22570
HLA	rs9267992	A	Strong T1D association[S55]	0.001	1.09 (1.03-1.15)	22570
HLA	rs660895	G	HLA-DR4 tag	0.007	1.07 (1.02-1.12)	22570
HLA	rs3129941	G	Strongest T1D association in WTCCC [S56]	0.015	1.06 (1.01-1.11)	22570
HLA	rs926070	A	Strong T1D association [S55]	0.036	1.05 (1.00-1.09)	22570
HLA	rs2187668	T	HLA-DR3 tag	0.75	1.01 (0.95-1.07)	22570
HLA	rs1270942	G	Tags rs1048709 (see text)	0.28	1.04 (0.97-1.11)	22570
HLA	rs630379	A	Tags rs1048709 (see text)	0.18	0.96 (0.91-1.02)	12595
INS	rs3842748	C	Best available tag for INS rs689	0.70	0.99 (0.92-1.06)	13971
PTPN22	rs2476601	A	Strongest signal in ref [S57]	0.45	1.02 (0.96-1.10)	22570
CD25	rs2104286	T	T1D signal in ref [S21]	0.15	0.97 (0.82-1.01)	22570
CTLA4	rs1427676	T	Strongest signal in in ref [S57]	0.75	1.01 (0.97-1.05)	22570

Supplementary Note Table SN2: DIAGRAM+ Stage 1 OR and P-values for variants tagging the strongest signals for T1D (and LADA).

As the table shows, we did find a weak association between several of the SNPs that have the strongest T1D associations and/or which tag classical HLA risk alleles (e.g. rs9267992, P=0.001). However, none of these SNPs was sufficiently associated in DIAGRAM+ Stage 1 to reach the threshold for follow-up. No evidence of T2D association was seen at other loci.

One SNP in the HLA region worth comment is rs1048709: this is a SNP in the HLA Class 2 that did meet our Stage 1 threshold ($p=6x10^{-6}$; **supplementary table 3**) and which showed some evidence of replication, but did not reach our threshold for overall significance (combined P-value ~ $1x10^{-7}$). This variant has not been typed directly in T1D association studies, though imputation analyses conducted by John Todd's group in Cambridge do indicate a strong association with T1D (J. Howson, personal communication). However, imputation within the HLA is not straightforward and neither of the two SNPs that were used to impute rs1048709 in the T1D analyses (rs630379 and rs1270942, the latter even more strongly associated with T1D than rs1048709) has any evidence of a T2D association in Stage 1 analyses (see table above). Consequently we are unable to say whether the (not quite genome wide significant) T2D signal at rs1048709 really reflects autoimmune diabetes misclassification.

All in all, the data in the table indicate weak HLA associations that mirror those seen in autoimmune forms of diabetes, and these would be consistent with a small degree of misclassification in our cases. This is not altogether surprising, since, as the sample descriptions make clear, the extent to which the various case sets had been purged of possible autoimmune cases differed from sample to sample. In some case samples, there was rigorous exclusion of apparent autoimmune diabetes, but in others the relevant phenotypic information (eg GAD antibodies) had never been collected. In many of these, clinical parameters (e.g. truncation on age of onset, BMI and/or clinical therapy) will have had some (unquantifiable) effect on minimizing LADA and T1D admixture. In designing this study, we took the pragmatic view that there was more to be gained in terms of power by including all available cohorts (whether or not they had been able to exclude all autoimmune diabetes) than would be lost by possible admixture. First, we reasoned that the success of T1D genetic discovery (~45 loci described, several with large effects) would allow us to pick out any signals of autoimmune diabetes that emerged from our analysis. Second, we reconsidered the ongoing debate about the exact etiological relationship between LADA and T2D, prompted by recent studies that have shown that LADA subjects have associations with T2D-risk alleles such as *TCF7L2*[refS54], as well as T1D-risk alleles such as HLA and *INS*. This implies that there is no strict diagnostic dividing line between LADA and T2D[S58].

Based on the strong T1D association with HLA (typically single-SNP ORs of 2-3) and the weak effects seen in the Stage 1, we estimate that the extent of any admixture of autoimmune diabetes amongst our T2D cases is at most a few percent. Crucially, this degree of admixture was insufficient to drive even the relatively large HLA signals for autoimmune diabetes to genome wide significance in our study and none of the other T1D/LADA signals (despite allelic ORs far bigger than seen for any T2D signal to date) showed any signal whatsoever. In this context it seems highly unlikely that any of the signals identified by DIAGRAM+ represent loci for autoimmune diabetes that have been missed by the systematic large-scale efforts at gene discovery in T1D.

DIAGRAM+ association signals identified through imputation

As shown in **Supplementary Table 3,** Stage 1 associations for the lead SNPs at six of our new loci were derived wholly or predominantly from imputed data. To ensure that these associations signals were not being spuriously inflated we carried out the following checks (see **Supplementary Note Table SN3**).

- for each of these SNPs, imputation quality was good (all were obliged to have r^2>0.3 or proper_info >0.5 and all were far better than that [at least 0.77]);
- for five of the six SNPs, we had directly typed proxies that showed similar levels of association in Stage 1 data (see figure 2 and table below): four of these five would have met the criteria for Stage 1 follow-up had we not chosen instead the imputed SNP.
- directly genotyped data for the lead SNP were available from many Stage 2 cohorts (typically ~ 30,000 cases and ~40,000 controls), and all of these show strong evidence of directionally-consistent replication in those directly-genotyped samples.

SNP	Locus	Sample size weighted average imputation quality	Proxy SNP	r^2 of proxy to lead SNP	Proxy stage 1 P value	Stage 2 P-value for directly genotyped data only
rs4457053	*ZBED3*	0.777	rs7708285	0.86	$1.0x10^{-7}$	$4.8x10^{-8}$
rs972283	*KLF14*	0.965	rs4731702	1.0	$2.1x10^{-6}$	$1.6x10^{-4}$
rs13292136	*CHCHD9*	0.984	rs10512085	0.88	$7.2x10^{-6}$	$1.9x10^{-4}$
rs231362	*KCNQ1*	0.787	None	NA	NA	$1.2x10^{-8}$
rs7957197	*HNF1A*	0.952	rs7965349	0.82	$1.5x10^{-6}$	$1.7x10^{-4}$
rs11634397	*ZFAND6*	0.884	rs4778582	0.90	$1.3x10^{-5}$	$2.5x10^{-5}$

Supplementary Note Table SN3: Further information on the 6 lead SNPs for which the Stage 1 data was predominantly or exclusively based on imputation.

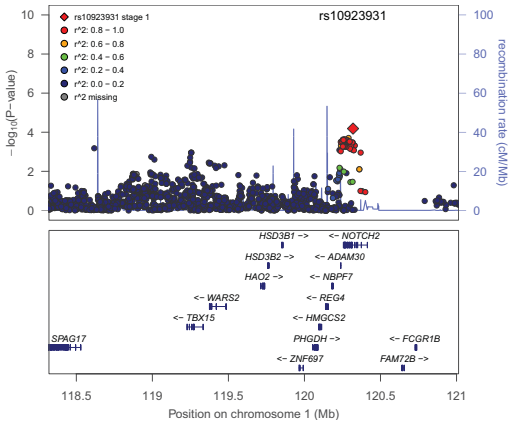
All in all, we see no evidence to suggest that the use of imputed data has led to false positive claims of association at any of the new loci. All of the new loci have substantial confirmation based on direct Stage 2 genotyping and/or directly typed Stage 1 proxies. It is notable that for the two loci with the least convincing imputation quality (rs4457053; rs231362) the directly typed Stage 2 samples are genome-wide significant in their own right, suggesting that if anything, the use of imputed data in Stage 1 has led to an underestimate of effect size rather than spurious associations.

Supplementary Note Figure SN2: Locus plots for previously-reported and DIAGRAM+ genome-wide significant signals. Stage 1 meta-analysis data only. All loci shown as 2Mb either side of lead SNP. Lead SNP denoted with red diamond: colors of other SNPs denote LD relationships with the lead SNP based on CEU HapMap information (see key). Recombination rates in CEU HapMap shown in blue traces (right-hand axis). The names of some genes are not given due to reasons of space.

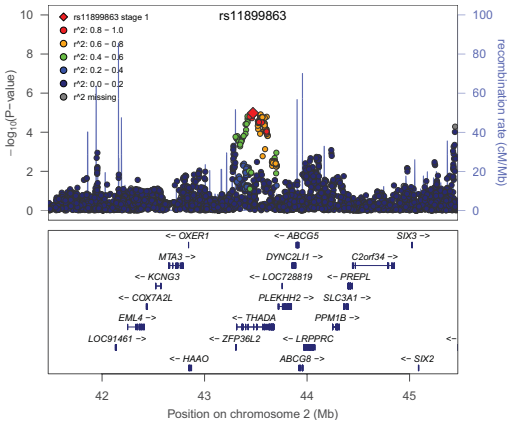
See following three pages for plots

Nature Genetics: doi:10.1038/ng.609

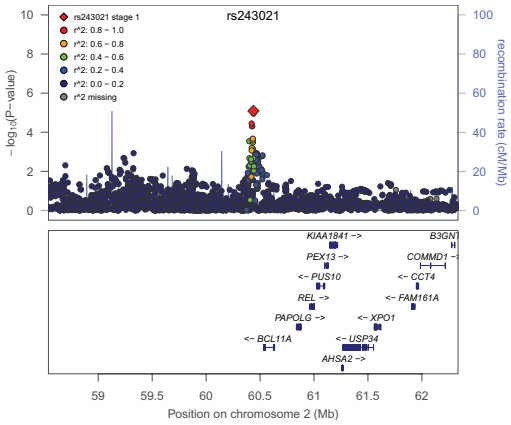
NOTCH2 Region



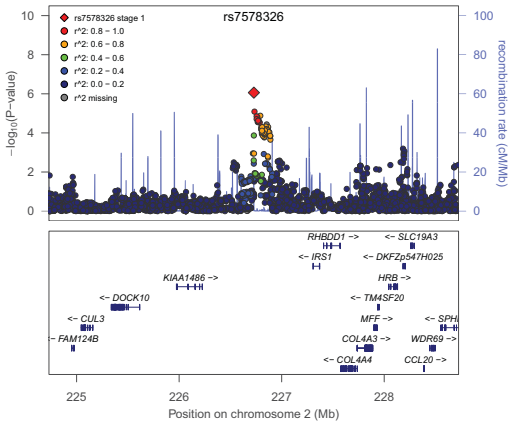
THADA Region



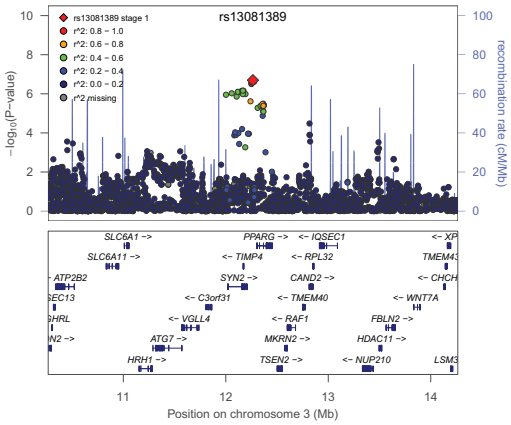
BCL11A Region



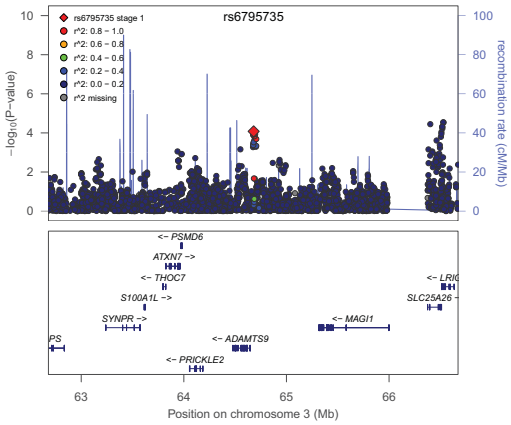
IRS1 Region



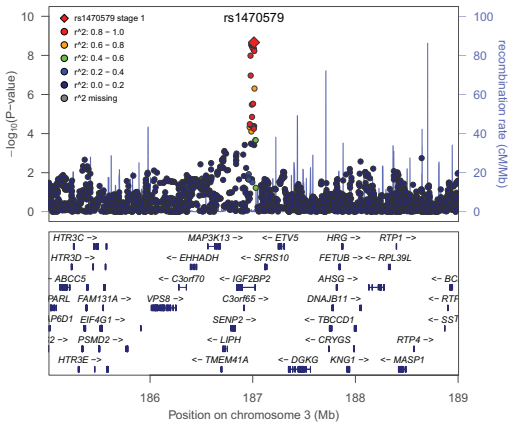
PPARG Region



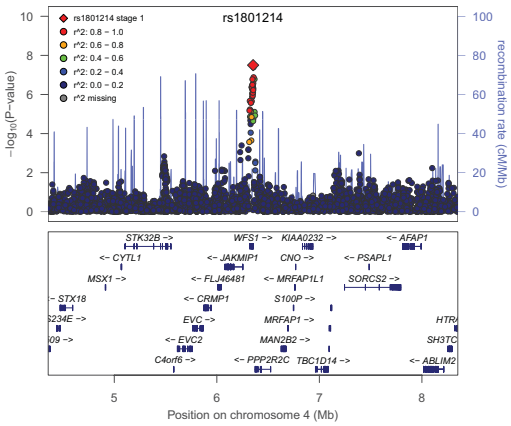
ADAMTS9 Region



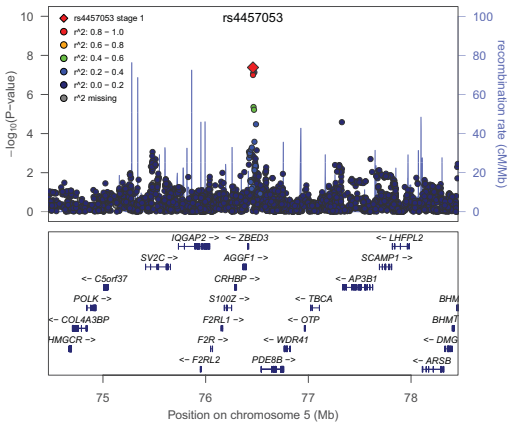
IGF2BP2 Region



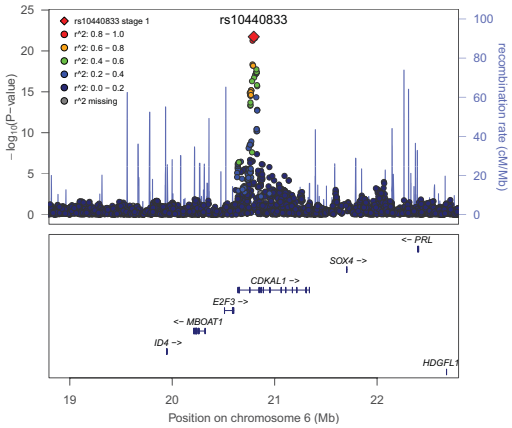
WFS1 Region



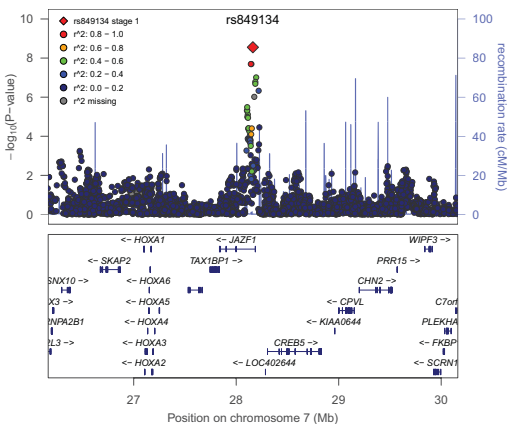
ZBED3 Region



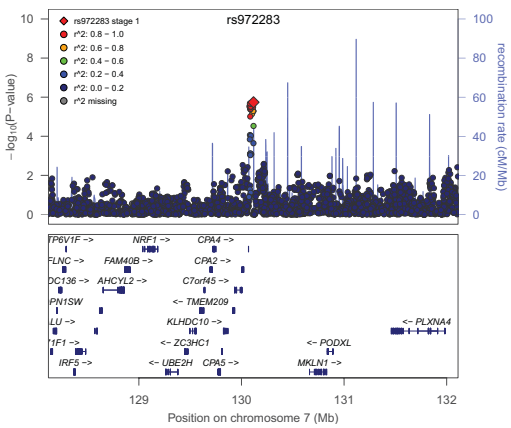
CDKAL1 Region



JAZF1 Region

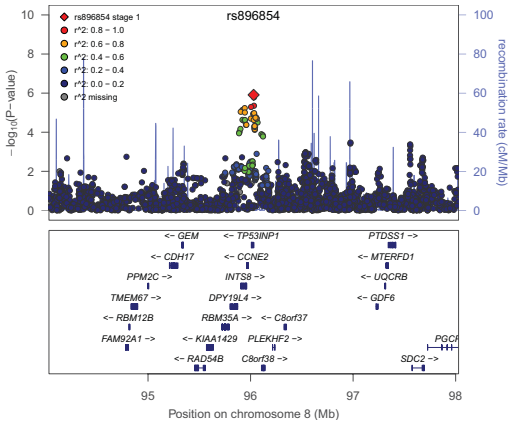


KLF14 Region

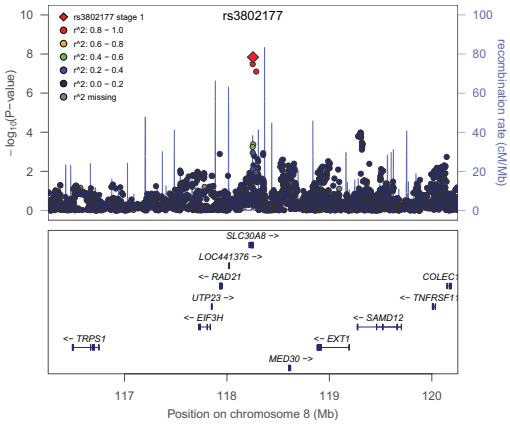


Nature Genetics: doi:10.1038/ng.609

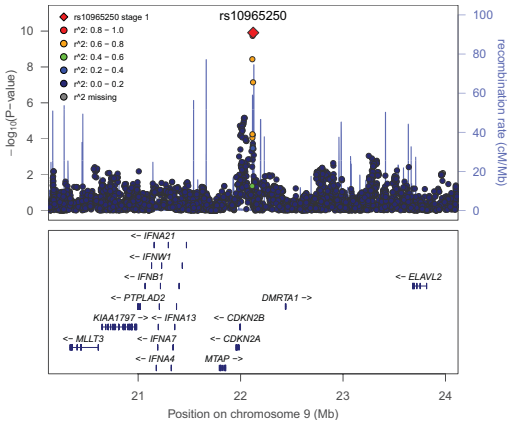
TP53INP1 Region



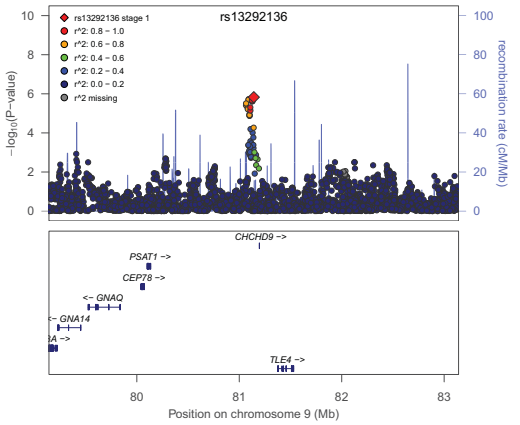
SLC30A8 Region



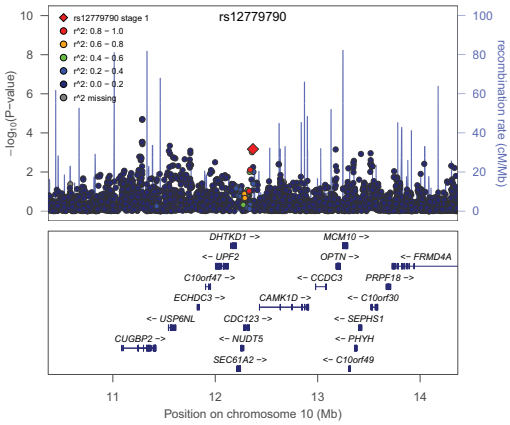
CDKN2A/B Region



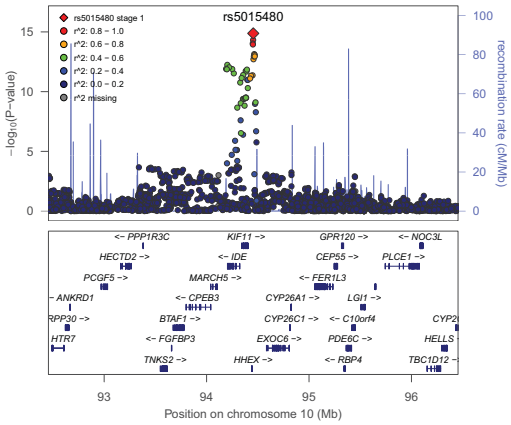
CHCHD9 Region



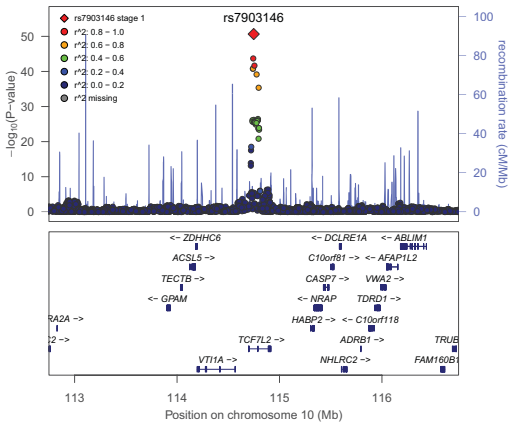
CDC123/CAMK1D Region



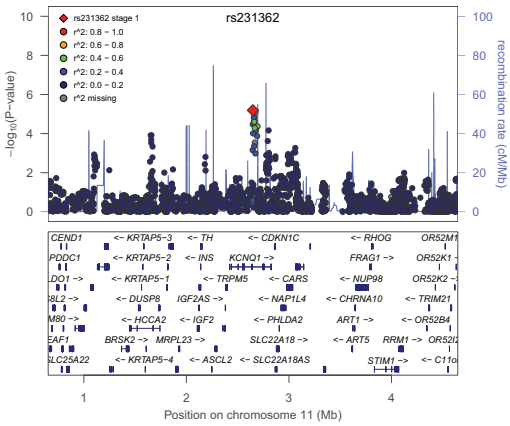
HHEX/IDE Region



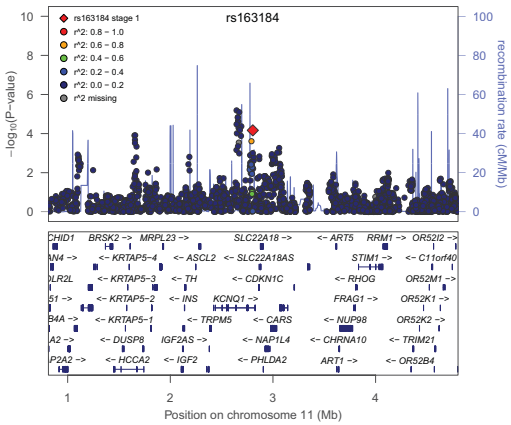
TCF7L2 Region



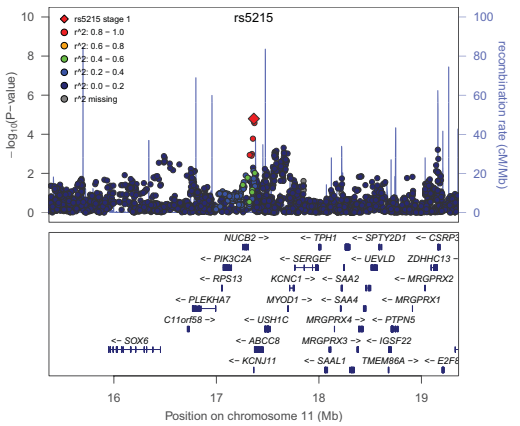
KCNQ1 Region



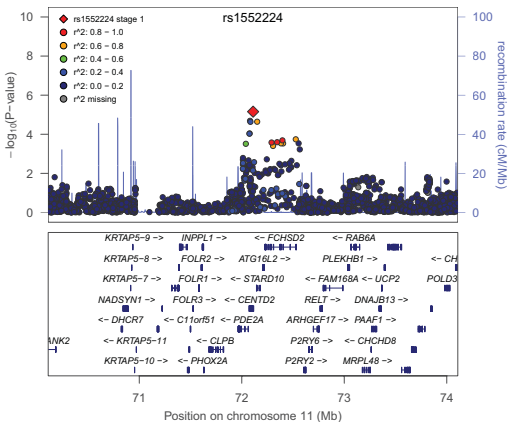
KCNQ1 Region



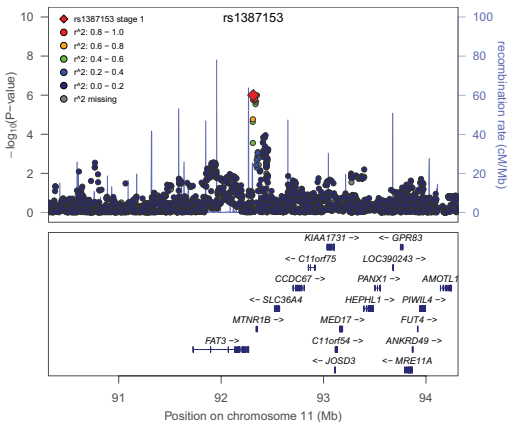
KCNJ11 Region



CENTD2 Region

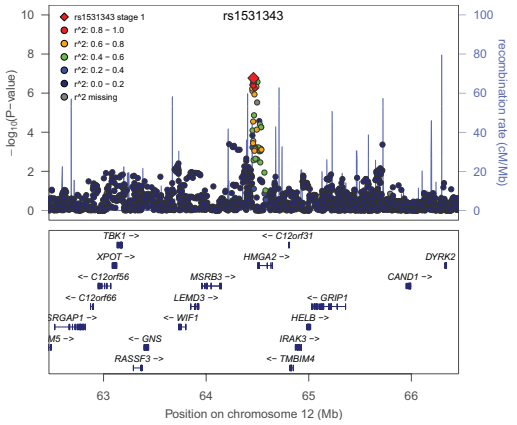


MTNR1B Region

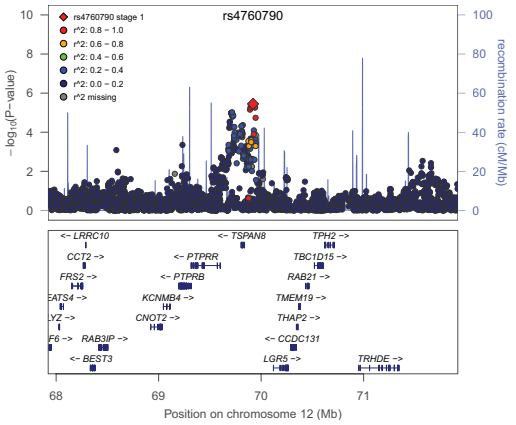


Nature Genetics: doi:10.1038/ng.609

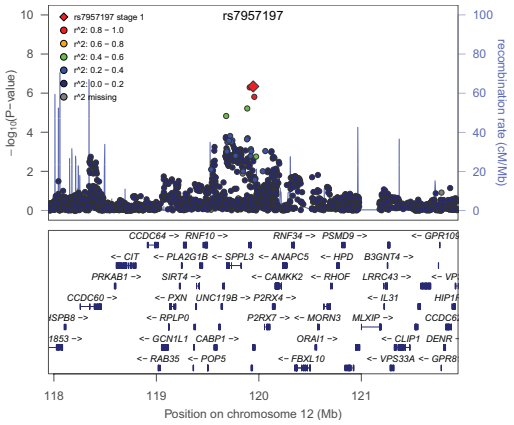
HMGA2 Region



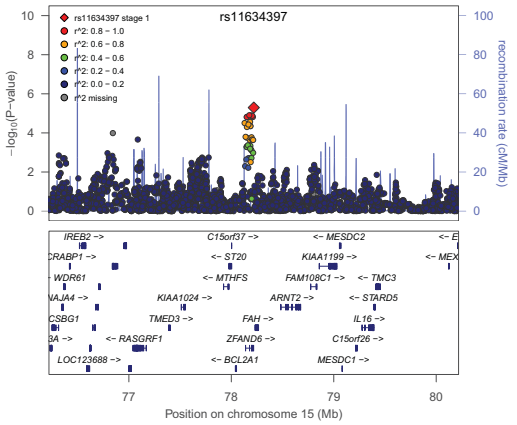
TSPAN8/LGR5 Region



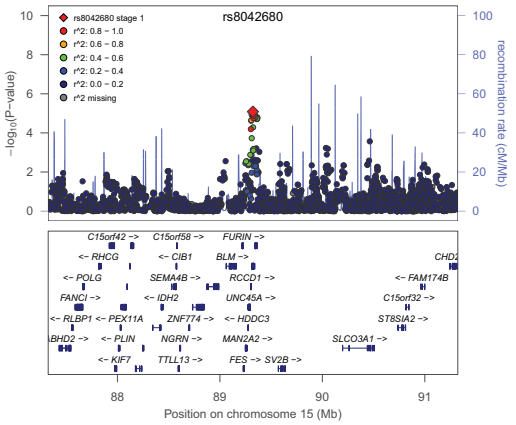
HNF1A Region



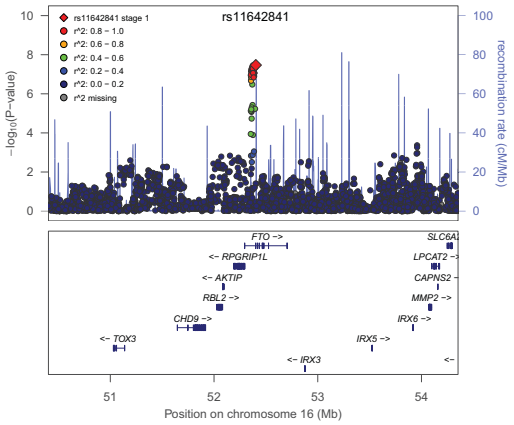
ZFAND6 Region



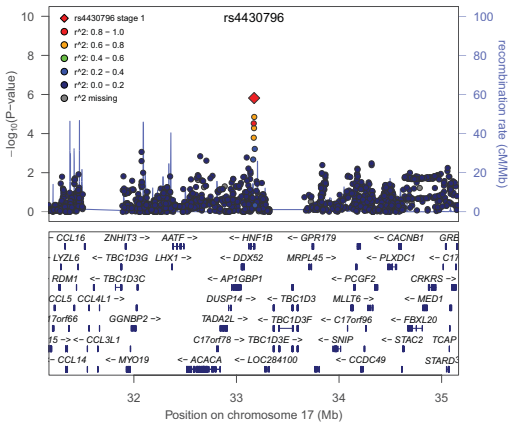
PRC1 Region



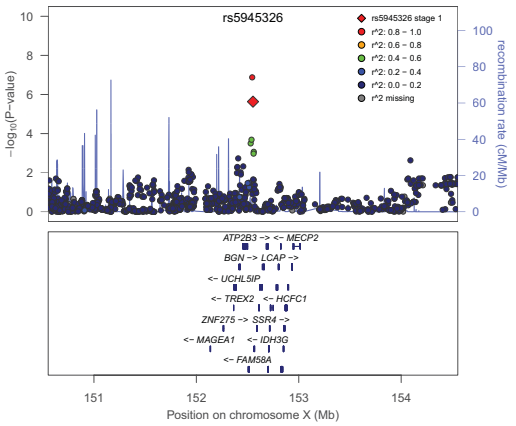
FTO Region



HNF1B(TCF2) Region



DUSP9 Region



							Stage 1					Stage2					Stages 1 and 2 combined				
SNP	Chr	Position B36 (basepair)	Risk Allele[c]	Non-risk Allele[c]	Frequency Risk Allele [Hapmap CEU]	Nearby Gene[e]	P-value[f]	Cases (N)	Controls (N)	I^2 (95% CI)	P_{het}-value	P-value[f]	Cases (N)	Controls (N)	I^2 (95% CI)	P_{het}-value	P-value[f]	OR	95% CI	Cases (N)	Controls (N)
Achieving genome wide significance in joint analysis																					
rs243021	2	60,438,323	A	G	0.46	*BCL11A*	8.1×10^{-6}	8,130	38,987	0.11 (0 - 0.71)	0.34	6.2×10^{-11}	27,340	51,999	0.49 (0.11 - 0.71)	0.012	2.9×10^{-15}	1.08	(1.06-1.10)	35,470	90,986
rs7578326	2	226,728,897	A	G	0.64	*IRS1*	8.7×10^{-7}	8,130	38,987	0 (0 - 0.67)	0.45	2.2×10^{-15}	29,299	53,310	0.24 (0 - 0.57)	0.18	5.4×10^{-20}	1.11	(1.08-1.13)	37,429	92,297
rs4457053	5	76,460,705	G	A	0.26	*ZBED3*	4.2×10^{-8}	7,108	37,912	0.20 (0 - 0.63)	0.28	2.7×10^{-7}	29,292	48,166	0 (0 - 0.42)	0.67	2.8×10^{-12}	1.08	(1.06-1.11)	36,400	86,078
rs972283	7	130,117,394	G	A	0.55	*KLF14*	1.8×10^{-6}	8,130	38,987	0 (0 - 0.34)	0.84	6.4×10^{-6}	25,281	42,018	0.16 (0 - 0.54)	0.28	2.2×10^{-10}	1.07	(1.05-1.10)	33,411	81,005
rs896854	8	96,029,687	T	C	0.48	*TP53INP1*	1.2×10^{-6}	8,130	38,987	0 (0 - 0.31)	0.86	2.2×10^{-5}	28,023	51,144	0 (0 - 0.55)	0.46	9.9×10^{-10}	1.06	(1.04-1.09)	36,153	90,131
rs13292136	9	81,141,948	C	T	0.93	*CHCHD9*	1.5×10^{-6}	8,130	38,987	0 (0 - 0.57)	0.62	2.4×10^{-4}	34,412	59,925	0 (0 - 0.36)	0.70	2.8×10^{-8}	1.11	(1.07-1.15)	42,542	98,912
rs231362	11	2,648,047	G	A	0.52	*KCNQ1*	6.4×10^{-6}	8,130	38,987	0 (0 - 0.32)	0.85	3.2×10^{-9}	32,456	54,275	0.14 (0 - 0.5)	0.30	2.8×10^{-13}	1.08	(1.06-1.10)	40,586	93,262
rs1552224	11	72,110,746	A	C	0.88	*CENTD2*	7.0×10^{-6}	8,130	38,987	0 (0 - 0.67)	0.45	3.2×10^{-18}	34,412	59,925	0.21 (0 - 0.55)	0.20	1.4×10^{-22}	1.14	(1.11-1.17)	42,542	98,912
rs1387153[a]	11	92,313,476	T	C	0.28	*MTNR1B*	1.0×10^{-6}	8,130	38,987	0 (0 - 0.65)	0.49	4.4×10^{-10}	34,412	59,925	0.14 (0 - 0.5)	0.10	7.8×10^{-15}	1.09	(1.06-1.11)	42,542	98,912
rs1531343	12	64,461,161	C	G	0.10	*HMGA2*	1.7×10^{-7}	8,130	38,987	0 (0 - 0.56)	0.65	1.1×10^{-4}	29,724	53,381	0.23 (0 - 0.57)	0.28	3.6×10^{-9}	1.10	(1.07-1.14)	37,854	92,368
rs7957197	12	119,945,069	T	A	0.85	*HNF1A*	4.6×10^{-7}	8,130	38,987	0 (0 - 0.56)	0.64	4.6×10^{-4}	30,032	49,788	0.13 (0 - 0.51)	0.31	2.4×10^{-8}	1.07	(1.05-1.10)	38,162	88,775
rs11634397	15	78,219,277	G	A	0.60	*ZFAND6*	5.1×10^{-6}	8,130	38,987	0 (0 - 0.59)	0.60	1.2×10^{-5}	34,412	59,925	0.3 (0 - 0.6)	0.10	2.4×10^{-9}	1.06	(1.04-1.08)	42,542	98,912
rs8042680	15	89,322,341	A	C	0.22	*PRC1*	8.2×10^{-6}	8,130	38,987	0.16 (0 - 0.59)	0.31	1.6×10^{-6}	34,412	59,925	0.52 (0.19 - 0.72)	0.01	2.4×10^{-10}	1.07	(1.05-1.09)	42,542	98,912
rs5945326	X	152,553,116	A	G	0.21	*DUSP9*	2.3×10^{-6}	8,130	38,987	0.77 (0.70 - 0.83)	0.0002	2.3×10^{-5}	8,535	12,326	0.50 (0.13 - 0.71)	0.11	3.0×10^{-10}	1.27	(1.18-1.37)	16,665	51,313
Failing to reach 5×10^{-8} on joint analysis																					
rs6709268	2	3,058,146	A	C	0.11	*TSSC1*	8.6×10^{-6}	8,130	38,987	0 (0 - 0.52)	0.70	0.80	30,032	49,788	0 (0 - 0.28)	0.83	0.02	1.04	(1.01-1.07)	38,162	88,775
rs6442037	3	46,904,550	A	G	0.65	*PTH1R*	4.7×10^{-6}	8,130	38,987	0 (0 - 0.001)	0.94	0.27	34,412	59,925	0.41 (0 - 0.66)	0.03	0.0022	1.03	(1.01-1.05)	42,542	98,912
rs7674212	4	104,208,348	G	T	0.57	*NHEDC2*	1.7×10^{-7}	8,130	38,987	0.35 (0 - 0.71)	0.15	0.30	21,889	39,568	0.38 (0 - 0.69)	0.10	1.8×10^{-4}	1.04	(1.02-1.07)	30,019	78,555
rs1048709	6	32,022,914	G	A	0.75	*CFB*	6.0×10^{-6}	8,130	38,987	0 (0 - 0.64)	0.50	4.0×10^{-4}	29,983	54,591	0.05 (0 - 0.55)	0.40	1.1×10^{-7}	1.07	(1.04-1.10)	38,113	93,578
rs2789681[b]	10	81,898,110	C	G	0.09	*ANXA11*	2.8×10^{-6}	7,862	35,277	0.38 (0 - 0.74)	0.14	0.997	27,212	49,518	0.17 (0 - 0.52)	0.20	0.020	1.04	(1.01-1.08)	35,074	84,795
rs960078	11	10,380,683	A	T	(0.98)[d]	*AMPD3*	2.7×10^{-6}	3,379	5,451	0.34 (0 - 0.78)	0.22	0.992[g]	25,148	43,092	0.19 (0 - 0.57)	0.25	0.089	1.08	(0.99-1.19)	28,527	48,543
rs7118472	11	86,833,159	A	G	0.05	*TMEM135*	6.2×10^{-6}	6,665	15,793	0 (0 - 0.30)	0.87	0.78[g]	30,032	49,788	0 (0 - 0.13)	0.92	0.057	1.04	(1.00-1.08)	36,697	65,581
rs2288232	12	27,836,125	A	G	0.63	*KLHDC5*	3.9×10^{-6}	8,130	38,987	0.19 (0 - 0.62)	0.28	8.5×10^{-4}	33,721	53,500	0.17 (0 - 0.52)	0.25	5.4×10^{-7}	1.05	(1.03-1.07)	41,851	92,487
rs17795982	14	24,602,523	T	C	0.87	*STXBP6*	1.7×10^{-6}	8,130	38,987	0 (0 - 0.54)	0.67	0.13	31,293	55,549	0.02 (0 - 0.53)	0.43	2.7×10^{-4}	1.05	(1.02-1.08)	39,423	94,536
rs8057749	16	77,015,972	C	G	0.02	*WWOX*	9.4×10^{-6}	3,076	29,393	0 (0 - 0.79)	0.61	0.051	26,923	51,358	0.07 (0 - 0.43)	0.37	0.0014	1.14	(1.05-1.23)	29,999	80,751

Supplementary Note Table SN4: Detailed Stage 1 and Stage 2 results for the 24 SNPs with Stage 2 follow-up.

[a] Proxy SNP rs2612067 was genotyped in METSIM, FUSION2 and HUNT Stage 2 samples; [b] Formerly rs7088994; [c] Alleles are indexed to the forward strand of NCBI Build 36; [d] rs960078 is reported as monomorphic in HapMap CEU but has frequency of 0.98 in our study; [e] Gene regions are named for the nearest gene (except where there is a very strong positional candidate); [f] All P-values are reported two-sided and based on an inverse-variance weighted meta-analysis model (fixed effects); [g] Stage 2 effect size estimate is in opposite direction to Stage 1 for these two SNPs.

OR, Odd Ratio; CI, Confidence Interval. Lower sample sizes for some SNPs simply reflect variable performance in stage 1 imputation or stage 2 typing.

SNP	Chr	Position B36 (basepair)	Risk Allele[b]	Nonrisk Allele[b]	Frequency Risk Allele [Hapmap CEU]	Nearby Gene[c]	Stage 1 [2nd GC] OR (95%CI)	Stage 1 [2nd GC] P-value[d]	Stage 2 OR (95%CI)	Stage 2 P-value[d]	Stage 1 [2nd GC] + 2 OR (95%CI)	Stage 1 [2nd GC] + 2 P-value[d]
							up to 8,130 cases and 38,987 controls		up to 34,412 cases and 59,925 controls		up to 42,542 cases and 98,912 controls	
Novel T2D-susceptibility loci												
rs243021	2	60,438,323	G	A	0.46	*BCL11A*	1.09 (1.05-1.14)	1.6×10^{-5}	1.08 (1.06-1.10)	6.2×10^{-11}	1.08 (1.06-1.10)	5.8×10^{-15}
rs4457053	5	76,460,705	A	G	0.26	*ZBED3*	1.16 (1.10-1.23)	1.1×10^{-7}	1.07 (1.04-1.10)	2.8×10^{-7}	1.08 (1.06-1.11)	6.1×10^{-12}
rs972283	7	130,117,394	A	G	0.55	*KLF14*	1.10 (1.06-1.15)	3.9×10^{-6}	1.06 (1.03-1.09)	6.4×10^{-6}	1.07 (1.05-1.10)	4.4×10^{-10}
rs896854	8	96,029,687	C	T	0.48	*TP53INP1*	1.10 (1.06-1.15)	2.7×10^{-6}	1.05 (1.03-1.08)	2.2×10^{-5}	1.06 (1.04-1.09)	2.0×10^{-9}
rs13292136	9	81,141,948	T	C	0.93	*CHCHD9*	1.20 (1.11-1.30)	3.3×10^{-6}	1.08 (1.04-1.13)	2.4×10^{-4}	1.11 (1.07-1.15)	5.2×10^{-8}
rs231362	11	2,648,047	A	G	0.52	*KCNQ1*	1.11 (1.06-1.16)	1.3×10^{-5}	1.07 (1.05-1.09)	3.2×10^{-9}	1.08 (1.06-1.10)	4.7×10^{-13}
rs1552224	11	72,110,746	C	A	0.88	*CENTD2*	1.13 (1.07-1.20)	1.4×10^{-5}	1.14 (1.11-1.18)	3.2×10^{-18}	1.14 (1.11-1.17)	2.9×10^{-22}
rs1531343[a]	12	64,461,161	G	C	0.10	*HMGA2*	1.20 (1.12-1.29)	4.3×10^{-7}	1.08 (1.04-1.12)	1.1×10^{-4}	1.10 (1.07-1.14)	7.2×10^{-9}
rs7957197	12	119,945,069	A	T	0.85	*HNF1A*	1.14 (1.08-1.20)	1.0×10^{-6}	1.05 (1.02-1.09)	4.6×10^{-4}	1.07 (1.05-1.10)	4.9×10^{-8}
rs11634397	15	78,219,277	A	G	0.60	*ZFAND6*	1.11 (1.06-1.16)	1.0×10^{-5}	1.05 (1.03-1.08)	1.2×10^{-5}	1.06 (1.04-1.08)	4.5×10^{-9}
rs8042680	15	89,322,341	C	A	0.22	*PRC1*	1.10 (1.05-1.15)	1.6×10^{-5}	1.06 (1.03-1.08)	1.6×10^{-6}	1.07 (1.05-1.09)	4.1×10^{-10}
Previously Known												
rs7578326	2	226,728,897	G	A	0.64	*IRS1*	1.12 (1.07-1.17)	2.0×10^{-6}	1.10 (1.08-1.13)	2.2×10^{-15}	1.11 (1.08-1.13)	1.3×10^{-19}
rs1387153	11	92,313,476	C	T	0.28	*MTNR1B*	1.12 (1.07-1.17)	2.2×10^{-6}	1.08 (1.05-1.10)	4.4×10^{-10}	1.09 (1.06-1.11)	1.6×10^{-14}

Supplementary Note Table SN5: Association results for Stage 1 and 2, for those SNPs which exceeded genome-wide significance threshold (overall $P<5 \times 10^{-8}$ on single-GC correction), but here showing double GC-corrected P-values. Genomic control correction for X chromosome signals was not performed so the *DUSP9* (ChrX) SNP is not included

[a] Proxy SNP rs2612067 was genotyped in METSIM, FUSION2 and HUNT Stage 2 samples
[b] Alleles are indexed to the forward strand of NCBI Build 36
[c] Gene regions are named for the nearest gene (except where there is a very strong positional candidate)
[d] All P-values are reported two-sided and based on an inverse-variance weighted meta-analysis model (fixed effects)
OR: odds ratio; CI: confidence interval.

CNV analyses

Additional methods: Given interest in the potential roles of common copy number variants with respect to T2D risk, we re-examined the Stage 1 meta-analysis data looking for evidence of association with SNPs known to provide robust, high-LD tags for (autosomal) CNVs in European-descent samples. We combined CNV-tagging SNPs from four inventories:

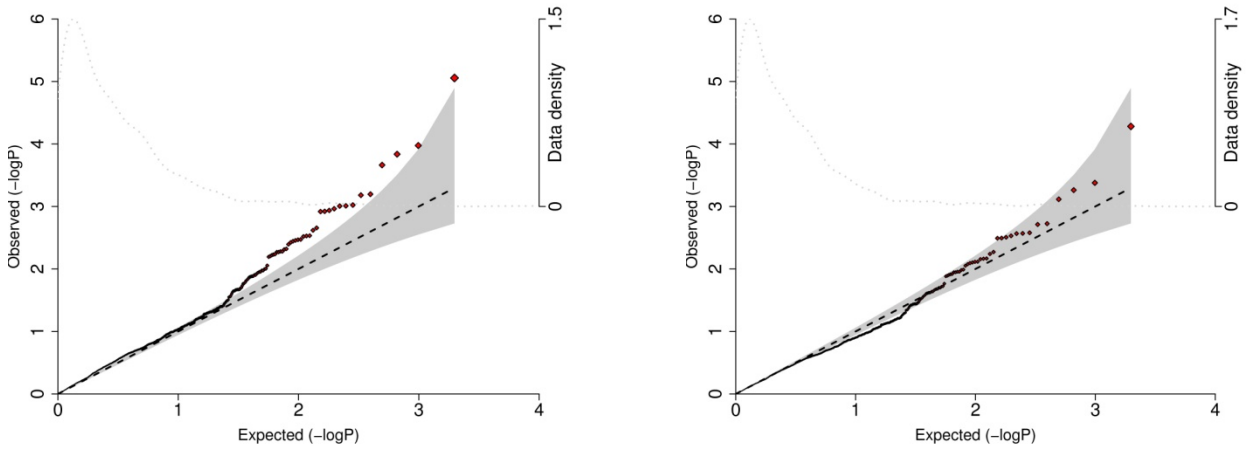
- A list of 261 CNV-tagging SNPs (r^2>0.8) generated by Steve McCarroll and colleagues at the Broad based on typing HapMap samples on the Affymetrix 6.0 array [S59];
- A list of 2174 multiethnic CNV-tagging SNPs (r^2>0.8, of which 1168 are polymorphic in CEU) recently made available by the Genomic Structural Variation consortium and based largely on typing 450 HapMap samples on a bespoke Agilent 105K array capable of genotyping ~3320 CNPs in CEU [S60];
- A list of 3113 multiethnic CNV-tagging SNPs, generated on HapMap 3 samples using Affymetrix 6.0 and Illumina 1M arrays, generated by the HapMap 3 project (www.hapmap.org);
- A list of 2905 CNV-tagging SNPs, generated on a custom Agilent 105K array (as per Conrad *et al*, 2009 [S60]), but using ~19,000 samples (all European-descent, 3000 controls and 2000 cases for each of 8 diseases) typed by the Wellcome Trust Case Control Consortium[S61].

The union of these lists provided a total of 5219 unique CNV-tagging SNPs for which we had DIAGRAM+ data in at least 17,000 Stage 1 samples. This list of SNPs was not further "pruned" for mutual LD and will contain some instances where multiple SNPs that are tagging the same CNV.

Additional Results: Quantile-quantile plots for this combined set of CNV-tagging SNPs are shown in **Figure SN3,** and the top hits (those where the Stage 1 association P-value for the tagged SNP was $<7\times10^{-4}$) are listed in **Table SN6.**

Apart from the HLA association (which is hard to interpret given the extent and complexity of the MHC region), the only notable tag-SNP association is that with SNPs in the *TSPAN8* region. SNPs in this region have been shown to be conclusively associated with T2D [S7] and this tag-SNP association with existing published GWAS hits was noted in the GSV paper [S60]. However, comparison of the P-values obtained for the CNV-tagging SNPs (rs1705261, rs1355371) with those for the best SNP-association observed in the region in the same Stage 1 data (rs4760790, $P=3.6\times10^{-6}$, OR 1.11 (1.06-1.16): see **Supplementary Table 3**) suggests that the CNV is unlikely to be causal.

We conclude on the basis of this systematic though incomplete survey of common CNVs (estimated to cover at least 40% of common CNVs >1kb in size) [S60,S61] that common CNVs are unlikely to make a substantial contribution to T2D-susceptibility.



Supplementary Note Figure SN3: QQ plots for the 4 combined CNV-tagging SNP inventories (n=5219 unique SNPs). The left panel is without genomic control correction whereas the right one is GC corrected. The grey band in both plots is the confidence band indicating the range of results consistent with a 95% interval around the null.

SNP	Chr	Position B36 (basepair)	Risk Allele	Nonrisk Allele	P-value	OR (95%CI)	Nearby Genes
rs9270986	6	32,682,038	C	A	8.8×10^{-6}	1.129(1.070-1.091)	*C6orf10, BTNL2, HLA-DRA, HLA-DRB5, HLA-DRB1, HLA-DQA1, HLA-DQB1, HLA-DQA2, HLA-OB, TAP2, PSMB8, TAP1, PSMB9*
rs930668	4	61,833,099	A	G	6.4×10^{-4}	1.224(1.090-1.374)	*LPHN3*
rs701831	6	32,657,379	C	T	1.1×10^{-4}	1.115(1.055-1.178)	*C6orf10, BTNL2, HLA-DRA, HLA-DRB5, HLA-DRB1, HLA-DQA1, HLA-DQB1, HLA-DQA2, HLA-DOB, TAP2*
rs1705261	12	69,813,913	T	A	1.5×10^{-4}	1.082(1.039-1.128)	*PTPRR, TSPAN8*
rs1355371	12	69,899,139	T	C	2.2×10^{-4}	1.079(1.036-1.123)	*TSPAN8, LGR5*
rs2596568	6	31,458,211	G	A	5.7×10^{-4}	1.084(1.036-1.135)	*PSORS1C1, CCHCR1, TCF19, POU5F1, HCG27, HLA-C, HLA-B, HCP5, MICB, BAT1, ATP6V1G2, NFKBIL1, LTA, TNF, LTB, LST1, NCR3, AIF1, BAT2*
rs3117117	6	32,429,250	A	C	6.4×10^{-4}	1.099(1.041-1.161)	*CREBL1, FKBPL, PRRT1, PPT2, EGFL8, AGPAT1, RNF5, AGER, PBX2, GPSM3, NOTCH4, C6orf10, BTNL2, HLA-DRA, HLA-DRB5, HLA-DRB1*
rs2101247	3	47,469,307	G	A	6.5×10^{-4}	1.073(1.031-1.117)	*KIF9, KLHL18, PTPN23, SCAP, TMEM103, CSPG5, SMARCC1*
rs6792461	3	47,448,172	C	T	6.6×10^{-4}	1.073(1.030-1.117)	*KIF9, KLHL18, PTPN23, SCAP, TMEM103, CSPG5, SMARCC1*

Supplementary Note Table SN6: Most significant associations seen in DIAGRAM+ Stage 1 meta-analysis for CNV tagging SNPs. All those with $P<7x10^{-4}$ are shown here.

Conditional analyses

SNP	Chr	Nearby Gene
rs10923931	1	*NOTCH2*
rs11899863	2	*THADA*
rs243021	2	*BCL11A*
rs7578326	2	*IRS1*
rs13081389	3	*PPARG*
rs6795735	3	*ADAMTS9*
rs1470579	3	*IGF2BP2*
rs1801214	4	*WFS1*
rs4457053	5	*ZBED3*
rs10440833	6	*CDKAL1*
rs849134	7	*JAZF1*
rs972283	7	*KLF14*
rs896854	8	*TP53INP1*
rs3802177	8	*SLC30A8*
rs10965250	9	*CDKN2A/B*
rs13292136	9	*CHCHD9*
rs12779790	10	*CDC123/CAMK1D*
rs5015480	10	*HHEX/IDE*
rs7903146	10	*TCF7L2*
rs231362	11	*KCNQ1*
rs5215	11	*KCNJ11*
rs1552224	11	*CENTD2*
rs1387153	11	*MTNR1B*
rs1531343	12	*HMGA2*
rs4760790	12	*TSPAN8*
rs7957197	12	*HNF1A*
rs11634397	15	*ZFAND6*
rs8042680	15	*PRC1*
rs11642841	16	*FTO*
rs4430796	17	*HNF1B*

Supplementary Note Table SN7: Markers (n=30) used for the conditional analyses reported in this paper. Conditional analyses did not include the *DUSP9* X-chromosomal signal (since this was identified some time after the other loci) nor the previously-described signal at *KCNQ1* (to allow us to detemine if it was independent of our newly identified signal), nor other contemporaneous reports of further T2D signals from the MAGIC consortium[S62,S63] and deCODE[S64] genetics.

Heterogeneity analyses

Study	T2D cases AOD < 45 years of age	T2D cases AOD ≥ 45 years of age	T2D cases for continuous AOD
deCODE	111	641	752
DGDG	353	381	734
DGI	167	1,236	1,403
FUSION	157	630	787
KORAgen	NA	NA	356
Rotterdam	NA	NA	1,148
WTCCC	529	1,395	1,924
Total	1,317	4,283	7,104

NA: Not analyzed

Study	Obese (BMI > 30kgm^{-2}) Cases	Obese (BMI > 30kgm^{-2}) Controls	Non-Obese (BMI ≤ 30kgm^{-2}) Cases	Non-Obese (BMI ≤ 30kgm^{-2}) Controls
deCODE	625	5,064	840	18,130
DGDG	NA	NA	697	625
DGI	225	295	719	858
FUSION	529	242	564	930
KORAgen	381	219	210	1,051
Rotterdam	NA	NA	NA	NA
WTCCC	1,011	228	899	1,194
EUROSPAN	106	716	119	2,308
Total	2,877	6,764	4,048	25,096

NA: Not analyzed

Supplementary Note Table SN8: Numbers of subjects from each Stage 1 study included in the AOD- and BMI-stratified analyses (top and bottom panels respectively)

SNP	Chr	Nearby Gene[a]	Non-Obese OR (95% CI)	Non-Obese P-value	Obese OR (95% CI)	Obese P-value	Non-obese OR > obese OR?	P(Het)
rs10923931	1	*NOTCH2*	1.17 (1.09-1.26)	3.4×10^{-4}	1.13 (1.00-1.26)	0.07	Yes	0.619
rs11899863	2	*THADA*	1.23 (1.09-1.41)	8.42×10^{-5}	1.14 (0.97-1.38)	0.093	Yes	0.489
rs243021	**2**	***BCL11A***	**1.13 (1.08-1.19)**	**1.03×10^{-5}**	**1.01 (0.93-1.09)**	**0.84**	**Yes**	**0.016**
rs7578326	2	*IRS1*	1.12 (1.05-1.18)	5.29×10^{-4}	1.19 (1.10-1.28)	1.4×10^{-4}	No	0.192
rs13081389	3	*PPARG*	1.31 (1.19-1.43)	1.48×10^{-5}	1.18 (1.01-1.36)	0.06	Yes	0.267
rs6795735	3	*ADAMTS9*	1.10 (1.03-1.17)	0.0013	1.07 (0.98-1.18)	0.11	Yes	0.676
rs1470579	3	*IGF2BP2*	1.13 (1.06-1.21)	5.22×10^{-5}	1.17 (1.06-1.30)	7.78×10^{-4}	No	0.628
rs1801214	4	*WFS1*	1.16 (1.10-1.22)	7.42×10^{-7}	1.06 (0.97-1.15)	0.19	Yes	0.06
rs4457053	5	*ZBED3*	1.18 (1.09-1.30)	1.3×10^{-5}	1.09 (0.97-1.23)	0.12	Yes	0.251
rs10440833	6	*CDKAL1*	1.26 (1.20-1.32)	2.09×10^{-13}	1.21 (1.11-1.30)	8.12×10^{-5}	Yes	0.362
rs849134	7	*JAZF1*	1.14 (1.08-1.19)	9.38×10^{-6}	1.10 (1.01-1.18)	0.033	Yes	0.424
rs972283	7	*KLF14*	1.12 (1.05-1.19)	1.41×10^{-4}	1.11 (1.02-1.22)	0.013	Yes	0.921
rs896854	8	*TP53INP1*	1.12 (1.06-1.17)	9.86×10^{-5}	1.12 (1.03-1.20)	0.0084	No	0.993
rs3802177	8	*SLC30A8*	1.16 (1.07-1.27)	7.42×10^{-5}	1.22 (1.08-1.4)0	2.70×10^{-4}	No	0.536
rs10965250	9	*CDKN2A/B*	1.31 (1.18-1.46)	4.96×10^{-11}	1.18 (1.04-1.37)	0.0051	Yes	0.246
rs13292136	9	*CHCHD9*	1.29 (1.13-1.50)	3.74×10^{-6}	1.13 (0.96-1.36)	0.13	Yes	0.23
rs12779790	10	*CDC123/CAMK1D*	1.14 (1.05-1.24)	3.81×10^{-4}	1.05 (0.95-1.18)	0.31	Yes	0.275
rs5015480	10	*HHEX/IDE*	1.20 (1.12-1.29)	3.5×10^{-10}	1.16 (1.05-1.28)	6.68×10^{-4}	Yes	0.515
rs7903146	**10**	***TCF7L2***	**1.55 (1.49-1.61)**	**9.72×10^{-45}**	**1.34 (1.25-1.43)**	**4.34×10^{-10}**	**Yes**	**< 0.001**
rs231362	11	*KCNQ1*	1.08 (1.01-1.16)	0.022	1.05 (0.96-1.16)	0.32	Yes	0.637
rs5215	11	*KCNJ11*	1.14 (1.07-1.22)	7.15×10^{-6}	1.08 (0.99-1.19)	0.063	Yes	0.366
rs1552224	11	*CENTD2*	1.22 (1.14-1.30)	5.74×10^{-7}	1.12 (1.01-1.23)	0.052	Yes	0.144
rs1387153	11	*MTNR1B*	1.13 (1.06-1.19)	3.05×10^{-4}	1.13 (1.03-1.22)	0.013	No	0.994
rs1531343	12	*HMGA2*	1.25 (1.15-1.35)	7.05×10^{-6}	1.19 (1.04-1.33)	0.02	Yes	0.477
rs4760790	12	*TSPAN8/LGR5*	1.11 (1.04-1.17)	0.002	1.17 (1.08-1.26)	8.38×10^{-4}	No	0.256
rs7957197	12	*HNF1A*	1.19 (1.09-1.30)	2.92×10^{-6}	1.05 (0.95-1.17)	0.37	Yes	0.072
rs11634397	15	*ZFAND6*	1.13 (1.05-1.23)	5.10×10^{-4}	1.10 (1.00-1.23)	0.042	Yes	0.712
rs8042680	15	*PRC1*	1.14 (1.08-1.20)	2.83×10^{-5}	1.06 (0.98-1.15)	0.17	Yes	0.185
rs11642841	16	*FTO*	1.04 (0.98-1.10)	0.18	1.12 (1.04-1.21)	0.0083	No	0.125
rs4430796	17	*HNF1B*	1.21 (1.10-1.33)	2.71×10^{-6}	1.11 (0.99-1.27)	0.054	Yes	0.312

Supplementary Note Table SN9: BMI-stratified analyses of T2D association at known and newly-identified autosomal loci. For details see text. Grey lines denote newly-identified loci, bold the two loci (*TCF7L2, BCL11A*) for which there was nominal evidence of heterogeneity between obese and non-obese strata.

					AOD continuous analysis			Early- vs late-onset cases		
SNP	Chr	Position B36 (basepair)	Risk Allele	Nearby Gene	Beta	95% CI	P-value	OR	95% CI	P-value
rs10923931	1	120,319,482	T	*NOTCH2*	-0.63	(-1.11, -0.15)	0.011	1.14	(0.99, 1.31)	0.06
rs11899863	2	43,472,323	C	*THADA*	-0.03	(-0.62, 0.55)	0.915	1.06	(0.89, 1.26)	0.49
rs243021	2	60,438,323	A	*BCL11A*	0.01	(-0.30, 0.32)	0.935	0.98	(0.89, 1.07)	0.65
rs7578326	2	226,728,897	A	*IRS1*	0.06	(-0.29, 0.42)	0.734	1.00	(0.89, 1.12)	0.96
rs13081389	3	12,264,800	A	*PPARG*	0.74	(0.06, 1.42)	0.034	**0.80**	(0.65, 0.99)	**0.036**
rs6795735	3	64,680,405	C	*ADAMTS9*	-0.38	(-0.71, -0.06)	0.020	1.00	(0.91, 1.10)	0.95
rs1470579	3	187,011,774	C	*IGF2BP2*	-0.26	(-0.59, 0.07)	0.126	1.09	(0.99, 1.20)	0.08
rs1801214	4	6,353,923	T	*WFS1*	-0.52	(-0.85, -0.18)	0.002	1.05	(0.95, 1.18)	0.33
rs4457053	5	76,460,705	G	*ZBED3*	-0.13	(-0.58, 0.33)	0.577	0.96	(0.85, 1.10)	0.57
rs10440833	6	20,796,100	A	*CDKAL1*	-0.28	(-0.61, 0.06)	0.103	1.09	(0.99, 1.20)	0.87
rs849134	7	28,162,747	A	*JAZF1*	-0.35	(-0.66, -0.03)	0.032	1.03	(0.94, 1.13)	0.54
rs972283	7	130,117,394	G	*KLF14*	-0.52	(-0.84, -0.20)	0.002	1.08	(0.98, 1.20)	0.14
rs896854	8	96,029,687	T	*TP53INP1*	-0.16	(-0.48, 0.15)	0.311	1.03	(0.94, 1.13)	0.53
rs3802177	8	118,254,206	G	*SLC30A8*	-0.03	(-0.42, 0.36)	0.882	0.92	(0.82, 1.03)	0.13
rs10965250	9	22,123,284	G	*CDKN2A/B*	-0.07	(-0.53, 0.39)	0.759	0.98	(0.84, 1.14)	0.78
rs13292136	9	81,141,948	C	*CHCHD9*	-0.09	(-0.68, 0.49)	0.759	0.95	(0.76, 1.17)	0.62
rs12779790	10	12,368,016	G	*CAMK1D*	-0.07	(-0.47, 0.32)	0.712	1.00	(0.89, 1.12)	0.98
rs5015480	10	94,455,539	C	*HHEX/IDE*	-0.21	(-0.53, 0.11)	0.193	1.05	(0.95, 1.16)	0.34
rs7903146	10	114,748,339	T	*TCF7L2*	-0.38	(-0.72, -0.04)	0.028	1.00	(0.91, 1.11)	0.94
rs231362	11	2,648,047	G	*KCNQ1*	-0.53	(-0.89, -0.16)	0.005	**1.17**	(1.04, 1.32)	**0.0079**
rs5215	11	17,365,206	C	*KCNJ11*	0.02	(-0.30, 0.34)	0.890	0.97	(0.88, 1.06)	0.5
rs1552224	11	72,110,746	A	*CENTD2*	-0.11	(-0.54, 0.33)	0.630	1.00	(0.88, 1.13)	0.98
rs1387153	11	92,313,476	T	*MTNR1B*	-0.37	(-0.74, -0.01)	0.047	1.04	(0.93, 1.16)	0.47
rs1531343	12	64,461,161	C	*HMGA2*	-0.12	(-0.65, 0.41)	0.655	0.94	(0.80, 1.10)	0.42
rs4760790	12	69,921,061	A	*TSPAN8*	-0.08	(-0.43, 0.27)	0.651	1.02	(0.92, 1.14)	0.71
rs7957197	12	119,945,069	A	*HNF1A*	0.18	(-0.22, 0.59)	0.376	1.01	(0.90, 1.14)	0.86
rs11634397	15	78,219,277	G	*ZFAND6*	-0.03	(-0.39, 0.33)	0.871	0.91	(0.82, 1.02)	0.12
rs8042680	15	89,322,341	A	*PRC1*	-0.44	(-0.77, -0.1)	0.010	**1.15**	(1.04, 1.26)	**0.0072**
rs11642841	16	52,402,988	A	*FTO*	-0.39	(-0.73, -0.05)	0.024	1.02	(0.91, 1.13)	0.78
rs4430796	17	33,172,153	G	*HNF1B*	0.22	(-0.37, 0.82)	0.458	1.00	(0.86, 1.16)	0.99

Supplementary Note Table SN10: Age-of-diagnosis-stratified analyses of T2D association at known and newly-identified autosomal loci. For details see text. Beta values are relative to the T2D risk allele. Grey lines denote newly-identified loci, bold (in last column) the three loci for which there was nominal evidence of heterogeneity between young-onset and older-onset strata (two [*KCNQ1, PRC1*] more strongly associated with T2D in young-onset cases, the other [*PPARG*] with older-onset cases).

Significant associations to other phenotypes at T2D-susceptibility loci

To collect the information shown in **Supplementary Table 5,** we identified, for established and novel regions of T2D association, association signals ($P<10^{-6}$) for other traits mapping within 1Mb of the index T2D SNP using the NHGRI GWA association catalog[S65] complemented by the published literature. We excluded studies of bodyweight/size and glucose-related phenotypes, as well as GWA studies that considered multiple serum protein or metabolite levels.

To assess the significance of the co-occurrence of T2D loci in regions of the genome containing SNPs associated with other unrelated traits, we performed a simple simulation study (autosomal loci only). For each permutation, we selected 30 independent SNPs (i.e. not occurring within 1Mb of each other) from the ~2.4 million autosomal SNPs reported in the DIAGRAM+ meta-analysis. We then determined, for each of the 30 randomised T2D "pseudo"loci, if an associated SNP for an unrelated trait occurred within a given distance (100kb-1Mb) up- or down-stream.

The associations of SNPs with unrelated traits were taken from the NHGRI disease association catalogue [S65] restricting ourselves to those associations listed with $P<10^{-6}$ and excluding anthropometric and glycemic phenotype outcomes (adiposity, obesity, obesity related traits [including body mass index, waist circumference, other waist measurements and weight], fasting plasma glucose, glycated hemoglobin levels and incident diabetes) on the basis that co-localising signals for these T2D-related phenotypes are likely to reflect the same causal variant (as with the association of *FTO* variants with T2D and obesity; and of *MTNR1B* variants with T2D and fasting glucose). We also excluded studies involving GWAS conducted against multiple protein or metabolite levels (such as Melzer *et al,* 2008 [S66]; Gieger *et al*, 2009 [S67]). Note that these simulations did not include a number of association signals within T2D-susceptibility regions (such as those at *ZBED3* and *KLF14*) which were for one reason or another not cited in the NHGRI catalog, but are, for completeness, shown in **Supplementary Table 5**.

Over 1 million permutations, the proportion of simulations in which the number of flanking secondary signals was equal to or exceeded 13, the number observed in our study, was just 1.6×10^{-3}. Our results remained robust to the size of interval considered around each T2D locus ($P=7.0\times10^{-5}$ for 8 loci with flanking secondary signals within 100kb, $P=2.4\times10^{-5}$ for 11 loci within 200kb, $P=1.6\times10^{-3}$ for 13 within 500kb and 1.3×10^{-2} for 16 loci within 1Mb).

References for Supplementary Table 5

1. Kathiresan, S. *et al.* Common variants at 30 loci contribute to polygenic dyslipidemia. *Nat Genet.* 2009;**41**:56-65
2. Aulchenko, Y.S. *et al.* Loci influencing lipid levels and coronary heart disease risk in 16 European population cohorts. *Nat Genet*. 2009;**41**:47-55.
3. Buch, S. *et al.* A genome-wide association scan identifies the hepatic cholesterol transporter ABCG8 as a susceptibility factor for human gallstone disease. *Nat Genet* 2007;**39**:995-999
4. Uda, M. *et al.* Genome-wide association study shows BCL11A associated with persistent fetal hemoglobin and amelioration of the phenotype of beta-thalassemia. *Proc Natl Acad Sci U S A.* 2008;**105**:1620-1625.
5. Menzel,S. *et al.* A QTL influencing F cell production maps to a gene encoding a zinc-finger protein on chromosome 2p15. *Nat Genet*. 2007;**39**:1197-1199.
6. Samani, N.J. *et al.* Genomewide association analysis of coronary artery disease. *N Engl J Med.* 2007;**357**:443-453.
7. Arnaud-Lopez, L. *et al* . Phosphodiesterase 8B gene variants are associated with serum TSH levels and thyroid function. *Am J Hum Genet.* 2008;**82**:1270-1280.
8. Barrett, J.C. *et al.* Genome-wide association defines more than 30 distinct susceptibility loci for Crohn's disease. *Nat Genet.* 2008;**40**:955-962.
9. Quaranta, M. *et al.* Differential contribution of *CDKAL1* variants to psoriasis, Crohn's disease and type II diabetes. *Genes Immun.* 2009 Jul 9. [Epub ahead of print]
10. Thomas, G. *et al.* Multiple loci identified in a genome-wide association study of prostate cancer. *Nat Genet*. 2008;**40**:310-315.
11. Johansson, A. *et al.* Common variants in the *JAZF1* gene associated with height identified by linkage and genome-wide association analysis. *Hum Mol Genet.* 2009;**18**:373-380.
12. Soranzo, N. *et al.* Meta-analysis of genome-wide scans for human adult stature identifies novel Loci and associations with measures of skeletal frame size. *PLoS Genet.* 2009;**5**:e1000445.
13. Stacey, S.N. *et al.* New common variants affecting susceptibility to basal cell carcinoma. *Nat Genet*. 2009;**41**:909-914.
14. Helgadottir, A. *et al.* A common variant on chromosome 9p21 affects the risk of myocardial infarction. *Science*. 2007;**316**:1491-1493.
15. McPherson, R. *et al.* A common allele on chromosome 9 associated with coronary heart disease. *Science*. 2007;**316**:1488-1491.
16. Helgadottir, A. *et al.* The same sequence variant on 9p21 associates with myocardial infarction, abdominal aortic aneurysm and intracranial aneurysm. *Nat Genet.* 2008;**40**:217-224
17. Bishop, D.T. *et al.* Genome-wide association study identifies three loci associated with melanoma risk. *Nat Genet.* 2009;**41**:920-925
18. Falchi, M. *et al.* Genome-wide association study identifies variants at 9p21 and 22q13 associated with development of cutaneous nevi. *Nat Genet.* 2009;**41**:915-919
19. Shete, S. *et al.* Genome-wide association study identifies five susceptibility loci for glioma. *Nat Genet*. 2009;**41**:899-904.
20. Bilguvar, K. *et al.* Susceptibility loci for intracranial aneurysm in European and Japanese populations. *Nat Genet.* 2008;**40:**1472-1477.
21. Hancock, D.B. *et al.* Genome-wide association study implicates chromosome 9q21.31 as a susceptibility locus for asthma in mexican children. *PLoS Genet.* 2009;**5**:e1000623.
22. Stolk, L. *et al.* Loci at chromosomes 13, 19 and 20 influence age at natural menopause. *Nat Genet.* 2009 May 17. [Epub ahead of print]
23. Newton-Cheh, C. *et al.* Common variants at ten loci influence QT interval duration in the QTGEN Study. *Nat Genet.* 2009;**41**:399-406.
24. Pfeufer, A. *et al.* Common variants at ten loci modulate the QT interval duration in the QTSCD Study. *Nat Genet.* 2009;**41**:407-414.
25. Pfeufer, A. *et al.* Common variants in myocardial ion channel genes modify the QT interval in the general population: results from the KORA study. *Circ Res.* 2005;**96**:693-701.
26. Johnson, A.D. *et al.* Genome-wide association meta-analysis for total serum bilirubin levels. *Hum Mol Genet*. 2009;**18**:2700-2710.
27. Levy, D. *et al.* Genome-wide association study of blood pressure and hypertension. *Nat Genet.* 2009 May 10. [Epub ahead of print]
28. Vasan, R.S. *et al.* Genetic variants associated with cardiac structure and function: a meta-analysis and replication of genome-wide association data. *JAMA.* 2009;**302**:168-178.
29. Weedon, M.N. *et al.* A common variant of *HMGA2* is associated with adult and childhood height in the general population. *Nat Genet.* 2007;**39**:1245-1250.
30. Reiner, A.P. *et al.* Polymorphisms of the *HNF1A* gene encoding hepatocyte nuclear factor-1 alpha are associated with C-reactive protein. *Am J Hum Genet.* 2008;**82**:1193-1201.
31. Ridker, P.M. *et al.* Loci related to metabolic-syndrome pathways including *LEPR*, *HNF1A*, *IL6R*, and *GCKR* associate with plasma C-reactive protein: the Women's Genome Health Study. *Am J Hum Genet.* 2008;**82**:1185-1192.

32. Yuan, X. *et al.* Population-based genome-wide association studies reveal six loci influencing plasma levels of liver enzymes. *Am J Hum Genet.* 2008;**83**:520-528.
33. Elliott, P. *et al.* Genetic Loci associated with C-reactive protein levels and risk of coronary heart disease. *JAMA*. 2009;**302**:37-48.
34. Erdmann, J. *et al.* New susceptibility locus for coronary artery disease on chromosome 3q22.3. *Nat Genet.* 2009;**41**:280-282.
35. Gudmundsson, J. *et al.* Two variants on chromosome 17 confer prostate cancer risk, and the one in TCF2 protects against type 2 diabetes. *Nat Genet*. 2007;**39**:977-983.
36. Lesch, K.P. *et al.*. Molecular genetics of adult ADHD: converging evidence from genome-wide association and extended pedigree linkage studies. *J Neural Transm.* 2008;**115**:1573-1585

Physiological analyses

Locus	SNP	Nearby Gene	Risk Allele	Nonrisk Allele	BMI Z-score[a]	P-value
1	rs10923931	*NOTCH2*	T	G	0.014	0.99
2	rs11899863	*THADA*	C	T	-0.125	0.90
3	rs243021	*BCL11A*	A	G	0.927	0.35
4	rs7578326	*IRS1*	A	G	-1.835	0.07
5	rs13081389	*PPARG*	A	G	-0.109	0.91
6	rs6795735	*ADAMTS9*	C	T	-1.94	0.05
7	rs1470579	*IGF2BP2*	C	A	-0.207	0.84
8	rs1801214	*WFS1*	T	C	-0.105	0.92
9	rs4457053	*ZBED3*	G	A	-1.808	0.07
10	rs10440833	*CDKAL1*	A	T	1.105	0.27
11	rs849134	*JAZF1*	A	G	-0.292	0.77
12	rs972283	*KLF14*	G	A	-3.661	2.51×10^{-4}
13	rs896854	*TP53INP1*	T	C	1.289	0.20
14	rs3802177	*SLC30A8*	G	A	-0.913	0.36
15	rs10965250	*CDKN2A*	G	A	-1.278	0.20
16	rs13292136	*CHCHD9*	C	T	-0.347	0.73
17	rs12779790	*CAMK1D*	G	A	0.271	0.79
18	rs5015480	*HHEX*	C	T	-0.218	0.83
19	rs7903146	*TCF7L2*	T	C	-0.83	0.41
20	rs231362	*KCNQ1*	G	A	-1.301	0.19
21	rs163184	*KCNQ1*	G	T	0.909	0.36
22	rs5215	*KCNJ11*	C	T	-2.361	0.02
23	rs1552224	*CENTD2*	A	C	-0.52	0.60
24	rs1387153	*MTNR1B*	T	C	0.585	0.56
25	rs1531343	*HMGA2*	C	G	-0.365	0.72
26	rs4760790	*TSPAN8*	A	G	0.524	0.60
27	rs7957197	*HNF1A*	T	A	0.603	0.55
28	rs11634397	*ZFAND6*	G	A	-0.308	0.76
29	rs8042680	*PRC1*	A	C	-0.977	0.33
30	rs11642841	*FTO*	A	C	7.952	1.83×10^{-15}
31	rs4430796	*HNF1B*	G	A	-0.304	0.76

Supplementary Note Table SN11. Analysis of all T2D-associated variants for association with BMI in population-based samples from the GIANT meta-analysis[S68]. Apart from *HNF1B* (for which the SNP of interest was imputed well in only a subset of samples) estimates of effect size are based on sample sizes between 28,223 and 32,529. Loci identified in the present study are in grey. No GIANT data were available for chromosome X, so the *DUSP9* signal is not included.

[a]Z-score is reported for the T2D risk allele; positive (negative) z-score means the direction of the T2D risk allele effect is toward higher (lower) BMI.

							Fasting glucose (mmol/l)			Fasting insulin ln(pmol/l)			HOMA-B[c]			HOMA-IR[c]			2 hour Glucose (mmol/l)		
SNP	Chr	Position B36	Risk Allele	Non risk Allele	Nearby Gene	T2D OR	Beta[a]	SE[b]	P-value	Beta[a]	SE[b]	P-value	Beta[a]	SE[b]	P-value	Beta[a]	SE[b]	P-value	Beta[a]	SE[b]	P-value
rs10923931	1	120,319,482	T	G	*NOTCH2*	1.14	0.0107	0.0058	0.068	-0.0005	0.0061	0.93	-0.0045	0.0052	0.38	0.0011	0.0064	0.87	0.0575	0.0299	0.055
rs11899863	2	43,472,323	C	T	*THADA*	1.17	0.0298	0.0063	2.58×10^{-6}	-0.0115	0.0068	0.091	-0.0241	0.0060	6.98×10^{-5}	-0.0088	0.0071	0.22	0.0197	0.0328	0.55
rs243021	2	60,438,323	A	G	*BCL11A*	1.09	0.0061	0.0037	0.10	0.0014	0.0039	0.72	-0.0002	0.0033	0.96	0.0010	0.0040	0.80	-0.0311	0.0185	0.093
rs7578326	2	226,728,897	A	G	*IRS1*	1.12	-0.0008	0.0040	0.85	0.0088	0.0041	0.033	0.0054	0.0036	0.13	0.0071	0.0043	0.099	0.0344	0.0202	0.088
rs13081389	3	12,264,800	A	G	*PPARG*	1.24	0.0092	0.0074	0.21	0.0207	0.0076	0.0067	0.0085	0.0063	0.18	0.0208	0.0079	0.0089	0.0645	0.0362	0.074
rs6795735	3	64,680,405	C	T	*ADAMTS9*	1.09	0.0084	0.0037	0.024	0.0065	0.0039	0.092	0.0011	0.0033	0.75	0.0069	0.0040	0.086	0.0327	0.0188	0.082
rs1470579	3	187,011,774	C	A	*IGF2BP2*	1.14	0.0093	0.0040	0.019	-0.0068	0.0041	0.099	-0.0113	0.0035	0.0014	-0.0085	0.0043	0.047	0.0501	0.0200	0.012
rs1801214	4	6,353,923	T	C	*WFS1*	1.13	0.0180	0.0039	4.09×10^{-6}	0.0058	0.0041	0.15	-0.0039	0.0035	0.27	0.0080	0.0042	0.058	-0.0125	0.0194	0.52
rs4457053	5	76,460,705	G	A	*ZBED3*	1.16	0.0175	0.0045	9.15×10^{-5}	0.0033	0.0047	0.48	-0.0071	0.0040	0.078	0.0049	0.0049	0.31	0.0467	0.0238	0.050
rs10440833	6	20,796,100	A	T	*CDKAL1*	1.25	0.0102	0.0040	0.011	-0.0104	0.0041	0.012	-0.0101	0.0035	0.0043	-0.0091	0.0043	0.035	0.0415	0.0214	0.053
rs849134	7	28,162,747	A	G	*JAZF1*	1.13	0.0063	0.0036	0.086	-0.0020	0.0038	0.61	-0.0036	0.0033	0.27	-0.0012	0.0040	0.75	0.0357	0.0187	0.056
rs972283	7	130,117,394	G	A	*KLF14*	1.10	0.0029	0.0037	0.44	0.0085	0.0039	0.029	0.0039	0.0033	0.24	0.0095	0.0041	0.019	0.0323	0.0192	0.093
rs896854	8	96,029,687	T	C	*TP53INP1*	1.10	0.0122	0.0037	8.79×10^{-4}	-0.0003	0.0038	0.94	-0.0046	0.0033	0.16	0.0003	0.0040	0.94	-0.0086	0.0185	0.64
rs3802177	8	118,254,206	G	A	*SLC30A8*	1.15	0.0273	0.0043	2.03×10^{-10}	-0.0036	0.0045	0.43	-0.0160	0.0038	1.96×10^{-5}	-0.0005	0.0047	0.92	0.0942	0.0219	1.73×10^{-5}
rs10965250	9	22,123,284	G	A	*CDKN2A*	1.20	0.0181	0.0049	2.08×10^{-4}	0.0031	0.0051	0.54	-0.0076	0.0045	0.089	0.0061	0.0054	0.25	0.0436	0.0255	0.087
rs13292136	9	81,141,948	C	T	*CHCHD9*	1.20	0.0098	0.0068	0.15	0.0001	0.0069	0.99	0.0003	0.0056	0.96	0.0022	0.0072	0.76	0.0079	0.0318	0.80
rs12779790	10	12,368,016	G	A	*CAMK1D*	1.09	0.0158	0.0049	0.0012	-0.0004	0.0052	0.94	-0.0108	0.0044	0.015	-0.0010	0.0054	0.85	-0.0054	0.0246	0.82
rs5015480	10	94,455,539	C	T	*HHEX*	1.18	0.0095	0.0037	0.010	0.0013	0.0039	0.73	-0.0047	0.0033	0.17	0.0025	0.0040	0.53	0.0307	0.0187	0.099
rs7903146	10	114,748,339	T	C	*TCF7L2*	1.40	0.0229	0.0041	2.80×10^{-8}	-0.0122	0.0043	0.0046	-0.0200	0.0038	1.39×10^{-7}	-0.0096	0.0045	0.034	0.1178	0.0212	2.57×10^{-8}
rs231362	11	2,648,047	G	A	*KCNQ1*	1.11	0.0187	0.0041	5.29×10^{-6}	0.0048	0.0043	0.27	-0.0062	0.0037	0.089	0.0043	0.0045	0.34	0.0612	0.0208	0.0033
rs163184	11	2,803,645	G	T	*KCNQ1*	1.09	0.0153	0.004	1.27×10^{-4}	-0.002	0.0042	0.64	-0.0086	0.0035	0.016	0.0007	0.0044	0.87	0.0424	0.02	0.034
rs5215	11	17,365,206	C	T	*KCNJ11*	1.09	-0.0032	0.0038	0.40	-0.0021	0.0039	0.59	0.0009	0.0033	0.78	-0.0018	0.0041	0.66	0.0263	0.0189	0.16
rs1552224	11	72,110,746	A	C	*CENTD2*	1.13	0.0193	0.0050	1.20×10^{-4}	-0.0126	0.0052	0.015	-0.0166	0.0043	9.39×10^{-5}	-0.0092	0.0054	0.085	0.0527	0.0246	0.032
rs1387153	11	92,313,476	T	C	*MTNR1B*	1.12	0.0602	0.0043	6.59×10^{-45}	-0.0034	0.0045	0.44	-0.0286	0.0037	1.13×10^{-14}	0.0057	0.0046	0.22	0.0529	0.0215	0.014
rs1531343	12	64,461,161	C	G	*HMGA2*	1.20	0.0110	0.0061	0.069	0.0057	0.0063	0.36	0.0006	0.0056	0.91	0.0094	0.0066	0.16	-0.0152	0.0321	0.64
rs4760790	12	69,921,061	A	G	*TSPAN8*	1.11	0.0069	0.0041	0.091	-0.0023	0.0043	0.58	-0.0010	0.0037	0.79	0.0012	0.0045	0.78	0.0142	0.0212	0.50
rs7957197	12	119,945,069	T	A	*HNF1A*	1.14	-0.0033	0.0045	0.46	-0.0077	0.0047	0.10	-0.0042	0.0040	0.30	-0.0067	0.0049	0.17	0.0474	0.0234	0.043
rs11634397	15	78,219,277	G	A	*ZFAND6*	1.11	0.0003	0.0040	0.94	-0.0051	0.0041	0.22	-0.0034	0.0035	0.33	-0.0038	0.0043	0.38	-0.0015	0.0206	0.94
rs8042680	15	89,322,341	A	C	*PRC1*	1.10	0.0097	0.0039	0.013	0.0011	0.0041	0.79	-0.0019	0.0035	0.58	0.0033	0.0043	0.44	0.0106	0.0199	0.59
rs11642841	16	52,402,988	A	C	*FTO*	1.13	0.0066	0.0039	0.091	0.0170	0.0041	3.13×10^{-5}	0.0094	0.0035	0.0075	0.0168	0.0043	8.17×10^{-5}	0.0110	0.0199	0.58
rs4430796	17	33,172,153	G	A	*HNF1B*	1.14	0.0002	0.0052	0.97	-0.0063	0.0056	0.26	-0.0089	0.0043	0.040	-0.0110	0.0057	0.055	0.0574	0.0375	0.13

Supplementary Note Table SN12: Analysis of all T2D-associated variants for continuous glycemic phenotypes using data from the MAGIC investigators [S62,S63].
All effects are those of the T2D-risk allele. MAGIC meta-analyses included up to 46,186 individuals for fasting glucose, up to 38,238 individuals for fasting Insulin, HOMA-B and HOMA-IR, and up to 15,234 individuals for 2h post-oral glucose tolerance test glucose. MAGIC has not performed a meta-analysis for the X chromosome, so data for the T2D-signal near *DUSP9* are not available: however, as per main text, analysis in a subset of MAGIC samples found no significant association with these continuous glycemic parameters. [a]Beta (coefficient from the linear regression) is reported for the T2D-risk allele; [b]Standard error of beta; [c]HOMA-B and HOMA-IR are indices for homeostasis model assessment. For HOMA-B, negative beta implies that the T2D risk allele is associated with decreased beta-cell function; for HOMA-S, positive beta implies that the T2D-risk allele is associated with decreased insulin sensitivity.

Expression QTL analyses

Expression QTL data are widely used in the follow-up of genome wide association analysis in the expectation that these will link the most strongly-associated (and putatively causal) disease/trait associated variants to one or more genes in the region and thereby provide a clue to the mechanisms responsible for predisposition. However, for such inferences to be reliable, the disease/trait signal of interest and the cis-eQTL signal identified need to be coincident: if they are not, then no conclusions regarding the biology can be reached. Crucially, cis-eQTL association signals are often far more powerful than disease/trait association signals. This can mean that, even when the disease/trait association and a particular cis-eQTL association are driven by distinct sets of "lead" SNPs that have only very modest mutual LD, the cis-eQTL signal can still achieve compelling levels of significance at the disease/trait signal. It is therefore not sufficient merely to report which cis-eQTL signals can be observed at the disease/trait SNPs of interest – one needs to establish whether or not those signals are coincident.

In our analyses (in **table 2)** we address this issue using a number of related approaches.

- First, we report our findings when we take the "lead" T2D-associated variant at each locus and search for interesting ($P<0.001$) cis-eQTL signals for transcripts that map within 2Mb. We find, for example, that the lead SNP at the *HNF1A* locus shows a number of modest cis-eQTL signals to nearby genes including *ACADS, PSMD9, OASL, UNC119B, CAMKK2* and *P2RX4*. Of these, the *CAMKK2* association appears (on superficial analysis) the most convincing reaching $P\sim10^{-12}$.
- Next, we ask, for each of the transcripts for which an interesting cis-eQTL signal was detected, whether there is another SNP in the neighborhood that has a more convincing association with expression of that transcript. Returning to the *HNF1A* locus again, in the case of *CAMKK2*, it is clear that rs11065504 has a far greater effect on *CAMKK2* expression than the T2D lead SNP rs7957197 (P of $3x10^{-117}$ vs $1x10^{-12}$). The lead disease SNP and the *CAMKK2* cis-eQTL lead SNP have low LD ($r^2\sim0.08$). Clearly, if the T2D signal and the cis-eQTL signal here were coincident, one would expect the T2D signal at rs11065504 to be enormous and for that rather than rs7957197 to have been the lead SNP for T2D.
- Finally, to explore the relationships between the cis-eQTL and disease signals, we perform mutual conditional studies. We see what happens to the cis-eQTL signal at the T2D-SNP when we condition on the best cis-eQTL SNP; and we see what happens to the cis-eQTL signal at the best cis-eQTL SNP when we condition on the T2D-risk variant. In the case of *CAMKK2*, these results confirm our suspicion that variation in *CAMKK2* expression is driven by rs11065504: the cis-eQTL signal at rs7957197 evaporates when conditioning on rs11065504, whereas the reverse conditional analysis has little impact.

In the example above, it is clear then that the cis-eQTL signal for *CAMKK2* observed at the *HNF1A* lead SNP is likely to be the result of rs7957197 picking up the "shadow" of the strong *CAMKK2* cis-eQTL centered around rs11065504. The T2D signal and this cis-eQTL signal are distinct and there is no reason to believe that *CAMKK2* is involved in the mediation of the T2D signal at rs7957197. Contrast this with the examples of *IRS1*, *JAZF1* and *CAMK1D*, where the T2D-lead SNPs and the best cis-eQTL SNPs are highly correlated (r^2 0.9-1.0), and conditioning on either abolishes the cis-eQTL signal at the other. For these three loci, one can safely infer that the T2D and cis-eQTL signals are identical and that the transcripts concerned (*IRS1*, *JAZF1* and *CAMK1D*) are probably implicated in disease predisposition.

Expression Profile data

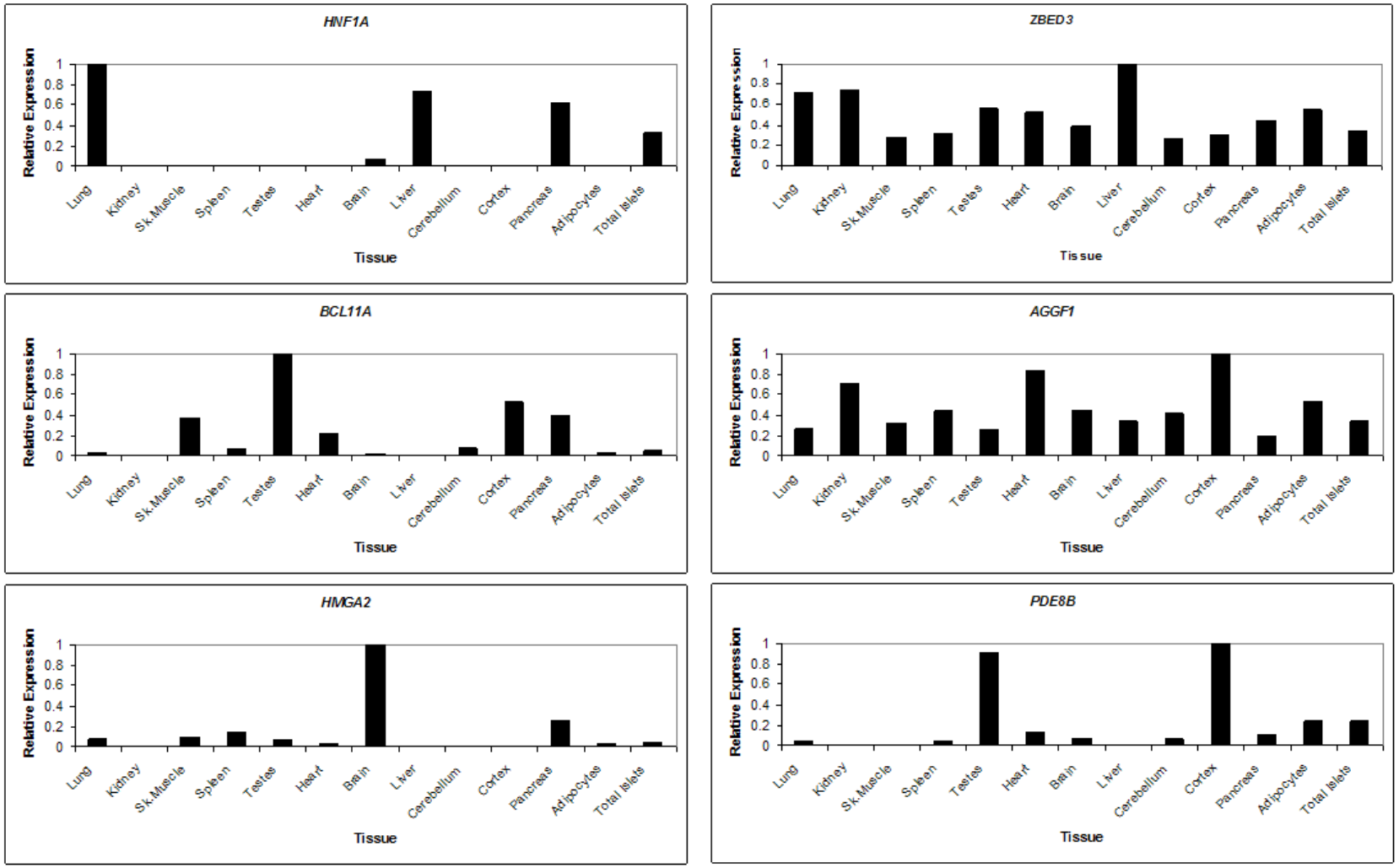
For methods, see Online Methods. The list of probes which we included in our expression profiling studies is listed in **Table SN13.**

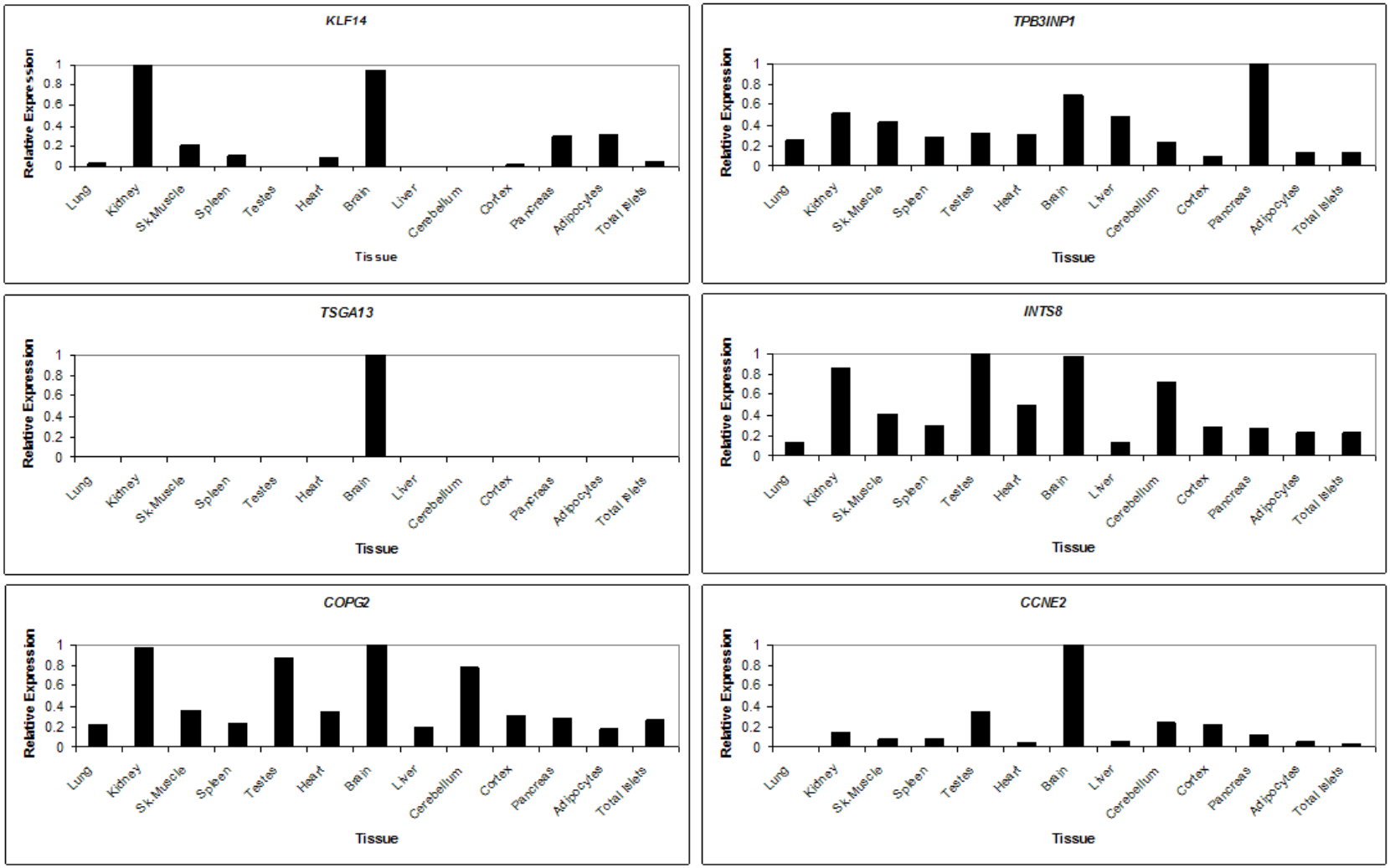
Broad expression of many of the transcripts, including 24 with evidence of beta-cell transcription (**Figures SN4** and **SN5)** limited our ability to prioritise amongst the candidate transcripts, though integration of physiological and expression data should help to direct future functional studies. For example, given the *CCNE2* eQTL signal at the *TP53INP1* locus, the fact that we could detect *CCNE2* expression in islets but not flow-sorted beta-cells may indicate that the beta-cell dysfunction associated with this locus reflects either a temporally-restricted phenotypic effect (e.g. on early beta-cell development) or mediation through islet cell types other than the beta-cell.

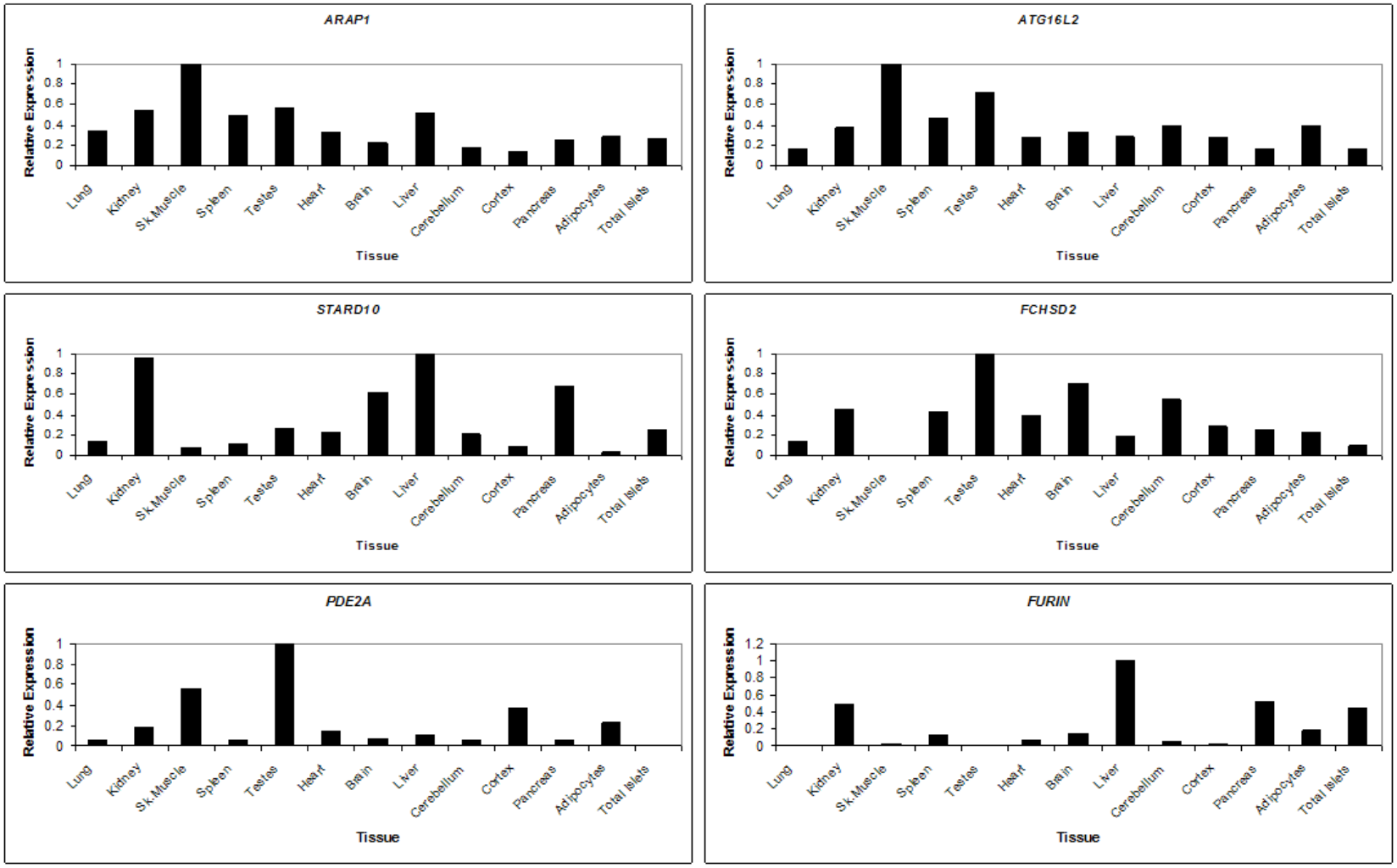
Locus	Gene	Chromosome	Probe	Source	Comments
HNF1A	*HNF1A*	12	Hs00167041_m1	Inventoried	
BCL11A	*BCL11A*	2	Hs00256254_m1	Inventoried	Covers all 3 transcripts
HMGA2	*HMGA2*	12	Hs00171569_m1	Inventoried	Covers both transcripts
ZBED3	*ZBED3*	5	Hs00260688_m1	Inventoried	
	AGGF1	5	HS00203293_m1	Inventoried	
	PDE8B	5	Hs00405493_m1	Inventoried	Covers all 5 transcripts
KLF14	*KLF14*	7	Hs00370951_s1	Inventoried	This is a single exon gene and s1 probes are the only ones available
	TSGA13	7	Hs00364691_m1	Inventoried	
	COPG2	7	Hs00273295_m1	Inventoried	
TP53INP1	*TP53INP1*	8	Hs01003820_m1	Made to order	
	INTS8	8	Hs00215334_m1	Inventoried	
	CCNE2	8	Hs01051894_m1	Inventoried	
ZFAND6	*ZFAND6*	15	Hs00375275_m1	Inventoried	
	FAH	15	Hs00164611_m1	Inventoried	
CENTD2	*STARD10*	11	Hs00246405_m1	Inventoried	
	ARAP1/CENTD2	11	Hs00373707_m1	Made to order	
	FCHSD2	11	Hs00207952_m1	Inventoried	
	ATG16L2	11	Hs01057324_m1	Made to order	
	PDE2A	11	Hs01042255_m1	Inventoried	
PRC1	*FURIN*	15	Hs00159829_m1	Inventoried	
	FES	15	Hs01120751_m1	Inventoried	
	PRC1	15	Hs01597831_m1	Inventoried	
	UNC45A	15	Hs00218751_m1	Inventoried	
	RCCD1	15	Hs00364408_m1	Inventoried	
	MAN2A2	15	Hs00196172_m1	Inventoried	
	HDDC3	15	Hs00826828_g1	Inventoried	Failed QC
	VPS33B	15	Hs00218719_m1	Inventoried	
CHCHD9	*TLE4*	9	Hs00419101_m1	Inventoried	Failed QC
	CHCHD9	9	Hs01892561_sH	Made to order	
Housekeeping genes					
	B2M		Hs99999907_m1	Inventoried	Housekeeping gene
	Beta-Actin		Hs99999903_m1	Inventoried	Housekeeping gene
	HPRT		Hs99999909_m1	Inventoried	Housekeeping gene
	TOP1		Hs01052828_m1	Made to Order	Housekeeping gene

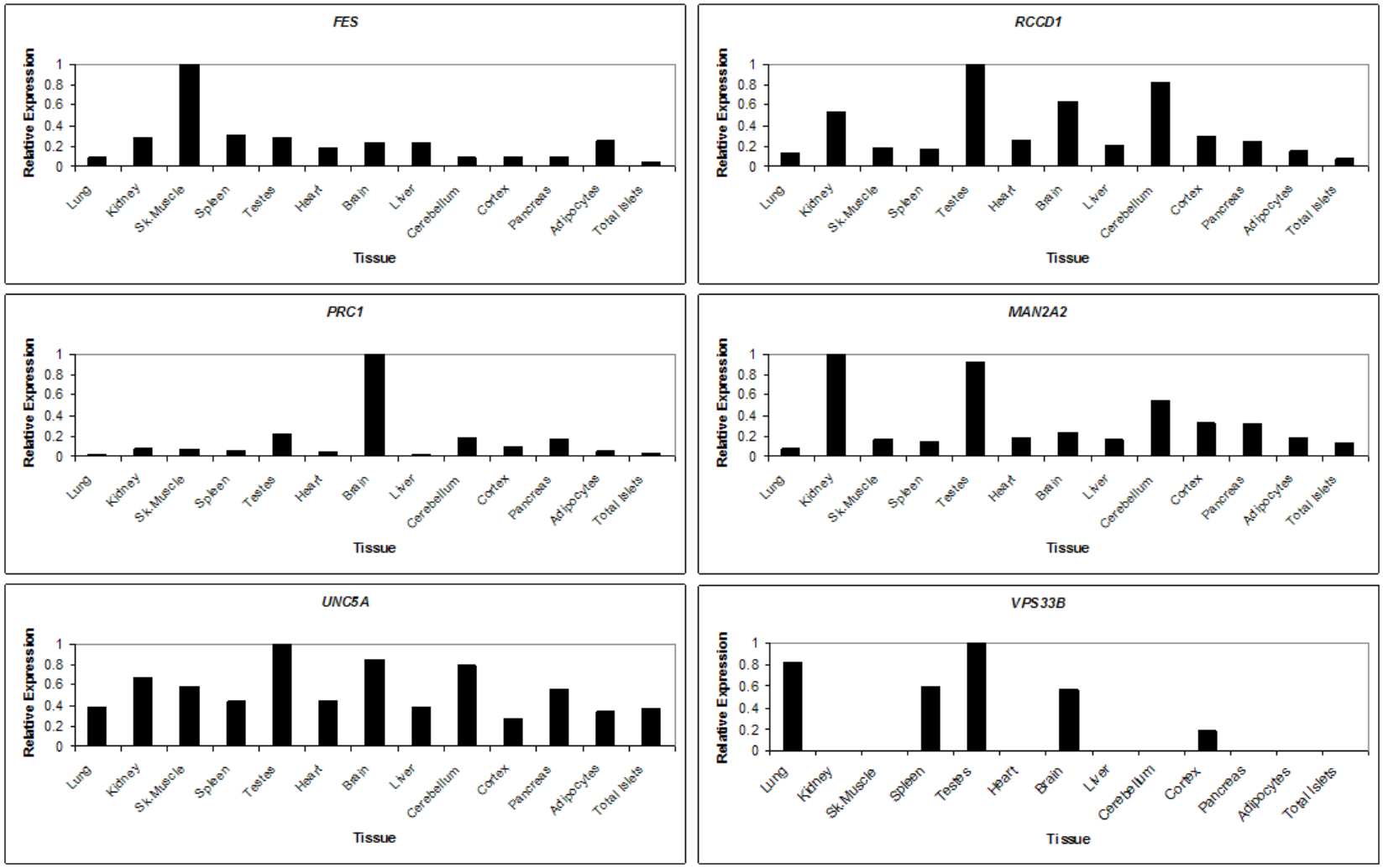
Supplementary Note Table SN13: List of probes used for tissue expression profiling studies. Genes at each locus were considered on the basis of proximity to the lead SNP and possible biological credibility.

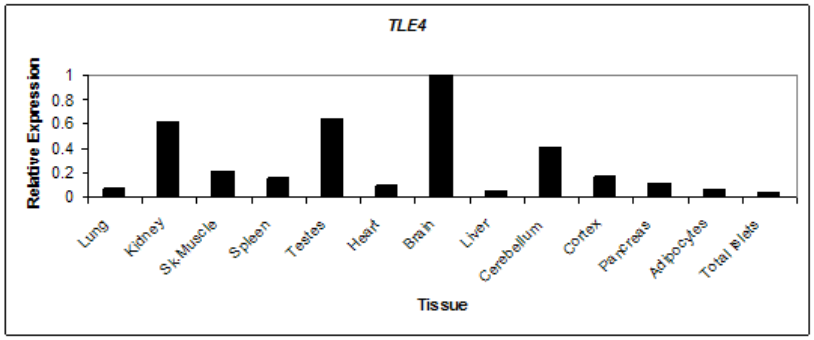
Supplementary Note Figure SN4: Expression profiles of selected positional candidates in tissues including several relevant to T2D phenotype (from left to right, the tissues tested are lung, kidney, skeletal muscle, spleen, testis, heart, brain, liver, cerebellum, cerebral cortex, whole pancreas, adipocyte and pancreatic islets)

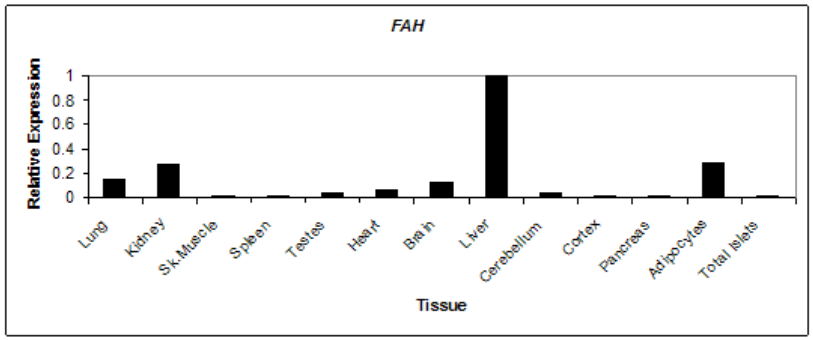


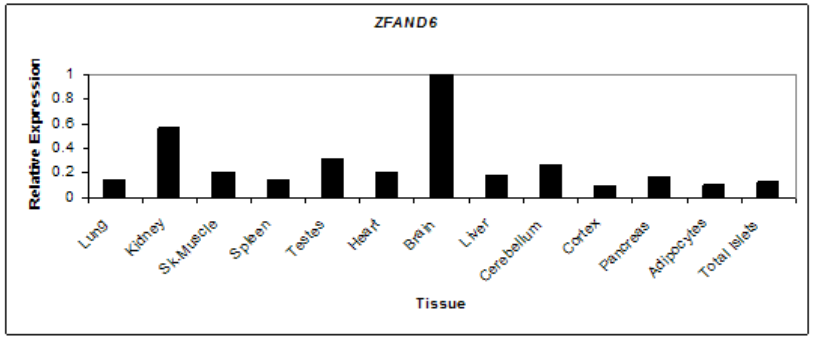

KLF14
TPB3INP1
TSGA13
INTS8
COPG2
CCNE2
Relative Expression
Tissue
Lung
Kidney
Sk.Muscle
Spleen
Testes
Heart
Brain
Liver
Cerebellum
Cortex
Pancreas
Adipocytes
Total Islets

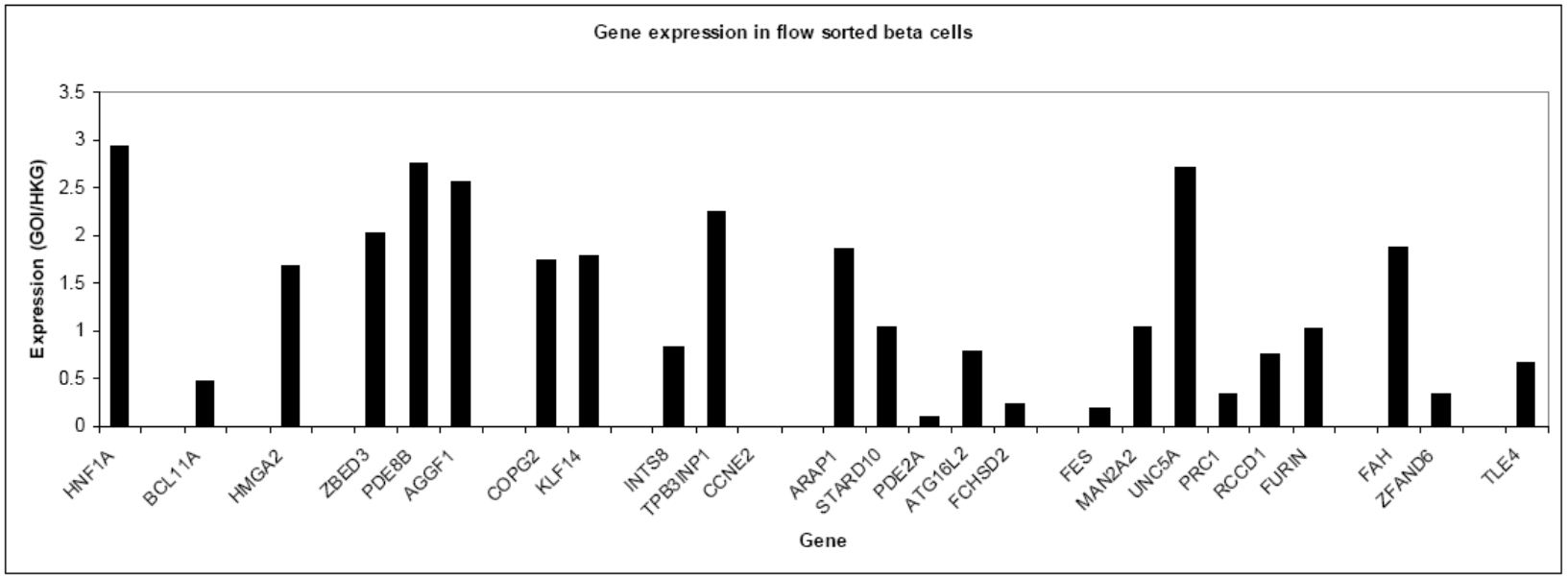

ARAP1
ATG16L2
STARD10
FCHSD2
PDE2A
FURIN
Relative Expression
Tissue
Lung
Kidney
Sk.Muscle
Spleen
Testes
Heart
Brain
Liver
Cerebellum
Cortex
Pancreas
Adipocytes
Total Islets


FES
RCCD1
PRC1
MAN2A2
UNC5A
VPS33B
Relative Expression
Tissue
Lung
Kidney
Sk.Muscle
Spleen
Testes
Heart
Brain
Liver
Cerebellum
Cortex
Pancreas
Adipocytes
Total Islets


TLE4
Relative Expression
1
0.8
0.6
0.4
0.2
0
Lung
Kidney
Sk.Muscle
Spleen
Testes
Heart
Brain
Liver
Cerebellum
Cortex
Pancreas
Adipocytes
Total Islets
Tissue


FAH
Relative Expression
1
0.8
0.6
0.4
0.2
0
Lung
Kidney
Sk.Muscle
Spleen
Testes
Heart
Brain
Liver
Cerebellum
Cortex
Pancreas
Adipocytes
Total Islets
Tissue


ZFAND6
Relative Expression
1
0.8
0.6
0.4
0.2
0
Lung
Kidney
Sk.Muscle
Spleen
Testes
Heart
Brain
Liver
Cerebellum
Cortex
Pancreas
Adipocytes
Total Islets
Tissue



Supplementary Note Figure SN5: Expression of selected positional candidates in flow-sorted beta-cells (average of 2 samples)

Pathway and protein interaction analyses

To search for evidence of enrichment of particular pathways or networks amongst the associated loci, we used four complementary strategies: GRAIL[S69]; PANTHER [S70,S71] and Reactome [S72,S73] (pathways and networks curated to various levels of detail); GWAS-PPI (examining protein-protein interactions). Alternatively to test for enrichment of association signal in prespecified pathways we used MAGENTA (Meta-Analysis Gene-set Enrichment of variaNT Associations), a modified gene-set enrichment analysis method (Segre, A.V., personal communication).

GWAS-BASED GENE LISTS

For all methods except MAGENTA, we generated a limited set of positional candidate gene-lists based on the GWA meta-analysis association P-values (ranging from the confirmed genome-wide significant signals [30 loci] to all signals with $P<10^{-4}$ [111 loci]).

These loci included in these lists were

- **"CONFIRMED LOCUS SET":** All 30 autosomal GWAS signals (including 20 previously reported, and 11 novel loci reported here, with *KCNQ1* represented once, and not including the X-chromosomal signal) ;
- **"EXPANDED LOCUS SET":** All signals with $P<10^{-4}$ on the Stage 1 meta-analysis, including the 30 loci in the "CONFIRMED LOCUS SET" plus another 81 loci for a total of 111;
- **"EXPANDED LOCUS SET MINUS HLA":** Because of concerns that the inclusion of loci close to the HLA region could distort findings (given the possibility of including a large number of functionally-related genes within the cognate LD region), we generated a third locus set which excluded the 1 signal near HLA.

For each of these locus sets, we generated a list of (coding) genes based on the following algorithm:

- Starting from the index SNP at each locus, go out to the flanking recombination hotspots as defined by the CEU Hapmap sample and a further 50kb beyond; take the interval so defined and include all coding genes which are, either partially or completely, included;
- If the above maneuver yields no coding genes for a given locus, extend a further 250kb in each direction and repeat the exercise, but if no coding transcripts are found at that point, desist.

These efforts yielded a total of 82 (confirmed), 393 (expanded) and 320 (expanded minus HLA) genes.

In addition, for selected GWAS-PPI analyses only, we generated an equivalent list ("**FURTHER EXPANDED LIST")** comprising all genes within loci containing SNPs attaining $P<10^{-3}$ on the Stage 1 meta-analysis. This list comprised 696 loci and 2024 genes.

Note that PANTHER and Reactome analyses test for functional category enrichment in the input gene list compared to all annotated genes and/or randomly-generated gene lists. As such they do not account formally for the potential for aggregation of functionally-related genes in particular genomic locations,

and may therefore overestimate the significance of enrichments. GRAIL and, to a lesser extent, GWAS-PPI and MAGENTA testing procedures are robust to this issue.

GRAIL – Gene Relationships Among Implicated Loci

We utilized a bioinformatics method designed to highlight genes using similarity of text in Pubmed abstracts or similarity in Gene-Ontology associated codes (GRAIL[S69]).

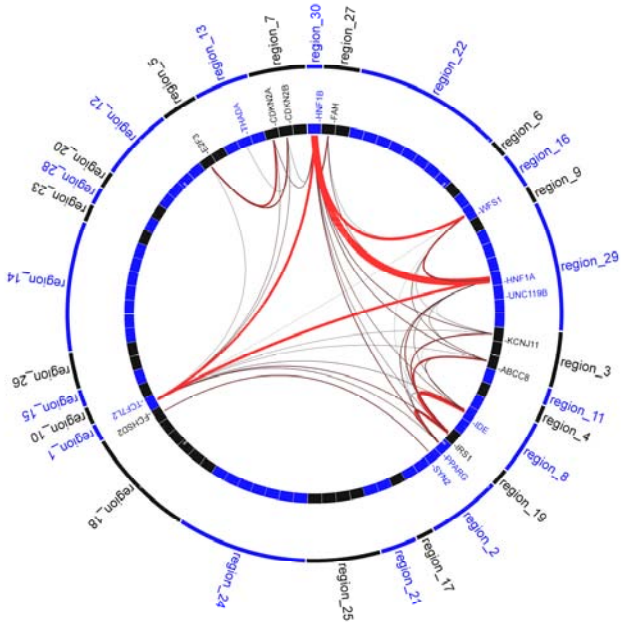
We used GRAIL to look for similarity of text in Pubmed abstracts (connectivity) across the T2D GWAS gene lists [S69]. To reduce confounding by published T2D GWA analyses, we used only abstracts published prior to December 2006. For each gene, we identified the most connected gene from each T2D associated region and calculated a P_{gene}-value based on these connectivities[S69]. Within each region, we selected the smallest P_{gene}-value and performed a Bonferroni correction for the number of genes tested in the region (P_{region}). To ask if the number of genes with $P_{region}<0.05$ was greater than expected, we randomly-selected n SNPs (where n was the number of SNPs underlying the particular T2D GWAS gene list) and identified genes in the region surrounding each SNP in the same way as for the T2D GWAS gene list. We retained only those "random" gene lists with similar (within 10) gene counts to those in the relevant T2D GWAS gene list and tested the genes for connectivity. We compared the observed number of genes with $P_{region}<0.05$ to the empirical distribution of number of genes with $P_{region}<0.05$ based on 1000 retained gene lists.

We first used GRAIL to analyze the list of genes contained within the 30 regions harboring genome-wide significant associations to T2D (the **CONFIRMED LOCUS SET**, above), using a database of abstracts obtained from December 2006 (a date chosen specifically to lie prior to the mainstream application of genome-wide association studies, since abstracts describing the loci identified in those discovery efforts would be expected to confound the analysis).

Among these genes, we observed modest enrichment of connectivity among the established loci (9 out of 30 with GRAIL $P<0.05$, ~2 expected, permutation $P<0.021$). The specific connections between genes and loci were dominated by those for which a role in diabetes susceptibility (monogenic or multifactorial) was already established prior to 2006 (*HNF1A, HNF1B, WFS1, KCNJ11, PPARG)* or suspected through candidate gene studies (*IRS1, IDE*) (**Tables SN14, SN15; Figure SN6**). Many of these links may well be artefacts arising through abstracts invoking previous discoveries. In addition, there was some suggestion of connections among genes related to cell-cycle (*E2F3*, *CDKN2A/2B*) and the "word-set" (i.e. terms overrepresented in the connecting abstracts) included terms consistent with this (e.g., "cyclin", "tumors", "benign") as well as more generic terms ("pancreatic", "glucose", "insulin", "diabetes").

REGION_ID	GRAIL P_{Region}-value	CANDIDATE GENE	Keywords Describing Functional Connections
region_29	**1.2E-04**	***HNF1A***	'diabetes'
region_30	**4.2E-04**	***HNF1B***	'insulin'
region_8	**2.4E-03**	***IDE***	'cyclin'
region_13	**3.7E-03**	***THADA***	'tumors'
region_3	**6.0E-03**	***ABCC8***	'beta'
region_19	**0.016**	***IRS1***	'pancreatic'
region_7	**0.020**	***CDKN2B***	'glucose'
region_16	**0.025**	***WFS1***	'syndrome'
region_1	**0.045**	***TCF7L2***	'catenin'
region_9	0.054	*SLC30A8*	'cytogenetic'
region_15	0.056	*ADAMTS9*	'mutation'
region_27	0.075	*FAH*	'benign'
region_5	0.078	*E2F3*	'degrading'
region_17	0.093	*KCNQ1*	'cases'
region_18	0.098	*FCHSD2*	'patients'
region_11	0.11	*TSPAN8*	'tumor'
region_2	0.11	*PPARG*	'ppargamma'
region_10	0.15	*JAZF1*	'zinc'
region_21	0.17	*AGGF1*	'mutations'
region_20	0.17	*BCL11A*	'hepatocyte'
region_23	0.19	*HMGA2*	
region_25	0.21	*CCNE2*	
region_28	0.27	*TLE4*	
region_6	0.37	*IGF2BP2*	
region_14	0.39	*PPIAL4*	
region_22	0.45	*FURIN*	
region_24	0.54	*CPA5*	
region_12	0.64	*SEC61A2*	
region_4	0.65	*FTO*	
region_26	0.66	*FAT3*	

Supplementary Note Table SN14: List of regions included in the confirmed locus set (using published PubMed abstracts prior to December, 2006) with the gene identified by GRAIL analysis to be the "most connected" to other loci. GRAIL P_{region}-values in the second column represent the most significant P_{gene}-value adjusted for the multiple comparisons within a locus if the region contained more than one gene. The right panel includes the words over-represented in PubMed abstracts featuring the most connected genes.



Supplementary Note Figure SN6: GRAIL "circle" plot highlighting connections among the 30 implicated loci identified by genome-wide association studies using abstracts prior to December, 2006. The statistical strength of each connection are proportional to the thickness of the lines (see above Table SN14 for precise P_{gene}-values). Gene names are listed if individual p-values for the connection have a GRAIL P_{gene} < 0.05.

GENE	GRAIL P_{Gene}-value	SELECTED SIMILAR GENES (Rank in parantheses)
HNF1A	1.7E-05	HNF1B(2), TCF7L2(31), CPA1(112), WFS1(133), HMGA2(150), PPARG(177), IRS1(239), ACADS(334), ABCC8(417), KCNJ11(508)
HNF1B	4.2E-04	HNF1A(2), CPA1(90), WFS1(94), HMGA2(104), TCF7L2(112), PPARG(341), ACADS(430), HHEX(475), CDKN2A(609)
IDE	7.9E-04	IRS1(6), HNF1A(104), PPARG(161), ABCC8(167), CPA1(243), SLC30A8(279), FURIN(296), KCNJ11(368), HNF1B(408)
THADA	1.2E-03	HMGA2(4), JAZF1(24), MTAP(170), CDKN2B(276), BCL11A(301)
ABCC8	2.0E-03	KCNJ11(2), KCNQ1(114), IRS1(135), P2RY2(143), HNF1A(150), CPA1(241), PPARG(355), HNF1B(356), WFS1(378), IDE(448), SLC30A8(493)
CDKN2B	6.7E-03	CDKN2A(2), MTAP(5), HMGA2(73), CCNE2(93), E2F3(184), TP53INP1(274), CPA1(347), BCL11A(402), JAZF1(463)
CDKN2A	0.011	CDKN2B(2), MTAP(3), TP53INP1(65), HMGA2(73), E2F3(113), CENTD2(321), HNF1B(461), HNF1A(578), CCNE2(606), TCF7L2(607)
WFS1	0.013	HNF1B(62), HNF1A(63), CPA1(212), KCNQ1(240), SEC61A2(319), ABCC8(476), HMGA2(531), CPA5(578)
KCNJ11	0.013	ABCC8(4), KCNQ1(104), P2RY2(174), HNF1A(212), IRS1(230), CPA1(342), PPARG(366), HNF1B(377), WFS1(511)
IRS1	0.016	IDE(91), PPARG(103), HNF1A(193), ABCC8(235), KCNJ11(355), SLC30A8(610), TCF7L2(624)
FCHSD2	0.017	KIAA0152(3), HDDC3(94), PLEKHH2(115), RCCD1(167), C8orf38(202), CDKAL1(307)
UNC119B	0.028	TLE4(18), C8orf38(104), PLEKHH2(153), UNC45A(235), HDDC3(297), CDKAL1(367), KIAA0152(571)
PPARG	0.030	IRS1(72), HNF1A(160), HMGA2(364), HNF1B(372), TCF7L2(452)
SYN2	0.038	KIAA0152(43), FCHSD2(44), CAMK1D(82), TLE4(301), CPA1(500)
FAH	0.038	HNF1A(73), HNF1B(119), CPA1(123), HMGA2(174), ACADS(225), WFS1(517), MTAP(559)
E2F3	0.040	CCNE2(45), CDKN2A(110), TP53INP1(135), HMGA2(186), CDKN2B(272)
TCF7L2	0.045	HNF1A(92), HNF1B(193), HMGA2(242), TLE4(418), PPARG(506), HHEX(525), CPA1(546)

Supplementary Note Table SN15: List of connections from the GRAIL analysis for selected genes (P_{gene} < 0.05) for the analysis shown in Table SN14 and Figure SN6. GRAIL P_{gene}-values listed here are the results for the test of the gene itself, and thus are not adjusted for the number of genes tested within a locus. We note here that connections between previously established diabetes genes (through their role in monogenic or syndromic forms of diabetes) tend to dominate, as expected, based on their prior literature.

We also examined evidence of connectivity among the known 30 GWAS regions in conjunction with 81 independent regions putatively associated to T2D in our Stage 1 meta-analysis (the **EXPANDED LOCUS SET**, above) again using the 2006 abstract database. Overall we observed no strong evidence for specific direct connections, and though we observed a small excess of connected loci, this was not different to expectation (13/111 with GRAIL P<0.05 vs 6 expected, permutation P~0.54) (**Tables SN16, SN17**). Exclusion of the HLA (the **EXPANDED LOCUS SET MINUS HLA** set, above) in the GRAIL analysis did not quantitatively alter the results (data not shown).

REGION_ID	GRAIL P_{Region}-value	CANDIDATE GENE
region_30	5.6E-03	*HNF1B*
region_85	8.0E-03	*SPRY4*
region_29	0.016	*HNF1A*
region_13	0.016	*THADA*
region_94	0.017	*NRXN1*
region_53	0.020	*SPRY2*
region_52	0.030	*AP3B1*
region_38	0.035	*SATB2*
region_108	0.038	*NRXN3*
region_1	0.041	*TCF7L2*
region_27	0.041	*FAH*
region_83	0.042	*TBX18*
region_16	0.044	*WFS1*

Keywords Describing Functional Connections
'insulin'
'notch'
'cadherin'
'calcineurin'
'adhesion'
'factor'
'diabetes'
'zinc'
'signaling'
'growth'
'mesoderm'
'transcription'
'vegf'
'cell'
'tyrosine'
'expression'
'endothelial'
'cells'
'kinase'
'development'

Supplementary Note Table SN16: List of regions included in the expanded locus set (using published PubMed abstracts prior to December, 2006) with the gene identified by GRAIL analysis to be the "most connected" to other loci. GRAIL P_{region}-values in the second column represent the most significant P_{gene}-value adjusted for the multiple comparisons within a locus if the region contained more than one gene. Presented only are the top results where $P_{region} < 0.05$. The right panel includes the words over-represented in PubMed abstracts featuring the most connected genes.

GENE	GRAIL P_{Gene}-value	SELECTED SIMILAR GENES (Rank in parantheses)
HNF1A	2.2E-03	HNF1B(1), PCBD2(10), APOM(15), TCF7L2(30), TNF(66), RPL32(86), PITX1(104), CPA1(111), WFS1(132), HMGA2(149), PPARG(176), POU5F1(187), IRS1(238), PRKACB(253), BAT1(260), ACADS(333), ABCC8(416), LTA(463), KCNJ11(507), NFATC1(520), CFB(527), ITGA2(537), ETV1(574)
SPRY4	4.0E-03	SPRY2(1), FGF9(18), PITX1(52), RAF1(133), PRKACB(151), POU5F1(181), RPL32(188), SALL3(192), TCF7L2(202), LRIG1(262), CPA1(265), HMGA2(276), ETV1(284), VEGFA(521)
THADA	5.5E-03	HMGA2(3), WDR22(6), JAZF1(23), LHFP(24), MTAP(169), CDKN2B(275), BCL11A(300), MDS1(332), GALNTL1(358), RPL32(428), WDR3(458), KLHDC5(499), RICH2(575)
HNF1B	5.6E-03	HNF1A(1), PCBD2(12), APOM(24), RPL32(65), PITX1(86), CPA1(89), WFS1(93), TNF(100), HMGA2(103), TCF7L2(111), POU5F1(153), PRKACB(271), BAT1(317), PPARG(340), ELAC2(401), FAM148B(412), ACADS(429), HHEX(474), ETV1(492), CUGBP2(496), CFB(553), AGT(587), CDKN2A(608)
NOTCH4	0.010	NOTCH2(3), RPL32(60), HMGA2(90), CPA1(98), EGFL8(110), TLE4(150), SALL3(153), BAT3(194), ETV1(205), PRKACB(224), BAT2(232), POU5F1(244), PITX1(245), TNF(287), ELAC2(307), CUGBP2(348), VEGFA(398), PBX2(435), BAT5(437), ITGA2(440), BAT1(486), TNFSF15(554), TBX18(581)
AGER	0.013	TNF(15), VEGFA(25), TNFSF15(40), AGT(55), APOE(62), ITGA2(77), PPARG(85), HNF1B(117), LTA(119), ITGA1(161), HNF1A(168), IFNGR1(209), FPRL1(210), ELAC2(252), HTRA1(305), HMGA2(324), BAT2(353), HSPA1B(398), PRKCE(411), PRKACB(472), RPL32(587)
TBX18	0.014	MSGN1(6), PITX1(69), NOTCH2(105), SALL3(122), HHEX(271), ETV1(311), POU5F1(325), PRKACB(342), SPRY4(351), RPL32(366), CUGBP2(377), TLE4(384), CPA1(402), NOTCH4(447), HMGA2(483), BAT1(507)
AP3B1	0.015	MUTED(2), RAB38(16), AFTPH(54), SCARB2(77), RPL32(96), PRKACB(104), NEU1(107), VPS33B(115), COG2(139), CPA1(140), COG6(218), SAR1B(234), KIAA0152(266), BAT3(393), CSNK2B(480), HMGA2(491), PELO(511), ITGA2(545), BAT1(561), BAT5(619)
NRXN1	0.017	NRXN3(1), MAGI1(71), FCHSD2(156), SYN2(157), PRKACB(163), SYT16(189), RPL32(223), CUGBP2(228), HNT(325), CAMK1D(329), BAT1(348), CSNK2B(471), BAT3(499), ITGA1(535), PPFIBP1(578)
SPRY2	0.020	SPRY4(3), FGF9(23), RAF1(50), PRKACB(107), PTK2B(141), SALL3(236), LRIG1(255), PITX1(260), BCAR1(291), PRKCE(331), VEGFA(533), RPL32(553), CAMK1D(559), CSNK2B(607), HMGA2(614)
NOTCH2	0.021	NOTCH4(21), TLE4(63), SALL3(77), TBX18(82), POU5F1(161), PRKACB(188), PITX1(197), CPA1(232), RPL32(276), MSGN1(288), MSI2(369), ETV1(384), SPRY4(423), CUGBP2(447), SPRY2(507), PBX2(519), HHEX(527), HMGA2(595)
FAH	0.021	EPHX2(10), HIBCH(16), RPL32(65), MAT1A(71), HNF1A(72), NEU1(95), CYP21A2(106), HNF1B(118), CPA1(122), HMGA2(173), CTSC(201), ACADS(224), TNF(251), PRKACB(334), BAT3(398), ELAC2(411), CUGBP2(457), PELO(504), WFS1(516), BAT1(518), HADHA(523), MTAP(558)
WFS1	0.022	OTOF(7), HNF1B(61), HNF1A(62), CTSC(75), ELAC2(140), DFNB31(170), CPA1(211), RPL32(227), KCNQ1(239), IMPA2(304), SEC61A2(318), SSR1(328), CYP21A2(345), SAR1B(459), ABCC8(475), AP3B1(496), TNF(514), CREBL1(528), HMGA2(530), NEU1(533), TXNDC5(567), CPA5(577)
MDS1	0.022	BCL11A(24), RPL32(28), HMGA2(40), L3MBTL(86), MSI2(167), TACC2(207), CPA1(228), ETV1(267), SETD4(312), CYP21A2(323), PRKACB(386), CUGBP2(447), IFT52(457), BAT1(462), PBX2(497), WDR22(529), WDR3(569), SKIV2L(570)
FAT3	0.026	CDH4(12), PTPRT(137), LRIG1(156), RPL32(178), EGFL8(186), SALL3(207), CUGBP2(227), HNT(240), MAGI1(296), MSI2(404), SLIT3(440), TLE4(457), FTO(474), PRKACB(543), PHF15(565)
CDH4	0.026	FAT3(53), MAGI1(129), PTPRT(138), TLN2(218), ITGA2(231), HNT(269), RPL32(271), ETV1(277), CPA1(283), POU5F1(294), PITX1(311), SALL3(384), HMGA2(417), PRKACB(431), CAMK1D(513), ITGA1(520), TNC(566), TLE4(612), CUGBP2(622)
RAF1	0.029	RREB1(31), SPRY2(63), PRKCE(116), PRKACB(121), PPP2R2C(150), SPRY4(157), PTK2B(179), CSNK2B(206), MAP3K5(218), CAMK1D(264), HMGA2(294), SGK2(356), TNF(359), BCAR1(462), TNFSF15(514), IRS1(534), RPL32(535), NFATC1(562), BAT3(568)
ATP6V1G2	0.032	ATP6V1G1(1), ATP6V0E1(6), ATP6V1B2(8), LY6G6D(32), C6orf47(37), BAT5(40), LY6G6C(41), BAT1(45), C6orf26(47), ATP9B(49), LY6G5B(61), BAT4(70), C6orf25(76), BAT3(78), NFKBIL1(92), EGFL8(97), RPL32(106), BAT2(111), SKIV2L(113), RDBP(191), HSPA1L(232), LY6G5C(268), KIAA0152(279), VARS(320), WDR3(324), COG6(329), CPA1(393), GTPBP2(417), LTA(452), COG2(510), LST1(539), TCF19(554)
TNC	0.034	TNXB(1), CREBL1(6), ITGA1(23), ITGA2(32), HMGA2(105), TNF(183), RPL32(187), PRKACB(201), VEGFA(220), ETV1(225), PITX1(303), BAT3(313), HTRA1(321), CUGBP2(367), CPA1(377), LRIG1(442), POU5F1(466), TLN2(509)
SATB2	0.035	CLPTM1(12), PITX1(37), RPL32(65), HMGA2(90), SALL3(152), BAT3(225), BAT4(233), CPA1(249), BAT5(263), CUGBP2(270), WDR22(291), PELO(292), BAT1(296), CFDP1(307), PRKACB(327), LHFP(391), PBX2(422), FTO(433), ZMIZ1(441), RDBP(540), KIAA0152(558), TBX18(613)
RREB1	0.036	ZNF239(4), RAF1(44), ZNF32(52), ZFP1(86), ZFP36L2(135), BCL11A(269), RPL32(271), MYNN(284), ZNF175(301), CDKN2A(314), TAF4(475), HMGA2(520), PITX1(535), PRKACB(550), CREBL1(615), SALL3(623)
RAB38	0.036	MUTED(3), AP3B1(11), RAB10(25), SAR1B(88), CPA1(168), RPL32(180), COG6(185), VPS33B(192), COG2(228), GTPBP2(318), ATP9B(320), HMGA2(340), BAT1(442), SEC24A(527), NEU1(532), PRKACB(540), SCARB2(587), BAT3(609)
IDE	0.036	IRS1(5), APOE(46), HNF1A(103), RPL32(130), HTRA1(153), PPARG(160), TNF(165), ABCC8(166), APOC1(169), ELAC2(204), PRKACB(216), CSNK2B(225), CPA1(242), SLC30A8(278), FURIN(295), KCNJ11(367), HNF1B(407), BAT3(419), HIBCH(423), APOM(508), AGT(538)
NRXN3	0.038	NRXN1(2), SYN2(39), FCHSD2(136), PRKACB(163), CAMK1D(177), GRM5(200), KIAA0152(202), CHRNA2(220), RPL32(228), TNF(277), ETV1(312), ACTN1(323), ITGA1(352), CUGBP2(420), MAGI1(445), BAT1(559), ITGA2(560)
VEGFA	0.041	TNFSF15(24), TNF(47), AGGF1(110), ITGA2(120), SMOC2(149), AGT(185), SPRY4(266), HMGA2(276), TNC(289), PPARG(311), SPRY2(316), APOE(334), ITGA1(369), HTRA1(413), LTA(420), PRKCE(432), FGF9(531), CDKN2A(545), RPL32(557), PITX1(560), AGER(568), NOTCH4(601), NFATC1(614)
TCF7L2	0.041	CCDC88C(68), HNF1A(91), SPRY4(133), PITX1(142), HNF1B(192), PRKACB(194), POU5F1(206), HMGA2(241), RPL32(254), ELAC2(255), TNF(296), NFATC1(362), TLE4(417), ITGA2(459), CSNK2B(464), ETV1(473), RREB1(490), PPARG(505), BAT1(514), SALL3(520), HHEX(524), ZMIZ1(525), CPA1(545)
ZFP36L2	0.041	ZNF239(3), ZFP1(50), RREB1(53), ZNF32(62), RPL32(138), CUGBP2(252), BAT1(257), BCL11A(261), SLC30A8(390), JAZF1(392), MYNN(411), SKIV2L(415), PHF15(418), CFDP1(497), RDBP(557), RNF5(600), CPA1(611), ZNF175(613)
POU5F1	0.042	MSI2(57), HHEX(97), PITX1(102), RPL32(131), SALL3(233), TLE4(261), ETV1(294), PELO(303), HMGA2(305), PRKACB(353), CPA1(374), SPRY4(410), NOTCH2(567), CUGBP2(570)
CDKN2B	0.045	CDKN2A(1), MTAP(4), HMGA2(72), CCNE2(92), RPL32(123), TNF(180), L3MBTL(181), E2F3(183), TP53INP1(273), RREB1(295), PRKACB(308), CPA1(346), BAT3(363), TNFSF15(388), BCL11A(401), SPRY4(406), HTRA1(417), PITX1(429), ELAC2(454), JAZF1(462), PELO(491), MSI2(513)
FGF9	0.046	SPRY4(28), SPRY2(29), PITX1(88), POU5F1(112), ETV1(187), HMGA2(291), PRKACB(310), VEGFA(312), CPA1(325), SALL3(328), RPL32(399), TNC(401), TNFSF15(514), TNF(526), NOTCH2(549), HTRA1(576)
HDDC3	0.047	SRBD1(1), EXDL2(8), CDKAL1(17), C6orf206(48), C8orf38(121), TMEM170(149), RCCD1(165), C2orf39(322), PLEKHH2(348), INTS8(568)

Supplementary Note Table SN17: List of connections from the GRAIL analysis for selected genes (P_{gene} < 0.05) for the analysis shown in Table SN16. GRAIL P_{gene}-values listed here are the results for the test of the gene itself, and thus are not adjusted for the number of genes tested within a locus. We present here the results for the top GRAIL P_{gene} < 0.05.

PANTHER

We used the Panther database (http://www.pantherdb.org/)[S70,S71] to search for biological processes over-represented by genes at T2D associated loci by comparing gene lists (described above) from T2D associated loci to a reference list containing 25,431 human genes, 14,110 with available biological process annotation. We asked whether the proportion of input genes in a given biological pathway or function was greater than the proportion in the reference list of 14,110 biologically-annotated genes. The P-values were adjusted for multiple testing by generating 20,000 random sets of genes, each including the same number of genes as the set used in the original analysis. For each of the 20,000 gene sets the correlation with the 242 biological processes was re-tested and the P_{adj}-value was calculated as the fraction of sets that produced a P-value smaller than or equal to the original P value.

All 82 genes in the confirmed locus list were identified in the database and biological process annotation was available for 62 genes. No pathways were significant after adjustment for multiple testing but the highest ranking processes were "Cell adhesion-mediated signaling", "Signal transduction" and "Cell cycle control". A total of 310 out of 320 genes on the expanded (minus HLA) list (which was prioritized for this analysis) were identified and biological process annotation was available for 230 genes. This analysis yielded two significant biological processs (P<0.05 after correction) (**Table SN18**). The most prominent signal was obtained for signal transduction with 64 observed hits in the expanded (minus HLA) analysis and 19 in the confirmed locus set. Those 19 genes in the confirmed locus set map to a total of 13 of the confirmed loci and include the genes listed in parentheses after each locus name

- *NOTCH2* locus (*NOTCH2, ADAM30, FCGR1B)*
- *THADA* locus (*ZFP36L2*)
- *IRS1* locus (*IRS1*)
- *PPARG* locus (*PPARG*)
- *ADAMTS9* locus (*ADAMTS9*)
- *CDKAL1* locus (*E2F3*)
- *CAMK1D* locus (*CAMK1D)*
- *CENTD2* locus (*CENTD2, PDE2A, P2YR2*)
- *MTNR1B* locus (*MTNR1B, FAT3*)
- *TSPAN8* locus (*TSPAN8*)
- *HNF1A* locus (*SPPL3*)
- *ZFAND6* locus (*ZFAND6*)
- *PRC1* locus (*FURIN, FES*).

Biological Process	NCBI: H. sapiens genes (REF) Total	Expanded minus HLA 310 gene IDs Observed	Expected	P- value	P- value adj
Signal transduction	3406	64	41.52	0.000269	**0.017**
Protein metabolism & modification	3040	57	37.06	0.000657	**0.043**
Transport	1306	30	15.92	0.000748	0.051
Cell adhesion-mediated signaling	379	13	4.62	0.000913	0.062

Supplementary Note Table SN18: PANTHER analysis using the "expanded (minus HLA) locus list". Only the top two processes (in bold) were significant (P<0.05) after Bonferroni correction: "Signal transduction" and "Protein metabolism and modification".

Reactome

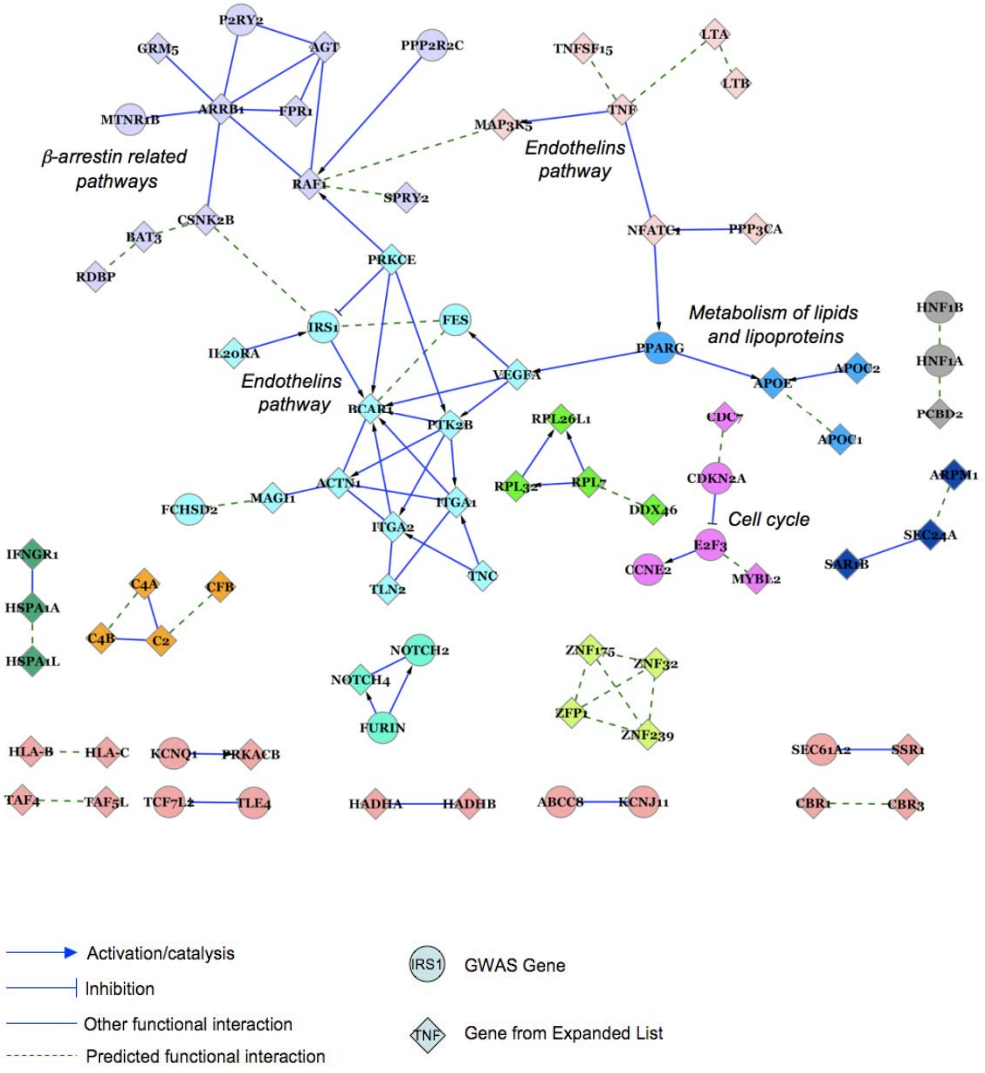
We used the Reactome database for pathway enrichment analysis[S72,S73]. Reactome provides author-curated validated summaries of key biological processes. To increase the protein/gene coverage of the Reactome pathways, these core curated pathways are decorated with annotations provided by importing information from four other databases: Panther (http://www.pantherdb.org), Nature-PID (http://pid.nci.nih.gov), KEGG (http://www.genome.jp/kegg), and Cell-Map (http://cancer.cellmap.org). In the enrichment analysis, we calculated P-values (comparing the proportion of genes in a given pathway represented in the T2D-associated gene list as compared to that in the total annotated gene list) based on a binomial test of proportion. For a given P-value cutpoint the false discovery rate (FDR) was calculated as $n_{empirical\ <P\text{-}value}/(n_{obs\ <P\text{-}value} * 1000)$[S74]. Pathways with FDR ≤ 0.20 were considered significant.

In **Table SN19** we show all pathways attaining FDR<0.20 for the "confirmed" and "expanded" locus sets (82 and 383 recognised genes included respectively). When the single HLA-related locus was removed to generate the "expanded minus HLA" list, the "natural killer cell mediated cytotoxicity" and "SLE" pathways were no longer significant (as might have been expected given their heavy dependence on genes mapping to the HLA LD interval). The "notch-signalling" pathway also became non-significant after excluding the HLA signal (presumably because one of the group of genes within this pathway maps to the HLA region [NOTCH4]).

In addition to these analyses, and to create and visualize gene set interaction networks, we used a functional interaction (FI) network generated by curated interactions extracted from reactions and complexes annotated in pathways, combined with bimolecular interactions predicted using a Naïve Bayes Classifier (NBC). This NBC was trained using features from protein-protein interactions in human, yeast, worm and fly, gene co-expression from DNA array experiments, protein-protein interactions from text mining, domain-domain interactions, and Gene Ontology biological annotations. To construct a network among a set of genes, we first hierarchically clustered genes based on shortest path among each pair of the genes, and then added linking genes between two child clusters if needed. For network clustering, we used an "edge betweenness" algorithm[S75]. The final diagrams were drawn in Cytoscape[S76] and the functional interactions annotated by a Cytoscape plug-in developed for the FI network[S77] (**Figure SN7).**

Pathway	# proteins in pathway	# proteins from sample	P-Value	FDR	IDs
CONFIRMED LOCUS LIST					
Maturity onset diabetes of the young	23	3	1.46E-04	0.02	*[HHEX, HNF1B, HNF1A]*
FOXA2 and FOXA3 transcription factor networks	39	3	6.81E-04	0.05	*[KCNJ11, ABCC8, HNF1A]*
Cell cycle	112	4	1.42E-03	0.08	*[E2F3, CDKN2A, CCNE2, CDKN2B]*
Signaling by Notch	16	2	2.32E-03	0.11	*[FURIN, NOTCH2]*
Regulation of retinoblastoma protein	61	3	2.45E-03	0.09	*[PPARG, E2F3, CDKN2A]*
cyclins and cell cycle regulation	19	2	3.24E-03	0.11	*[CDKN2A, CDKN2B]*
cell cycle: g1/s check point	22	2	4.31E-03	0.11	*[CDKN2A, CDKN2B]*
growth hormone signaling pathway	22	2	4.31E-03	0.11	*[IRS1, HNF1A]*
FOXA transcription factor networks	75	3	4.36E-03	0.10	*[KCNJ11, ABCC8, HNF1A]*
Thyroid cancer	25	2	5.52E-03	0.11	*[PPARG, TCF7L2]*
EXPANDED LOCUS LIST					
Vibrio cholerae infection	38	6	1.56E-04	0.03	*[ATP6V1G2, ATP6V0E1, ATP6V1B2, ATP6V1G1, PRKACB, KCNQ1]*
Natural killer cell mediated cytotoxicity	116	10	1.65E-04	0.02	*[NFATC1, PPP3CA, NCR3, TNF, IFNGR1, PTK2B, HLA-B, MICA, MICB, HLA-C]*
role of ß-arrestins in the activation and targeting of map kinases	28	5	3.21E-04	0.02	*[P2RY2, FPR1, ARRB1, RAF1, AGT]*
roles of ß arrestin dependent recruitment of src kinases in gpcr signaling	32	5	5.86E-04	0.03	*[P2RY2, FPR1, ARRB1, RAF1, AGT]*
Metabolism of lipids and lipoproteins	103	8	1.46E-03	0.06	*[APOE, ACADS, HADHA, PRKACB, CYP21A2, HADHB, SAR1B, APOC2]*
Systemic lupus erythematosus	11	3	1.64E-03	0.06	*[C4A, C2, C4B]*
ß-arrestins in gpcr desensitization	26	4	2.23E-03	0.07	*[P2RY2, FPR1, ARRB1, AGT]*
Endothelins	168	10	2.74E-03	0.08	*[NFATC1, GNAL, PPARG, PRKCE, IRS1, TNF, BCAR1, RAF1, PTK2B, MAP3K5]*
Signaling by Notch	16	3	4.69E-03	0.13	*[NOTCH4, NOTCH2, FURIN]*
FGF signaling pathway	102	7	5.62E-03	0.14	*[PRKCE, PPP2R2C, SPRY2, SPRY4, RAF1, MAP3K5, FGF9]*
sprouty regulation of tyrosine kinase signals	18	3	6.48E-03	0.14	*[SPRY2, SPRY4, RAF1]*

Supplementary Note Table SN19: Reactome analyses showing all pathways with FDR<20%. Note that these analyses do not adjust for regions with multiple related genes (e.g. *KCNJ11, ABCC8*) and therefore may overestimate significance where multiple genes within a pathway map to a single locus.



Supplementary Note Figure SN7: Functional interaction network generated for the EXPANDED locus set. Of the 383 genes that could be matched, 212 were represented in the FI network and 82 genes are displayed in the subnetworks shown. Genes in different colors are in different clusters except those in clusters containing two genes only, which are in brown at the bottom of the figure. Genes in the same network cluster are assumed to have same or similar function.

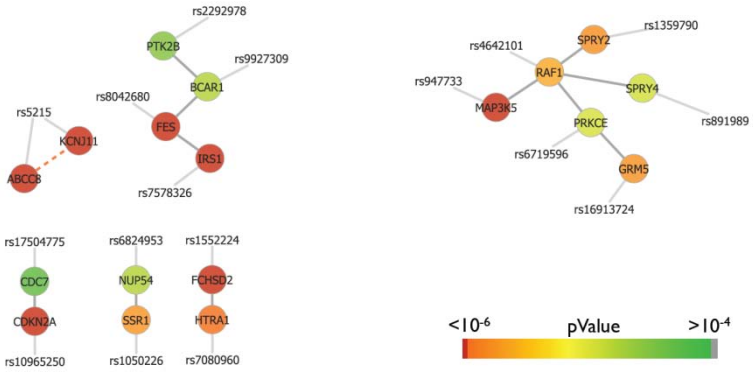
GWAS-PPI

The GWAS-PPI approach relies on mapping genes within regions of GWAS association to networks based on empirical protein-protein interaction data without direct recourse to pathway data. Thus, genes within the canonical gene lists were mapped to the Human Protein Reference Database (HPRD release 8)[S78] protein-protein interaction network by HUGO gene name. To construct networks without prior knowledge of the associated gene function, a protein-protein interaction network of SNP-associated gene products was built by stepwise addition of individual gene products (nodes) to the network, using the P-value of the maximally-associated SNP in each region to define the order of addition (starting from the most-significant). We initially focused on adding genes within the expanded locus list, but later extended this to the "further expanded locus list" which includes all loci containing SNPs with $p<10^{-3}$ on the Stage 1 meta-analysis.

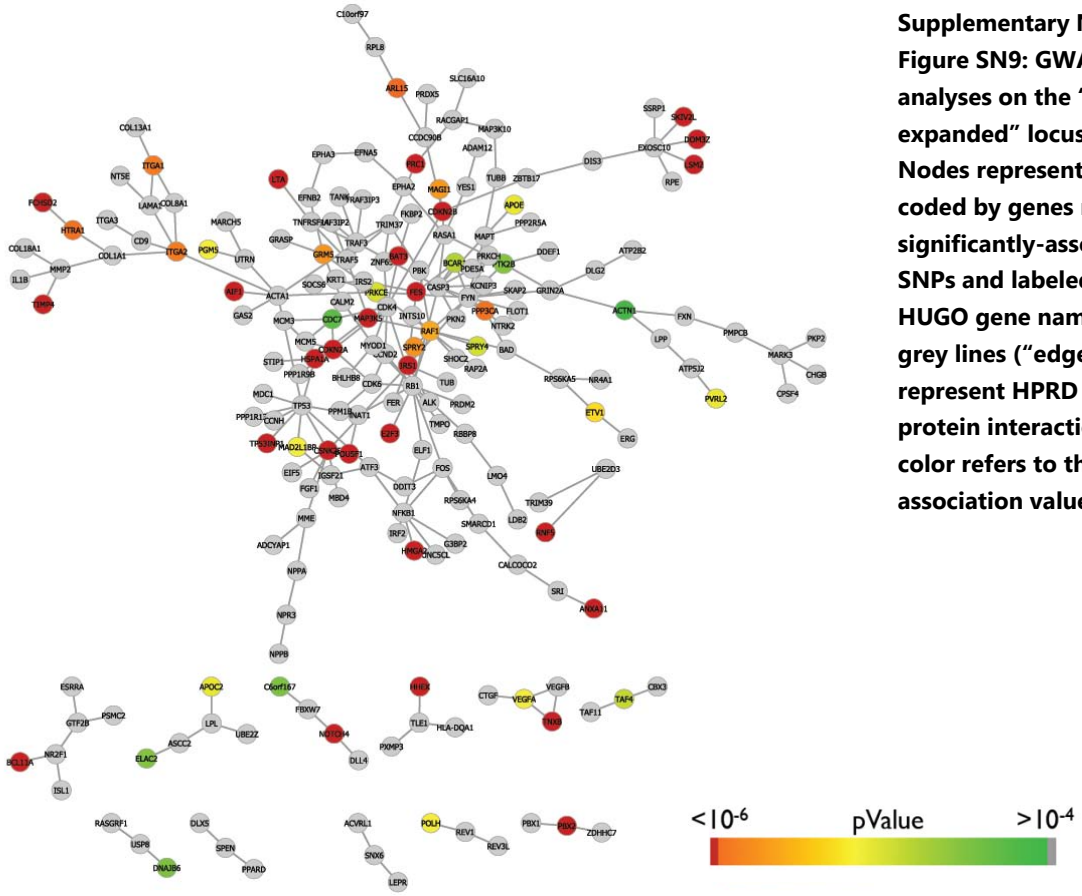
Edges between nodes (genes) were added if a protein-protein interaction between a newly added node and any node already present in the network was found to be recorded in HPRD. Multiple nodes associated with the same SNP, and hence the same significance value, were added in random order. To avoid a gene clustering effect during network construction, interactions between genes associated with the same SNP were flagged as likely false positives and removed during visualization. All nodes not connected to any other node were ignored.

Out of 82 genes in the "confirmed" gene list, 44 have known interaction partners in HPRD, but only 4 interact with other genes in the same list. These include the *KCNJ11-ABCC8* link which is flagged as a likely false positive, as both are associated with the same SNP (though in truth, the products of these two genes do indeed interact as the two components of the heteroctomeric beta-cell K_{ATP} channel). The other observed interaction was between FES (which maps to the *PRC1* locus) and IRS1. Equivalent analyses involving the expanded list (minus HLA) identified a total of 166 genes with products represented in HPRD. The previously-identified interactions between FES and IRS1 extend to include BCAR1 and PTK2B and a larger subnetwork centered on RAF1 emerges. Excluding KCNJ11-ABCC8, there are 16 genes with 11 binding interactions in the resulting network **(Figure SN8)**. Inclusion of the HLA signal led to expansion of some of these networks (and identification of additional networks based around known HLA interactions) but we regard these with some caution (because of the disproportionate size and content of the LD interval around HLA and the enhanced chance that we might be picking up subtle stratification effects at this signal which did not reach genome-wide significance in our combined Stage 1 and Stage 2 analyses (rs1048709: **Table SN4**; **Supplementary Table 3**).

We wanted to see if including a larger set of genes/proteins associated with less pronounced association P-values enriched network identification and extended these analyses to the "further expanded data set". This resulted in a large network which becomes difficult to disentangle manually (**Figure SN9**).



Supplementary Note Figure SN8: GWAS-PPI analyses on expanded locus set minus HLA. Nodes represent proteins coded by genes mapped to significantly-associated SNPs and labeled with HUGO gene names. Solid grey lines ("edges") represent HPRD protein-protein interactions. Node color refers to the T2D-association value. As described in the text, the ABCC8-KCNJ11 interaction is here labeled as a false positive, though in fact the products of both genes do interact to create the Kir6.2 channel.



Supplementary Note Figure SN9: GWAS-PPI analyses on the "further expanded" locus set. Nodes represent proteins coded by genes mapped to significantly-associated SNPs and labeled with HUGO gene names. Solid grey lines ("edges") represent HPRD protein-protein interactions. Node color refers to the T2D-association value

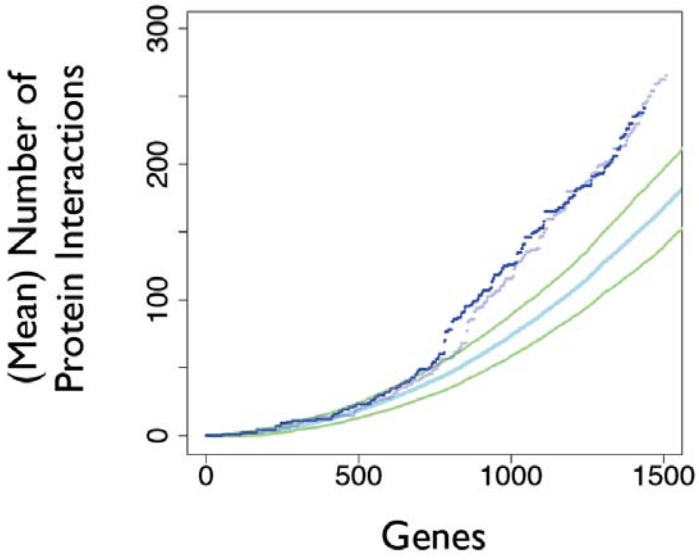
Suppl

To determine whether or not the extent of protein interaction observed was greater than expected by chance, this large network was compared to a set of "random" networks, generated by randomly selecting a number of gene identifiers from the HUGO gene name table (Ensembl, Release 55) equal to the number of genes contained in the GWAS correlated marker set. We compared edge counts found in the empirical T2D-association gene set network to the distribution of signals found in 100 such random networks. We repeated this exercise as we increased the number of proteins added to the network (in order of T2D-association P-value) and observed a higher than expected edge count in the T2D-associated network (whether or not we included HLA-associated loci). Notably, this increased signal only became evident as the number of genes/proteins included was increased to ~800 or more (ie beyond the number in the "expanded" gene list) (**Figure SN10).**

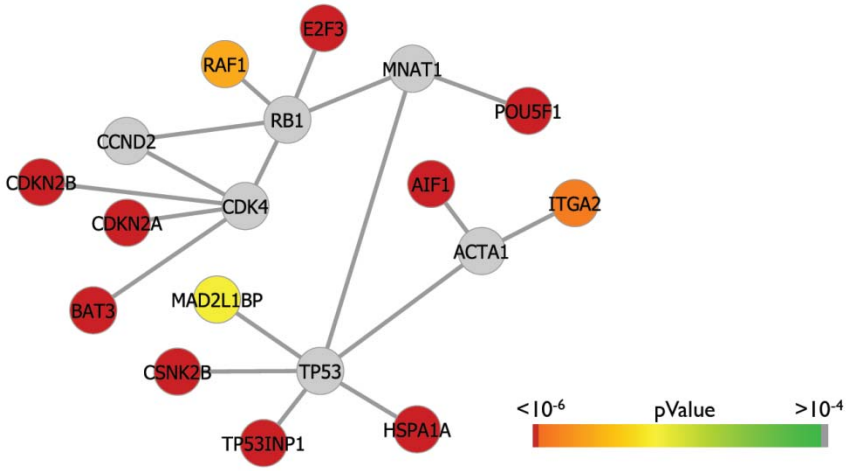
Finally, to tease out interesting functional subunits from the largest network (ie the "further expanded locus list") we searched for subnetworks characterized by particularly low T2D-association P-values SNPs. We applied the jActiveModules algorithm [S79,S80] utilizing the Cytoscape framework (http://www.cytoscape.org/). The standard search (search depth 2, scoring for size and regional scoring) identified one particularly notable subnetwork. This network, which includes 18 protein members, was the only large (>3 members) network to generate a normalized aggregate subnetwork T2D-association Z-score >2 (in this case 3.00, so equivalent to p=0.004). This network still emerged as the most significant network if we corrected for instances of multiple related genes within the same locus (eg at chr9) (**Figure SN11**).

The identified subnetwork was studied for functional enrichment using GeneGo's MetaCore system[S81], and demonstrated strong enrichment for components of cell-cycle regulation ($p \sim 2.8 \times 10^{-7}$) (**Figure SN12)**. This subnetwork is partly overlapping with the cell-cycle network revealed by the bimolecular interaction analyses described earlier (in the Reactome section): the two analyses also partly overlap in the databases used.

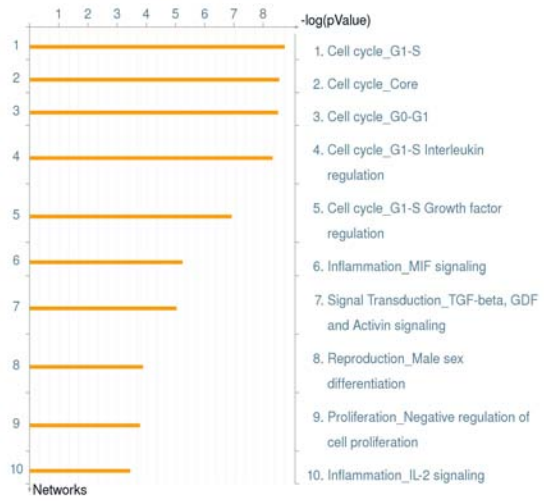
This cell-cycle regulation signal emerges strongly even though the PPI databases did not provide links to several additional genes within confirmed association intervals that have putative roles in cell-cycle regulation and which constitute excellent positional candidates. These include *CDKAL1* (a putative regulator of CDK5); *CCNE2* (within the *TP53INP1* locus); and *CDKN1C* (the likely effector gene for the *KCNQ1* association). We conclude that this analysis provides further evidence supporting the role of cell cycle regulation in the pathogenesis of T2D.



Supplementary Note Figure SN10: SNP-associated genes coding for proteins present in the "further expanded" data set (dark blue) plotted in the order of the T2D-association P-value versus the number of their protein-protein interactions present in HPRD. The number of protein-protein interactions is compared to the distribution of protein-protein interaction counts (mean in light blue, 1SD in green) obtained from conducting a similar analysis on 100 different "random" protein lists generated from protein identifiers in the ENSEMBL database. The empirical data plotted are the "further expanded" gene list before (light blue) and after (dark blue) removing genes mapping to HLA-associated loci.



Supplementary Note Figure SN11: Subnetwork characterized by highly-associated P-values. Nodes represent proteins derived from genes mapping to intervals in the "further expanded" locus list. Solid grey lines (edges) represent HPRD protein-protein interactions. Node colors refer to the T2D-association P-value of the maximally associated SNP in the cognate association interval



Supplementary Note Figure SN12: Functional assessment of the 18-protein subnetwork revealed within the "further expanded" gene list.

GENE SET ENRICHMENT ANALYSIS

Using a gene set enrichment analysis (GSEA) approach, we tested a number (30) of specific hypotheses relating to the etiology of T2D, proposed on the basis of previous genetic and/or functional studies. In particular, we tested whether annotated pathways related to insulin signaling, inflammation, cell cycle, fat oxidation, amyloid metabolism, or vitamin D metabolism were enriched for genes associated with T2D.

The GSEA approach used is based on one previously developed for expression data[S74,S82], and analyzes SNP data from GWA studies in the context of genes and sets of functionally related genes, testing whether given biological processes are enriched for genes associated with a chosen polygenic disease or trait. This method, called Meta-Analysis Gene-set Enrichment of variaNT Associations (MAGENTA)[S83] does not require genotype data as input but only SNP association statistics, making it especially amenable for analyzing GWA meta-analyses. Briefly, the method involves: (i) assigning to each gene in the chosen annotated pathway, and for the trait of interest, the most significant association score (z-score) amongst all SNPs within 110kb upstream and 40kb downstream of the gene's most extreme transcript start and end sites, respectively; (ii) correcting for confounders on the best-SNP-per-gene score using a step-wise multivariate linear regression model that regresses out the effects of five previously identified confounders (physical gene size, density of SNPs per gene, density of estimated number of independent SNPs per gene (i.e. SNPs in linkage equilibrium), linkage disequilibrium units per gene, and density of number of recombination hotspots spanning each gene) to generate a "normalized gene score"; (iii) Defining the sets of functionally related genes to be tested; if multiple genes within a given gene-set contained the same strongly associated SNP, the gene with the most significant normalized gene score was retained (this was done to prevent overestimation of the gene set enrichment P-value due to two or more genes in a gene set that lie next to each other in the genome); and (iv) Testing whether any of the given gene sets are enriched for highly ranked gene association scores (after adjustment for confounders and positional clustering) above a given gene score cutoff. An enrichment P-value was defined as the fraction of 10,000 gene sets, randomly sampled from the genome, whose number of genes above a given gene score cutoff ("highly ranked" genes) was similar to or larger than the number of highly-ranked genes in the tested gene set. The size of the randomly sampled gene sets was identical to that of the queried gene set, following adjustment for positional clustering of subsets of genes in each gene set. We used a Bonferroni correction to account for the testing of multiple functional categories. To improve the power of the analysis, and in contrast to the other analyses presented in this section, we used the 95th percentile of all gene scores calculated from the DIAGRAM+ study as the enrichment threshold. This resulted in a total of 920 genes above the 95^{th} percentile threshold.

Across the Gene Ontology (http://www.geneontology.org), KEGG (http://www.genome.jp/kegg), Ingenuity (http://www.ingenuity.com), and PANTHER (http://www.pantherdb.org) databases, we identified and tested 30 sets related to insulin signaling, inflammation, cell cycle, fat oxidation, amyloid metabolism, and vitamin D metabolism, using MAGENTA (**Table SN20**). Overall, we observed that gene sets related to cell cycle, inflammatory response, and fatty acid oxidation were nominally enriched for genes association scores that ranked in the top 5% most significant scores in the DIAGRAM+ meta-analysis (GSEA $p<0.05$). Vitamin metabolism, amyloid processes and insulin signaling were not significantly enriched for T2D gene

associations relative to genomic background (GSEA p>0.05). The results for the cell cycle and fatty acid oxidation groups, but not for inflammatory response, were primarily driven by association to confirmed loci. No gene set was significantly associated with T2D after a Bonferroni correction for multiple testing (nominal p<0.0017), though this threshold is quite conservative given the considerable overlap between gene sets within a given hypothesis. The cell cycle hypothesis was not rejected after a more modest adjustment for the number of hypotheses tested (6 hypotheses requires a nominal p < 0.0083).

Hypothesis	Database	Gene set	# genes in category tested by GSEA	# genes linked to confirmed T2D SNPs	Nominal GSEA P-value	Observed # genes above cutoff	Expected # genes above cutoff	Genes linked to confirmed T2D SNPs
Fat oxidation	GO, biological process	Fatty acid beta oxidation (GO:0006635)	11	1	**0.0144***	3	1	*ACADS*
	GO, biological process	Fatty acid oxidation (GO:0019395)	18	1	0.0580*	3	1	*ACADS*
	GO, biological process	Fatty acid metabolic process (GO:0006633)	57	1	0.1582	5	3	*ACADS*
	KEGG	Fatty acid elnogation in mitochondria	8	0	0.3323	1	<1	-
	KEGG	Fatty acid metabolism	45	1	0.3877	3	2	-
	GO, biological process	Fatty acid biosynthetic process (GO:0006633)	12	0	0.4601	1	1	-
Insulin signaling	KEGG	Insulin signaling pathway	127	1	0.2972	8	6	-
	Ingenuity	Insulin receptor signaling	33	1	0.497	2	2	-
	GO, biological process	Insulin receptor signaling pathway (GO:0008286)	18	1	0.6036	1	1	*IRS1*
	Panther	Insulin/IGF pathway-mitogen activated protein kinase kinase/MAP kinase cascade	19	0	0.6277	2	1	-
	Panther	Insulin/IGF pathway-protein kinase B signaling cascade	44	0	0.8937	1	2	-
Amyloid metabolism	Panther	Alzheimer_disease-amyloid_secretase_pathway	24	0	0.7088	1	1	-
	GO, biological process	Amyloid precursor protein metabolic process	10	0	1	0	1	-
Vitamin D metabolism	Panther	Vitamin D metabolism and pathway	13	0	1	0	1	-
	GO, biological process	Vitamin metabolic process (GO:0006766)	16	0	1	0	1	-
Inflammation	GO, biological process	Inflammatory response (GO:0006954)	113	0	**0.0107**	12	6	-
	Ingenuity	IL-6 signaling	27	0	0.1429	3	1	-
	Ingenuity	IL-4.Signaling	16	1	0.2164	3	1	*IRS1*
	GO, biological process	Acute inflammatory response (GO:0002526)	10	0	0.4026	1	1	-
	Ingenuity	IL-10.Signaling	24	0	0.4388	3	1	-
	Ingenuity	IL-2 signaling	16	0	0.5545	1	1	-
	Panther	Inflammation mediated by chemokine and cytokine signaling pathway	83	0	0.9285	2	4	-
Cell cycle	GO, biological process	Cell cycle (GO:0007049)	296	8	**0.0062***	25	15	*CDKN2A, CDKN2B, KIF11, NOTCH2, CDC123, PRC1, CCNE2, CENTD2*
	GO, biological process	M-Phase (GO:0000279)	109	2	**0.0454***	10	5	*CDKN2B, KIF11*
	GO, biological process	Cell cycle process (GO:0022403)	180	5	0.0605	14	9	*CDKN2A, CDKN2B, KIF11, PRC1, CENTD2*
	GO, biological process	Mitotic cell cycle (GO:0000278)	142	5	0.0967	11	7	*CDKN2A, CDKN2B, KIF11, CDC123, PRC1*
	GO, biological process	Cell cycle phase (GO:0022403)	158	3	0.1630	11	8	*CDKN2A, CDKN2B, KIF11*
	GO, biological process	Mitosis (GO:0007067)	78	2	0.1893	6	4	*CDKN2B, KIF11*
	GO, biological process	M phase of mitotic cell cycle (GO:0000087)	81	2	0.2025	6	4	*CDKN2B, KIF11*
	KEGG	Cell cycle	104	3	0.7759	4	5	*CDKN2A, CDKN2B, CCNE2*

Supplementary Note Table SN20: GSEA results from 30 gene sets tested for enrichment of association signals from the overall meta-analytical results. "Confirmed" refers to the 30 autosomal hits from this and previous studies attaining genome wide significance; "tested" refers to the number of genes in each gene set included in the analysis following removal of genes without SNPs in their extended gene boundaries, and removal of all but one gene from each subset of genes assigned the same best local SNP (the gene with the most significant score was retained); "cutoff" refers to the 95 percentile of all gene scores calculated from the DIAGRAM+ study. Results in bold highlight gene sets that were modestly enriched (p<0.05). A stringent Bonferroni cutoff correcting for the total number of gene sets tested is p<0.0017, and a less stringent Bonferroni cutoff correcting for the total number of hypotheses tested is p<0.0083. Results noted by asterisks (*) on the right become non-significant when associations at confirmed loci are removed from the set.

Supplementary Note References

S1. Sladek, R. *et al.* A genome-wide association study identifies novel risk loci for type 2 diabetes. *Nature* 2007;**445**:881-885.

S2. Report of the Expert Committee on the Diagnosis and Classification of Diabetes Mellitus. *Diabet Care* 1997;**20**:1183-1197.

S3. Expert Committee on the Diagnosis and Classification of Diabetes Mellitus.Report of the expert committee on the diagnosis and classification of diabetes mellitus. *Diabetes Care* 2003;**26** Suppl 1:S5-20.

S4. Steinthorsdottir, V. *et al.* A variant in *CDKAL1* influences insulin response and risk of type 2 diabetes. *Nat Genet*. 2007;**39**:770-775.

S5. Thorleifsson, G. *et al*. Genome-wide association yields new sequence variants at seven loci that associate with measures of obesity. *Nat Genet*. 2009;**41**:19-24.

S6. Diabetes Genetics Initiative. Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels. *Science* 2007;**316**:1331-1336.

S7. Zeggini, E. *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**:638-645.

S8. World Health Organization. Definition, diagnosis and classification of diabetes mellitus, report of a WHO consultation, part 1: diagnosis and classification of diabetes mellitus. 1999 WHO, Geneva.

S9. Rudan, I., Campbell, H., Rudan, P. Genetic epidemiological studies of eastern Adriatic Island isolates, Croatia: objective and strategies. *Coll Antropol*. 1999;**23**:531-546.

S10. Aulchenko, Y.S. *et al.* Linkage disequilibrium in young genetically isolated Dutch population. *Eur J Hum Genet*. 2004;**12**:527-534.

S11. Pattaro, C. *et al*. The genetic study of three population microisolates in South Tyrol (MICROS): study design and epidemiological perspectives. *BMC Medical Genetics*. 2007;**8**:29.

S12. Hicks, A.A. *et al*. Genetic determinants of circulating sphingolipid concentrations in European populations. *PloS Genet*. 2009;**5**:e1000672.

S13. Marroni, F. *et al.* Population Isolates in South Tyrol and Their Value for Genetic Dissection of Complex Diseases. *Ann Hum Genet. 2006;* **70:**812-821.

S14. Pardo, L.M., MacKay, I., Oostra, B., van Duijn, C.M. and Aulchenko, Y.S. The effect of genetic drift in a young genetically isolated population. *Ann Hum Genet.* 2005; **69:**288-295.

S15. Scott, L.J. *et al.* A genome-wide association study of type 2 diabetes in Finns detects multiple susceptibility variants. *Science* 2007;**316**:1341-1345.

S16. Saaristo, T. *et al.* Cross-sectional evaluation of the Finnish Diabetes Risk Score: a tool to identify undetected type 2 diabetes, abnormal glucose tolerance and metabolic syndrome. *Diab Vasc Dis Res* 2005;**2**:67-72

S17. Wichmann H.E. *et al*. KORA-gen- resource for population genetics, controls and a broad spectrum of disease phenotypes. *Gesundheitswesen* 2005;**67**:S26-S30.

S18. Herder, C. *et al.* RANTES/CCL5 gene polymorphisms, serum concentrations, and incident type 2 diabetes: results from the MONICA/KORA Augsburg case-cohort study, 1984-2002. *Eur J Endocrinol.* 2008;**158**:R1-5.

S19. Huth, C. *et al.* IL6 Promoter Polymorphisms and Type 2 Diabetes Mellitus:Joint Analysis of Individual Participants' Data from 21 Studies. Diabetes 2006;**55**: 2915-2921.

S20. Hofman, A. *et al.* The Rotterdam Study: objectives and design update. *Eur J Epidemiol.* 2007;**22**:819-829.

S21. The Wellcome Trust Case Control Consortium, Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* 2007;**447**:661-678.

S22. Zeggini, E. *et al.* Replication of genome-wide association signals in UK samples reveals risk loci for type 2 diabetes. *Science* 2007;**316**:1336-1341.

S23. Wiltshire S, *et al.* A genome-wide scan for loci predisposing to type 2 diabetes in a UK population (The Diabetes UK Warren 2 Repository): analysis of 573 pedigrees provides independent replication of a susceptibility locus on chromosome 1q. *Am. J. Human Genet.* 2001;**69**:553-569.

S24. Frayling, T.M. *et al.* Parent-offspring trios: a resource to facilitate the identification of type 2 diabetes genes *Diabetes* 1999;**48**:2475-2479.

S25. Groves, C.J. *et al.* Association analysis of 6736 UK subjects provides replication and confirms TCF7L2 as a type 2 diabetes susceptibility gene with a substantial effect on individual risk. *Diabetes* 2006;**55**:2640-2644.

S26. Strachan DP, *et al.* Lifecourse influences on health among British adults: Effects of region of residence in childhood and adulthood. *Int. J. Epidemiol.* 2007;**36**:522-531.

S27. The ARIC investigators. The Atherosclerosis Risk in Communities (ARIC) Study: design and objectives. *Am J Epidemiol.* 1989;**129**:687–702.

S28. Psaty, B.M. *et al*. Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) Consortium: design of prospective meta-analyses of genome-wide association studies from five cohorts. *Circ Cardiovasc Genet.* 2009;**2**:73–80.

S29. Dawber, T.R., Kannel, W.B., Lyell, L.P. An approach to longitudinal studies in a community: the Framingham Study. *Ann NY Acad Sci.* 1963;**107**:539–556.

S30. Kannel, W.B., Feinleib, M., McNamara, P.M., Garrison, R.J., Castelli, W.P. An investigation of coronary heart disease in families. The Framingham offspring study. *Am J Epidemiol.* 1979;**110**:281–290.

S31. Splansky, G.L. *et al*. The Third Generation Cohort of the National Heart, Lung, and Blood Institute's Framingham Heart Study: design, recruitment, and initial examination. *Am J Epidemiol.* 2007;**165**:1328–1335.

S32. Colditz G.A. *et al.* The Nurses' Health Study: lifestyle and health among women. *Nat Rev Cancer* 2005;**5**:388-396.

S33. National Diabetes Data Group. Classification and diagnosis of diabetes mellitus and other categories of glucose intolerance. *Diabetes* 1979;**28**:1039-1057.

S34. Lauritzen, T. *et al*. The ADDITION study: proposed trial of the cost-effectiveness of an intensive multifactorial intervention on morbidity and mortality among people with type 2 diabetes detected by screening. *Int J Obes.* 2000;**Suppl 3:**S6-S11.

S35. Loos, R.J. *et al*. TCF7L2 polymorphisms modulate proinsulin levels and beta-cell function in a British Europid population. *Diabetes* 2007;**56:**1943-1947.

S36. Van Tilburg, J.H.O. *et al.* A genome-wide scan in type 2 diabetes mellitus provides independent replication of a susceptibility locus on 18p11 and suggests the existence of novel loci on 2q12 and 19q13. *J Clin Endocrinol Metab* 2003;**88**:2223-2230.

S37. Monsuur, A.J., *et al*. Myosin IXB variant increases the risk of celiac disease and points toward a primary intestinal barrier defect. *Nat Genet.* 2005;**37**:1341-1344.

S38. Alberti K.G., Zimmet P.Z. Definition, diagnosis and classification of diabetes mellitus and its complications. Part 1: diagnosis and classification of diabetes mellitus provisional report of a WHO consultation. *Diabet Med.* 1998;**15**:539-553.

S39. Halsall, D.J. *et al*. Typical type 2 diabetes mellitus and HFE gene mutations: a population-based case-control study. *Hum Mol Genet.* 2003;**15:**1361-1365.

S40. Rung, J. *et al.* Genetic variant near IRS1 is associated with type 2 diabetes, insulin resistance and hyperinsulinemia. *Nat Genet.* 2009;**41**:1110-1115.

S41. Lafay, L. *et al.* Determinants and nature of dietary underreporting in a free-living population: the Fleurbaix Laventie Ville Sante (FLVS) Study. *Int J Obes Relat Metab Disord.* 1997;**21**:567-573.

S42. Stancáková. A., *et al.* Changes in insulin sensitivity and insulin release in relation to glycemia and glucose tolerance in 6,414 Finnish men. *Diabetes* 2009;**58**:1212-1221.

S43. Jørgensen, T. *et al*. A randomized non-pharmacological intervention study for prevention of ischaemic heart disease: baseline results Inter99 (1). *Eur J Cardiovasc Prevention Rehab*. 2003;**10**:377-386.

S44. Sparsø, T. *et al.* The *GCKR* rs780094 polymorphism is associated with elevated fasting serum triacylglycerol, reduced fasting and OGTT-related insulinaemia, and reduced risk of type 2 diabetes. *Diabetologia* 2008;**51**:70-75.

S45. Boehm, B.O. *et al.* Prevalence of the metabolic syndrome in southwest Germany. *Scand. J Clin Lab Invest Suppl.* 2005;**240**:122-128.

S46. Boehm, B.O., *et al.* Epidemiology and immunogenetic background of islet cell antibody--positive nondiabetic schoolchildren. Ulm-Frankfurt population study. *Diabetes*. 1991;**40**:1435-1439.

S47. Powell, M., *et al.* Glutamic acid decarboxylase autoantibody assay using 125I-labelled recombinant GAD65 produced in yeast. *Clin. Chim. Acta.* 1996;**256**:175-188.

S48. Marchini, J., Howie, B., Myers, S., McVean, G. & Donnelly, P. A new multipoint method for genome-wide association studies by imputation of genotypes. *Nat Genet.* 2007;**39**:906-913.

S49. Purcell, S. *et al*. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet.* 2007:**81**;559-575.

S50. Aulchenko, Y.S., Ripke, S., Isaacs, A., van Duijn, C.M. GenABEL: an R library for genome-wide association analysis. *Bioinformatics* 2007;**23**:1294-1296.

S51. Desai, M. *et al.* The variable number of tandem repeats upstream of the insulin gene is a susceptibility locus for latent autoimmune diabetes in adults. *Diabetes*. 2006;**55**:1890-1894.

S52. Desai, M. *et al.* An association analysis of the HLA gene region in latent autoimmune diabetes in adults. *Diabetologia*. 2007;**50**:68-73

S53. Haller, K *et al.* Insulin gene VNTR, CTLA-4 +49A/G and HLA-DQB1 alleles distinguish latent autoimmune diabetes in adults from type 1 diabetes and from type 2 diabetes group. *Tissue Antigens.* 2007;**69**:121-127.

S54. Cervin, C. *et al.* Genetic similarities between latent autoimmune diabetes in adults, type 1 diabetes, and type 2 diabetes. *Diabetes* 2008;**57**:1433-1437.

S55. Barrett JC *et al.* Genome-wide association study and meta-analysis find that over 40 loci affect risk of type 1 diabetes. *Nat Genet*. 2009;**41**,703-707

S56. Nejentsev, S. *et al.* Localization of type 1 diabetes susceptibility to the MHC class I genes HLA-B and HLA-A. *Nature*. 2007;**450**:887-892.

S57. Howson, J.M., Walker, N.M., Smyth, D.J., Todd, J.A., Type I Diabetes Genetics Consortium. Analysis of 19 genes for association with type I diabetes in the Type I Diabetes Genetics Consortium families. *Genes Immun.* 2009;**10 Suppl 1**:S74-84.

S58. Owen, K.R. and McCarthy, M.I. Type 1 and type 2 diabetes-chalk and cheese? *Diabetologia*. 2009;**52**:1983-1986.

S59. McCarroll, S.A. *et al.* Integrated detection and population-genetic analysis of SNPs and copy number variation. *Nat Genet*. 2008;**40**:1166-1174.

S60. Conrad, D.F. *et al.* Origins and functional impact of copy number variation in the human genome. *Nature* (Epub ahead of print 08 Oct 2009).

S61. Craddock, N. *et al.* Genome-wide association study of copy number variation in 16,000 cases of eight common diseases and 3,000 shared controls. *Nature* (in press)

S62. Dupuis, J. *et al.* New genetic loci implicated in fasting glucose homeostasis and their impact on type 2 diabetes risk. *Nat Genet* 2010;**42:**105-116.

S63. Saxena, R., *et al.* Genetic Variation in Gastric Inhibitory Polypeptide Receptor (GIPR) Impacts the Glucose and Insulin Responses to an Oral Glucose Challenge. *Nat Genet* 2010;**42:**142-148

S64. Kong, A., *et al.* Parental origin of sequence variants associated with complex diseases. *Nature* 2009;**462:**868-874.

S65. Hindorff, L.A. *et al.* Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc Natl Acad Sci U S A*. 2009;**106**:9362-9367.

S66. Melzer, D. *et al.* A genome-wide association study identifies protein quantitative trait loci (pQTLs). *PLoS Genet*. 2008;**4**:e1000072.

S67. Gieger, C. *et al.* Genetics meets metabolomics: a genome-wide association study of metabolite profiles in human serum. *PLoS Genet*. 2008;**4**:e1000282.

S68. Willer, C.J. *et al.* Six new loci associated with body mass index highlight a neuronal influence on body weight regulation. *Nat Genet*. 2009;**41**:25-34.

S69. Raychaudhuri, S. *et al.* Identifying relationships among genomic disease regions: predicting genes at pathogenic SNP associations and rare deletions. *PLoS Genet.* 2009;**5**:e1000534.

S70. Mi, H. *et al.* The PANTHER database of protein families, subfamilies, functions and pathways. *Nucleic Acids Res.* 2005;**33**:D284-288.

S71. Thomas, P.D. *et al.* PANTHER: a library of protein families and subfamilies indexed by function. *Genome Res.* 2003;**13**:2129-2141.

S72. Vastrik, I. *et al.* Reactome: a knowledge base of biologic pathways and processes. *Genome Biol.* 2007;**8**:R39.

S73. Matthews, L. *et al.* Reactome knowledgebase of human biological pathways and processes. *Nucleic Acids Res.* 2009;**37**:D619-622.

S74. Subramanian, A. *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A*. 2005;**102**:15545-15550.

S75. Girvan, M., Newman, M.E.J. Community structure in social and biological networks. *Proc Natl Acad Sci. USA* 2002;**99**:7821-7826.

S76. Shannon, P. *et al.* Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* 2003;**13**: 2498-2504.

S77. Wu, G., Feng, X., Stein, L. A human functional protein interaction network and its application to cancer data analysis. *Genome Biology* 2010;**11:**R53

S78. Mishra, G.R. *et al.* Human protein reference database--2006 update. *Nucleic Acids Res.*, 2006;**34**:D411-414.

S79. Yeang, C.H. *et al.* Validation and refinement of gene-regulatory pathways on a network of physical interactions. *Genome Biol.* 2005;**6**:R62.

S80. Ideker, T., Ozier, O., Schwikowski, B., Siegel, A.F. Discovering regulatory and signalling circuits in molecular interaction networks. *Bioinformatics.* 2002;**18 Suppl 1**:S233-240

S81. Ekins, S. *et al.* Pathway mapping tools for analysis of high content data. *Methods Mol Biol.* 2007;**356**:319-350.

S82. Mootha, V.K. *et al.* PGC-1alpha-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. *Nat Genet.* 2003;**34**: 267-273.

S83. Segrè, A.V. *et al.* Common Inherited Variation in Mitochondrial Genes is not Enriched for Associations with Type 2 Diabetes or Related Glycemic Traits. *PLoS Genet* 2010 (in press)

Full acknowledgments

ARIC : The Atherosclerosis Risk in Communities Study is carried out as a collaborative study supported by National Heart, Lung, and Blood Institute contracts N01-HC-55015, N01-HC-55016, N01-HC-55018, N01-HC-55019, N01-HC-55020, N01-HC-55021, N01-HC-55022, R01HL087641, R01HL59367, R01HL086694 and RC2 HL102419; National Human Genome Research Institute contract U01HG004402; and National Institutes of Health contract HHSN268200625226C. The authors thank the staff and participants of the ARIC study for their important contributions. Infrastructure was partly supported by Grant Number UL1RR025005, a component of the National Institutes of Health and NIH Roadmap for Medical Research.

BREDA : The Breda study was financially supported by the Dutch Diabetes Foundation and SenterNovem (IOP Genomics grant IGE05012)

DGDG : This work was also supported by the French Government ("*Agence Nationale de la Recherche*"), the French Region of "*Nord Pas De Calais*" ("*Contrat de Projets État-Région*"), and the charities: "*Association Française des Diabétiques*", "*Programme National de Recherche sur le Diabète*" and "*Association de Langue Française pour l'Etude du Diabète et des Maladies Métaboliques*". This study was supported in part by a grant from the European Union (Integrated Project EuroDia LSHM-CT-2006-518153 in the Framework Programme 6 [FP6] of the European Community). The D.E.S.I.R. study has been supported by CNAMTS, Lilly, Novartis Pharma and Sanofi-Aventis, by INSERM ("*Réseaux en Santé Publique, Interactions entre les déterminants de la santé*"), by "*Association Diabète Risque Vasculaire*", "*Fédération Française de Cardiologie*", "*Fondation de France*", ALFEDIAM, ONIVINS, Ardix Medical, Bayer Diagnostics, Becton Dickinson, Cardionics, Merck Santé, Novo Nordisk, Pierre Fabre, Roche, Topcon. The D.E.S.I.R. Study Group: INSERM U780: B. Balkau, P. Ducimetière, E. Eschwège; INSERM U367: F. Alhenc-Gelas; CHU D'Angers: Y. Gallois, A. Girault; Bichat Hospital: F. Fumeron, M. Marre; Medical Examination Services: Alençon, Angers, Caen, Chateauroux, Cholet, Le Mans, and Tours; Research Institute for General Medicine: J. Cogneau; General practitioners of the region; Cross-Regional Institute for Health: C. Born, E. Caces, M. Cailleau, J. G. Moreau, F. Rakotozafy, J. Tichet, S. Vol. We thank M. Deweider and F. Allegaert for the DNA bank management. We are sincerely indebted to all subjects who participated in this study.

DGI: We thank the study participants who made this research possible. We thank colleagues in the Broad Genetic Analysis and Biological Samples Platforms for their expertise and contributions to genotyping, data and sample management, and analysis. The initial GWAS genotyping was supported by Novartis (to D.A.); support for additional analysis and genotyping in this report was provided by funding from the Broad Institute of Harvard and MIT, by the Richard and Susan Smith Family Foundation/American Diabetes Association Pinnacle Program Project Award (to D.A.), and by a Freedom to Discovery award of the Foundation of Bristol Myers Squibb (to D.A.). M.J.D. and D.A. acknowledge support from US National Institutes of Health/National Heart, Lung, and Blood Institute grant (U01 HG004171). D.A. was a Burroughs Wellcome Fund Clinical Scholar in Translational Research and is a Distinguished Clinical Scholar of the Doris Duke Charitable Foundation. D.A. and B.F.V. acknowledges sponsored research funding from Pfizer, Inc. S.R. is supported by an NIH Career Development Award (1K08AR055688). A.V.S acknowledges support from the American Diabetes Association (Award No.: 7-08-MN-OK) A.L.E. was supported as a research fellow by the Sarnoff Cardiovascular Research Fellowship. L.G., T.T., B.I. and the Botnia Study are principally supported by the Sigrid Juselius Foundation, the Finnish Diabetes Research Foundation, The Folkhalsan Research Foundation and Clinical Research Institute HUCH Ltd; work in Malmö, Sweden was also funded by a Linne grant from the Swedish Research Council (349 2006-237P) and the Skaraborg Institute, Skövde. Additional genotyping for Stage 2 samples in Malmö was supported by grants from the Swedish Research Council (LG) and the Wallenberg Foundation. We thank the Botnia and Skara research teams for clinical contributions.

deCODE: We thank the Icelandic study participants whose contribution made this work possible. We also acknowledge the staff at the Clinical Research Centre (Iceland) and the deCODE Genetics biological materials and genotyping facilities for their work. The research performed at deCODE Genetics was part funded through the European Community's Seventh Framework Programme (FP7/2007-2013), ENGAGE project, grant agreement HEALTH-F4-2007-201413.

DENMARK : The Danish study was supported by grants from the Lundbeck Foundation Centre of Applied Medical Genomics for Personalized Disease Prediction, Prevention and Care (LUCAMP), the European Union (EUGENE2)-grant LSHM-CT-2004-512013, EXGENESIS-grant LSHM-CT-2004-005272, the Danish Diabetes Association and the Danish Health Research Council.

EUROSPAN : EUROSPAN (European Special Populations Research Network) was supported by European Commission FP6 STRP grant number 018947 (LSHG-CT-2006-01947). High-throughput genome-wide association analysis of the data was supported by joint grant from Netherlands Organisation for Scientific Research and the Russian Foundation for Basic Research (NWO-RFBR 047.017.043). The ERF study was supported by grants from the NWO, Erasmus MC and the Centre for Medical Systems Biology (CMSB). We are grateful to all patients and their relatives, general practitioners and neurologists for their contributions and to P. Veraart for her help in genealogy, Jeannette Vergeer for the supervision of the laboratory work and P. Snijders for his help in data collection. For the MICROS study, we thank the primary care practitioners Raffaela Stocker, Stefan Waldner, Toni Pizzecco, Josef Plangger, Ugo Marcadent and the personnel of the Hospital of Silandro (Department of Laboratory Medicine) for their participation and collaboration in the research project. In South Tyrol, the study was supported by the Ministry of Health and Department of Educational Assistance, University and Research of the Autonomous Province of Bolzano and the South Tyrolean Sparkasse Foundation. ORCADES was supported by the Scottish Government Chief Scientist Office and the Royal Society. DNA extractions were performed at the Wellcome Trust Clinical Research Facility in Edinburgh. We would like to acknowledge the invaluable contributions of Lorraine Anderson and the research nurses in Orkney, the administrative team in Edinburgh and the people of Orkney. The VIS study in the Croatian island of Vis was supported through the grants from the Medical Research Council UK to H.C., A.F.W. and I.R.; and Ministry of Science, Education and Sport of the Republic of Croatia to I.R. (number 108-1080315-0302). The authors collectively thank a large number of individuals for their individual help in organizing, planning and carrying out the field work related to the project and data management: Professor Pavao Rudan and the staff of the Institute for Anthropological Research in Zagreb, Croatia (organization of the field work, anthropometric and physiological measurements, and DNA extraction); Professor Ariana Vorko-Jovic and the staff and medical students of the Andrija Stampar School of Public Health of the Faculty of Medicine, University of Zagreb, Croatia (questionnaires, genealogial reconstruction and data entry); Dr Branka Salzer from the biochemistry lab "Salzer", Croatia (measurements of biochemical traits); local general practitioners and nurses (recruitment and communication with the study population); and the employees of several other Croatian institutions who participated in the field work, including but not limited to the University of Rijeka and Split, Croatia; Croatian Institute of Public Health; Institutes of Public Health in Split and Dubrovnik, Croatia. SNP Genotyping of the Vis samples was carried out by the Genetics Core Laboratory at the Wellcome Trust Clinical Research Facility, WGH, Edinburgh. J.W. receives personal funding from the Royal Society.

Framingham Heart Study: 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 No. N01-HC-25195) and its contract with Affymetrix, Inc for genotyping services (Contract No. 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 Drs. Meigs, Dupuis and Florez, NIDDK K24 DK080140 to Dr. Meigs, a Massachusetts General Hospital Physician Scientist Development Award and both NIDDK K23-DK65978 and a Doris Duke Charitable Foundation Clinical Scientist Development Award to Dr. Florez.

FUSION: We would like to thank the many Finnish volunteers who generously participated in the FUSION, D2D, Health 2000, Finrisk 1987, Finrisk 2002, and Savitaipale studies from which we chose our FUSION GWA and replication cohorts (no overlap with Health 2000 replication cohort). We also thank Terry Gliedt and Peggy White for informatics and analysis support and Randy Pruim for graphical support. The Center for Inherited Disease Research performed the GWA genotyping. Support for this study was provided by the following: American Diabetes Association grant 1-05-RA-140 (R.M.W.) and 7-06-MN-05 (M.Boe.) and NIH grants DK069922 (R.M.W.), U54 DA021519 (R.M.W.), DK062370 (M.Boe.), and DK072193 (K.L.M.). Additional support comes from the National Human Genome Research Institute intramural project number 1 Z01 HG000024 (F.S.C.) and from a Pathway to Independence Award (1K99HL094535-01A1) to C.W.

HUNT: The Nord-Trøndelag Health Study (The HUNT Study) is a collaboration between HUNT Research Centre, Faculty of Medicine, the Norwegian University of Science and Technology (NTNU), The Norwegian Institute of Public Health, The Nord-Trøndelag Hospital Trust and the Nord-Trøndelag County Council.

KORA : The MONICA/KORA Augsburg studies were financed by the Helmholtz Zentrum München- Research Center for Environment and Health, Neuherberg, Germany and supported by grants from the German Federal Ministry of Education and Research (BMBF, FKZ01GS0823 and the German Center for Diabetes Research [DZD e.V.]) the Federal Ministry of Health (Berlin, Germany), the Ministry of Innovation, Science, Research and Technology of the state North Rhine-Westphalia (Düsseldorf, Germany), the German National Genome Research Network (NGFN), the Munich Center of Health Sciences (MC Health) and Deutsche Forschungsgemeinschaft (DFG RA 459) as part of LMU innovativ. We thank all members of field staffs who were involved in the planning and conduct of the MONICA/KORA Augsburg studies.
METSIM: The METSIM study is supported by the Academy of Finland (no. 124243).
NHS : The Nurses' Health Study diabetes GWAS was supported by NIH grants U01HG004399 and DK58845.
Rotterdam: We acknowledge the work of Dr F Rivadeneira, Dr M Moorhouse, P Arp and M Jhamai, who created the Rotterdam Study genotypic database. This research was supported by grants from the Netherlands Foundation for Scientific Research (NWO), and by the Centre for Medical Systems Biology (CMSB) Genome-wide genotyping of Rotterdam Study is supported by NWO (175.010.2005.011). The Rotterdam Study (I,I, and II) 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), the Netherlands Genomics Initiative (NGI)-sponsored Netherlands Consortium for Healthy Aging (NCHA) and the Municipality of Rotterdam. Erasmus Rucphen Family (ERF) was supported by grants from The Netherlands Organization for Scientific Research (NWO; Pionier Grant), Erasmus MC and the Netherlands Genomics Initiative (NGI)-sponsored Center of Medical Systems Biology (CMSB).
UK (Cambridge: GEM): The work on the Cambridgeshire case-control, Ely, ADDITION and EPIC-Norfolk studies was funded by support from the Wellcome Trust and MRC. The Norfolk Diabetes study is funded by the MRC with support from NHS Research & Development and the Wellcome Trust and we are grateful to the study team. We thank the technical teams, in particular Emma Gray, Jackie Bryant, Nicholas Smith and Sarah Edkins, at the Wellcome Trust Sanger Institute and the MRC Epidemiology laboratory for genotyping and informatics support. IB acknowledges support from the Wellcome Trust 077016/Z/05/Z. We thank all study participants.
UK (Oxford/Exeter): 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). The GWA genotyping was supported by the Wellcome Trust (076113) and replication genotyping by the European Commission (EURODIA LSHG-CT-2004-518153), MRC (Project Grant G0601261), Wellcome Trust (083270), Peninsula Medical School, and Diabetes UK. Additional support has come through the NIDDK (DK073490) and Wellcome Trust Centre for Human Genetics Core Award (075491). CML and EZ are Wellcome Trust Career Development Fellows (#086596, #079557, #088885), APM a Wellcome Trust Senior Fellow (#081682), MNW a Vandervell Fellow, ALG an MRC New Investigator (#81696) and LJMcC is supported by a Diabetes UK Studentship. KRO is an NIHR Clinician Scientist, and MIMcC receives funding from the Oxford NIHR Biomedical Research Centre. We are grateful to Philippe Halban (Geneva) for providing training in FAC sorting of human pancreatic beta-cells and the European Foundation for the Study of Diabetes for financial support to LJMcC for this training. We acknowledge the contribution of Dr Michael Sampson, and our team of research nurses.
ULM: Grant support to BOB from Deutsche Forschungsgemeinschaft (DFG GrK 1041, SFB 518) and Centre of Excellence "Metabolic Disorders" Baden-Wuerttemberg.
Other: We would like to acknowledge Jonathan Marchini and Peter Donnelly (Oxford) for providing the imputed CCC datasets, and Gil McVean (Oxford) for providing X-chromosome recombination data and Joanne Howson (Cambridge) for assistance with interpretation of T1D associations in the HLA region.

Members of the GIANT (Genetic Investigation of Anthropometric Traits) Consortium are:
Cristen J. Willer[1], Elizabeth K. Speliotes[2,3], Ruth J.F. Loos[4,5], Shengxu Li[4,5], Cecilia M. Lindgren[6], Iris M. Heid[7], Sonja I. Berndt[8], Amanda L. Elliott[9,10], Anne U. Jackson[1], Claudia Lamina[7], Guillaume Lettre[9,11], Noha Lim[12], Helen N. Lyon[3,11], Steven A. McCarroll[9,10], Konstantinos Papadakis[13], Lu Qi[14,15], Joshua C. Randall[6], Rosa Maria Roccasecca[16], Serena Sanna[17], Paul Scheet[18], Michael N. Weedon[19], Eleanor Wheeler[16], Jing Hua Zhao[4,5], Leonie C. Jacobs[20], Inga Prokopenko[6,21], Nicole Soranzo[16,22], Toshiko Tanaka[23], Nicholas J. Timpson[24], Peter Almgren[25], Amanda Bennett[26], Richard N. Bergman[27], Sheila A. Bingham[28,29], Lori L. Bonnycastle[30], Morris Brown[31], Noël P. Burtt[9], Peter Chines[30], Lachlan Coin[32], Francis S. Collins[30], John M. Connell[33], Cyrus Cooper[34], George Davey Smith[24], Elaine M. Dennison[34], Parimal Deodhar[30], Paul Elliott[32], Michael R. Erdos[30], Karol Estrada[20], David M. Evans[24], Lauren Gianniny[9], Christian Gieger[7], Christopher J. Gillson[4,5], Candace Guiducci[9], Rachel Hackett[9], David Hadley[13], Alistair S. Hall[35], Aki S. Havulinna[36], Johannes Hebebrand[37], Albert Hofman[38], Bo Isomaa[39], Kevin B. Jacobs[40], Toby Johnson[41,42,43], Pekka Jousilahti[36], Zorica Jovanovic[5,44], Kay-Tee Khaw[45], Peter Kraft[46], Mikko Kuokkanen[9,47], Johanna Kuusisto[48], Jaana Laitinen[49], Edward G. Lakatta[50], Jian'an Luan[4,5], Robert N. Luben[45], Massimo Mangino[51], Wendy L. McArdle[52], Thomas Meitinger[53,54], Antonella Mulas[17], Patricia B. Munroe[55], Narisu Narisu[30], Andrew R. Ness[56], Kate Northstone[52], Stephen O'Rahilly[5,44], Carolin Purmann[5,44], Matthew G. Rees[30], Martin Ridderstråle[57], Susan M. Ring[52], Fernando Rivadeneira[20,38], Aimo Ruokonen[58], Manjinder S. Sandhu[4,45], Jouko Saramies[59], Laura J. Scott[1], Angelo Scuteri[60], Kaisa Silander[47], Matthew A. Sims[4,5], Kijoung Song[12], Jonathan Stephens[61], Suzanne Stevens[51], Heather M. Stringham[1], Y.C. Loraine Tung[5,44], Timo T. Valle[62], Cornelia M. Van Duijn[38], Karani S. Vimaleswaran[4,5], Peter Vollenweider[63], Gerard Waeber[63], Chris Wallace[55], Richard M. Watanabe[64], Dawn M. Waterworth[12], Nicholas Watkins[61], The Wellcome Trust Case Control Consortium[65], Jacqueline C.M. Witteman[38], Eleftheria Zeggini[6], Guangju Zhai[22], M. Carola Zillikens[20], David Altshuler[9,10], Mark J. Caulfield[55], Stephen J. Chanock[8], I. Sadaf Farooqi[5,44], Luigi Ferrucci[23], Jack M. Guralnik[66], Andrew T. Hattersley[67], Frank B. Hu[14,15], Marjo-Riitta Jarvelin[32], Markku Laakso[48], Vincent Mooser[12], Ken K. Ong[4,5], Willem H. Ouwehand[16,61], Veikko Salomaa[36], Nilesh J. Samani[51], Timothy D. Spector[22], Tiinamaija Tuomi[68,69], Jaakko Tuomilehto[62], Manuela Uda[17], André G. Uitterlinden[20,38], Nicholas J. Wareham[4,5], Panagiotis Deloukas[16], Timothy M. Frayling[19], Leif C. Groop[25,70], Richard B. Hayes[8], David J. Hunter[9,14,15,46], Karen L. Mohlke[71], Leena Peltonen[9,16,72], David Schlessinger[73], David P. Strachan[13], H-Erich Wichmann[7,74], Mark I. McCarthy[6,21,75], Michael Boehnke[1], Inês Barroso[16], Gonçalo R. Abecasis[18], Joel N. Hirschhorn[3,11,76]

1. Department of Biostatistics, School of Public Health, University of Michigan, Ann Arbor, MI 48109, USA
2. Division of Gastroenterology, Massachusetts General Hospital, Boston, MA 02114, USA
3. Metabolism Initiative and Program in Medical and Population Genetics, Broad Institute of Harvard and MIT, Boston, MA 02142, USA
4. MRC Epidemiology Unit, Addenbrooke's Hospital, Cambridge CB2 0QQ, UK
5. Institute of Metabolic Science, Addenbrooke's Hospital, Cambridge CB2 0QQ, UK
6. Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford OX3 7BN, UK
7. Institute of Epidemiology, Helmholtz Zentrum München, Ingolstaedter Landstr. 1, 85764 Neuherberg, Germany
8. Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Department of Health and Human Services, Bethesda, MD 20892, USA
9. Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, MA 02142, USA
10. Department of Molecular Biology, Massachusetts General Hospital, Cambridge, MA 02144, USA
11. Program in Genomics and Divisions of Endocrinology and Genetics, Children's Hospital, Boston, MA 02115, USA
12. Medical Genetics/Clinical Pharmacology and Discovery Medicine, PA 19406, USA
13. Division of Community Health Sciences, St George's, University of London, London SW17 0RE, UK

14. Department of Nutrition, Harvard School of Public Health, Boston, MA 02115, USA
15. Channing Laboratory, Department of Medicine, Brigham and Women's Hospital, Boston, MA 02115, USA
16. Wellcome Trust Sanger Institute, Hinxton, Cambridge CB10 1SA, UK
17. Istituto di Neurogenetica e Neurofarmacologia, Consiglio Nazionale delle Ricerche, Cagliari, Italy
18. Center for Statistical Genetics, Department of Biostatistics, University of Michigan, Ann Arbor, MI 48109, USA
19. Genetics of Complex Traits, Peninsula Medical School, Exeter EX1 2LU, UK
20. Department of Internal Medicine, Erasmus MC, PO Box 2400, NL-3000-CA Rotterdam, The Netherlands
21. Oxford Centre for Diabetes, Endocrinology and Metabolism, University of Oxford, Churchill Hospital, Oxford OX3 7LJ, UK
22. Department of Twin Research and Genetic Epidemiology, King's College London, London SE1 7EH, UK
23. National institute of Aging, Clinical research branch - Longitudinal Studies Section, Baltimore, MD 21225, USA
24. MRC Centre for Causal Analyses in Translational Epidemiology, Department of Social Medicine, University of Bristol, Bristol BS8 2PR, UK
25. Lund University Diabetes Centre, Department of Clinical Sciences, Lund University, 20502 Malmö, Sweden
26. DRL, OCDEM, Churchill Hospital, Headington, Oxford OX3 7LJ, UK
27. Physiology and Biophysics, University of Southern California School of Medicine, Los Angeles, CA 90033, USA
28. MRC Dunn Human Nutrition Unit, Wellcome Trust/MRC Building, Cambridge CB2 0XY, UK
29. MRC Centre for Nutritional Epidemiology in Cancer Prevention and Survival, Cambridge CB1 8RN, UK
30. National Human Genome Research Institute, Bethesda, MD 20892, USA
31. Clinical Pharmacology Unit, University of Cambridge, Addenbrooke's Hospital, Cambridge, UK
32. Department of Epidemiology and Public Health, Imperial College London, St Mary's Campus, Norfolk Place, W2 1PG London, UK
33. BHF Glasgow Cardiovascular Research Centre, Faculty of Medicine, University of Glasgow, Glasgow G12 8TA, UK
34. MRC Epidemiology Resource Centre, (University of Southampton), Southampton General Hospital, Southampton SO16 6YD, UK.
35. Yorkshire Heart Centre, Leeds General Infirmary, Leeds LS1 3EX, UK
36. KTL-National Public Health Institute, FI-00300 Helsinki, Finland
37. Department of Child and Adolescent Psychiatry, University of Duisburg-Essen, Virchowstr. 174, 45147 Essen, Germany
38. Department of Epidemiology, Erasmus MC, PO Box 2400, NL-3000-CA Rotterdam, The Netherlands
39. Folkhalsan Research Center, Malmska Municipal Health Center and Hospital, Jakobstad, Finland
40. Bioinformed Consulting Services, Gaithersburg, MD 20877, USA
41. Department of Medical Genetics, University of Lausanne, 1005 Lausanne, Switzerland
42. University Institute for Social and Preventative Medicine, Centre Hospitalier Universitaire Vaudois (CHUV), 1005 Lausanne, Switzerland
43. Swiss Institute of Bioinformatics, Switzerland
44. University of Cambridge Metabolic Research Laboratories, Addenbrooke's Hospital, Cambridge CB2 0QQ, UK
45. Department of Public Health and Primary Care, Institute of Public Health, University of Cambridge, Cambridge CB2 0SR, UK
46. Program in Molecular and Genetic Epidemiology, Harvard School of Public Health, Boston, MA 02115, USA
47. Department of Molecular Medicine, National Public Health Institute, FIN-00300 Helsinki, Finland
48. Department of Medicine, University of Kuopio, 70210 Kuopio, Finland
49. Finnish Institute of Occpational Health, Aapistie 1, Fin-90220 Oulu, Finland
50. Laboratory of Cardiovascular Science, Gerontology Research Center, National Institute on Aging,Baltimore, MD 21224, USA
51. Department of Cardiovascular Sciences, University of Leicester, Clinical Sciences, Glenfield General Hospital, Leicester LE3 9QP, UK
52. ALSPAC, Department of Social Medicine, University of Bristol, Bristol BS8 1TQ, UK
53. Institute of Human Genetics, Helmholtz Zentrum München, Ingolstaedter Landstr. 1, 85764 Neuherberg, Germany

54. Institute of Human Genetics, Technical University Munich, D-81765, Munich, Germany
55. Clinical Pharmacology, The William Harvey Research Institute, Bart's and The London, Queen Mary's School of Medicine and Dentistry, Charterhouse Square, London EC1M 6BQ, UK
56. Department of Oral & Dental Science, Bristol BS1 2LY, UK
57. Department of Clinical Sciences, Lund University, 20502 Malmö, Sweden
58. Department of Clinical Chemistry, University of Oulu, Fin-90220, Oulu, Finland
59. Savitaipale Health Center, Savitaipale, Finland
60. Unita' Operativa Geriatria - Istituto Nazionale Ricovero e Cura Anziani - Rome, Italy
61. Department of Haematology, University of Cambridge/NHS Blood & Transplant, Cambridge CB2 2PR, UK
62. National Public Health Institute, Department of Epidemiology and Health Promotion, Mannerheimintie 166, FIN-00300 Helsinki, FINLAND
63. Department of Internal Medicine, BH-10 Centre Hospitalier Universitaire Vaudois (CHUV), 1011 Lausanne, Switzerland
64. Department of Preventive Medicine, Division of Biostatistics, Keck School of Medicine, University of Southern California, CHP-220, Los Angeles, CA 90089, USA
65. The membership of this consortium is listed in the supplementary material.
66. Laboratory of Epidemiology, Demography, and Biometry; Gerontology Research Center, National Institute on Aging, Bethesda, MD 20892, USA
67. Peninsula Medical School, Exeter EX5 2DW, UK
68. 3 Department of Medicine, Helsinki University Central Hospital, Helsinki, Finland
69. Research Program of Molecular Medicine, University of Helsinki, Helsinki, Finland
70. Department of Medicine, Helsinki University, Helsinki, Finland
71. Department of Genetics, University of North Carolina, CB #7264, Chapel Hill, NC 27599, USA
72. Institute of Molecular Medicine, University of Helsinki, Finland
73. Laboratory of Genetics, NIH Biomedical Research Center, National Institute on Aging, Baltimore, MD 21224, USA
74. Institute of Medical Information Processing, Biometry, and Epidemiology, Ludwig-Maximilians-University München, Marchioninistr. 15, 81377 München, Germany
75. NIHR Oxford Biomedical Research Centre, University of Oxford, Old Road, Headington, Oxford OX3 7LJ, UK
76. Department of Genetics, Harvard Medical School, Boston, MA 02115, USA

MAGIC (Meta-Analysis of Glucose and Insulin-Related traits) Consortium investigators are:

Josée Dupuis[1,2], Claudia Langenberg[3], Inga Prokopenko[4,5], Richa Saxena[6,7], Nicole Soranzo[8,9], Anne U Jackson[10], Eleanor Wheeler[11], Nicole L Glazer[12], Nabila Bouatia-Naji[13], Anna L Gloyn[4], Cecilia M Lindgren[4,5], Reedik Mägi[4,5], Andrew P Morris[5], Joshua Randall[5], Toby Johnson[14–16], Paul Elliott[17,176], Denis Rybin[18], Gudmar Thorleifsson[19], Valgerdur Steinthorsdottir[19], Peter Henneman[20], Harald Grallert[21], Abbas Dehghan[22], Jouke Jan Hottenga[23], Christopher S Franklin[24], Pau Navarro[25], Kijoung Song[26], Anuj Goel[5,27], John R B Perry[28], Josephine M Egan[29], Taina Lajunen[30], Niels Grarup[31], Thomas Sparsø[31], Alex Doney[32], Benjamin F Voight[6,7], Heather M Stringham[10], Man Li[33], Stavroula Kanoni[34], Peter Shrader[35], Christine Cavalcanti-Proença[13], Meena Kumari[36], Lu Qi[37], Nicholas J Timpson[38], Christian Gieger[21], Carina Zabena[39], Ghislain Rocheleau[40,41], Erik Ingelsson[42,43], Ping An[44], Jeffrey O'Connell[45], Jian'an Luan[3], Amanda Elliott[6,7], Steven A McCarroll[6,7], Felicity Payne[11], Rosa Maria Roccasecca[11], François Pattou[46], Praveen Sethupathy[47], Kristin Ardlie[48], Yavuz Ariyurek[49], Beverley Balkau[50], Philip Barter[51], John P Beilby[52,53], Yoav Ben-Shlomo[54], Rafn Benediktsson[55,56], Amanda J Bennett[4], Sven Bergmann[14,16], Murielle Bochud[15], Eric Boerwinkle[57], Amélie Bonnefond[13], Lori LBonnycastle[47], Knut Borch-Johnsen[58,59], Yvonne Böttcher[60], Eric Brunner[36], Suzannah J Bumpstead[8], Guillaume Charpentier[61], Yii-Der Ida Chen[62], Peter Chines[47], Robert Clarke[63], Lachlan J MCoin[17], Matthew N Cooper[64], Marilyn Cornelis[37], Gabe Crawford[6], Laura Crisponi[65], Ian N M Day[38], Eco J C de Geus[23], Jerome Delplanque[13], Christian Dina[13], Michael R Erdos[47], Annette C Fedson[64,66], Antje Fischer-Rosinsky[67,68], Nita G Forouhi[3], Caroline S Fox[2,69], Rune Frants[70], Maria Grazia Franzosi[71], Pilar Galan[72], Mark O Goodarzi[62], Jürgen Graessler[73], Christopher J Groves[4], Scott Grundy[74], Rhian Gwilliam[8], Ulf Gyllensten[75], Samy Hadjadj[76], Göran Hallmans[77], Naomi Hammond[8], Xijing Han[10], Anna-Liisa Hartikainen[78], Neelam Hassanali[4], Caroline Hayward[25], Simon C Heath[79], Serge Hercberg[80], Christian Herder[81], Andrew A Hicks[82], David R Hillman[66,83], Aroon D Hingorani[36], Albert Hofman[22], Jennie Hui[52,84], Joe Hung[85,86], Bo Isomaa[87,88], Paul R V Johnson[4,89], Torben Jørgensen[90,91], Antti Jula[92], Marika Kaakinen[93], Jaakko Kaprio[94–96], Y Antero Kesaniemi[97], Mika Kivimaki[36], Beatrice Knight[98], Seppo Koskinen[99], Peter Kovacs[100], Kirsten Ohm Kyvik[101], G Mark Lathrop[79], Debbie A Lawlor[38], Olivier Le Bacquer[13], Cécile Lecoeur[13], Yun Li[10], Valeriya Lyssenko[102], Robert Mahley[103], Massimo Mangino[9], Alisa K Manning[1], María Teresa Martínez-Larrad[39], Jarred B McAteer[6,104,105], Laura J McCulloch[4], Ruth McPherson[106], Christa Meisinger[21], David Melzer[28], David Meyre[13], Braxton D Mitchell[45], Mario A Morken[47], Sutapa Mukherjee[66,83], Silvia Naitza[65], Narisu Narisu[47], Matthew J Neville[4,107], Ben A Oostra[108], Marco Orrù[65], Ruth Pakyz[45], Colin N A Palmer[109], Giuseppe Paolisso[110], Cristian Pattaro[82], Daniel Pearson[47], John F Peden[5,27], Nancy L Pedersen[42], Markus Perola[96,111,112], Andreas F H Pfeiffer[67,68], Irene Pichler[82], Ozren Polasek[113], Danielle Posthuma[23,114], Simon C Potter[8], Anneli Pouta[115], Michael A Province[44], Bruce M Psaty[116,117], Wolfgang Rathmann[118], Nigel W Rayner[4,5], Kenneth Rice[119], Samuli Ripatti[96,111], Fernando Rivadeneira[22,120], Michael Roden[81,121], Olov Rolandsson[122], Annelli Sandbaek[123], Manjinder Sandhu[3,124], Serena Sanna[65], Avan Aihie Sayer[125], Paul Scheet[126], Laura J Scott[10], Udo Seedorf[127], Stephen J Sharp[3], Beverley Shields[98], Gunnar Sigurðsson[55,56], Eric J G Sijbrands[22,120], Angela Silveira[128], Laila Simpson[64,66], Andrew Singleton[129], Nicholas L Smith[130,131], Ulla Sovio[17], Amy Swift[47], Holly Syddall[125], Ann-Christine Syvänen[132], Toshiko Tanaka[133,134], Barbara Thorand[21], Jean Tichet[135], Anke Tönjes[60,136], Tiinamaija Tuomi[87,137], André G Uitterlinden[22,120], Ko Willems van Dijk[70,138], Mandy van Hoek[120], Dhiraj Varma[8], Sophie Visvikis-Siest[139], Veronique Vitart[25], Nicole Vogelzangs[140], Gérard Waeber[141], Peter J Wagner[96,111], Andrew Walley[142], G Bragi Walters[19], Kim L Ward[64,66], Hugh Watkins[5,27], Michael N Weedon[28], Sarah H Wild[24], Gonneke Willemsen[23], Jaqueline C M Witteman[22], John W G Yarnell[143], Eleftheria Zeggini[5,8], Diana Zelenika[79], Björn Zethelius[43,144], Guangju Zhai[9], Jing Hua Zhao[3], M Carola Zillikens[120], Ingrid B Borecki[44], Ruth J F Loos[3], Pierre Meneton[80], Patrik K E Magnusson[42], David M Nathan[104,105], Gordon H Williams[69,105], Andrew T Hattersley[98], Kaisa Silander[96,111], Veikko Salomaa[145], George Davey Smith[38], Stefan R Bornstein[73], Peter Schwarz[73], Joachim Spranger[67,68], Fredrik Karpe[4,107], Alan R Shuldiner[45], Cyrus Cooper[125], George V Dedoussis[34], Manuel Serrano-Ríos[39], Andrew D Morris[109], Lars

Lind[132], Lyle J Palmer[64,66,84], Frank B Hu[146,147], Paul W Franks[148], Shah Ebrahim[149], Michael Marmot[36], WH Linda Kao[33,150,151], James S Pankow[152], Michael J Sampson[153], Johanna Kuusisto[154], Markku Laakso[154], Torben Hansen[31,155], Oluf Pedersen[31,59,156], Peter Paul Pramstaller[82,157,158], H Erich Wichmann[21,159,160], Thomas Illig[21], Igor Rudan[24,161,162], Alan F Wright[25], Michael Stumvoll[60], Harry Campbell[24], James F Wilson[24], Anders Hamsten on behalf of Procardis Consortium[128], Richard N Bergman[163], Thomas A Buchanan[163,164], Francis S Collins[47], Karen L Mohlke[165], Jaakko Tuomilehto[94,166,167], Timo T Valle[166,167], David Altshuler[6,7,104,105], Jerome I Rotter[62], David S Siscovick[168], Brenda WJ H Penninx[140], Dorret I Boomsma[23], Panos Deloukas[8], Timothy D Spector[8,9], Timothy M Frayling[28], Luigi Ferrucci[169], Augustine Kong[19], Unnur Thorsteinsdottir[19,170], Kari Stefansson[19,170], Cornelia M van Duijn[22], Yurii S Aulchenko[22], Antonio Cao[65], Angelo Scuteri[65,171], David Schlessinger[47], Manuela Uda[65], Aimo Ruokonen[172], Marjo-Riitta Jarvelin[17,93,173], Dawn M Waterworth[26], Peter Vollenweider[141], Leena Peltonen[8,48,96,111,112], Vincent Mooser[26], Goncalo R Abecasis[10], Nicholas J Wareham[3], Robert Sladek[40,41], Philippe Froguel[13,142], Richard M Watanabe[163,174], James B Meigs[35,105], Leif Groop[102], Michael Boehnke[10], Mark I McCarthy[4,5,107], Jose C Florez[6,7,104,105], Inês Barroso[11].

[1] Department of Biostatistics, Boston University School of Public Health, Boston, Massachusetts, USA.
[2] National Heart, Lung, and Blood Institute's Framingham Heart Study, Framingham, Massachusetts, USA.
[3] Medical Research Council (MRC), Epidemiology Unit, Institute of Metabolic Science, Addenbrooke's Hospital, Cambridge, UK.
[4] Oxford Centre for Diabetes, Endocrinology and Metabolism, University of Oxford, Oxford, UK.
[5] Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK.
[6] Program in Medical and Population Genetics, Broad Institute, Cambridge, Massachusetts, USA.
[7] Center for Human Genetic Research, Massachusetts General Hospital, Boston, Massachusetts, USA.
[8] Wellcome Trust Sanger Institute, Hinxton, Cambridge, UK.
[9] Twin Research and Genetic Epidemiology Department, King's College London, St Thomas' Hospital Campus, London, UK.
[10] Center for Statistical Genetics, Department of Biostatistics, University of Michigan School of Public Health, Ann Arbor, Michigan, USA.
[11] Metabolic Disease Group, Wellcome Trust Sanger Institute, Hinxton, Cambridge, UK. [12]Cardiovascular Health Research Unit and Department of Medicine, University of Washington, Seattle, Washington, USA.
[13] Centre National de la Recherche Scientifique–Unité Mixte de Recherche 8090, Pasteur Institute, Lille 2–Droit et Santé University, Lille, France.
[14] Department of Medical Genetics, University of Lausanne, Lausanne, Switzerland.
[15] University Institute of Social and Preventative Medicine, Centre Hospitalier Universitaire Vaudois (CHUV) and University of Lausanne, Lausanne, Switzerland.
[16] Swiss Institute of Bioinformatics, Lausanne, Switzerland.
[17] Department of Epidemiology and Public Health, Imperial College London, Faculty of Medicine, Norfolk Place, London, UK.
[18] Boston University Data Coordinating Center, Boston, Massachusetts, USA.
[19] deCODE Genetics, Reykjavik, Iceland.
[20] Department of Human Genetics, Leiden University Medical Centre, Leiden, The Netherlands.
[21] Institute of Epidemiology, Helmholtz Zentrum Muenchen, German Research Center for Environmental Health, Neuherberg, Germany.
[22] Department of Epidemiology, Erasmus Medical College, Rotterdam, The Netherlands. [23]Department of Biological Psychology, VU University Amsterdam, Amsterdam, The Netherlands.
[24] Centre for Population Health Sciences, University of Edinburgh, Edinburgh, UK.
[25] MRC Human Genetics Unit, Institute of Genetics and Molecular Medicine, Edinburgh, UK.
[26] Division of Genetics, Research and Development, GlaxoSmithKline, King of Prussia, Pennsylvania, USA.
[27] Department of Cardiovascular Medicine, University of Oxford, Oxford, UK.
[28] Genetics of Complex Traits, Institute of Biomedical and Clinical Sciences, Peninsula College of Medicine and Dentistry, University of Exeter, Exeter, UK.

[29] National Institute of Aging, Baltimore, Maryland, USA.
[30] Unit for Child and Adolescent Health and Welfare, National Institute for Health and Welfare, Biocenter Oulu, University of Oulu, Oulu, Finland.
[31] Hagedorn Research Institute, Gentofte, Denmark.
[32] Department of Medicine and Therapeutics, Level 7, Ninewells Hospital and Medical School, Dundee, UK.
[33] Department of Epidemiology, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, Maryland, USA.
[34] Department of Nutrition–Dietetics, Harokopio University, Athens, Greece.
[35] General Medicine Division, Massachusetts General Hospital, Boston, Massachusetts, USA.
[36] Department of Epidemiology and Public Health, University College London, London, UK.
[37] Departments of Nutrition and Epidemiology, Harvard School of Public Health, Boston, Massachusetts, USA.
[38] MRC Centre for Causal Analyses in Translational Epidemiology, University of Bristol, Bristol, UK.
[39] Fundación para la Investigación Biomédica del Hospital Clínico San Carlos, Madrid, Spain.
[40] Departments of Medicine and Human Genetics, McGill University, Montreal, Canada. [41]Genome Quebec Innovation Centre, Montreal, Canada.
[42] Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden.
[43] Department of Public Health and Caring Sciences, Uppsala University, Uppsala, Sweden.
[44] Division of Statistical Genomics, Department of Genetics, Washington University School of Medicine, St. Louis, Missouri, USA.
[45] Division of Endocrinology, Diabetes and Nutrition, University of Maryland School of Medicine, Baltimore, Maryland, USA.
[46] INSERM U859, Universite de Lille-Nord de France, Lille, France.
[47] Genome Technology Branch, National Human Genome Research Institute, Bethesda, Maryland, USA.
[48] The Broad Institute, Cambridge, Massachusetts, USA.
[49] Leiden Genome Technology Center, Leiden University Medical Center, Leiden, The Netherlands.
[50] INSERM U780, Paris Sud University, Villejuif, France.
[51] The Heart Research Institute, Sydney, New South Wales, Australia.
[52] PathWest Laboratory of Western Australia, Department of Molecular Genetics, J Block, QEII Medical Centre, Nedlands West Australia, Australia.
[53] School of Surgery and Pathology, University of Western Australia, Nedlands West Australia, Australia.
[54] Department of Social Medicine, University of Bristol, Bristol, UK.
[55] Landspitali University Hospital, Reykjavik, Iceland.
[56] Icelandic Heart Association, Kopavogur, Iceland.
[57] The Human Genetics Center and Institute of Molecular Medicine, University of Texas Health Science Center, Houston, Texas, USA.
[58] Steno Diabetes Center, Gentofte, Denmark.
[59] Faculty of Health Science, University of Aarhus, Aarhus, Denmark.
[60] Department of Medicine, University of Leipzig, Leipzig, Germany.
[61] Endocrinology–Diabetology Unit, Corbeil-Essonnes Hospital, Essonnes, France.
[62] Medical Genetics Institute, Cedars-Sinai Medical Center, Los Angeles, California, USA.
[63] Clinical Trial Service Unit and Epidemiological Studies Unit, University of Oxford, Oxford, UK.
[64] Centre for Genetic Epidemiology and Biostatistics, University of Western Australia, Perth, Australia.
[65] Istituto di Neurogenetica e Neurofarmacologia (INN), Consiglio Nazionale delle Ricerche, c/o Cittadella Universitaria di Monserrato, Monserrato, Cagliari, Italy.
[66] Western Australian Sleep Disorders Research Institute, Queen Elizabeth Medical Centre II, Perth, Australia.
[67] Department of Endocrinology, Diabetes and Nutrition, Charite-Universitaetsmedizin Berlin, Berlin, Germany.
[68] Department of Clinical Nutrition, German Institute of Human Nutrition Potsdam-Rehbruecke, Nuthetal, Germany.
[69] Division of Endocrinology, Diabetes, and Hypertension, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts, USA.
[70] Department of Human Genetics, Leiden University Medical Centre, Leiden, The Netherlands.
[71] Department of Cardiovascular Research, Istituto di Ricerche Farmacologiche 'Mario Negri', Milan, Italy.

[72] Institut National de la Santé et de la Recherche Médicale, Institut National de la Recherche Agronomique, Université Paris 13, Bobigny Cedex, France.
[73] Department of Medicine III, Division Prevention and Care of Diabetes, University of Dresden, Dresden, Germany.
[74] Center for Human Nutrition, University of Texas Southwestern Medical Center, Dallas, Texas, USA.
[75] Department of Genetics and Pathology, Rudbeck Laboratory, Uppsala University, Uppsala, Sweden.
[76] Centre Hospitalier Universitaire, de Poitiers, Endocrinologie Diabetologie, CIC INSERM 0802, INSERM U927, Université de Poitiers, Unité de Formation et de Recherche, Médecine Pharmacie, Poitiers, France.
[77] Department of Public Health and Clinical Medicine, Section for Nutritional Research, Umeå University, Umeå, Sweden.
[78] Department of Clinical Sciences, Obstetrics and Gynecology, University of Oulu, University of Oulu, Finland.
[79] Centre National de Génotypage/Institut de génomique/Commissariat à l'énergie atomique, Evry Cedex, France.
[80] INSERM U872, Faculté de Médecine Paris Descartes, Paris Cedex, France.
[81] Institute for Clinical Diabetology, German Diabetes Center, Leibniz Center for Diabetes Research at Heinrich Heine University Düsseldorf, Düsseldorf, Germany.
[82] Institute of Genetic Medicine, European Academy Bozen/Bolzano (EURAC), Viale Druso, Bolzano, Italy, Affiliated Institute of the University Lübeck, Lübeck, Germany.
[83] Department of Pulmonary Physiology, Sir Charles Gairdner Hospital, Perth, Australia.
[84] Busselton Population Medical Research Foundation, Sir Charles Gairdner Hospital, Perth, Australia.
[85] Heart Institute of Western Australia, Sir Charles Gairdner Hospital, Nedlands West Australia, Australia.
[86] School of Medicine and Pharmacology, University of Western Australia, Nedlands West Australia, Australia.
[87] Folkhalsan Research Centre, Helsinki, Finland.
[88] Malmska Municipal Health Care Center and Hospital, Jakobstad, Finland.
[89] Nuffield Department of Surgery, University of Oxford, Oxford, UK.
[90] Research Centre for Prevention and Health, Glostrup University Hospital, Glostrup, Denmark.
[91] Faculty of Health Science, University of Copenhagen, Copenhagen, Denmark.
[92] National Institute for Health and Welfare, Unit of Population Studies, Turku, Finland.
[93] Institute of Health Sciences and Biocenter Oulu, University of Oulu, Oulu, Finland.
[94] Department of Public Health, Faculty of Medicine, University of Helsinki, Helsinki, Finland.
[95] National Institute for Health and Welfare, Unit for Child and Adolescent Mental Health, Helsinki, Finland.
[96] Institute for Molecular Medicine Finland (FIMM), University of Helsinki, Helsinki, Finland.
[97] Department of Internal Medicine and Biocenter Oulu, Oulu, Finland.
[98] Diabetes Genetics, Institute of Biomedical and Clinical Science, Peninsula College of Medicine and Dentistry, University of Exeter, Exeter, UK.
[99] National Institute for Health and Welfare, Unit of Living Conditions, Health and Wellbeing, Helsinki, Finland.
[100] Interdisciplinary Centre for Clinical Research, University of Leipzig, Leipzig, Germany.
[101] The Danish Twin Registry, Epidemiology, Institute of Public Health, University of Southern Denmark, Odense, Denmark.
[102] Department of Clinical Sciences, Diabetes and Endocrinology, Lund University, University Hospital Malmö, Malmö, Sweden.
[103] Gladstone Institute of Cardiovascular Disease, University of California, San Francisco, California, USA.
[104] Diabetes Research Center, Diabetes Unit, Massachusetts General Hospital, Boston, Massachusetts, USA.
[105] Department of Medicine, Harvard Medical School, Boston, Massachusetts, USA.
[106] Division of Cardiology, University of Ottawa Heart Institute, Ottawa, Ontario, Canada.
[107] Oxford National Institute for Health Research, Biomedical Research Centre, Churchill Hospital, Oxford, UK.
[108] Department of Clinical Genetics, Erasmus Medical College, Rotterdam, The Netherlands.
[109] Biomedical Research Institute, University of Dundee, Ninewells Hospital and Medical School, Dundee, UK.
[110] Department of Geriatric Medicine and Metabolic Disease, Second University of Naples, Naples, Italy.
[111] National Institute for Health and Welfare, Unit of Public Health Genomics, Helsinki, Finland.
[112] Department of Medical Genetics, University of Helsinki, Helsinki, Finland.
[113] Department of Medical Statistics, Epidemiology and Medical Informatics, Andrija Stampar School of Public Health, Medical School, University of Zagreb, Rockefellerova, Zagreb, Croatia.
[114] Department of Clinical Genetics, VU University and Medical Center, Amsterdam, The Netherlands.

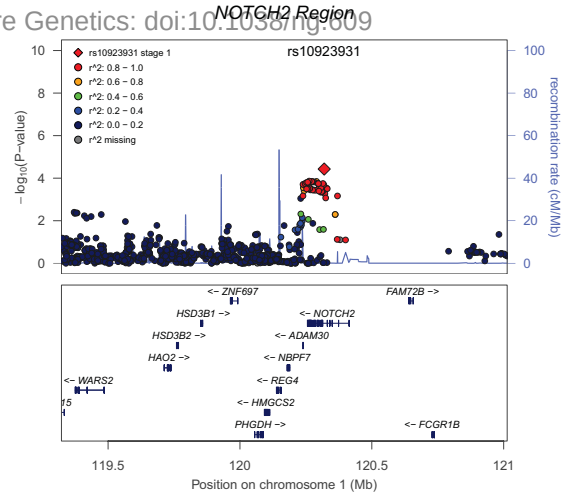
[115] Department of Obstetrics and Gynaecology, Oulu University Hospital, Oulu, Finland.
[116] Departments of Medicine, Epidemiology and Health Services, University of Washington, Seattle, Washington, USA.
[117] Group Health Research Institute, Group Health Cooperative, Seattle, Washington, USA.
[118] Institute of Biometrics and Epidemiology, German Diabetes Centre, Leibniz Centre at Heinrich Heine University Düsseldorf, Düsseldorf, Germany.
[119] Department of Biostatistics, University of Washington, Seattle, Washington, USA.
[120] Department of Internal Medicine, Erasmus Medical College, Rotterdam, The Netherlands.
[121] Department of Metabolic Diseases, Heinrich Heine University Düsseldorf, Düsseldorf, Germany.
[122] Department of Public Health and Clinical Medicine, Section for Family Medicine, Umeå University, Umeå, Sweden.
[123] School of Public Health, Department of General Practice, University of Aarhus, Aarhus, Denmark.
[124] Department of Public Health and Primary Care, Strangeways Research Laboratory, University of Cambridge, Cambridge, UK.
[125] MRC Epidemiology Resource Centre, University of Southampton, Southampton General Hospital, Southampton, UK.
[126] Department of Epidemiology, University of Texas, M.D. Anderson Cancer Center, Houston, Texas, USA.
[127] Leibniz-Institut für Arteriosklerosеforschung an der Universität Münster, Münster, Germany.
[128] Atherosclerosis Research Unit, Department of Medicine, Karolinska Institutet, Stockholm, Sweden.
[129] Laboratory of Neurogenetics, National Institute on Aging, Bethesda, Maryland, USA.
[130] Department of Epidemiology, University of Washington, Seattle, Washington, USA.
[131] Seattle Epidemiologic Research and Information Center, Department of Veterans Affairs Office of Research and Development, Seattle, Washington, USA.
[132] Department of Medical Sciences, Uppsala University, Uppsala, Sweden.
[133] Medstar Research Institute, Baltimore, Maryland, USA.
[134] Clinical Research Branch, National Institute on Aging, Baltimore, Maryland, USA.
[135] Institut interrégional pour la santé (IRSA), La Riche, France.
[136] Coordination Centre for Clinical Trials, University of Leipzig, Leipzig, Germany.
[137] Department of Medicine, Helsinki University Hospital, University of Helsinki, Helsinki, Finland.
[138] Department of Internal Medicine, Leiden University Medical Centre, Leiden, The Netherlands.
[139] Research Unit, Cardiovascular Genetics, Nancy University Henri Poincaré, Nancy, France.
[140] EMGO Institute for Health and Care Research, Department of Psychiatry, VU University Medical Center, Amsterdam, The Netherlands.
[141] Department of Internal Medicine, Centre Hospitalier Universitaire Vaudois, Lausanne, Switzerland.
[142] Genomic Medicine, Imperial College London, Hammersmith Hospital, London, UK.
[143] Epidemiology and Public Health, Queen's University Belfast, Belfast, UK.
[144] Medical Products Agency, Uppsala, Sweden.
[145] National Institute for Health and Welfare, Unit of Chronic Disease Epidemiology and Prevention, Helsinki, Finland.
[146] Departments of Nutrition and Epidemiology, Harvard School of Public Health, Boston, Massachusetts, USA.
[147] Channing Laboratory, Brigham and Women's Hospital and Harvard Medical School, Boston, Massachusetts, USA.
[148] Genetic Epidemiology and Clinical Research Group, Department of Public Health and Clinical Medicine, Section for Medicine, Umeå University Hospital, Umeå, Sweden.
[149] London School of Hygiene and Tropical Medicine, London, UK.
[150] Department of Medicine, School of Medicine, Johns Hopkins University, Baltimore, Maryland, USA.
[151] The Welch Center for Prevention, Epidemiology, and Clinical Research, School of Medicine and Bloomberg School of Public Health, Johns Hopkins University, Baltimore, Maryland, USA.
[152] Division of Epidemiology and Community Health, School of Public Health, University of Minnesota, Minneapolis, Minnesota, USA.
[153] Department of Endocrinology and Diabetes, Norfolk and Norwich University Hospital National Health Service Trust, Norwich, UK.
[154] Department of Medicine, University of Kuopio and Kuopio University Hospital, Kuopio, Finland.
[155] Faculty of Health Science, University of Southern Denmark, Odense, Denmark.
[156] Institute of Biomedical Science, Faculty of Health Science, University of Copenhagen, Copenhagen, Denmark.
[157] Department of Neurology, General Central Hospital, Bolzano, Italy.

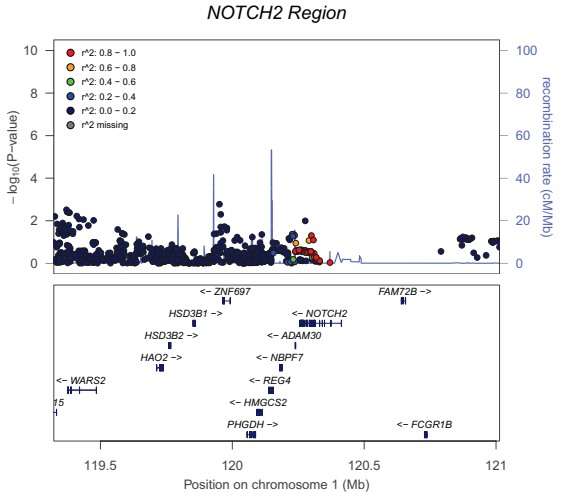
[158] Department of Neurology, University of Lübeck, Lübeck, Germany.
[159] Institute of Medical Informatics, Biometry and Epidemiology, Ludwig-Maximilians-Universität, Munich, Germany.
[160] Klinikum Grosshadern, Munich, Germany.
[161] School of Medicine, University of Split, Split, Croatia.
[162] Gen-Info Ltd., Zagreb, Croatia.
[163] Department of Physiology and Biophysics, Keck School of Medicine, University of Southern California, Los Angeles, California, USA.
[164] Department of Medicine, Division of Endocrinology, Keck School of Medicine, University of Southern California, Los Angeles, California, USA.
[165] Department of Genetics, University of North Carolina, Chapel Hill, North Carolina, USA.
[166] National Institute for Health and Welfare, Unit of Diabetes Prevention, Helsinki, Finland.
[167] South Ostrobothnia Central Hospital, Seinajoki, Finland.
[168] Departments of Medicine and Epidemiology, University of Washington, Seattle, Washington, USA.
[169] Longitudinal Studies Section, Clinical Research Branch, National Institute on Aging, NIH, Baltimore, Maryland, USA.
[170] Faculty of Medicine, University of Iceland, Reykjavík, Iceland.
[171] Lab of Cardiovascular Sciences, National Institute on Aging, National Institutes of Health, Baltimore, Maryland, USA.
[172] Department of Clinical Sciences/Clinical Chemistry, University of Oulu, University of Oulu, Oulu, Finland.
[173] National Institute of Health and Welfare, Oulu, Finland.
[174] Department of Preventive Medicine, Keck School of Medicine, University of Southern California, Los Angeles, California, USA.

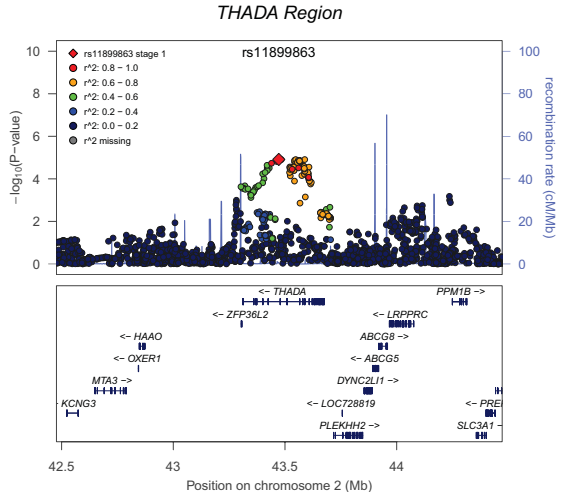
SUPPLEMENTARY FIGURE LEGENDS

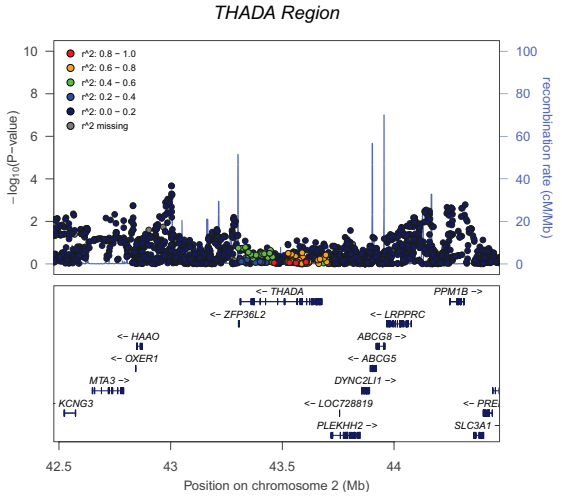
Supplementary Figure 1: **Regional plots of the 30 established T2D loci before (left panel) and after (right panel) conditioning on the index associated variant.** Genotyped and imputed SNPs passing QC across all Stage 1 studies are plotted with their meta-analysis P-values (as -log10 values) as a function of genomic position (NCBI Build 36). In each panel, the index association SNP is represented by a diamond, with Stage 1 meta-analysis results denoted by a red diamond, and the combined Stage 1 and Stage 2 meta-analysis result with a clear symbol. Estimated recombination rates (taken from HapMap CEU) are plotted to reflect the local LD structure. SNP color indicates LD with the index SNP according to a scale from r^2=0 to r^2=1 based on pairwise r^2 values from HapMap CEU. Gene annotations were taken from the University of California-Santa Cruz genome browser. Five of the genome-wide significant regions (*TP53INP1, CDKN2A, HHEX/IDE, KCNJ11, KCNQ1*) show preliminary evidence for secondary signals ($P<10^{-4}$).

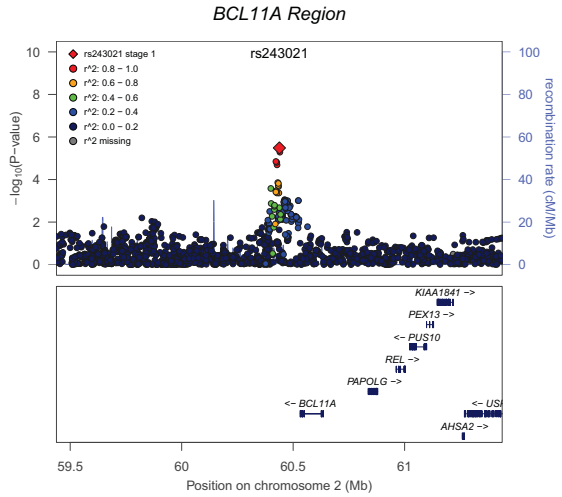
Supplementary Figure 2: **Comparison of T2D and fasting glucose effect sizes for 31 autosomal signals.** For each of the 31 autosomal signals shown, in this and previous papers, to be genome-wide significant for T2D (and excluding six recently reported signals[11-13]) we plot the T2D OR obtained from the Stage 1 DIAGRAM+ analysis (y-axis) against the fasting glucose effect size obtained from the ~46,000 non-diabetic individuals included in the MAGIC stage 1 analysis[12] (x-axis). The colors denote the association P-value for fasting glucose in the MAGIC analysis (red: $P<10^{-3}$, orange: $10^{-3}<P<10^{-2}$, yellow $0.01<P<0.05$, green $0.05<P<0.20$, blue: $P>0.20$). All effect directions are shown for the T2D-risk allele. Whilst some variants have an effect on both T2D-risk and fasting glucose in non-diabetic subjects (e.g. *MTNR1B, SLC30A8*), other T2D-risk variants (e.g. *IRS1, KCNJ11, HNF1A*) have no detectable impact on glucose homeostasis in non-diabetic individuals

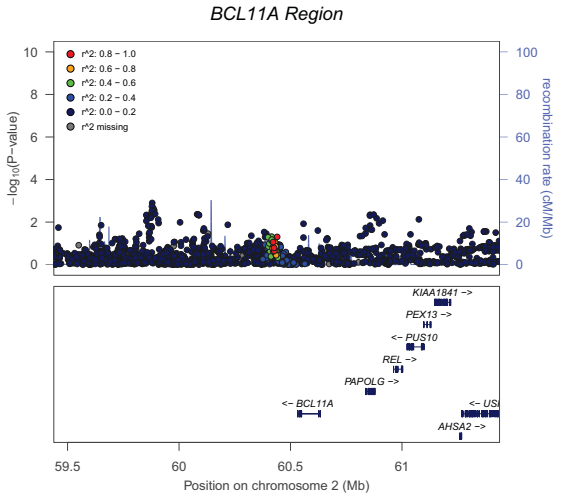


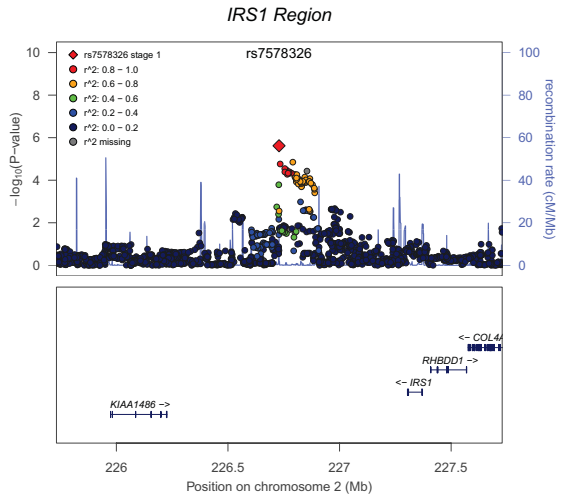


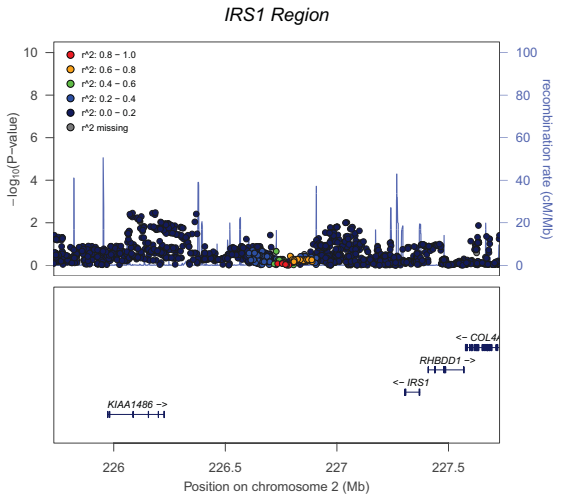




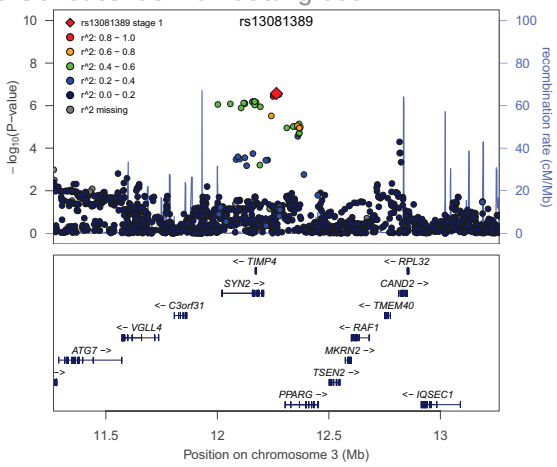




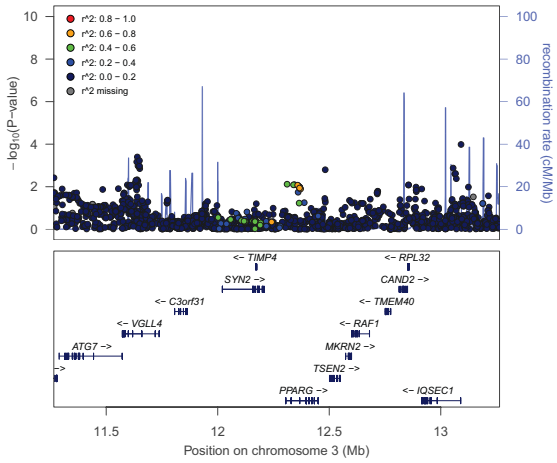




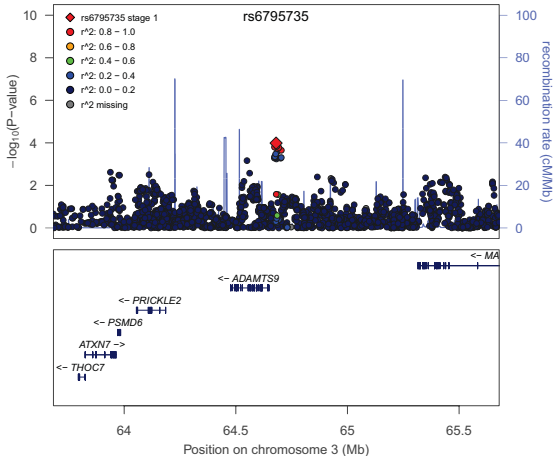
PPARG Region



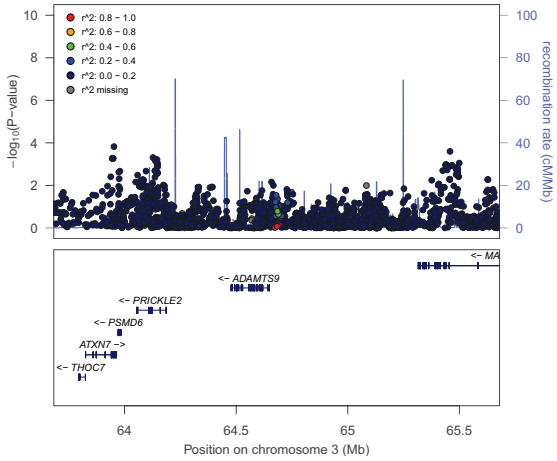
PPARG Region



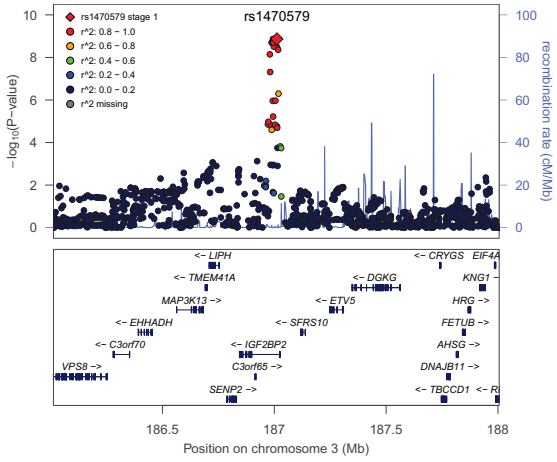
ADAMTS9 Region



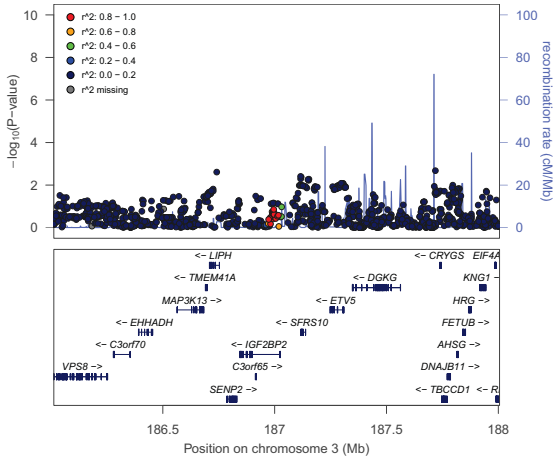
ADAMTS9 Region



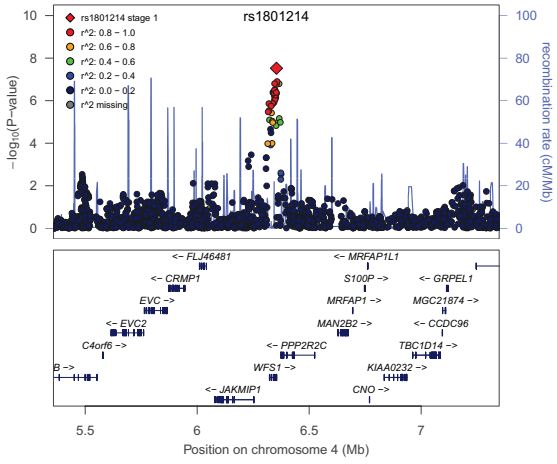
IGF2BP2 Region



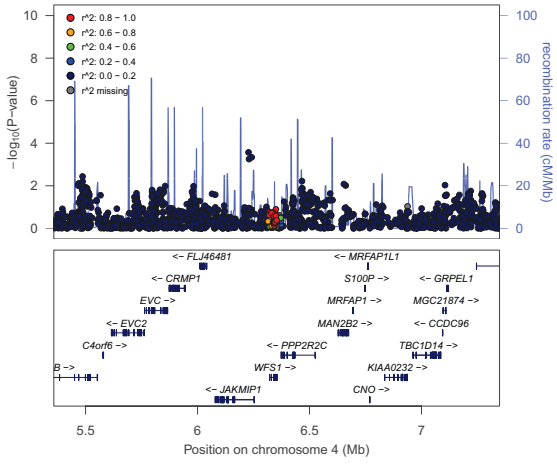
IGF2BP2 Region

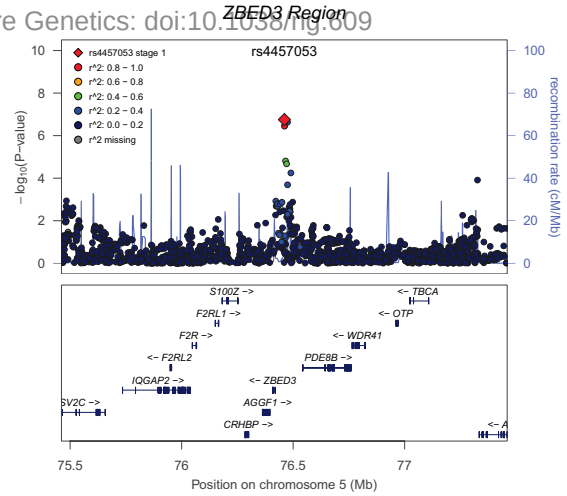


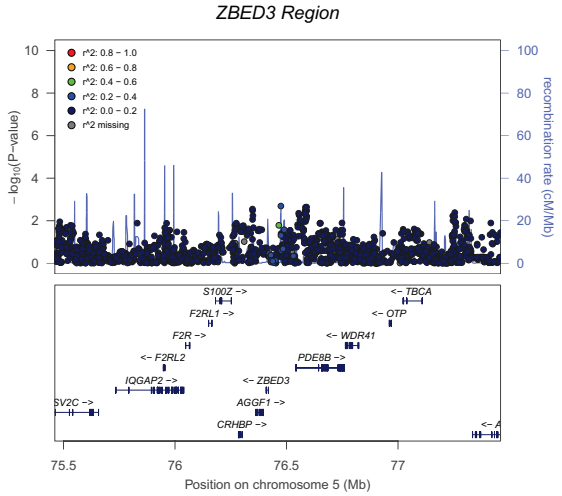
WFS1 Region

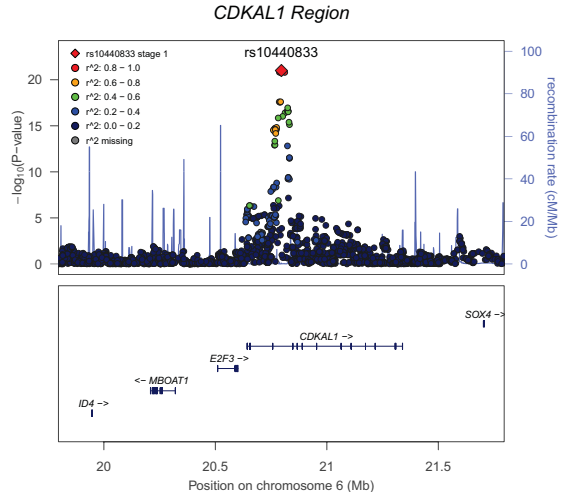


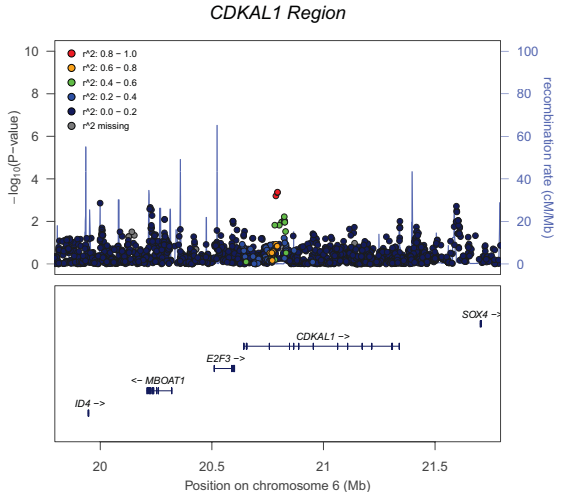
WFS1 Region

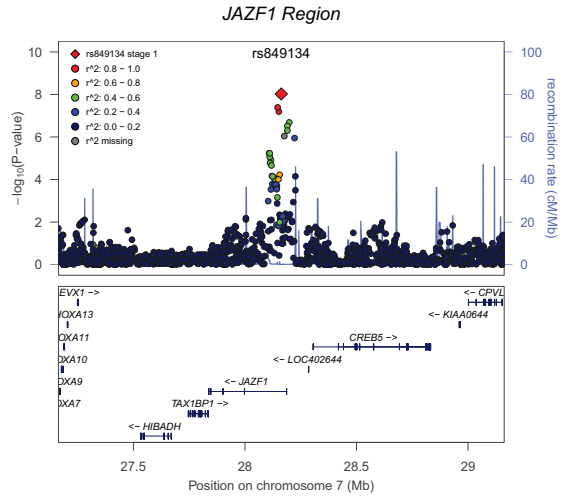


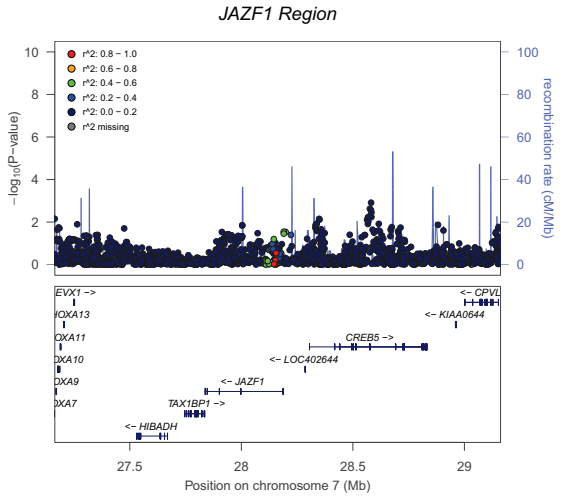


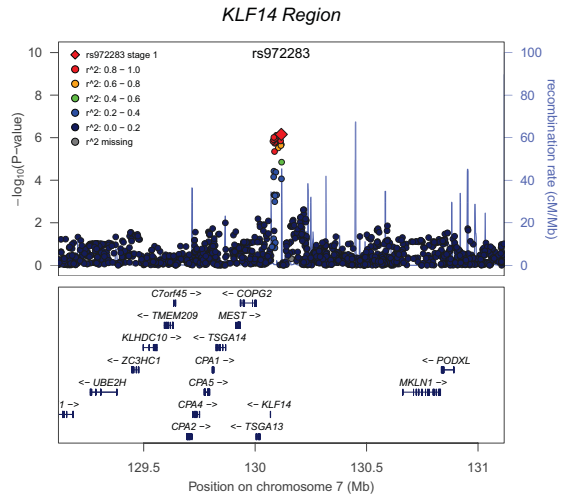


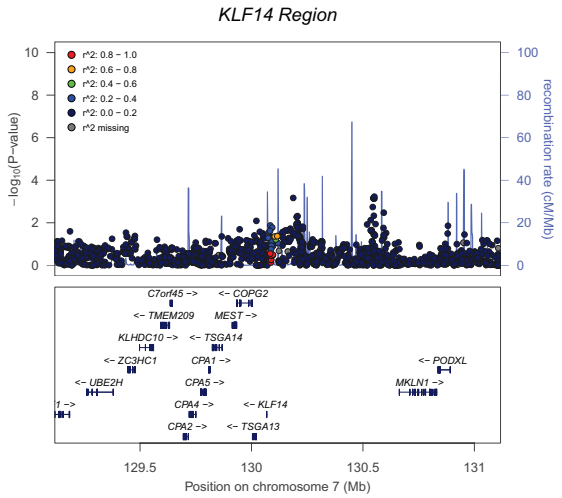


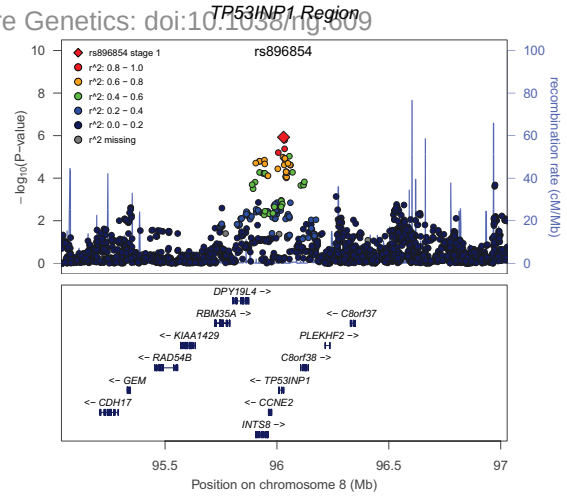


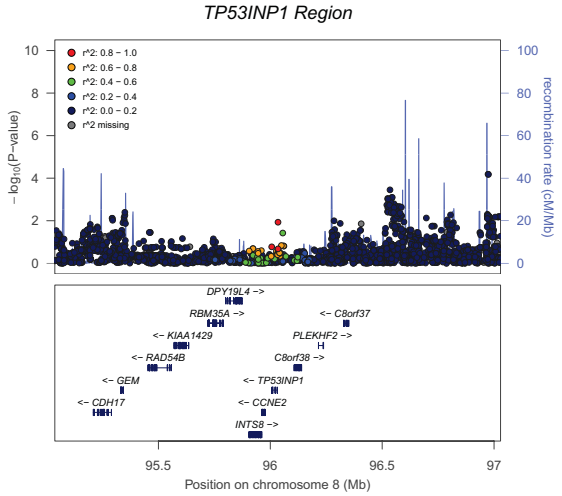


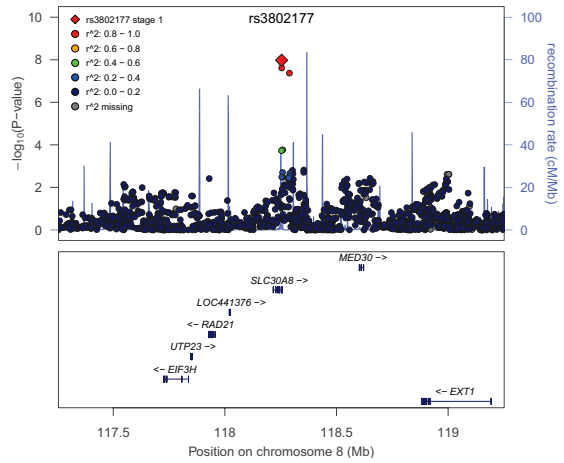


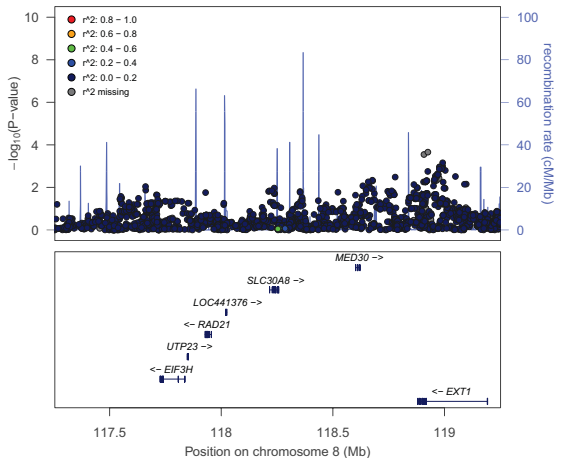


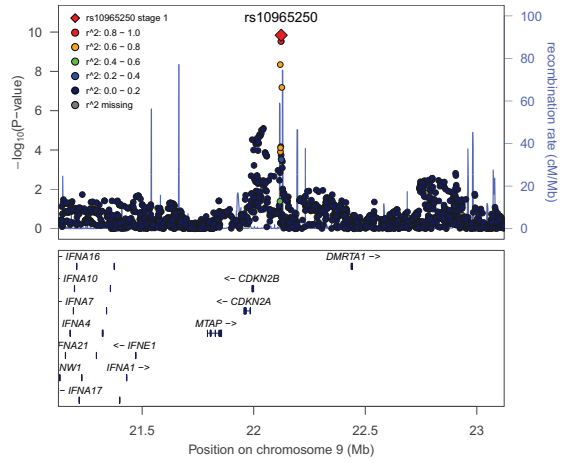


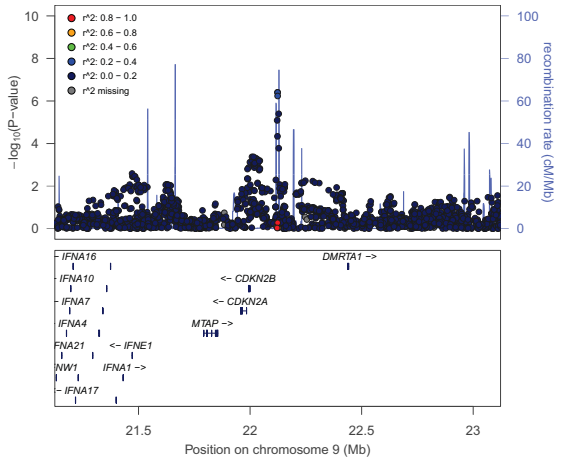


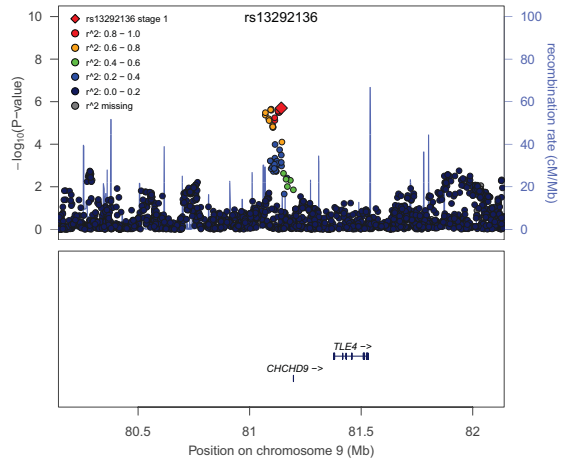


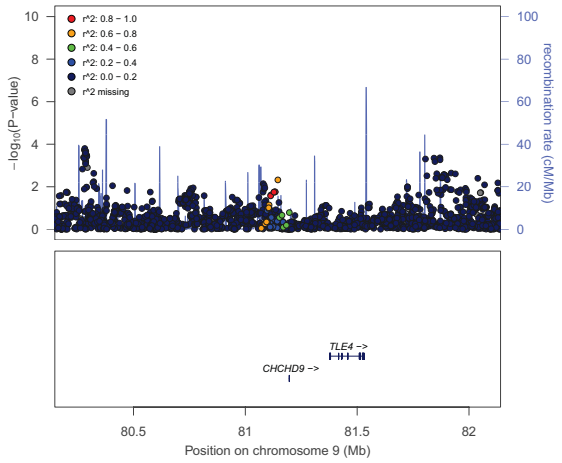


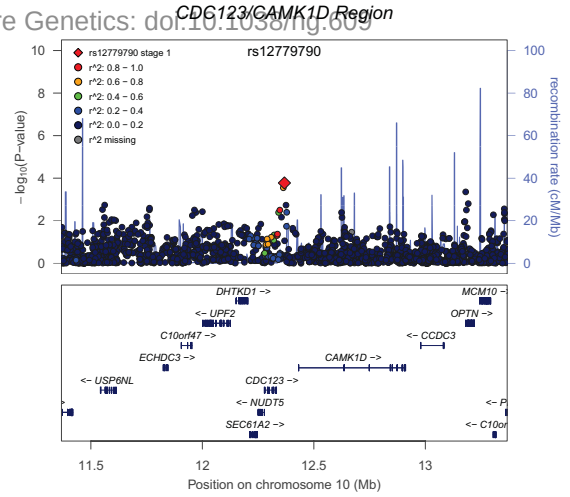


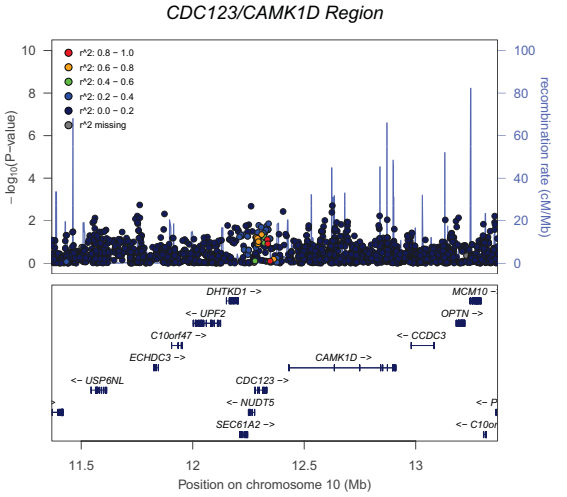


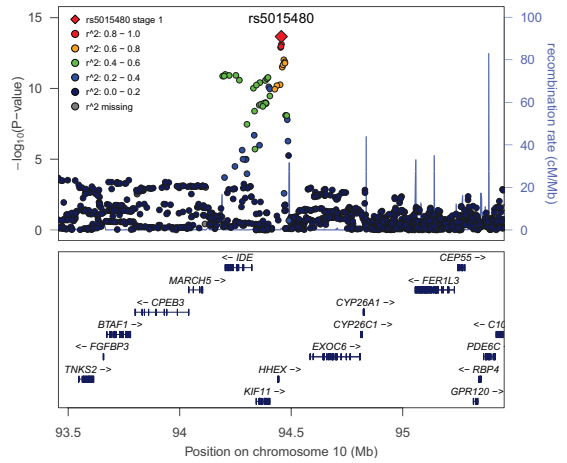


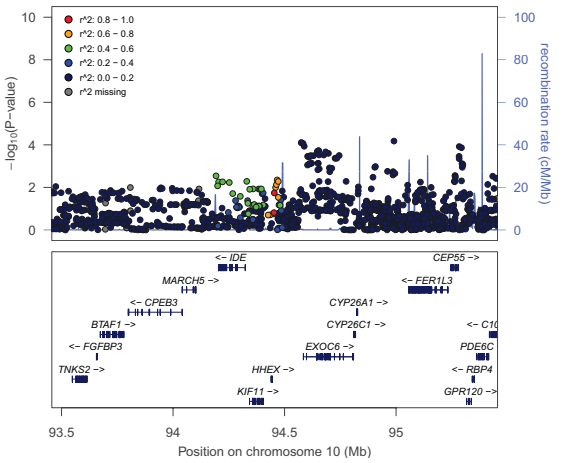


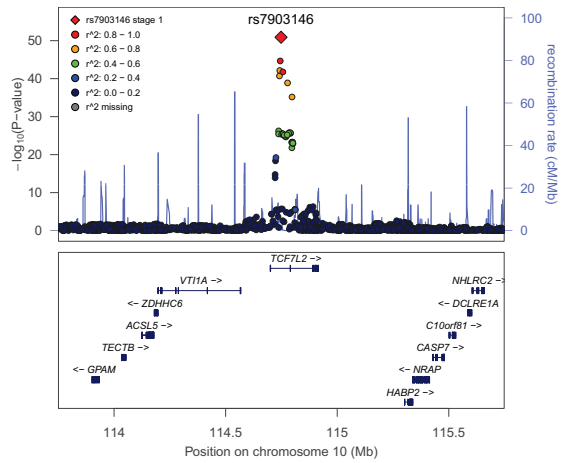


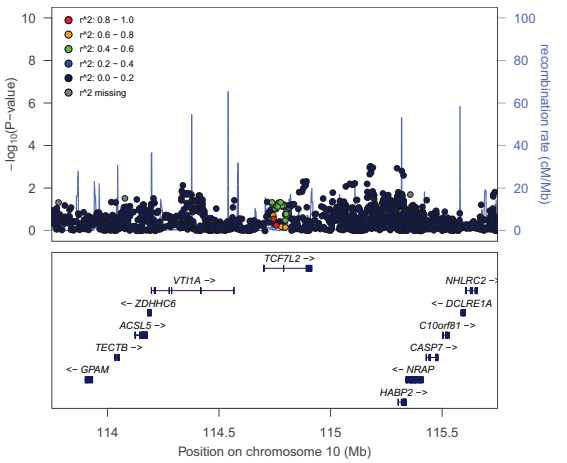


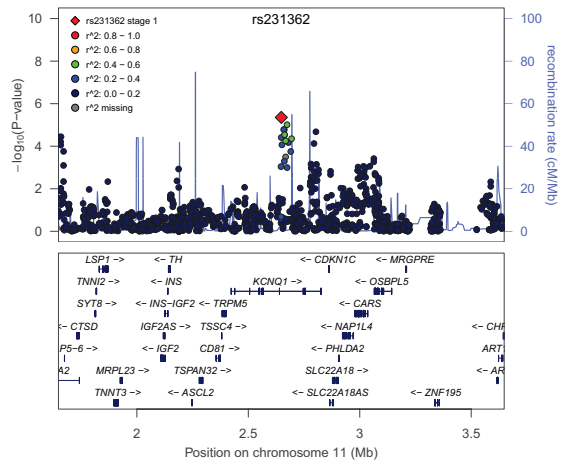


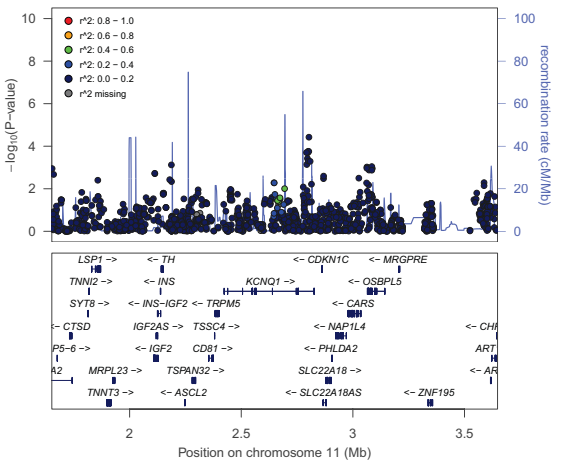






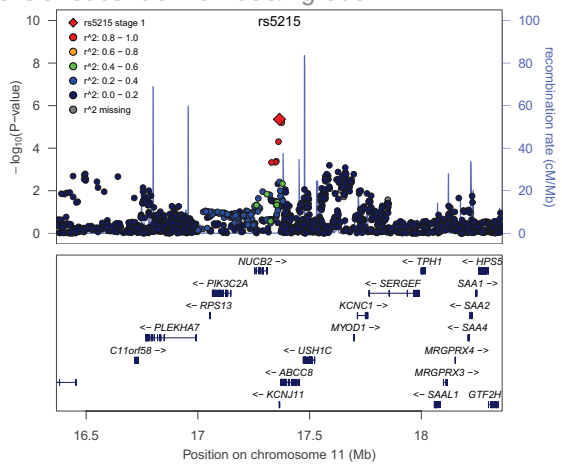




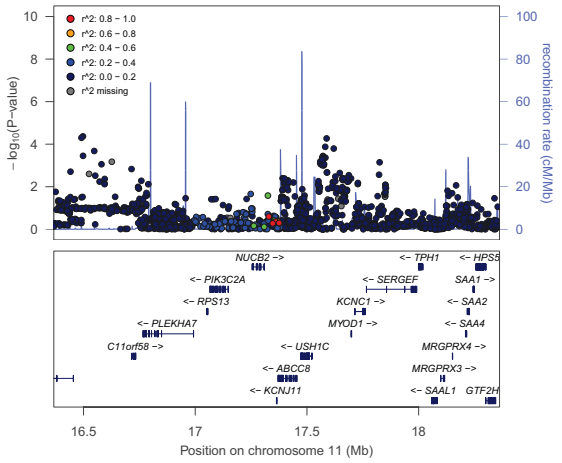


Nature Genetics: doi:10.1038/ng.609

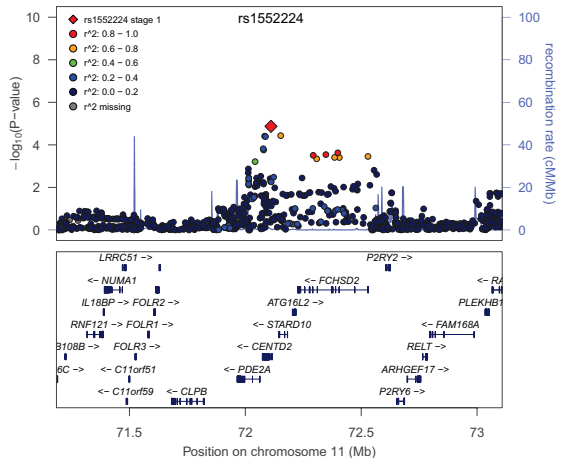
KCNJ11 Region



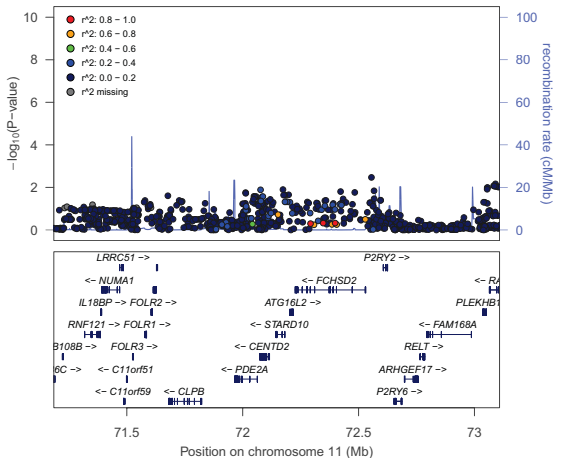
KCNJ11 Region



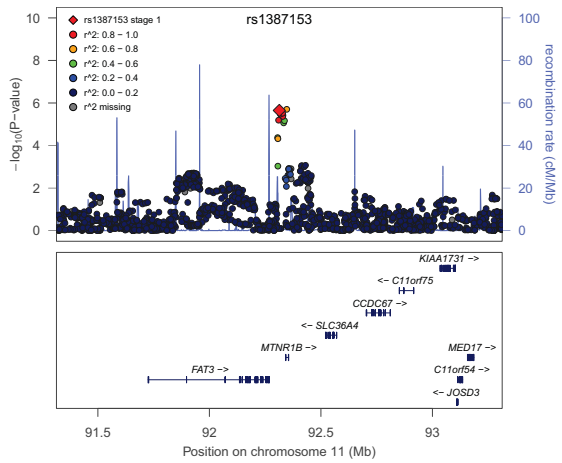
CENTD2 Region



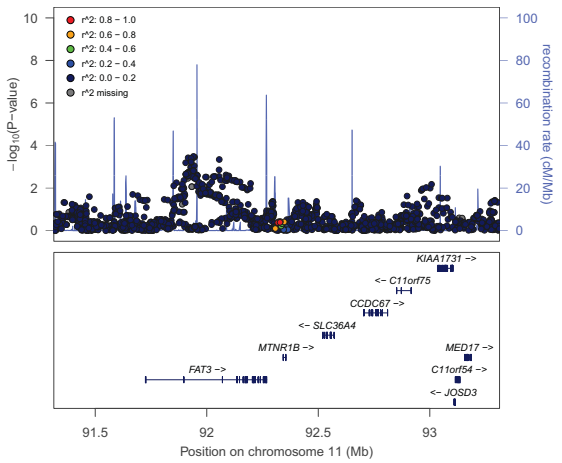
CENTD2 Region



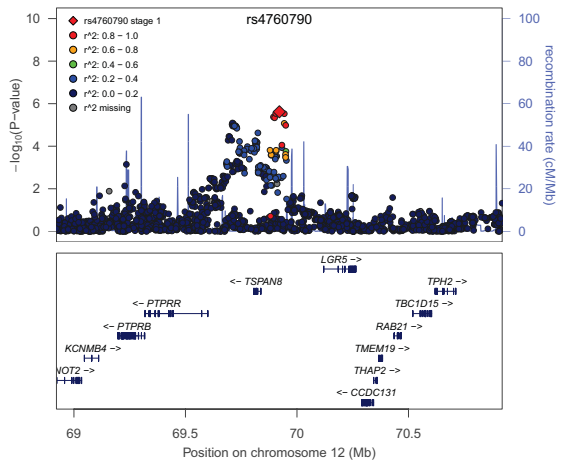
MTNR1B Region



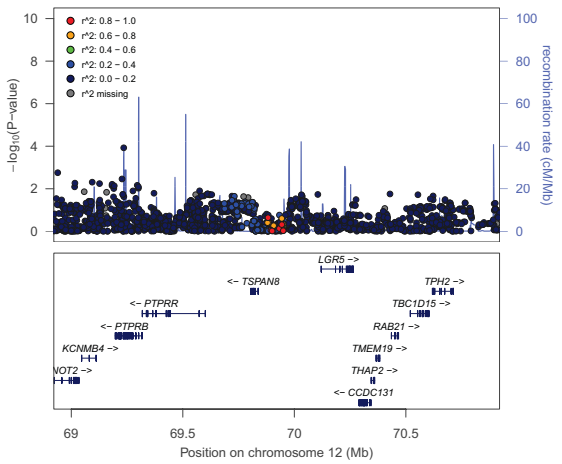
MTNR1B Region

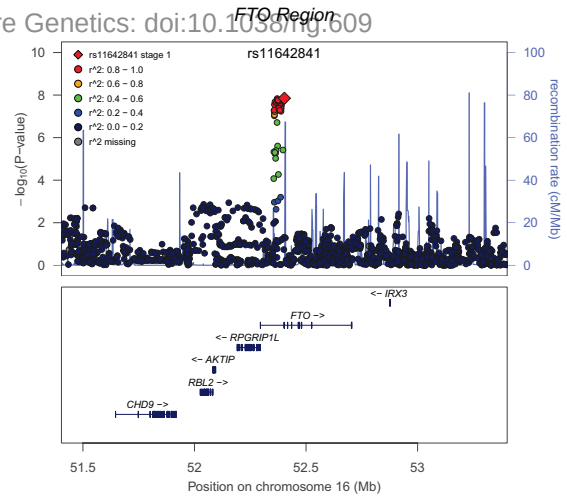


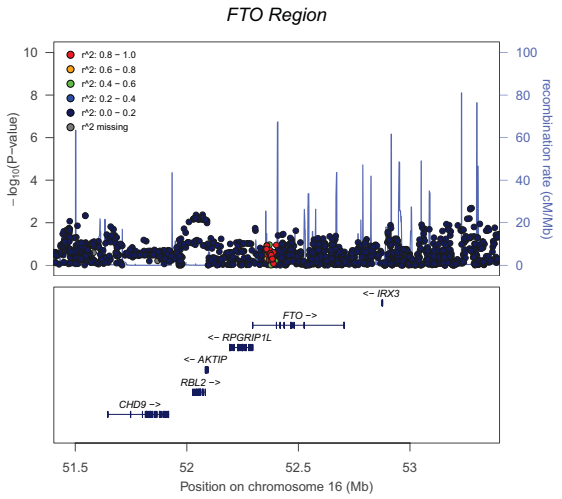
TSPAN8/LGR5 Region

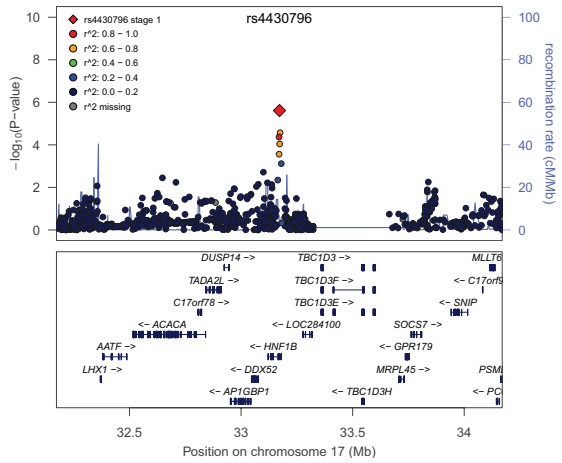


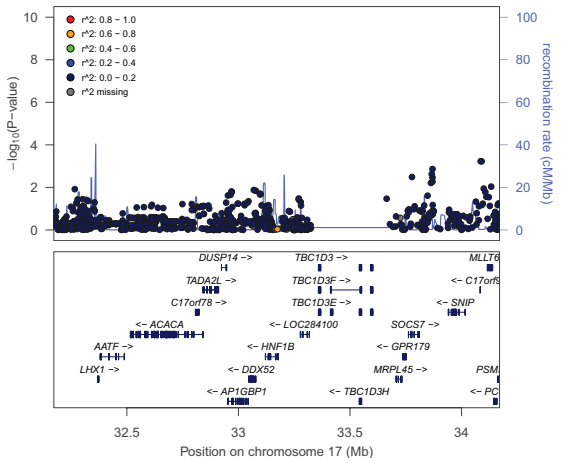
TSPAN8/LGR5 Region

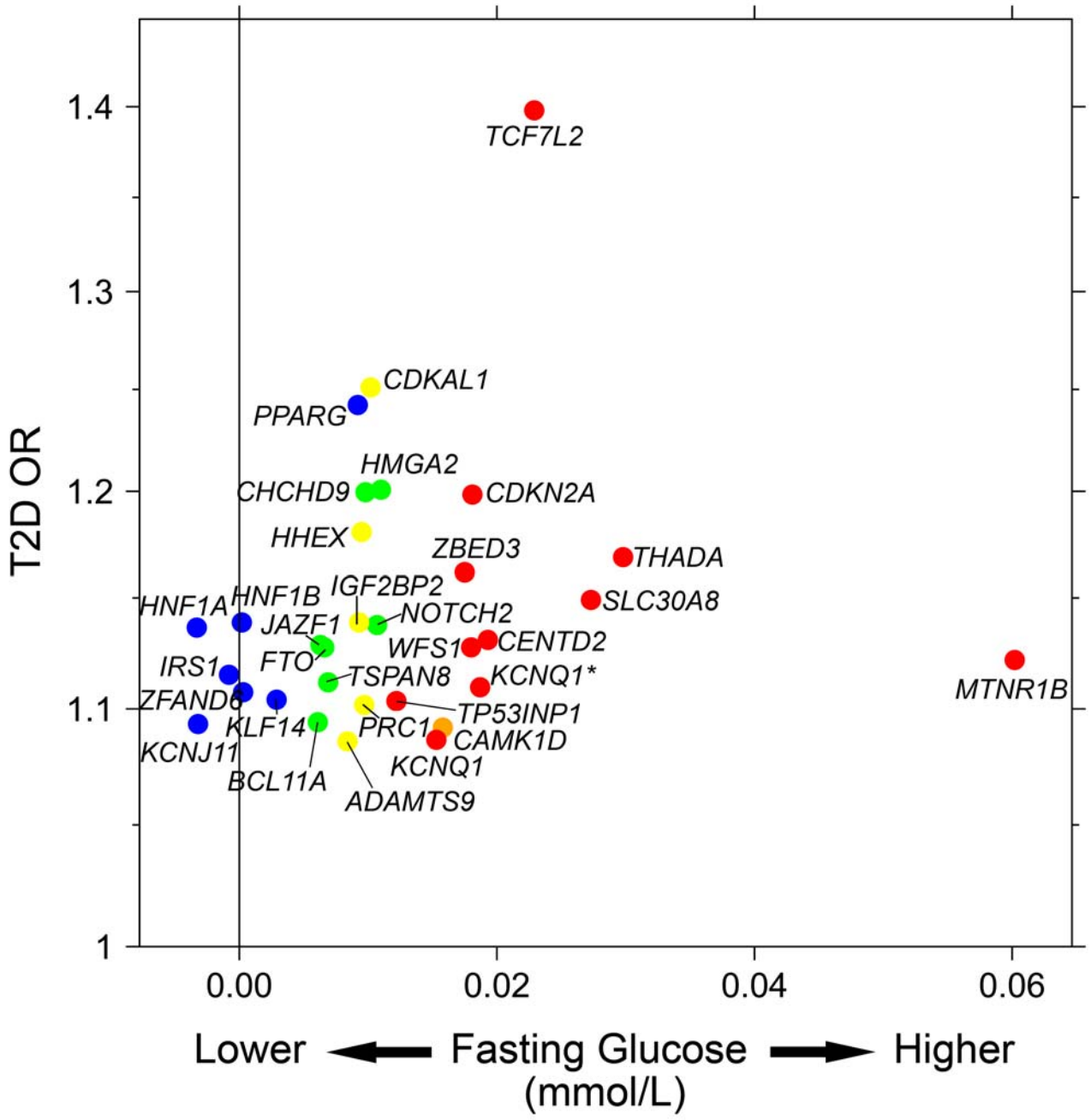











T2D OR
1.4
1.3
1.2
1.1
1
0.00
0.02
0.04
0.06
Lower
Fasting Glucose
Higher
(mmol/L)
TCF7L2
CDKAL1
PPARG
HMGA2
CHCHD9
CDKN2A
HHEX
ZBED3
THADA
HNF1A
HNF1B
IGF2BP2
SLC30A8
JAZF1
NOTCH2
WFS1
CENTD2
IRS1
FTO
TSPAN8
KCNQ1*
MTNR1B
ZFAND6
TP53INP1
KLF14
PRC1
CAMK1D
KCNJ11
KCNQ1
BCL11A
ADAMTS9

Supplementary Table 1: Summary of samples included in Stages 1 and 2

Study	Origin	Cases	Controls	Effective size
Stage 1				
Diabetes Gene Discovery Group (DGDG)	**France**	**679**	**697**	**1,376**
deCODE	**Iceland**	**1,465**	**23,194**	**5,512**
Diabetes Genetics Initiative (DGI)	**Sweden, Finland**	**1,022**	**1,075**	**2,096**
EUROSPAN (ERF, MICROS, ORCADES, VIS)	**Croatia, Italy, Netherlands, UK**	**268**	**3,710**	**966**
Finland US Investigation of NIDDM genetics (FUSION)	**Finland**	**1,161**	**1,174**	**2,334**
KORAgen	**Germany**	**433**	**1,438**	**1,332**
Rotterdam	**Netherlands**	**1,178**	**4,761**	**3,778**
Wellcome Trust Case Control Consortium (WTCCC)	**UK**	**1,924**	**2,938**	**4,650**
Combined Stage 1		**8,130**	**38,987**	**22,044**
Stage 2				
In silico replication [Autosomal]				
Atherosclerosis Risk in Communities (ARIC: European subset)	**US**	**691**	**6,425**	**2,496**
Framingham Heart Study (FHS)	**US**	**674**	**7,664**	**2,478**
Nurses Health Study (NHS)	**US**	**1,467**	**1,754**	**3,196**
		2,832	**15,843**	**8,170**
De novo genotyping				
ADDITION/Ely	**UK**	**892**	**1,612**	**2,296**
Breda[a,b]	**Netherlands**	**(545)**	**(880)**	**(1,346)**
Cambridgeshire Case Control Study (CCCS)	**UK**	**541**	**527**	**1,068**
DARTS (UK T2D Genetics Consortium)[a]	**UK**	**5,113**	**6,615**	**11,536**
Diabetes Gene Discovery Group (Stage 2)[a]	**France**	**3,200**	**3,125**	**6,324**
DiaGene/Rotterdam	**Netherlands**	**1,699**	**1,656**	**3,354**
Finland US Investigation of NIDDM genetics (FUSION2 data set)	**Finland**	**1,211**	**1,266**	**2,476**
GCI-Polish	**Poland**	**790**	**803**	**1,592**
GCI-US	**US**	**1,190**	**1,168**	**2,358**
HUNT	**Norway**	**1,213**	**3,221**	**3,524**
KORAgen	**Germany**	**1,016**	**1,498**	**2,422**
Malmo Diabetes Registry/Diet and Cancer Study	**Sweden**	**2,449**	**3,363**	**5,668**
METSIM	**Finland**	**940**	**4,152**	**3,066**
Norfolk Diabetes Case Control Study and EPIC (NDCCS)	**UK**	**6,056**	**6,428**	**12,472**
Steno	**Denmark**	**3,672**	**5,046**	**8,502**
Ulm[a]	**Germany**	**944**	**1,949**	**2,544**
W2C/UKBS (UK T2D Genetics Consortium, RS2 set)	**UK**	**654**	**1,653**	**1,874**
		31,580	**44,082**	**71,076**
Combined Stage 2 [Autosomal only]		**34,412**	**59,925**	**79,246**
Total Overall [Autosomal]		**42,542**	**98,912**	**101,290**

[a] Only these studies contributed samples to the follow-up of the X-chromosome signal, totalling 8,535 cases and 12,362 controls.

[b] Note that the Breda cohort contributed to replication of the X chromosome result only, so is not included in the combined sample count.

Supplemental Table 2: Sample characteristics for Stage 1 and Stage 2 samples

Study Sample	Status	Total	N (M/F)	Mean Age (Years)	Age at Onset (Years)[b]	BMI (kg/m^2)	Fasting Glucose (mmol/l)[b]
Stage 1							
DGDG	Case	679	413/266	59.5 ± 10.1	45.1 ± 8.4	25.9 ± 2.8	9.2 ± 3.1
	Control	697	281/416	53.9 ± 5.6	-	23.2 ± 1.8	5.1 ± 0.4
deCODE	Case	1,456	868/597	68.4 ± 12.7	55.1 ± 12.7 (n=1,009)	30.1 ± 5.4	8.5 ± 2.7 (n=1,318)
	Control	23,194	7,316/15,878	59.7 ± 18.1	-	26.8 ± 5.0	5.3 ± 0.7 (n=5,906)
DGI	Case	1,022	529/493	65.0 ± 10.0	58.0 ± 10.0	28.1 ± 4.1	9.5 ± 3.1
	Control	1,075	540/535	58.0 ± 10.0	-	27.6 ± 3.7	5.3 ± 0.5
EUROSPAN – VIS	Case	123	52/71	64.5 ± 13.3	n/a	28.4 ± 4.2	8.0 ± 2.5
	Control	660	276/384	55.0 ± 15.4	-	27.2 ± 4.2	5.3 ± 0.7
EUROSPAN – ERF	Case	74	34/40	60.2 ± 12.4	n/a	30.3 ± 5.4	7.2 ± 2.2
	Control	1,327	523/804	49.9 ± 15.8	-	26.6 ± 4.6	4.5 ± 0.7
EUROSPAN – ORCADES	Case	27	19/9	62.4 ± 10.6	n/a	33.1 ± 5.8	8.9 ± 2.5
	Control	688	313/375	53.2 ± 15.7	-	27.5 ± 4.7	5.3 ± 0.5
EUROSPAN – MICROS	Case	44	22/22	63.1 ± 12.5	n/a	30.6 ± 8.6	8.9 ± 3.4
	Control	1,035	445/590	44.3 ± 15.7	-	25.4 ± 4.4	4.6 ± 0.6
FUSION	Case	1,161	653/508	62.9 ± 7.6	53.7 ± 9.1	30.2 ± 4.7	9.4 ± 3.1
	Control	1,174	574/600	63.6 ± 7.4	-	27.1 ± 3.9	5.3 ± 0.5
KORAgen	Case	433	255/178	65.2 ± 8.3	58.2 ± 10.3	30.9 ± 5.0	n/a
	Control	1,438	693/745	61.9 ± 10.2	-	27.7 ± 4.3	n/a
Rotterdam	Case	1,178	488/690	71.1 ± 8.9[a]	71.5 ± 8.9	27.4 ± 4.0[a]	n/a
	Control	4,761	1,928/2,833	69.0 ± 9.1	-	26.0 ± 3.6	n/a
WTCCC	Case	1,924	1,118/806	58.6 ± 10.1	50.3 ± 9.2	30.7 ± 6.1	n/a
	Control	1,480	741/739	44.5 ± 0.3	-	27.4 ± 4.8	n/a
	Control	1,458	705/753	43.4 ± 12.4	-	26.2 ± 4.3	n/a
Stage 2							
***In silico* replication**							
ARIC	Case	691	375/316	56.1 ± 5.6	50.4 ± 10.8 n=454	30.4 ± 5.5	9.3 ± 3.6
	Control	6,425	2,848/3,577	54.0 ± 5.7	-	26.4 ± 4.6	5.4 ± 0.4
FHS	Case	674	386/288	63.7 ± 12.4	58[c]	31.4 ± 6.5	8.6 ± 2.8
	Control	7,664	3,443/4,221	52.3 ± 16.0	-	27.0 ± 5.1	5.3 ± 0.5
NHS	Case	1,467	0/1,467	43.5 ± 6.7[a]	59.4 ± 9.9	27.4 ± 5.0[a]	n/a
	Control	1,754	0/1,754	43.1 ± 6.8	-	23.5 ± 3.9	n/a
***De novo* genotyping or combined *de novo* genotyping/*in silico* replication**							
Addition/Ely	Case	892	339/553	61.2 ± 7.2	n/a	33.1 ± 5.7	n/a
	Control	1,612	886/726	60.9 ± 9.1	-	27.1 ± 4.7	5.0 ± 0.5
Breda	Case	545	245/300	70.4 ± 10.0	58.6 ± 11.5	27.9 ± 4.3	8.3 ± 2.4
	Control	880	341/539	47.7 ± 12.8	-	n/a	n/a
CCCS	Case	541	198/343	63.4 ± 7.9	55.8 ± 8.8	29.8 ± 5.3	n/a
	Control	527	189/338	63.6 ± 7.9	-	27.4 ± 4.1	n/a
DARTS	Case	5,113	2,918/2,195	63.6 ± 9.3	56.2 ± 8.6	31.9 ± 6.3	n/a
	Control	6,615	3,254/3,361	58.6 ± 11.8	-	27.2 ± 4.6	4.9 ± 0.5
DGDG (Stage 2)	Case	3,200	2,013/1,185	62.2 ± 11.0	50.9 ± 11.0	29.2 ± 3.9	8.4 ± 2.9 (n=1,711)
	Control	3,125	1,320/1,805	49.3 ± 11.7	-	24.5 ± 3.6	5.1 ± 0.5
DiaGene / Rotterdam	Case	1,699	920/779	66.9 ± 10.2	n/a	30.4 ± 5.3	n/a
	Control	1,656	727/929	64.6 ± 7.9	-	27.0 ± 3.8	n/a
FUSION2	Case	1,211	722/489	59.5 ± 8.5	55.5 ± 8.9	30.8 ± 5.4	7.7 ± 2.4
	Control	1,266	773/493	58.7 ± 7.7	-	26.9 ± 3.9	5.4 ± 0.4
GCI-Polish	Case	790	326/464	61.9 ± 9.6	n/a	29.5 ± 4.8	n/a
	Control	803	300/503	58.2 ± 7.1	-	26.1 ± 3.7	n/a
GCI-US	Case	1,190	624/566	62.1 ± 11.0	n/a	32.9 ± 7.0	n/a
	Control	1,168	615/553	60.2 ± 9.7	-	27.4 ± 5.1	n/a
HUNT	Case	1,213	575/638	68.5 ± 11.4	n/a	29.5 ± 4.8	n/a
	Control	3,221	1,574/1,647	65.0 ± 14.5	-	26.8 ± 4.1	n/a
KORAgen	Case	1,016	561/455	60.5 ± 9.3	n/a	30.5 ± 5.3	n/a
	Control	1,498	821/677	60.6 ± 9.2	-	28.0 ± 4.4	n/a
Malmo	Case	2,449	1,448/1,001	58.4 ± 11.5	n/a	29.7 ± 5.5	10.6 ± 5.3
	Control	3,363	1,254/2,109	57.5 ± 6.0		25.0 ± 3.6	5.4 ± 0.4
METSIM	Case	940	940/0	60.4 ± 6.3	57.1 ± 7.8	30.2 ± 5.2	7.7 ± 2.1
	Control	4,152	4,152/0	56.9 ± 6.8	-	26.2 ± 3.4	5.5 ± 0.4
NDCCS2	Case	6,056	2,481/3,575	66.7 ± 10.9	60.4 ± 11.7	29.9 ± 5.6	n/a
	Control	6,428	3,390/3,038	58.4 ± 9.3	-	26.1 ± 3.6	n/a
Steno	Case	3,672	2,183/1,489	60.4 ± 9.7	52.3 ± 11.0	30.1 ± 5.5	9.2 ± 3.5
	Control	5,046	2,349/2,697	46.4 ± 8.8	-	25.5 ± 4.0	5.3 ± 0.4
Ulm	Case	944	525/419	58.5 ± 10.4	56.1 ± 10.3	27.6 ± 4.1	6.8 ± 2.5
	Control	1,949	885/1,064	43.7 ± 14.7	-	25.9 ± 5.0	4.6 ± 0.6
W2C/UKBS	Case	654	397/252	55.4 ± 8.1	51.3 ± 7.7	32 ± 6.3	n/a
	Control	1,653	852/795	43.6 ± 12.4	-	26 ± 4.5	n/a

[a]Age/Phenotype at study entry, T2D present at study entry or detected during follow-up

[b]Number of samples with trait if >10% of trait values are missing

[c]Median

Supplementary Table 3: DIAGRAM+ Stage 1 for this study (Genome-wide Meta-analysis results for SNPs of Interest)

Locus	SNP	Chr	Position B36 (basepair)	Risk Allele[b]	Nonrisk Allele[b]	Frequency Range for Risk Allele in Stage 1 Studies	Imputed (i) or Genotyped (g)[c]	P-value[d]	Cases (N)[e]	Controls (N)[e]	OR	CI (Lower 95%)	CI (Upper 95%)	Nearby Gene[f]	Rationale for Inclusion in this Table	r^2 Between Pairs[g]
Loci already established as genome wide significant based on previous reports																
1	rs10923931	1	120,319,482	T	G	[0.091 - 0.135]	iigiiigg	6.52×10^{-5}	8130	38987	1.138	1.068	1.212	*NOTCH2*	DIAGRAM+ most significant association P-value	-
2	rs11899863	2	43,472,323	C	T	[0.885 - 0.946]	iiiiiiii	1.04×10^{-5}	8130	38987	1.169	1.091	1.252	*THADA*	DIAGRAM+ most significant association P-value	0.786
	rs7578597	2	43,586,327	T	C	[0.881 - 0.944]	ggigigii	4.47×10^{-5}	8130	38987	1.151	1.076	1.231		other previously reported lead SNP	
3	rs13081389	3	12,264,800	A	G	[0.902 - 0.952]	iiiiiiii	2.01×10^{-7}	8130	38987	1.242	1.145	1.349	*PPARG*	DIAGRAM+ most significant association P-value	0.536
	rs1801282	3	12,368,125	C	G	[0.817 - 0.901]	iigiiigg	8.01×10^{-6}	8130	38987	1.148	1.080	1.219		other previously reported lead SNP	
4	rs6795735	3	64,680,405	C	T	[0.546 - 0.649]	ggggiggg	8.35×10^{-5}	8130	38987	1.086	1.042	1.131	*ADAMTS9*	DIAGRAM+ most significant association P-value	0.280
	rs4607103	3	64,686,944	C	T	[0.706 - 0.814]	iigiiigg	2.34×10^{-4}	8130	38987	1.096	1.044	1.150		other previously reported lead SNP	
5	rs1470579	3	187,011,774	C	A	[0.277 - 0.333]	ggggiggg	2.17×10^{-9}	8130	38987	1.139	1.091	1.188	*IGF2BP2*	DIAGRAM+ most significant association P-value	-
6	rs1801214	4	6,353,923	T	C	[0.545 - 0.644]	iiiiiiii	3.16×10^{-8}	8130	38987	1.128	1.080	1.176	*WFS1*	DIAGRAM+ most significant association P-value	0.917
	rs10010131	4	6,343,816	G	A	[0.563 - 0.634]	iiiiiiii	4.59×10^{-7}	8130	38987	1.110	1.065	1.155		other previously reported lead SNP	
7	rs10440833	6	20,796,100	A	T	[0.229 - 0.342]	iiiiiiii	1.84×10^{-22}	8130	38987	1.251	1.197	1.307	*CDKAL1*	DIAGRAM+ most significant association P-value	0.677
	rs7754840	6	20,769,229	C	G	[0.293 - 0.373]	iigiiigg	3.11×10^{-15}	8130	38987	1.185	1.135	1.236		other previously reported lead SNP	
8	rs849134	7	28,162,747	A	G	[0.492 - 0.534]	iiiiiiii	2.80×10^{-9}	8130	38987	1.129	1.085	1.175	*JAZF1*	DIAGRAM+ most significant association P-value	0.967
	rs864745	7	28,147,081	T	C	[0.493 - 0.528]	iigiiigg	2.06×10^{-8}	8130	38987	1.121	1.077	1.165		other previously reported lead SNP	
9	rs3802177	8	118,254,206	G	A	[0.605 - 0.724]	iiiiii0i	1.45×10^{-8}	7697	37549	1.149	1.095	1.206	*SLC30A8*	DIAGRAM+ most significant association P-value	1.000
	rs13266634	8	118,253,964	C	T	[0.604 - 0.724]	ggigig0i	1.52×10^{-8}	7697	37549	1.149	1.095	1.206		other previously reported lead SNP	
10	rs10965250	9	22,123,284	G	A	[0.816 - 0.855]	iiiiiiii	1.23×10^{-10}	8130	38987	1.198	1.134	1.266	*CDKN2A/B*	DIAGRAM+ most significant association P-value	0.950
	rs10811661	9	22,124,094	T	C	[0.810 - 0.844]	iigiiigg	1.45×10^{-10}	8130	38987	1.191	1.130	1.257		other previously reported lead SNP	
11	rs12779790	10	12,368,016	G	A	[0.168 - 0.226]	iiiiiiii	6.75×10^{-4}	8130	38987	1.092	1.038	1.149	*CDC123/CAMK1D*	DIAGRAM+ most significant association P-value	-
12	rs5015480	10	94,455,539	C	T	[0.526 - 0.624]	iiggiigg	1.33×10^{-15}	8130	38987	1.181	1.133	1.230	*HHEX/IDE*	DIAGRAM+ most significant association P-value	1.000
	rs1111875	10	94,452,862	C	T	[0.526 - 0.626]	ggiggggi	9.10×10^{-15}	8130	38987	1.172	1.126	1.221		other previously reported lead SNP	
13	rs7903146	10	114,748,339	T	C	[0.179 - 0.378]	ggigigii	2.21×10^{-51}	8130	38987	1.398	1.339	1.459	*TCF7L2*	DIAGRAM+ most significant association P-value	-
14	rs163184	11	2,803,645	G	T	[0.470 - 0.633]	ggigigii	6.82×10^{-5}	8130	38987	1.087	1.043	1.132	*KCNQ1*	DIAGRAM+ most significant association P-value	0.064
	rs2237892	11	2,796,327	C	T	[0.928 - 0.952]	ggigigii	0.0027	8130	38987	1.141	1.047	1.244		genome wide significant association in East Asians only	
15	rs5215	11	17,365,206	C	T	[0.351 - 0.461]	ggggiggg	1.60×10^{-5}	8130	38987	1.093	1.050	1.139	*KCNJ11*	DIAGRAM+ most significant association P-value	-
16	rs4760790	12	69,921,061	A	G	[0.212 - 0.300]	iiiiiiii	3.56×10^{-6}	8130	38987	1.112	1.063	1.162	*TSPAN8/LGR5*	DIAGRAM+ most significant association P-value	0.909
	rs7961581	12	69,949,369	C	T	[0.200 - 0.292]	iigiiigg	1.82×10^{-5}	8130	38987	1.106	1.056	1.158		other previously reported lead SNP	
17	rs11642841	16	52,402,988	A	C	[0.391 - 0.445]	iiiiiiii	3.40×10^{-8}	8130	38987	1.128	1.080	1.176	*FTO*	DIAGRAM+ most significant association P-value	0.870
	rs9939609	16	52,378,028	A	T	[0.373 - 0.420]	iigiiigg	8.65×10^{-8}	8130	38987	1.116	1.072	1.162		other previously reported lead SNP	
18	rs4430796	17	33,172,153	G	A	[0.512 - 0.657]	ii0gii00	1.52×10^{-6}	4751	33536	1.139	1.080	1.201	*HNF1B (TCF2)*	DIAGRAM+ most significant association P-value	0.615
	rs757210	17	33,170,628	T	C	[0.269 - 0.382]	ii0iii00	1.60×10^{-4}	4751	33536	1.115	1.054	1.179		other previously reported lead SNP	
Loci taken forward to replication but recent reports establish as genome wide significant																
19	rs1387153	11	92,313,476	T	C	[0.254 - 0.316]	ggigigii	9.96×10^{-7}	8130	38987	1.122	1.071	1.175	*MTNR1B*	DIAGRAM+ most significant association P-value	0.705
	rs10830963	11	92,348,358	G	C	[0.275 - 0.338]	iiiiiiii	1.01×10^{-6}	8130	38987	1.129	1.075	1.185		other previously reported lead SNP	
20	rs7578326	2	226,728,897	A	G	[0.622 - 0.679]	iiigiiii	8.72×10^{-7}	8130	38987	1.115	1.068	1.165	*IRS1*	DIAGRAM+ most significant association P-value	0.740
	rs2943641	2	226,801,989	C	T	[0.615 - 0.657]	ggigigii	6.62×10^{-5}	8130	38987	1.087	1.043	1.133		other previously reported lead SNP	
Genome wide significant associations reported for the first time here																
21	rs243021	2	60,438,323	A	G	[0.442 - 0.488]	ggigigii	8.14×10^{-6}	8130	38987	1.094	1.052	1.139	*BCL11A*	DIAGRAM+ most significant association P-value	-
22	rs4457053	5	76,460,705	G	A	[0.193 - 0.292]	ii0iiiii	4.15×10^{-8}	7108	37912	1.162	1.102	1.226	*ZBED3*	DIAGRAM+ most significant association P-value	-
23	rs972283	7	130,117,394	G	A	[0.443 - 0.557]	iiiiiiii	1.82×10^{-6}	8130	38987	1.104	1.060	1.150	*KLF14*	DIAGRAM+ most significant association P-value	-
24	rs896854	8	96,029,687	T	C	[0.472 - 0.539]	ggigigii	1.23×10^{-6}	8130	38987	1.104	1.060	1.148	*TP53INP1*	DIAGRAM+ most significant association P-value	-
25	rs13292136	9	81,141,948	C	T	[0.859 - 0.944]	iiiiiiii	1.49×10^{-6}	8130	38987	1.200	1.114	1.292	*CHCHD9*	DIAGRAM+ most significant association P-value	-
26	rs231362	11	2,648,047	G	A	[0.441 - 0.491]	iiiiiiii	6.41×10^{-6}	8130	38987	1.110	1.060	1.161	*KCNQ1*	DIAGRAM+ most significant association P-value	-
27	rs1552224	11	72,110,746	A	C	[0.759 - 0.872]	ggigigii	6.99×10^{-6}	8130	38987	1.131	1.072	1.194	*CENTD2*	DIAGRAM+ most significant association P-value	-
28	rs1531343 [a]	12	64,461,161	C	G	[0.065 - 0.120]	iigiiigg	1.70×10^{-7}	8130	38987	1.201	1.121	1.285	*HMGA2*	DIAGRAM+ most significant association P-value	-
29	rs7957197	12	119,945,069	T	A	[0.728 - 0.832]	iiiiiiii	4.56×10^{-7}	8130	38987	1.136	1.082	1.195	*HNF1A*	DIAGRAM+ most significant association P-value	-
30	rs11634397	15	78,219,277	G	A	[0.607 - 0.679]	iiiiiiii	5.08×10^{-6}	8130	38987	1.107	1.060	1.157	*ZFAND6*	DIAGRAM+ most significant association P-value	-
31	rs8042680	15	89,322,341	A	C	[0.279 - 0.351]	ggigigii	8.19×10^{-6}	8130	38987	1.102	1.056	1.150	*PRC1*	DIAGRAM+ most significant association P-value	-
32	rs5945326	X	152,553,116	A	G	[0.723 - 0.801]	ggigggii	2.33×10^{-6}	8130	38987	1.246	1.137	1.365	*DUSP9*	Proxy for DIAGRAM+ most significant association (rs3020789)	-
Others taken to replication																
	rs6709268	2	3,058,146	A	C	[0.111 - 0.255]	iiiiiiii	8.61×10^{-6}	8130	38987	1.148	1.080	1.330	*TSSC1*	DIAGRAM+ most significant association P-value	-
	rs6442037	3	46,904,550	A	G	[0.641 - 0.689]	ggigigii	4.72×10^{-6}	8130	38987	1.110	1.061	1.160	*PTH1R*	DIAGRAM+ most significant association P-value	-
	rs7674212	4	104,208,348	G	T	[0.558 - 0.608]	ggigigii	1.70×10^{-7}	8130	38987	1.114	1.070	1.161	*NHEDC2*	DIAGRAM+ most significant association P-value	-
	rs1048709	6	32,022,914	G	A	[0.734 - 0.870]	iiiiiiii	5.96×10^{-6}	8130	38987	1.125	1.069	1.183	*CFB*	DIAGRAM+ most significant association P-value	-
	rs2789681 [h]	10	81,898,110	C	G	[0.068 - 0.120]	iiiiOiii	2.78×10^{-6}	7862	35277	1.176	1.099	1.257	*ANXA11*	DIAGRAM+ most significant association P-value	-
	rs960078	11	10,380,683	A	T	[0.976 - 0.988]	00g000gg	2.71×10^{-6}	3379	5451	1.846	1.429	2.385	*AMPD3*	DIAGRAM+ most significant association P-value	-
	rs7118472	11	86,833,159	A	G	[0.059 - 0.078]	g0igigii	6.15×10^{-6}	6665	15793	1.220	1.120	1.330	*TMEM135*	DIAGRAM+ most significant association P-value	-
	rs2288232	12	27,836,125	A	G	[0.488 - 0.618]	iiiiiiii	3.92×10^{-6}	8130	38987	1.112	1.063	1.163	*KLHDC5*	DIAGRAM+ most significant association P-value	-
	rs17795982	14	24,602,523	T	C	[0.847 - 0.886]	iiiiiiii	1.72×10^{-6}	8130	38987	1.155	1.088	1.224	*STXBP6*	DIAGRAM+ most significant association P-value	-
	rs8057749	16	77,015,972	C	G	[0.011 - 0.021]	0i0i00g0	9.36×10^{-6}	3076	29393	1.811	1.392	2.354	*WWOX*	DIAGRAM+ most significant association P-value	-

OR, odds ratio; CI, confidence interval

[a] Proxy SNP rs2612067 was genotyped in METSIM, FUSION2 and HUNT Stage 2 samples

[b] Alleles are indexed to the forward strand of NCBI Build 36.

[c] Denotes type of assay for each study contributing to the result, either (i) for imputed, (g) for genotyped, or (0) for missing. Order of the list (left to right): DGDG, deCODE, DGI, Rotterdam, EUROSPAN, FUSION, KORAgen, and WTCCC.

[d] All P-values reported are two-sided and based on an inverse-variance weighted meta-analysis model (fixed effects).

[e] Reduced sample numbers for *HNF1B, AMPD3* and *WWOX* reflect the fact that these SNPs (or proxies thereof) were not represented on all arrays used for stage 1 GWAS and/or were poorly imputed

[f] Gene region named for nearest gene (except where there is a very strong positional candidate).

[g] r^2 between SNPs is based on CEU HapMap data, release 22.

[h] formerly rs7088994

Supplementary Table 4: Results of conditional analyses: potential secondary hits with association P-value<1x10^{-4} within 1MB of index associations

Conditioning SNP			Secondary signal										
SNP	Chr	Nearby Gene[a]	SNP	Position B36 (basepair)	Risk Allele[b]	Nonrisk Allele[b]	Frequency Risk Allele [HapMap CEU]	OR (95% CI)	P-value[c]	Cases (N)	Controls (N)	Unadjusted P-value[c]	r^2 with Conditioning SNP	r^2 with other SNP(s)
rs896854	8	*TP53INP1*	**rs10106654**	96,973,161	C	A	0.167	1.11 (1.06-1.70)	6.47 x 10^{-5}	8,130	38,987	3.80 x 10^{-4}	0.009	1.000 (rs10111313)
			rs10111313	96,974,875	G	A	0.177	1.11 (1.06-1.17)	6.62 x 10^{-5}	8,130	38,987	3.37 x 10^{-4}	0.006	1.000 (rs10106654)
rs1096525	9	*CDKN2A/B*	rs2891169	22,121,825	G	A	0.42	1.14 (1.08-1.21)	8.17 x 10^{-6}	6,397	12,083	6.15 x 10^{-5}	0.156	0.753 (rs10217762), 0.753 (rs10757282), 0.753 (rs10757283)
			rs10217762	22,123,645	C	T	0.491	1.17 (1.10-1.25)	3.91 x 10^{-7}	6,397	12,083	2.72 x 10^{-4}	0.276	0.753 (rs2891169), 1.000 (rs10757282), 1.000 (rs10757283)
			rs10757282	22,123,984	C	T	0.456	1.17 (1.10-1.25)	4.00 x 10^{-7}	6,397	12,083	3.14 x 10^{-4}	0.276	0.753 (rs2891169), 1.000 (rs10217762), 1.000 (rs10757283)
			rs10757283	22,124,172	T	C	0.455	1.17 (1.10-1.24)	5.82 x 10^{-7}	6,397	12,083	3.39 x 10^{-4}	0.276	0.753 (rs2891169), 1.000 (rs10217762), 1.000 (rs10757282)
rs5015480	10	*HHEX/IDE*	**rs7907142**	94,575,536	T	C	0.424	1.09 (1.04-1.13)	7.76 x 10^{-5}	8,130	38,987	0.0013	0.000	1.000 (rs11187186), 1.000 (rs7916038), 0.006 (rs11596005)
			rs11187186	94,576,008	C	T	0.425	1.09 (1.04-1.13)	8.03 x 10^{-5}	8,130	38,987	0.0014	0.000	1.000 (rs7907142), 1.000 (rs7916038), 0.006 (rs11596005)
			rs7916038	94,583,387	A	G	0.429	1.09 (1.04-1.13)	8.62 x 10^{-5}	8,130	38,987	0.0011	0.000	1.000 (rs7907142), 1.000 (rs11187186), 0.006 (rs11596005)
			rs11596005	94,990,952	T	C	0.915	1.17 (1.08-1.27)	6.62 x 10^{-5}	8,130	38,987	0.012	0.000	0.006 (rs7907142), 0.006 (rs11187186), 0.006 (rs7916038)
rs231362	11	*KCNQ1*	**rs163184**	2,803,645	G	T	0.438	1.09 (1.05-1.14)	3.73 x 10^{-5}	8,130	38,987	6.82 x 10^{-5}	0.010	
rs5215	11	*KCNJ11*	rs12802656	16,490,991	C	A	0.473	1.09 (1.04-1.13)	4.94 x 10^{-5}	8,130	38,987	0.0013	0.014	0.967 (rs12799418)
			rs12799418	16,497,625	T	C	0.465	1.09 (1.04-1.13)	4.24 x 10^{-5}	8,130	38,987	0.0016	0.041	0.967 (rs12802656)

OR, odds ratio; CI, confidence interval

[a] Gene region named for nearest gene (except where there is a very strong positional candidate).

[b] Alleles are indexed to the forward strand of NCBI Build 36.

[c] All P-values reported are two-sided and based on an inverse-variance weighted meta-analysis model (fixed effects)

Supplementary Table 5: T2D loci with published associations with other traits mapping within 500kb of the lead T2D SNP

Nearby Gene[a]	T2D SNP	Other SNP	Other Trait	r^2	D'	Data set	Distance apart (kb)	Reference	DIAGRAM unadjusted P-value at "Other SNP"	DIAGRAM unadjusted P-value at "Other SNP" after conditioning on T2D SNP
THADA (ABCG5/ABCG8)	rs11899863	rs6544713	Plasma LDL	0.007	0.52	CEU	455	Kathiresan et al, NG, 2009	0.23	0.60
		rs6756629	Cholesterol	0.002	0.045	CEU	446	Aulchenko et al, NG, 2009	0.61	0.46
		rs11887534	Gallstones	0.002	0.045	CEU	447	Buch et al, NG, 2007	0.41	0.36
BCL11A	rs243021	rs11886868	Fetal Hemoglobin Level	0.021	0.25	CEU	135	Uda et al, PNAS, 2008	0.30	0.72
		rs1427407	F-cell Distribution	0.017	0.31	CEU	133	Menzel et al, NG, 2008	0.15	0.21
IRS1	rs7578326	rs2943634	Coronary disease	0.93	0.96	CEU	47	Samani et al. NEJM , 2007	2.7×10^{-5}	0.98
ZBED3	rs4457053	rs4704397	Thyroid function	0.034	0.31	CEU	93	(Arnaud-Lopez et al, AJHG, 2008)	0.096	0.053
CDKAL1	rs10440833	rs6908425	Crohn's	0.055	1.00	CEU	41	Barrrett et al, NG, 2008	4.5 x 10-4	0.57
		rs6908425	Psoriasis	0.055	1.00	CEU	41	Quaranta et al, Genes Immun., 2009		0.57
JAZF1	rs849134	rs10486567	Prostate Cancer	0.025	0.29	CEU	220	(Thomas et al, NG, 2008)	5.7×10^{-4}	0.49
		rs1635852	Stature	1.00	1.00	CEU	7	Johansson et al, HMG, 2009	3.1×10^{-9}	(0.29)[b]
		rs849141	Skeletal Frame Size	0.37	1.00	CEU	11	Soranzo et al, PLoS Genet, 2009	0.016	NA
KLF14	rs972283	rs157935	Basal cell carcinoma	0.019	0.19	CEU	119	(Stacey et al, NG, 2009)	0.94	(0.23)[b]
CDKN2A/B	rs10965250	rs1333049	CAD	0	0.016	CEU	8	Samani et al. NEJM , 2007	0.052	0.096
		rs10757274	CAD	NA	NA	CEU	37	McPherson et al, Science, 2007	NA	NA
		rs10757278	CAD	0	0.016	CEU	9	Helgadottir et al, Science, 2007	0.052	0.096
		rs10757278	Aortic Aneurysm	0	0.016	CEU	9	Helgadottir et al, NG, 2008	0.052	0.096
		rs7023329	Melanoma	0.003	0.12	CEU	317	Bishop et al, NG, 2009	0.077	0.69
		rs2151280	Basal cell carcinoma	0	0.018	CEU	99	(Stacey et al, NG, 2009)	0.0045	0.064
		rs1333040	Intracranial aneurysm	0.023	0.38	CEU	50	Bilguvar et al, NG, 2008	0.15	0.28
		rs4977574	Early onset MI	0	0	CEU	35	Kathiresan et al, NG, 2009	0.095	0.60
		rs4636294	Cutaneous nevi	0.003	0.091	CEU	39	Falchi et al, NG, 2009	1.00	0.55
		rs4977756	Glioma (Brain Cancer)	0.001	0.051	CEU	65	Shete et al, NG, 2009	1.3×10^{-4}	0.013
CHCHD9 (TLE4)	rs13292136	rs2378383	Asthma	0.16	0.57	Spanish/Mexican	87	Hancock et al, PLoS Genet, 2009	0.029	0.62
		rs2151145	Menopause (age at onset)	0.005	0.11	CEU	394	Stolk et al, NG, 2009	0.22	0.64
KCNQ1	rs231362	rs2074238	QT interval	0.006	0.34	CEU	207	Newton-Cheh et al, NG, 2009	NA	NA
		rs12576239	QT interval	0.007	0.21	CEU	189	Newton-Cheh et al, NG, 2009	NA	0.78
		rs12296050	QT interval	0.014	0.24	CEU	202	Pfeufer et al, NG, 2009	0.94	0.83
		rs757092	QT interval	0.012	0.15	CEU	192	(Pfeufer et al, Circulation Research, 2005)	0.20	0.23
		rs16928809	Serum bilirubin levels	0.003	0.16	CEU	245	Johnson et al, HMG, 2009	0.18	0.06
		rs7111341	T1D	0.002	0.068	CEU	478	Barrett et al, NG, 2009	0.59	0.57
KCNJ11	rs5215	rs11024074	Diastolic blood pressure	0.002	0.073	CEU	491	Levy et al. NG, 2009	0.91	0.83
HMGA2	rs1531343[c]	rs4026608	Aortic Root Size	0.002	0.15	CEU	220	Vasan et al, JAMA, 2009	0.13	0.50
		rs8756	Skeletal Frame Size	0.002	0.12	CEU	185	Soranzo et al, PLoS Genet, 2009	0.16	0.83
		rs1042725	Stature	0	0.016	CEU	183	Weedon et al, NG, 2007	0.13	0.87
HNF1A	rs7957197	rs1169310	CRP	0.11	1.00	CEU	21	Reiner et al, AJHG, 2008	0.019	0.84
		rs7310409	CRP	0.12	1.00	CEU	36	Ridker et al, AJHG, 2008	0.0049	0.21
		rs1169313	Liver enzymes	0.11	1.00	CEU	18	Yuan et al, AJHG, 2008	0.016	0.85
		rs1183910	CAD, CRP	0.073	1.00	CEU	40	Elliott et al, JAMA, 2009	0.0061	0.20
		rs2259816	CAD	0.11	1.00	CEU	25	Erdmann et al, NG, 2009	0.019	0.37
		rs2650000	Plasma LDL	0.05	0.82	CEU	72	Kathiresan et al, NG, 2009	0.0033	0.088
PRC1	rs8042680	rs2677744	ADHD	0.36	0.73	CEU	71	Lesch et al. J Neural Transm, 2008	0.0034	0.37
HNF1B	rs4430796	rs7501939	Prostate Cancer	0.766	1.00	CEU	3	Gudmundsson et al, NG, 2007	1.4×10^{-5}	0.92

[a] Gene region named for nearest gene (except where there is a very strong positional candidate).

[b] Based on available sample size below the 17,000 cut off used elsewhere

[c] Proxy SNP rs2612067 was genotyped in METSIM, FUSION2 and HUNT Stage 2 samples

Rows highlighted in grey are those for which the T2D and "other disease" SNPs show strong correlations (R2>0.5) and which are likely to reflect the same causal locus

Other disease hits which involve closely related phenotypes on the same causal pathway (eg diabetes and BMI for *FTO,* diabetes and fasting glucose for *MTNR1B* are not listed).

Those references in parentheses are not cited in the NHGRI catalog, and were therefore not included in simulations designed to evaluate the significance of co-localisation (see METHODS)

The full list of references for this table is provided in the Supplementary Note

Supplementary Table 6: Results of conditional analyses: genome wide signals at P-value<1x10^{5}

							Conditional Analysis			Unconditional Analysis	Lowest Unconditional P-value if another Proxy SNP[i]		
SNP	Chr	Position B36 (basepair)	Risk Allele[f]	Nonrisk Allele[f]	Frequency Risk Allele [Hapmap CEU]	Nearby Gene[g]	OR (95% CI)	P-value[h]	Cases (N)	Controls (N)	P-value[h]	SNP	Position B36 (basepair)	P-value in Unconditional Analysis[h]
rs2755263	1	67,301,831	G	A	0.58	*SLC35D1*	1.11 (1.06-1.16)	5.48×10^{-6}	8,130	39,897	7.58×10^{-4}		none	
rs6752494	2	200,021,568	C	T	0.17	*SATB2*	1.15 (1.08-1.22)	8.63×10^{-6}	8,130	39,897	1.35×10^{-5}		none	
rs11708067[a]	3	124,548,468	A	G	0.77	*ADCY5*	1.13 (1.07-1.19)	2.17×10^{-6}	8,130	39,897	1.69×10^{-4}		none	
rs6805662	3	133,942,594	A	C	0.20	*ACAD11*	1.14 (1.08-1.20)	5.08×10^{-6}	7,697	37,549	0.023	rs11712049	134,045,094	0.011
rs9825140	3	159,918,068	T	C	0.18	*RARRES1*	1.14 (1.07-1.20)	9.35×10^{-6}	8,130	39,897	0.0010		none	
rs1481279[b]	4	104,197,387	T	A	0.61	*NHEDC2*	1.13 (1.08-1.18)	8.39×10^{-9}	8,130	39,897	2.83×10^{-6}	rs7674212	104,208,348	1.70×10^{-7}
rs17541441	5	142,434,665	T	C	0.20	*ARHGAP26*	1.15 (1.09-1.23)	2.93×10^{-6}	6,665	15,793	0.061	rs6869553	142,458,137	9.35×10^{-4}
rs2248804	5	174,491,104	G	A	0.50	*DRD1*	1.10 (1.06-1.14)	5.17×10^{-6}	8,130	39,897	0.0032	rs4868491	174,533,465	8.22×10^{-4}
rs2207720[c]	6	7,964,196	T	C	0.50	*TXNDC5*	1.11 (1.06-1.15)	7.88×10^{-7}	8,130	39,897	0.0057	rs2743993	7,999,412	8.62×10^{-5}
rs2815137[c]	6	7,981,949	T	A	0.31		1.10 (1.06-1.15)	5.21×10^{-6}	8,130	39,897	4.27×10^{-4}			
rs1267481	6	14,667,059	C	A	0.75	*CD83*	1.12 (1.07-1.18)	5.60×10^{-6}	8,130	39,897	2.27×10^{-4}	rs1267486	14,672,187	2.19×10^{-4}
rs851971	6	152,019,166	C	T	0.70	*ESR1*	1.10 (1.06–1.15)	7.86×10^{-6}	8,130	39,897	0.0049	rs851998	152,025,031	0.004
rs3935165	7	55,682,899	G	A	0.30	*LANCL2*	1.12 (1.07-1.18)	1.06×10^{-6}	7,697	37,549	0.0027	rs11766085[e]	55,667,750	0.0017
rs10217762	9	22,123,645	C	T	0.49	*CDKN2B*	1.17 (1.10-1.25)	3.91×10^{-7}	6,397	12,083	2.71×10^{-4}		none	
rs7129859[d]	11	126,295,515	T	C	0.81	*PRR10*	1.12 (1.07-1.17)	3.22×10^{-6}	8,130	39,897	0.028	rs10790824	126,107,867	0.0031
rs669765[d]	11	126,363,371	T	C	0.80		1.13 (1.07-1.18)	5.92×10^{-6}	7,697	37,549	0.014			
rs12580303	12	44,468,076	A	G	0.22	*ARID2*	1.13 (1.07-1.18)	4.53×10^{-6}	8,130	39,897	0.0010		none	
rs7297141	12	73,319,748	T	A	0.12	*NA*	1.17 (1.09-1.26)	9.75×10^{-6}	8,130	39,897	0.0026	rs17181816	73,271,259	0.0018
rs9565546	13	79,608,955	A	T	0.73	*SPRY2*	1.11 (1.06-1.16)	2.99×10^{-6}	8,130	39,897	4.53×10^{-5}	rs1359790	79,615,157	2.67×10^{-5}
rs8020602	14	98,099,065	T	C	0.74	*C14orf177*	1.12 (1.07-1.17)	2.08×10^{-6}	8,130	39,897	0.0030	rs8005287	98,092,855	0.0029
rs275746	15	37,942,230	C	G	0.07	*GPR176*	1.25 (1.15-1.37)	3.72×10^{-7}	8,130	39,897	0.0014		none	

OR, odds ratio; CI, confidence interval

Minimum effective sample size for inclusion in these analysis is 15000

[a] Recently confirmed as novel T2D gene in Dupuis et al, Nature Genetics 2010

[b] Already followed-up in Stage 2 but did not replicate

[c] rs2207720 and rs2815137 have r^2=0.431

[d] rs7129859 and rs669765 have r^2=0.364

[e] rs11766085 has merged with rs9986837

[f] Alleles are indexed to the forward strand of NCBI Build 36.

[g] Gene region named for nearest gene.

[h] All P-values reported are two-sided and based on an inverse-variance weighted meta-analysis model (fixed effects)

[i] These columns indicate whether there is a nearby SNP that generated a stronger association P-value in the original unconditioned analysis than the lead regional SNP emerging from the conditional analysis