SUPPLEMENTARY METHODS

Stage 1 samples, genome-wide genotyping and quality control

UK: The WTCCC stage 1 sample consists of 1,924 T2D cases and 2,938 population controls from the UK[1,2]. These samples were genotyped on the Affymetrix GeneChip Human Mapping 500k Array Set. The call frequency of included samples was >0.97. 393,143 autosomal SNPs passed quality control (QC) criteria: Hardy-Weinberg equilibrium [HWE] $p>10^{-4}$ in T2D cases and controls, call frequency >0.95, minor allele frequency (MAF)>0.01, and good clustering, as defined in[1,2].

DGI: DGI: The DGI stage 1 Swedish and Finnish sample consists of 1,464 T2D cases and 1,467 normoglycemic controls. Of these, 2,097 are population-based T2D cases and controls matched for BMI, gender, and geographic origin, and 834 are T2D cases and controls in 326 sibships discordant for T2D[3]. These samples were genotyped on the Affymetrix GeneChip Human Mapping 500k Array Set. Genotype calls were made using the Bayesian Robust Linear Model with Mahalanobis Distance (BRLMM) algorithm, which is a development on a method proposed by Rabbee & Speed[4]. In brief, genotype calls are initially made using the Dynamic Model (DM) method[5]. Separately, data are quantile normalized and transformed onto a contrast-strength two dimensional plane. DM genotype estimates are used to construct a prior which is updated in a Bayesian fashion to fit the data for each SNP. The posteriors are used to determine genotypes by computing the Mahalanobis distances of each point to its nearest cluster and next-nearest cluster. The ratio of the two distances is the confidence score (where 0.5 represents the transition from call to no-call). We included all samples that had a genotype call rate >0.95. 378,860 autosomal SNPs passed QC criteria (call frequency >0.95, HWE p>10–6 in controls and MAF>0.01 in both population and familial components)[3].

FUSION: The FUSION stage 1 sample consists of 1,161 Finnish T2D cases and 1,174 Finnish normal glucose tolerant controls[6]. 789 T2D cases came from families with at least one reported T2D sibling and 372 were population-based. Controls were approximately frequency matched to cases based on 5-year age category, sex, and birth province; they included 304 spouses of FUSION subjects, 219 older individuals, and 651 individuals from a population-based study. In addition, 122 FUSION offspring with genotyped parents were included for quality control purposes and quantitative trait analysis. Samples were genotyped with the Illumina HumanHap300 BeadChip (version 1.1). All samples included had a call frequency >0.975. 306,222 autosomal SNPs passed QC[6] and had a HWE $p\geq10^{-6}$ in the total sample, ≤3 combined duplicate or non-Mendelian inheritance errors (out of 79 duplicate samples and 122 parent-offspring sets), call frequency ≥0.90, and MAF>0.01.

Analysis of stage 1 genotype data

In combining data across the three studies, we did not attempt, given differences in study design and implementation, to harmonize every aspect of individual study analysis and QC. For the UK, DGI and FUSION studies

respectively, 393,143, 378,860 and 306,222 SNPs were analyzed under an additive model.

UK: 393,143 autosomal SNPs were analyzed using the Cochran-Armitage test for trend (additive model) using the software package SNPTEST[7]. The genomic control (GC) value for directly-typed SNPs was estimated to be 1.08; after correcting for population structure, by including the two ancestry informative principal components (PC)[1] as covariates, the GC value was estimated to be 1.06 (Supplementary Note).

DGI: 378,860 autosomal SNPs were analyzed using (a) logistic regression (additive model) with covariates sex, age, BMI, and ascertainment locale in the 2097 population-based cases and controls (r^2= 0.95 to results from the previously described Cochran-Mantel-Haenszel stratified test[3] and (b) a modified Cochran-Mantel-Haenszel stratified test in the 326 discordant sibships (DFAM; both tests implemented in PLINK[3,8]). Results were combined by weighted meta-analysis of the two components of the study by way of 2-sided p values converted to z statistics. The GC value (directly-typed SNPs) for the population-based portion of the study was 1.06. We re-analyzed each of the 11 SNPs in Table 2 using logistic regression (additive model) with covariates as above in the population-based sample, but we then considered either the first 2 or 10 principal components as further covariates in the association analysis (Supplementary Note).

FUSION: 306,222 autosomal SNPs were analyzed using logistic regression with adjustment for sex, 5-year age category, and birth province and an additive model for the genetic effect. The GC value for directly-typed SNPs was 1.03. We estimated the top 10 ancestry PCs in the stage 1 data based on genotype similarity between each pair of individuals using 10% of the GWA SNP data[9]. We reanalyzed each of the SNPs in Table 2, with the inclusion of these 10 PCs in addition to our standard covariates of age, sex and birthplace (Supplementary Note).

Stage 1 imputation and T2D analysis

For each stage 1 sample set, we imputed genotypes for autosomal SNPs that were present in HapMap Phase II but were not present in the genome-wide chip or did not pass direct genotyping QC. In each sample, genotypes were imputed using the genotype data from the GWA chips and phased HapMap II genotype data from the 60 CEU HapMap founders. We retained SNPs that had an estimated MAF>0.01 in the control samples. Imputed SNPs were then tested for T2D association.

UK: Genotypes were imputed using the program IMPUTE[7], which determines the probability distribution of missing genotypes conditional on a set of known haplotypes and an estimated fine-scale recombination map. Imputation was based on 387,667 stage 1 autosomal SNPs with MAF>0.01 that appear on the HapMap (and excluding SNPs that demonstrated poor genotype clustering upon manual inspection). 1,966,876 imputed SNPs, excluding the directly typed SNPs passing QC, had MAF>0.01 in controls. We analyzed imputed SNPs using SNPTEST, which implements frequentist tests that calculate p values,

parameter estimates and their standard errors that properly account for the uncertainty due to the probability distributions of the imputed genotypes[7]. SNPTEST also calculates an information measure for the test at each SNP that is a measure of the relative statistical information about the additive genetic effect being estimated. This measure has a direct relationship to the power of the test. For example, a value of 0.8 indicates that the imputed data at the SNP is equivalent to a dataset 80% of the full sample size with precisely known genotypes. For T2D analysis, we included 1,915,393 imputed autosomal SNPs with an imputation information score ≥0.5. The GC value for imputed SNPs was 1.08.

DGI: Genotypes were imputed using the program MACH 1 (Y.L., C.J.W., J.D, P.S., G.R.A. Markov Model for Rapid Haplotyping and Genotype Imputation in Genome Wide Studies. Submitted, 2007; http://www.sph.umich.edu/csg/abecasis/MaCH/download/), which determines the probability distribution of missing genotypes conditional on a set of known haplotypes while simultaneously estimating the fine-scale recombination map. Imputation was based on 371,186 stage 1 autosomal SNPs with MAF>0.01, combined with phased CEU chromosomes from the HapMap. 1,916,622 imputed SNPs had an estimated MAF>0.01 in controls. SNPTEST (described above) with covariates sex, age, BMI and clinical center was used to analyse imputed genotypes for the population-based component. Best guess genotypes of imputed SNPs were analyzed using DFAM in the sibship component of the study, and results from both sub-studies were combined by meta-analysis. 1,853,222 SNPs with SNPTEST information scores >0.5 were included in further analysis. The GC value for imputed SNPs was 1.07.

FUSION: Genotypes were imputed using MACH 1 (Y.L., C.J.W., J.D, P.S., G.R.A. Markov Model for Rapid Haplotyping and Genotype Imputation in Genome Wide Studies. Submitted, 2007; http://www.sph.umich.edu/csg/abecasis/MaCH/download/) as described above. Imputation was based on 303,174 autosomal SNPs with MAF>0.01, combined with phased CEU chromosomes from the HapMap. 2,110,199 imputed SNPs had an estimated MAF>0.01 in the sample and predicted r^2>0.3 (imputation accuracy measure) between the true and imputed genotypes. To test for T2D-SNP association with the imputed data we used logistic regression (additive model), in which SNPs were represented by the expected allele count, an approach that takes into account the degree of uncertainty of genotype imputation[6] (Y.L., C.J.W., J.D, P.S., G.R.A. Markov Model for Rapid Haplotyping and Genotype Imputation in Genome Wide Studies. Submitted, 2007; http://www.sph.umich.edu/csg/abecasis/MaCH/download/). The GC value for imputed SNPs was 1.04.

Stage 1 meta-analysis

We used meta-analysis to combine the T2D association results for the stage 1 WTCCC, DGI and FUSION samples. The combined stage 1 data are comprised of 10,128 samples: 4,549 T2D cases and 5,579 controls. We used association results from directly genotyped SNPs, where available, and imputed genotype

association results at all other positions. 2,168,847 genotyped and imputed autosomal SNPs passed QC and had MAF>0.01 in each of the three samples (44,750 were genotyped in all three samples, 308,628 were genotyped in two samples, 245,158 were genotyped in one sample, 1,570,311 were imputed in all samples). All association results were expressed relative to the forward strand of the reference genome based on dbSNP125. For our initial analysis, which was used to select signals for stage 2 genotyping, for each SNP we combined the ORs for a given reference allele weighted by the confidence intervals using a fixed effects model. We investigated evidence for heterogeneity of ORs using two commonly used statistics: Cochrans's Q statistic and $I^{2\,(10)}$.
The OR estimate for the DGI scan did not incorporate the discordant sibship component (as the methodology to do this had not been developed); thus, information from this portion of the DGI study was not incorporated into the initial meta-analysis. To subsequently add the contribution of association information from the discordant sibship data in the final reported association, we repeated the meta-analysis combining evidence for association based solely on the p value. Specifically, for each study we converted the two-sided p value to a z-statistic which was signed to reflect the direction of the association given the reference allele. Each z-score was then weighted; the squared weights were chosen to sum to 1 and each sample-specific weight was proportional to the square root of the effective number of individuals in the sample. To determine the effective number of individuals, we first calculated the power of the study based on the observed ratio of cases to controls and on the familial relationships using parameters that gave approximately 50% power. Using the same parameters and power, we calculated the sample size for a study with equal numbers of cases and controls. We used the resulting case/control sample size as the effective sample size for the study. Weighted z-statistics were summed across studies and the summary z-score converted to a two-sided p value.

SNP prioritisation for stage 2 genotyping
We prioritized 69 SNPs for replication in stage 2 based on the results from the three-study stage 1 meta-analysis, using a set of criteria we developed as part of a heuristic approach to the prioritization of loci for follow-up. Briefly, we considered SNPs with a meta-analysis p value $<10^{-4}$ and a meta-analysis heterogeneity p value $>10^{-4}$. These selections were largely made using the initial OR-based version of the meta-analysis, in which the best-guess genotype and summed allele counts (the latter taking into account genotype probability distributions) were used to test for association with T2D in DGI and WTCCC stage 1 samples respectively. All data presented throughout the manuscript report results of the final meta-analysis, so the p values are not always congruent with those used for initial prioritization of signals for replication.
We allowed some exceptions to the above follow-up criteria. First, SNPs that had $p<0.05$ in the same direction across all three stage 1 studies were included. Second, we included any non-synonymous SNPs with meta-analysis p values between 10^{-3} and 10^{-4} that adhered to the above criteria and represented a

signal for which a SNP with $p<10^{-4}$ had not already been selected. When a signal consisted of a cluster of similarly-associated SNPs, we selected the SNP that had been directly genotyped in at least one study. For SNPs imputed across all three studies, we selected the SNP with the highest imputation quality scores. Five additional SNPs were selected for replication genotyping on the basis of their strong association with T2D in the DGI GWA study (2 SNPs), association with T2D and with insulinogenic index in the DGI study (1 SNP), and overlap with FUSION or WTCCC ($p<0.05$ in DGI and one or both studies; 2 SNPs). DGI stage 2 results for three of these SNPs (rs10923931, rs6698181 and rs17044137) have been reported and were encouraging enough to warrant examination in the UK and FUSION stage 2 samples[3]. For 6 signals, we already had no evidence for association ($p>0.30$) in at least one of the three studies for SNPs in high LD with the strongest signal[2,3,6]. For known T2D loci (*TCF7L2*, *CDKAL1*, *IGF2BP2*, *KCNJ11*, *HHEX/IDE*, *SLC30A8*, *CDKN2A/2B* region, *WFS1*, *TCF2*, and *FTO*) we excluded from follow-up all SNPs that resided within the surrounding region, with region boundaries defined by the furthest neighboring SNPs with p values remaining ~0.01 (n=1,981). For the *PPARG* region, we identified a SNP, rs17036101, with a p value two orders of magnitude lower than the established Pro12Ala susceptibility variant, rs1801282, and took this signal forward to replication. A total of 69 SNPs were taken forward to stage 2 genotyping.
Of the 69 signals selected for follow-up, a total of 65 were successfully genotyped in stage 2, and represented loci that were independent of each other and of previously established susceptibility loci. There were a total of 71 independent loci (including the 6 loci not followed-up based on prior evidence of lack of stage 2 association). Nine of these 71 loci had a p value≤0.01 with association in the same direction as the original signal, far in excess of 0.36 expected under the null ($p=1.1x10^{-10}$, binomial test), and two SNPs had $p<10^{-4}$ as compared to an expectation of 0.0036 ($p=6.2x10^{-6}$, binomial test) (Supplementary Table 5).

Stage 2 samples, genotyping and analysis

UK: We genotyped the prioritized SNPs in cases and controls from three UK replication sets (RS1, RS2 and RS3, described in[2]; Supplementary Table 1). RS1 comprised 2,022 cases and 2,037 population-based controls from the UK Type 2 Diabetes Genetics Consortium collection (UKT2DGC) (all from Tayside, Scotland). RS2 included 632 additional T2D cases and 1,750 population controls from the Exeter Family Study of Child Health (EFSOCH). RS3 comprised a further 1,103 cases and 1,559 population-based controls from the UKT2DGC (Supplementary Table 1). Genotyping of prioritized SNPs in RS1, RS2 and RS3 was performed by Kbiosciences (Herts., UK). Kbiosciences designed and used assays based on either their proprietary competitive allele specific PCR system (KASPar) method or modified TaqMan assays, details of which are available on their website (http://www.kbioscience.co.uk/chemistry/index.htm). All assays were validated prior to use, using a standard 96-well validation plate used by Kbiosciences and up to 296 samples from the WTCCC study (see *Comparison of genotypes from imputation and direct genotyping*, below). SNPs rs11178531,

rs4493865, rs7961581, rs10817674, rs1153188, rs10516948, rs864745, and rs9472138 had already been genotyped in RS1, RS2 and/or RS3 (up to 3,757 cases and 5,346 controls[2]). The remaining signals prioritized for follow-up were genotyped in a slightly different subset of the replication samples (n=9,017), due to DNA availability (RS1: 1,962 cases, 1,962 controls; RS2: 631 cases, 1,829 controls; RS3: 1,094 cases, 1,539 controls). Concordance rates between the Affymetrix and KASPar/TaqMan genotypes (based on up to 296 replicate stage 1 samples) were 97.5% on average. SNPs rs2391592, rs7812465 and rs7333961 were not genotyped in the UK stage 2 samples because of assay failure. All other SNPs had genotype call frequency rates >94% in the replication sets and no SNPs had HWE p value<0.001 in cases or controls. We tested for association with T2D using the Cochran-Armitage test for trend. Results from the 3 replication sets were combined in a Cochran-Mantel-Haenszel meta-analysis framework.

DGI: We genotyped the prioritized SNPs in three stage 2 case-control samples[3] (Supplementary Table 1). The Sweden case control sample consisted of 2,830 cases from the Malmo Diabetes Registry and 3,550 normoglycemic controls from the Malmo Diet and Cancer study. Cases had age of onset >35 years, C-peptide >0.3 nmol/L and were GAD Ab negative. The US case-control sample comprised 1,226 cases of European ancestry from the United States individually matched to 1,226 control subjects by age, sex, and grandparental country of origin. The Poland case-control sample consisted of 1,009 diabetic cases and 1,009 control subjects, matched individually by age and sex. The US and Polish samples were obtained from Genomics Collaborative Inc and have no demonstrable population structure[11]. The prioritized SNPs were genotyped in all DGI stage 1 and 2 samples using the iPLEX Sequenom MassARRAY platform (http://www.sequenom.com/Assets/pdfs/appnotes/8876-006.pdf). 63 SNPs passing QC (>94% call rate, MAF>0.01 and HWE p value >0.001) were used for association testing. We tested for T2D association in each DGI stage 2 case-control set using a chi-squared analysis (assuming an additive genetic model). Results from the three DGI stage 2 samples were combined using Cochran-Mantel-Haenszel meta-analysis.

FUSION: We genotyped the prioritized SNPs in a Finnish sample of 1,215 T2D cases and 1,258 NGT controls (Supplementary Table 1) selected from the Dehko 2D, Health 2000, Finrisk 1987, Finrisk 2002, Savitaipale Diabetes, and Action LADA studies[6]. Genotyping was performed using the Sequenom Homogeneous Mass EXTEND or iPLEX Gold SBE assays and was carried out at the National Human Genome Research Institute (NHGRI). 59 SNPs had genotype call frequency >94% and HWE p value >0.001. The genotype consistency rate among 56 duplicate samples was 100% and the average call frequency of successfully genotyped SNPs was 97.3%. SNPs were analyzed using logistic regression with adjustment for sex, 5-year age category and birth province and an additive model for the genetic effect.

Comparison of genotypes from imputation and direct genotyping

A proportion of the prioritized imputed signals was genotyped in the stage 1 samples of the three studies and respective concordance rates were calculated (Supplementary Table 4). All results presented in the main manuscript text are based on directly-typed stage 1 data, expect rs7961581 in FUSION stage 1.
UK: As mentioned above (*Stage 2 samples, genotyping and analysis* section), SNPs prioritized for replication were re-genotyped in up to 296 of the WTCCC samples (at Kbiosciences, see above). Overall, there was a 97.5% concordance rate (range: 81.7% to 100%) between the directly assayed and best-guess imputed genotypes. We genotyped five of the imputed prioritized signals (rs1481279, rs7578597, rs17036101, rs319598, and rs12779790) in the full complement of 4,862 UK stage 1 samples. We compared direct genotypes to the best-guess imputed genotype at each SNP and observed high levels of genotype concordance (average 97.9%, range: 96.1%-99.6%). These rates are comparable with those obtained in a larger study of imputation data quality (98.6%), comparing the imputed genotypes for 10,180 SNPs with those obtained by directly genotyping 1,444 individuals (using an Illumina custom chip)[7].
DGI: We genotyped 58 SNPs from the prioritized signals in the 2,931 stage 1 DGI samples using the iPLEX Sequenom MassARRAY platform (http://www.sequenom.com/Assets/pdfs/appnotes/8876-006.pdf). 38 of 58 SNPs had not been directly genotyped on the Affymetrix platform or passed QC in the DGI scan. We compared the concordance rates between the best-guess imputed genotypes and the Sequenom genotypes for the 38 SNPs in 2,891 samples which passed quality control (in the replication genotyping stage), and found them to be consistent with the imputed data [average genotype concordance rate: 97.9%, range: 92.2–99.9%]. We found strong agreement in statistical support obtained using imputed SNP genotypes or using the Sequenom genotype data (r^2=0.84, p=7.6×10^{-12}).
FUSION: We genotyped seven of the imputed prioritized SNPs (rs864745, rs12779790, rs4607103, rs17036101, rs1153188, rs10490072, and rs10923931) in the FUSION stage 1 sample. We found a genotype concordance rate of 96.5% (range 86.2%-100%) between the best-guess imputed genotype and the genotyped SNP. In a larger-scale assessment of imputation data quality, we had previously compared the imputed genotypes for 510 SNPs with those obtained by directly genotyping 1,190 individuals using Illumina Golden Gate genotyping, as part of a different project, and observed a genotype concordance rate of 97.1% (Y.L., C.J.W., J.D, P.S., G.R.A. Markov Model for Rapid Haplotyping and Genotype Imputation in Genome Wide Studies. Submitted, 2007; http://www.sph.umich.edu/csg/abecasis/MaCH/download/).

Combined meta-analysis for stages 1 and 2
We combined stage 1 and stage 2 data using both the OR-based and the weighted z score-based meta-analysis approaches described above (*Stage 1 meta-analysis*). We also assessed our results using random effects meta-analysis to better account for any heterogeneity between the studies (Supplementary Table 6). Locus-specific and combined sibling relative risk estimates were calculated using sample size-weighted estimates of the effect size and risk-

allele frequency derived from stage 2 replication samples only, and under the assumption of allelic and locus independence, as described by[12,13].

Stage 3 sample, genotyping and association analysis

Eleven SNPs (rs2641348, rs10490072, rs7578597, rs17036101, rs4607103, rs9472138, rs864745, rs12779790, rs1153188, rs10923931, and rs7961581) were followed up in the stage 3 samples, from the from the deCODE, KORA, Danish, HUNT, NHS, GEM Consortium (CCC, EPIC, ADDITION/Ely, Norfolk) and METSIM studies (Supplementary Table 1).

deCODE study: For the deCODE stage 3 study, we used genotype data from the Icelandic GWA scan[14] for rs2641348, rs7578597 and rs9472138, and a perfect proxy (rs2793831, based on HapMap) for rs10923931. The GWA sample consisted of 1,520 Icelandic individuals with T2D and 25,235 controls[14,15] (Supplementary Table 1). The remaining SNPs had not been directly typed as part of this scan and were therefore genotyped separately, in a subset of the GWA scan samples: 1,422 of the 1,520 Icelandic T2D cases and 3,455 of the 23,235 Icelandic controls were genotyped using the Centaurus (Nanogen) platform[16]. The quality of each Centaurus assay was evaluated by genotyping each assay in the CEPH (Utah residents with ancestry from northern and western Europe) (CEU) and Yoruba in Ibadan, Nigeria (YRI) HapMap samples and comparing the results with the HapMap data. All SNPs had HWE p values >0.001 and call rates >94%. For association analysis we utilized a standard likelihood ratio statistic, implemented in the NEMO software[17] to calculate two-sided p values and odds ratio (OR) for each individual allele. We estimated the genomic control inflation factor for the extended and limited case-control analysis to be 1.347 and 1.287 respectively (inflated statistics reflect relatedness in the Icelandic sample). All deCODE stage 3 results are presented adjusted for GC.

KORA study: This study is based on data from participants of four independent cross-sectional surveys (S1–S4) of the KORA (Cooperative Health Research in the Region of Augsburg) project between 1984 and 2001[18], as well as from participants from KORA T2DM Family Study (T2DMFAM[19]), which was performed in 2001 / 2002. All probands were from the city or region of Augsburg. All participants were living in Germany and all were of European origin. For the analysis we used a case-control design including 630 T2D cases from all four KORA surveys, 611 cases from KORA T2DMFAM and 1,458 non-diabetic age- and sex-matched controls from KORA survey S4 (Supplementary Table 1). Genotyping was carried out using the Sequenom iPlex Gold SBE technique. Six of eleven attempted SNPs had genotype call frequency >94% and HWE p value>0.001, and were included in further analysis. The genotype consistency rate among 137 duplicate samples was 99.5% and the average call frequency of successfully genotyped SNPs was 96.6%. We tested for T2D association using logistic regression (assuming an additive model) adjusting for age and sex.

Danish study: The 11 prioritized polymorphisms were also genotyped in 9,132 Danes comprising the population-based Inter99 sample of middle-aged people sampled at Research Centre for Prevention and Health[20], T2D patients sampled through the out-patient clinic at Steno Diabetes Center, a population-based

group of middle-aged glucose-tolerant subjects recruited from Steno Diabetes Center, and the Danish ADDITION study group sampled through the Department of General Practice at University of Aarhus[21]. Detailed characteristics of study populations have been described elsewhere[22]. In total, 4,089 T2D patients and 5,043 normal glucose-tolerant control subjects (Supplementary Table 1) were genotyped using Taqman allelic discrimination (KBioscience, Herts, UK). Discordance between 1,090 random duplicate samples was <1% and the genotyping success rate was >94% for all 11 SNPs. All genotype groups obeyed Hardy-Weinberg equilibrium (all p>0.001). Logistic regression with adjustment for age, sex and BMI was performed using RGui version 2.5.0.

HUNT study: We genotyped eleven prioritized SNPs in a Norwegian sample from Nord Trøndelag county of 1,004 population-based T2D cases and 1,503 non-diabetic controls selected from the Nord-Trøndelag Health (HUNT) 2 study[23,24] (Supplementary Table 1). The individuals with diabetes were identified by their response to: "Do you have, or have you had diabetes - Yes/No". A validation study in HUNT 1 showed high reliability of this question compared to general practitioner medical records[25]. From diabetic individuals, T2D cases were identified by anti-GAD level <0.08 units and a report of treatment for diabetes of diet only, oral anti-diabetic medication, or insulin treatment started at least 12 months after the onset of diabetes. The controls were individuals that responded "No" to having diabetes. They were frequency-matched by 10-year age category to cases. Genotyping was performed using the TaqMan allelic discrimination assay (Applied Biosystems) at the University of North Carolina. All 11 attempted SNPs had genotype call frequency >94% and HWE p value >0.001. The genotype consistency rate among 72 duplicate samples was 100% and the average call frequency was 98.7%. We analyzed the data using an additive logistic regression model with adjustment for sex.

NHS study: The Nurses' Health Study (NHS) was established in 1976 when 121,700 female registered nurses aged 30–55 years and residing in 11 large U.S. states completed a mailed questionnaire on their medical history and lifestyle. The lifestyle factors, including smoking, menopausal status and postmenopausal hormone therapy, and body weight, have been updated by validated questionnaires every 2 years. 32,826 women provided blood samples between 1989 and 1990. Subjects for the present study were selected from those who were of European ancestry and provided blood samples. Diabetes cases were defined as self-reported diabetes confirmed by a validated supplementary questionnaire based on the National Diabetes Data Group (NDDG) and American Diabetes Association (ADA) criteria[26]. This study included 1,506 cases of type 2 diabetes and 2,014 non-diabetic controls. DNA was extracted from the buffy coat fraction of centrifuged blood using the QIAmp Blood Kit (Qiagen, Chatsworth, CA). SNP genotyping was performed at the Dana Farber/Harvard Cancer Center High Throughput Genotyping Core. All samples were genotyped using the Biotrove OpenArray SNP Genotyping Platform. Replicate quality control samples (5%) were included and genotyped with >99% concordance. Genotype frequencies for all SNPs in the controls were found to be in Hardy-Weinberg equilibrium (p>0.05). Associations between SNPs and diabetes risk

were tested with logistic regression under an additive model using SAS (Version 9.1 for UNIX; SAS Institute Inc., Cary, NC, USA).

GEM Consortium: The GEM Consortium contributed 4 case-control studies to stage 3 (CCC, EPIC, ADDITION/Ely and Norfolk).

CCC study: The Cambridgeshire case–control (CCC) study is a population-based study of T2D cases, aged 45-76 years, and age- and sex-matched controls. Cases were randomly selected from general practitioner diabetes registers in Cambridgeshire, UK, and T2D was defined as onset of diabetes after the age of 30 years and without insulin use in the first year after diagnosis[27]. Controls were recruited at random from the same population sampling frames, and individually matched to cases for age, sex and GP practice. Diabetes was excluded in controls by medical record search and by a glycated haemoglobin measurement of less than 6%. The study received ethical approval from the Cambridge Local Research Ethics Committee, and participants provided informed consent. In the current analyses, we include 547 cases and 533 controls, representing all white Europeans who had DNA and information on body mass index available.

EPIC study: The EPIC case-control study is nested within the EPIC Norfolk Study, a population-based cohort of men and women of European descent aged 40-78 years[28]. Incident cases of T2D were defined by a physician's diagnosis of T2D, with no insulin prescribed within the first year following diagnosis and/ or HbA1c >7%[(29)]. Two sets of controls, each matched in terms of age, sex, general practice, recruitment date, with one set additionally matched for BMI, were randomly selected from the EPIC-Norfolk cohort from among those without diabetes, cancer, stroke, or myocardial infarction at baseline and who had not developed diabetes during follow-up. Potential controls with measured HbA1c levels > 6% were excluded. The current analyses included 388 T2D cases and 774 matched controls. The EPIC-Norfolk study was approved by the Norfolk Local Research Ethics Committee.

ADDITION/Ely study: Previously undiagnosed prevalent cases of T2D, defined using WHO OGTT criteria, were identified via a population-based stepwise screening strategy among 40 to 69 year olds participating in the UK Cambridge arm of the ADDITION study. Current analyses include 800 white European men and women who had DNA available and information on body mass index[30]. Controls were identified from the MRC Ely study, a population-based cohort of white European men and women aged 35 to 79 years without diagnosed diabetes and from a similar sampling frame as the cases[31]. Based on WHO OGTT criteria, participants were confirmed as controls (n=1,610) or classified as cases (n=92), resulting in a total of 892 cases for the case-control comparison. The Cambridge Research Ethics Committee approved both studies.

Norfolk study: The Norfolk Diabetes Case-Control Study is an ongoing study of white European 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; a total of 2,311 white European cases were included in the current analyses. Participants with insulin use during the first year of

diagnosis, and those with cystic fibrosis, chronic pancreatitis or long term steroid use were excluded from the study. 2,400 controls free of known diabetes at baseline or during follow-up were randomly selected from EPIC-Norfolk participants. The Norfolk study was approved by the Norwich Local Research Ethics Committee.
Genotyping for the GEM Consortium studies was performed at the Wellcome Trust Sanger Institute using the Sequenom iPLEX Gold SBE assays for all SNPs except for rs10923931, which was genotyped using a custom TaqMan® SNP assay (Applied Biosystems, UK). Fill-in genotyping was performed on SNPs rs12779790, rs10490072, rs4607103, rs2641348, rs17036101 and rs7578597 to achieve >94% pass rates in each study. After fill-in genotyping, all 11 SNPs had genotype call frequency >94% and HWE p value>0.001 (except for SNP rs10923931 which deviated from HWE in the EPIC case-control population and thus was not included in the analysis). The experiment-wise genotype consistency rate among 6,853 duplicate samples across all SNPs was 99.3% and the average call frequency of successfully genotyped SNPs was 96.2%. Association between each SNP and T2D was tested using logistic regression, assuming an additive genetic model and adjusting for age, sex and BMI.
METSIM study: The METabolic Syndrome In Men (METSIM) study aims to investigate the metabolic syndrome, type 2 diabetes, cardiovascular disease, and cardiovascular risk factors, with the goal of collecting 7,000 men, randomly selected from the population of the town of Kuopio in Eastern Finland (population 95,000). Our sample included the first 659 T2D cases and 2,639 NGT controls, aged 50-72 years, and defined using the same WHO 1999 criteria as used for FUSION. Genotyping of the 11 SNPs was performed using iPLEX Gold SBE assays at the National Human Genome Research Institute (NHGRI). All markers have a HWE p value >0.001 and success rates >94% with an average success rate of 97.5%. The concordance rate was 99.95% for >4,000 duplicate samples for each of two SNPs genotyped as part of another study. We performed logistic regression adjusted for 5-year age category using an additive genetic model.

Combined meta-analysis for stages 1, 2, and 3
We combined stage 1, 2 and 3 data using both meta-analysis approaches (fixed-effects model to combine ORs and weighted p value-based z-statistic combination across all sample sets) described above (*Stage 1 meta-analysis*). All p values reported in the main text are based on the p value-based z-statistic meta-analysis. We also assessed our results using random effects meta-analysis (Supplementary Table 6). We observed some evidence for heterogeneity across studies (the I^2 statistic ranged from 0 to 57.8% depending on SNP), with rs7578597 and rs10923931 displaying the largest fold differences in association p value between the fixed- and random-effects model analyses. Differences in strength of association across studies (leading to evidence for heterogeneity) could reflect interesting biological associations that vary from study to study depending on subject ascertainment scheme.

Sample size calculations for Table 2

In Table 2, we present the sample size (sum of case and control samples) required for 80% power (to achieve nominal replication at p=0.05), calculated based on the stage 2 OR estimate, sample size-weighted risk allele frequency across the stage 2 studies and assuming an equal number of cases and controls. The p values for the z statistic-based meta-analysis method do not correspond exactly to those for the OR-based meta-analysis. This accounts for the discrepancy of the ranking of the stage 2 meta-analysis p values and the sample sizes.

Genomic control

Where the effect sizes to be detected are, as here, modest, it is particularly important to rule out sources of error and bias that could generate association signals of similar magnitude. Where GWA data are available, direct evaluation of the evidence for population stratification and cryptic relatedness is possible, and correction can be made, for example, using principal component and genomic control methods, although these can be quite conservative. For those follow-up samples where GWA data are not available, equivalent procedures are not possible and one is obliged to rely on knowledge of sample ascertainment procedures and prior focused association studies as guarantors of performance. We have deliberately adopted two strategies in reporting the findings from this study.

In the first, we performed GC-correction of data from DGI, FUSION and WTCCC prior to stage 1 meta-analysis. We corrected each individual study for the GC inflation observed (directly genotyped and imputed data separately), and combined results across studies. We present the genome-wide distribution of association statistics in Supplementary Figure 1. We note that, after study-specific genomic control adjustment, the estimated inflation factor for the stage-1 meta-analysis test statistic was 1.04.

In the second, we combined GC-uncorrected data from DGI, FUSION and WTCCC for stage 1 meta-analysis and did not correct the meta-analysis test statistics for the overall GC (to guard against over-conservativeness in the estimate of strength of association for interesting signals). We also present the genome-wide distribution of these statistics in Supplementary Figure 1.

For the combination of data across stages 1, 2 and 3, we also adopted these two strategies (of using GC-corrected and GC-uncorrected stage 1 data). In the first, we performed individual GC-correction of DGI, FUSION and WTCCC stage 1 data prior to meta-analysis with stage 2 and stage 3 data (an approach which may be over-conservative where, as here, none of the T2D-associated SNPs has particular hallmarks of stratification) (Supplementary Note). In the second, we combined only uncorrected data (except for the deCODE data, where we have applied GC correction given a more marked genomic control inflation [GC ~1.3] in that sample). We present the resulting data from both approaches (of using GC-corrected and GC-uncorrected stage 1 data for stage 1-3 meta-analysis) in Supplementary Table 6 and a comparison of results (showing very small

differences) in the Supplementary Note. All data presented elsewhere in the manuscript reflect the GC-corrected analysis strategy outcome.

Conditional analysis of T2D signals

For each SNP in Table 2, we assessed association using an additive genetic model in the stage 1 and 2 samples before and after including body mass index in the logistic regression model. For each genotyped and imputed SNP surrounding a specific T2D signal we assessed association using an additive genetic model in the stage 1 sample before and after including the Table 2 SNP from the same region in the model. We analyzed the data and adjusted for covariates for the stage 1 and stage 2 samples of each study. Data were combined across studies and stages as described above. The ORs and CIs were calculated using a fixed-effects model and p values were calculated using the weighted z-score method. For the UK stage 1 samples, we did not have BMI information available for ~1,500 of the population-based controls. We therefore carried out the conditional BMI analyses by using all T2D cases and only those controls for whom BMI data were available.

Quantitative trait analyses

Quantitative trait analyses were carried out in the UK, DGI and FUSION samples for the 11 SNPs taken forward to stage 3. We tested BMI, quantitative glycemic traits (fasting and 2 hour glucose and insulin, HOMA-IR [fasting glucose (mM) x fasting insulin (uUnits/ml)/22.5]), lipid traits (total, HDL and LDL cholesterol, and serum triglycerides) and blood pressure (systolic and diastolic), where available, for association using an additive genetic model. Data were combined across studies and stages using the weighted z score method for meta-analysis described above.

UK: BMI was transformed to normality (logBMI). All SNPs were analysed within T2D case and control groups separately (where BMI data were available, see above), by linear regression, adjusting for age and sex.

DGI: Glucose and insulin traits were analyzed in NGT individuals, whereas lipid and blood pressure traits were analyzed in case and control subjects. For lipid traits, we excluded subjects on lipid lowering medications, and for blood pressure traits, we excluded individuals over 60 years of age (for DBP) and with severe heart failure. Raw averaged SBP and DBP values were adjusted for treatment effect by imputation as previously described[32]. BMI, insulin traits, triglycerides and HOMA-IR were normalized by log transformation. Quantitative trait analysis was performed separately in cases and controls by linear regression, adjusting for covariates (age, sex, and logBMI for glycemic traits; age, age^2, sex, and diabetes status for lipid traits; age, sex, and logBMI for blood pressure traits; and, in the stage 1 sample, clinical site visited).

FUSION: Except where noted, each trait was analyzed in stage 1 and stage 2 T2D and NGT individuals. Insulin and HOMA were analyzed in stage 1 NGT individuals, and glucose in stage 1 and stage 2 NGT individuals. For glucose and insulin-related traits, we excluded subjects known to be on medications that could affect glucose concentration. For lipids, we excluded individuals known

to be on lipid-lowering medications[33], and for blood pressure we used all individuals but imputed the blood pressure for individuals known to be taking blood pressure medication[32]. We analyzed T2D and NGT individuals separately in each stage. We regressed the quantitative trait variables on age, age^2, sex, birth province, and study indicator. We transformed each quantitative trait to approximate normality using inverse normal scores, and then carried out association analysis on the residuals. To allow for relatedness of the 119 NGT offspring (genotyped in stage 1) with their NGT parents, regression coefficients were estimated in the context of a variance component model that also accounted for background polygenic effects[34]. We tested for association using the residuals under an additive genetic model. Genotype data were used for all analyses except for stage 1 results for rs7961581, where we used imputed data.

SUPPLEMENTARY METHODS REFERENCES

1. Wellcome Trust Case Control Consortium. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* **447**, 661-678 (2007).
2. Zeggini, E. *et al.* Replication of genome-wide association signals in UK samples reveals risk loci for type 2 diabetes. *Science* **316,** 1336-1341 (2007).
3. Diabetes Genetics Initiative. Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels. *Science* **316**, 1331-1336 (2007).
4. Rabbee, N. & Speed, T.P. A genotype calling algorithm for affymetrix SNP arrays. *Bioinformatics* **22**, 7-12 (2006).
5. Di, X. *et al.* Dynamic model based algorithms for screening and genotyping over 100 K SNPs on oligonucleotide microarrays. *Bioinformatics* **21**, 1958-63 (2005).
6. Scott, L.J. *et al.* A genome-wide association study of type 2 diabetes in Finns detects multiple susceptibility variants. *Science* **316**, 1341-1345 (2007).
7. 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* **39**, 906-913 (2007).
8. Purcell, S. *et al*. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559-575 (2007).
9. Price, A.L. *et al*. Principal components analysis corrects for stratification in genome-wide association studies. *Nat. Genet*. **38**, 904-909 (2006).
10. Higgins, J.P., Thompson, S.G., Deeks, J.J. & Altman, D.G. Measuring inconsistency in meta-analyses. *BMJ* **327**, 557-560 (2003).
11. Ardlie, K.G., Lunetta, K.L. & Seielstad, M. Testing for population subdivision and association in four case-control studies. *Am. J. Hum. Genet*. **71**, 304-311 (2002).
12. Risch, N. & Merikangas, K. The future of genetic studies of complex human diseases. *Science* **273**, 1516-1517 (1996).
13. Lin, S., Chakravarti, A. & Cutler, D.J. Exhaustive allelic transmission disequilibrium tests as a new approach to genome-wide association studies. *Nat. Genet*. **36**, 1181-1188 (2004).

14. Steinthorsdottir, V. *et al.* A variant in *CDKAL1* influences insulin response and risk of type 2 diabetes. *Nat. Genet.* **39**, 770-775 (2007).
15. Reynisdottir, I. *et al*. Localization of a susceptibility gene for type 2 diabetes to chromosome 5q34-q35.2. *Am. J. Hum. Genet.* **73**, 323-335 (2003).
16. Kutyavin, I.V. *et al*. A novel endonuclease IV post-PCR genotyping system. *Nucleic Acids Res.* **34**, e128 (2006).
17. Gretarsdottir, S. *et al*. The gene encoding phosphodiesterase 4D confers risk of ischemic stroke, *Nat. Genet.* **35**, 131-138 (2003).
18. Wichmann, H.E., Gieger, C. & Illig, T. KORA-gen- resource for population genetics, controls and a broad spectrum of disease phenotypes. *Gesundheitswesen* **67**, S26-S30 (2005).
19. Huth, C. e*t al*. IL6 Promoter Polymorphisms and Type 2 Diabetes Mellitus:Joint Analysis of Individual Participants' Data from 21 Studies. *Diabetes* **55**, 2915-2921 (2006).
20. Jørgensen, T. *et al*. A randomized non-pharmacological intervention study for prevention of ischaemic heart disease: baseline results Inter99 (1). *Eu.r J. Cardiovasc. Prevention Rehab.* **10**, 377-386 (2003).
21. 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.* **Suppl 3**, S6-S11 (2000).
22. 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* **51**, 70-75 (2008).
23. Midthjell, K. *et al*. Rapid changes in the prevalence of obesity and known diabetes mellitus in an adult Norwegian population. The Nord-Trøndelag Health Surveys: 1984-1986 and 1995-1997. *Diabetes Care* **22**, 1813-1820 (1999).
24. Holmen, J. *et al*. The Nørd-Trondelag Health Study 1995-97 (HUNT 2). Objectives, contents, methods and participation. *Norwegian Journal of Epidemiology* **13**, 19-32 (2003).
25. Midthjell, K., Holmen, J., Bjørndal, A. & Lund-Larsen, P.G. Is questionnaire information valid in the study of a chronic disease such as diabetes? The Nord-Trøndelag diabetes study. *J Epidemiol Community Health* **46**, 537-542 (1992).
26. Qi, L. *et al*. Genetic variation in IL6 gene and type 2 diabetes: Tagging-SNP haplotype analysis in large-scale case-control study and meta-analysis. *Hum. Mol. Genet*. **15**, 1914-1920 (2006).
27. Halsall, D.J., McFarlane, I., Luan, J., Cox, T.M. & Wareham, N.J. Typical type 2 diabetes mellitus and HFE gene mutations: a population-based case-control study. *Hum. Mol. Genet.* **15**, 1361-1365 (2003).
28. Day, N. *et al*. EPIC-Norfolk: study design and characteristics of the cohort. European Prospective Investigation of Cancer. Br. J. Cancer **80**, 95-103 (1999).
29. Harding, A.H. *et al.* Dietary fat and the risk of clinical type 2 diabetes: the European prospective investigation of Cancer-Norfolk study. Am *J. Epidemiol*. **159**, 73-82 (2004).

30. 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. Relat. Metab. Disord*. **24**, S6-11 (2000).
31. Loos, R.J. *et al*. TCF7L2 polymorphisms modulate proinsulin levels and beta-cell function in a British Europid population. *Diabetes* **56**, 1943-1947 (2007).
32. Levy, D. *et al*. Evidence for a gene influencing blood pressure on chromosome 17. Genome scan linkage results for longitudinal blood pressure phenotypes in subjects from the Framingham heart study. *Hypertension* **36**, 477-83 (2000).
33. Willer, C.J. *et al*. Genome-wide association scans identify novel loci that influence lipid levels and risk of coronary artery disease. *Nat. Genet.* **40**, 161-169 (2008).
34. Chen, W.M. & Abecasis, G.R. Family-based association tests for genomewide association scans. *Am. J. Hum. Genet.* **81**, 913-926 (2007).

SUPPLEMENTARY RESULTS

PPARG region conditional analyses

To assess whether rs17036101 represented an independent association signal from the known type 2 diabetes signal 115.3 kb downstream in the *PPARG* gene (rs1801282 (Pro12Ala)), we carried out conditional logistic regression in all stage 1 and stage 2 samples (Supplementary Table 7). Addition of rs1801282 to the model of rs17036101 association with T2D resulted in attenuation of the effect and significance of the signal (stage 1+2 OR 1.17[1.10-1.25], $p=1.1x10^{-6}$ before conditioning, and OR=1.11[1.03-1.20], p=0.02 after conditioning on rs1801282), suggesting that the two SNPs are not independent and represent the same association signal. Reciprocal conditional analysis yielded similar results (stage 1+2 OR=1.14[1.09-1.20], $p=3x10^{-8}$ for *PPARG* SNP rs1801282 as compared to combined OR=1.09[1.02-1.16], p=0.01 after including rs17036101 in the model). We performed meta-analysis of stage 1 imputed and genotyped p values for SNPs in the *SYN2/PPARG* region after including rs17036101 or rs1801282 in the logistic regression model (Supplementary Figure 3). Analysis of regional SNPs conditional on rs17036101 resulted in fewer residual signals than analysis conditional on rs1801282 (best residual SNP for both analyses was rs17035778 in *SYN2* ($r^2=0.004$ with rs17036101 and $r^2=0.007$ with rs1801282 in HapMap CEU; p=0.15 in stage 1 meta-analysis, $p=7.4 x10^{-3}$ after conditioning on rs17036101 and $p=1.1x10^{-3}$ after conditioning on rs1801282). This analysis suggests that more than one independent T2D signal may exist at this locus, but additional validation will be necessary.

Novel region conditional analyses

To identify independent signals at each locus, we repeated T2D association analyses in the vicinity (~1Mb) of novel regions after adding the index SNP in the logistic regression model. P values from conditional analyses at each locus of genotyped and best-guess imputed SNPs in DGI, FUSION and WTCCC were combined using a weighted z score meta-analysis. In all regions, no strong residual association signal remained with $p<10^{-5}$ after conditional analysis (Supplementary Figure 4); however, some residual association with $p<10^{-2}$ was observed at most loci. These included SNPs at (a) the *JAZF1* region (best residual SNP rs2391592 $r^2=0.75$ to rs864745 in HapMap CEU; $p=2x10^{-6}$ before conditioning and $p=4x10^{-3}$ after conditioning on rs864745, suggesting this may be a better proxy for this association signal) (b) the *CDC123/CAMK1D* region (best residual SNP rs928337 $r^2=0.03$ to rs12779790; p=0.01 before conditioning, and $p=5x10^{-4}$ after conditioning on rs12779790); (c) the *TSPAN8/LGR5* region (best residual SNP rs10784891 $r^2= 0.20$ to rs7961581; $p=2x10^{-6}$ before conditioning and $p=2x10^{-4}$ after conditioning on rs7961581, likely to be the same signal); (d) the *THADA* region (best residual SNP rs13382655 $r^2=0.006$ to rs7578597; p=0.06 before conditioning and p=0.005 after conditioning on rs7578597); (e) the *ADAMTS9* region (best residual SNP rs6762376 $r^2= 0.004$ to rs4607103; p=0.4 before conditioning and p=0.008 after conditioning on rs4607103); (f) the *NOTCH2/ADAM30* region (best residual SNP rs637868 $r^2=0$ to

rs10923931; p=0.03 before conditioning and $p=7x10^{-3}$ after conditioning on rs10923931); (g) the *DCD* region (best SNP rs11832807, $r^2=0.086$ to rs1153188; p=0.3 before conditioning and $p=7x10^{-3}$ after conditioning on rs1153188) and (h) the *VEGFA* region (best residual SNP rs9381299; p=0.04 before conditioning and p=0.019 after conditioning on rs9472138). Follow-up studies will be required to determine whether any of these SNPs represent independent association signals.

Investigation of the effect of population structure

UK: The WTCCC GWAS identified 13 chromosomal regions with SNPs showing large minor allele frequency differences between samples of individuals from 12 regions across the UK[1]. None of the replicating signals reported here reside in any of these 13 chromosomal regions. To further guard against the effects of population stratification on association results, we repeated analysis of genome-wide data in stage 1 by adjusting for the two ancestry-informative principal components arising from the WTCCC and find close correlation between adjusted and unadjusted p values ($r^2=0.94$, $p<10^{-4}$). In addition, we find little evidence for association between the replicating signals and the eigenvectors in stage 1 samples ($p>10^{-3}$). When adjusting the set of SNPs taken forward to replication in this study for the first two PCs, the p value correlation before and after adjustment is very high ($r^2>0.99$, $p<10^{-4}$). The largest fold change in p value for the SNPs presented in Table 2 was for rs2934381 (p=0.05 to p=0.12 following adjustment). In addition, we have genotyped 5 of the top 13 ancestry-informative variants (rs1042712, rs10774241, rs2143877, rs7696175 and rs9378805)[1] in the UK replications sets (RS) 1 and 3. We observe no association with case-control status and any of the 5 markers in these datasets.

DGI: To further guard against fine-scale population substructure as a potential confounder in our stage 1 results, we compared the statistical support for SNPs presented in Table 2 by adjustment using principal components as determined previously by EIGENSTRAT analysis in the population-based component of the DGI study[1]. We found that the -log10(p values) after adjustment for 2 or 10 principal components were strongly correlated with -log10(p values) without any principal component adjustment (for 2 PC: r=0.997, $p<4.0x10^{-11}$; for 10PC: r=0.994, $p<5.2x10^{-10}$). The largest fold change was observed for rs9472138 (p=0.015 before adjustment to p=0.021 and p=0.038 after adjustment for 2 and 10 PCs respectively). We concluded that our stage 1 results are therefore not confounded by cryptic fine-scale population substructure, and that treating the fine-scale ascertainment locales as covariates were sufficient to appropriately correct for population stratification in the stage 1 analysis.

FUSION: We saw very little change in stage 1 T2D association results of SNPs in Table 2 when including 10 ancestry principal components determined by EIGENSTRAT analysis of 10% of the stage 1 SNPs; the largest fold change in p value was from $1.4x10^{-4}$ to $2.8x10^{-4}$ for rs17036101. The strongest correlation between the SNP genotypes and a PC was r=-0.081 ($p=1.0x10^{-4}$) for rs12779790. The T2D association p value for this SNPs was only slightly less significant with the inclusion of the ancestry PCs (p=0.021 compared to p=0.015) than in our

standard analysis, suggesting our matching of case/control sample on birthplace and inclusion of the birthplace in the analysis was sufficient to prevent any pronounced effects of population stratification. In stage 2 samples, we matched on birthplace as in stage 1 and included birthplace in the analysis model. These samples are from a more diverse set of studies, but many of the sampling centers were in common with stage 1, suggesting that the birthplace matching may be similarly effective.

Comparison of meta-analysis results before and after genomic control correction of individual stage 1 studies

We included stage 1 data in the meta-analysis using two strategies. In our primary analysis, we corrected each set of stage 1 study results for genomic control. We also performed meta-analysis with stage 1 results that were not corrected for genomic control. Supplementary Table 6 provides details of the observed meta-analysis p values with and without stage 1 GC correction for the 11 SNPs taken forward to stage 3. We find very small differences between the results of the two analysis strategies. Specifically, the largest fold change in p value was observed for rs2641348 ($p=3.2x10^{-7}$ when combining GC-uncorrected data to $p=4.0x10^{-7}$ when combining GC-corrected stage 1 data). With the exception of Supplementary Figure 1 and Supplementary Table 6, where GC-uncorrected data are also presented, all other results in the manuscript reflect the outcome of the meta-analysis using GC-corrected stage 1 data.

SUPPLEMENTARY RESULTS REFERENCES

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

Membership of the Wellcome Trust Case Control Consortium

Management Committee: Paul R Burton[1], David G Clayton[2], Lon R Cardon[3], Nick Craddock[4], Panos Deloukas[5], Audrey Duncanson[6], Dominic P Kwiatkowski[3,5], Mark I McCarthy[3,7], Willem H Ouwehand[8,9], Nilesh J Samani[10], John A Todd[2], Peter Donnelly (Chair)[11]
Data and Analysis Committee: Jeffrey C Barrett[3], Paul R Burton[1], Dan Davison[11], Peter Donnelly[11], Doug Easton[12], David M. Evans[3], Hin-Tak Leung[2], Jonathan L Marchini[11], Andrew P Morris[3], Chris CA Spencer[11], Martin D Tobin[1], Lon R Cardon (Co-chair)[3], David G Clayton (Co-chair)[2]
UK Blood Services & University of Cambridge Controls: Antony P Attwood[5,8], James P Boorman[8,9], Barbara Cant[8], Ursula Everson[13], Judith M Hussey[14], Jennifer D Jolley[8], Alexandra S Knight[8], Kerstin Koch[8], Elizabeth Meech[15], Sarah Nutland[2], Christopher V Prowse[16], Helen E Stevens[2], Niall C Taylor[8], Graham R Walters[17], Neil M Walker[2], Nicholas A Watkins[8,9], Thilo Winzer[8], John A Todd[2], Willem H Ouwehand[8,9]
1958 Birth Cohort Controls: Richard W Jones[18], Wendy L McArdle[18], Susan M Ring[18], David P Strachan[19], Marcus Pembrey[18,20]
Bipolar Disorder (Aberdeen): Gerome Breen[21], David St Clair[21]; (**Birmingham):** Sian Caesar[22], Katherine Gordon-Smith[22,23], Lisa Jones[22]; **(Cardiff):** Christine Fraser[23], Elaine K Green[23], Detelina Grozeva[23], Marian L Hamshere[23], Peter A Holmans[23], Ian R Jones[23], George Kirov[23], Valentina Moskvina[23], Ivan Nikolov[23], Michael C O'Donovan[23], Michael J Owen[23], Nick Craddock[23]; **(London):** David A Collier[24], Amanda Elkin[24], Anne Farmer[24], Richard Williamson[24], Peter McGuffin[24]; **(Newcastle):** Allan H Young[25], I Nicol Ferrier[25]
Coronary Artery Disease (Leeds): Stephen G Ball[26], Anthony J Balmforth[26], Jennifer H Barrett[26], D Timothy Bishop[26], Mark M Iles[26], Azhar Maqbool[26], Nadira Yuldasheva[26], Alistair S Hall[26]; **(Leicester):** Peter S Braund[10], Paul R Burton[1], Richard J Dixon[10], Massimo Mangino[10], Suzanne Stevens[10], Martin D Tobin[1], John R Thompson[1], Nilesh J Samani[10]
Crohn's Disease (Cambridge): Francesca Bredin[27], Mark Tremelling[27], Miles Parkes[27]; **(Edinburgh):** Hazel Drummond[28], Charles W Lees[28], Elaine R Nimmo[28], Jack Satsangi[28]; **(London):** Sheila A Fisher[29], Alastair Forbes[30], Cathryn M Lewis[29], Clive M Onnie[29], Natalie J Prescott[29], Jeremy Sanderson[31], Christopher G Mathew[29]; **(Newcastle):** Jamie Barbour[32], M Khalid Mohiuddin[32], Catherine E Todhunter[32], John C Mansfield[32]; **(Oxford):** Tariq Ahmad[33], Fraser R Cummings[33], Derek P Jewell[33]
Hypertension (Aberdeen): John Webster[34]; **(Cambridge):** Morris J Brown[35], David G Clayton[2]; **(Evry, France): G** Mark Lathrop[36]; **(Glasgow):** John Connell[37], Anna Dominiczak[37]; **(Leicester):** Nilesh J Samani[10]; **(London):** Carolina A Braga Marcano[38], Beverley Burke[38], Richard Dobson[38], Johannie Gungadoo[38], Kate L Lee[38], Patricia B Munroe[38], Stephen J Newhouse[38], Abiodun Onipinla[38], Chris Wallace[38], Mingzhan Xue[38], Mark Caulfield[38]; **(Oxford):** Martin Farrall[39]
Rheumatoid Arthritis: Anne Barton[40], Ian N Bruce[40], Hannah Donovan[40], Steve Eyre[40], Paul D Gilbert[40], Samantha L Hider[40], Anne M Hinks[40], Sally L John[40],

Catherine Potter[40], Alan J Silman[40], Deborah PM Symmons[40], Wendy Thomson[40], Jane Worthington[40]
Type 1 Diabetes: David G Clayton[2], David B Dunger[2,41], Sarah Nutland[2], Helen E Stevens[2], Neil M Walker[2], Barry Widmer[2,41], John A Todd[2]
Type 2 Diabetes (Exeter): Timothy M Frayling[42,43], Rachel M Freathy[42,43], Hana Lango[42,43], John R B Perry[42,43], Beverley M Shields[43], Michael N Weedon[42,43], Andrew T Hattersley[42,43]; **(London):** Graham A Hitman[44]; **(Newcastle):** Mark Walker[45]; **(Oxford):** Kate S Elliott[3,7], Christopher J Groves[7], Cecilia M Lindgren[3,7], Nigel W Rayner[3,7], Nicholas J Timpson[3,46], Eleftheria Zeggini[3,7], Mark I McCarthy[3,7]
Tuberculosis (Gambia): Melanie Newport[47], Giorgio Sirugo[47]; **(Oxford):** Emily Lyons[3], Fredrik Vannberg[3], Adrian VS Hill[3]
Ankylosing Spondylitis: Linda A Bradbury[48], Claire Farrar[49], Jennifer J Pointon[48], Paul Wordsworth[49], Matthew A Brown[48,49]
AutoImmune Thyroid Disease: Jayne A Franklyn[50], Joanne M Heward[50], Matthew J Simmonds[50], Stephen CL Gough[50]
Breast Cancer: Sheila Seal[51], Michael R Stratton[51,52], Nazneen Rahman[51]
Multiple Sclerosis: Maria Ban[53], An Goris[53], Stephen J Sawcer[53], Alastair Compston[53]
Gambian Controls (Gambia): David Conway[47], Muminatou Jallow[47], Melanie Newport[47], Giorgio Sirugo[47]; **(Oxford):** Kirk A Rockett[3], Dominic P Kwiatkowski[3,5]
DNA, Genotyping, Data QC and Informatics (Wellcome Trust Sanger Institute, Hinxton): Suzannah J Bumpstead[5], Amy Chaney[5], Kate Downes[2,5], Mohammed JR Ghori[5], Rhian Gwilliam[5], Sarah E Hunt[5], Michael Inouye[5], Andrew Keniry[5], Emma King[5], Ralph McGinnis[5], Simon Potter[5], Rathi Ravindrarajah[5], Pamela Whittaker[5], Claire Widden[5], David Withers[5], Panos Deloukas[5]; **(Cambridge):** Hin-Tak Leung[2], Sarah Nutland[2], Helen E Stevens[2], Neil M Walker[2], John A Todd[2]
Statistics (Cambridge): Doug Easton[12], David G Clayton[2]; **(Leicester):** Paul R Burton[1], Martin D Tobin[1]; **(Oxford):** Jeffrey C Barrett[3], David M Evans[3], Andrew P Morris[3], Lon R Cardon[3]; **(Oxford):** Niall J Cardin[11], Dan Davison[11], Teresa Ferreira[11], Joanne Pereira-Gale[11], Ingeleif B Hallgrimsdóttir[11], Bryan N Howie[11], Jonathan L Marchini[11], Chris CA Spencer[11], Zhan Su[11], Yik Ying Teo[3,11], Damjan Vukcevic[11], Peter Donnelly[11]
PIs: David Bentley[5,54], Matthew A Brown[48,49], Lon R Cardon[3], Mark Caulfield[38], David G Clayton[2], Alistair Compston[53], Nick Craddock[23], Panos Deloukas[5], Peter Donnelly[11], Martin Farrall[39], Stephen CL Gough[50], Alistair S Hall[26], Andrew T Hattersley[42,43], Adrian VS Hill[3], Dominic P Kwiatkowski[3,5], Christopher G Mathew[29], Mark I McCarthy[3,7], Willem H Ouwehand[8,9], Miles Parkes[27], Marcus Pembrey[18,20], Nazneen Rahman[51], Nilesh J Samani[10], Michael R Stratton[51,52], John A Todd[2], Jane Worthington[40]

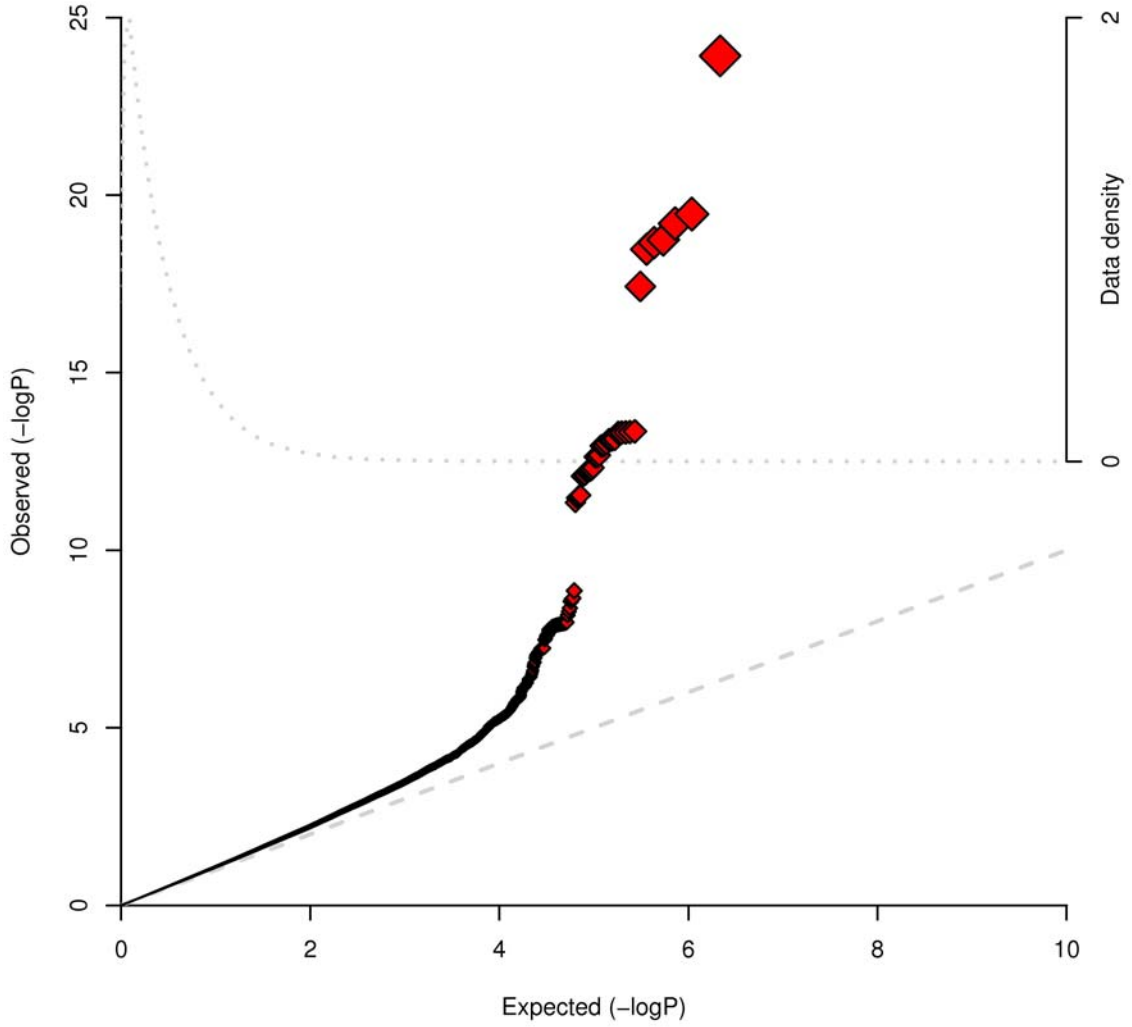
[1] Genetic Epidemiology Group, Department of Health Sciences, University of Leicester, Adrian Building, University Road, Leicester, LE1 7RH, UK; [2] Juvenile Diabetes Research Foundation/Wellcome Trust Diabetes and Inflammation Laboratory, Department of Medical Genetics, Cambridge Institute for Medical

Research, University of Cambridge, Wellcome Trust/MRC Building, Cambridge, CB2 0XY, UK; [3] Wellcome Trust Centre for Human Genetics, University of Oxford, Roosevelt Drive, Oxford OX3 7BN, UK; [4] Department of Psychological Medicine, Henry Wellcome Building, School of Medicine, Cardiff University, Heath Park, Cardiff CF14 4XN, UK; [5] The Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, UK; [6] The Wellcome Trust, Gibbs Building, 215 Euston Road, London NW1 2BE, UK; [7] Oxford Centre for Diabetes, Endocrinology and Medicine, University of Oxford, Churchill Hospital, Oxford, OX3 7LJ, UK; [8] Department of Haematology, University of Cambridge, Long Road, Cambridge, CB2 2PT, UK; [9] National Health Service Blood and Transplant, Cambridge Centre, Long Road, Cambridge, CB2 2PT, UK; [10] Department of Cardiovascular Sciences, University of Leicester, Glenfield Hospital, Groby Road, Leicester, LE3 9QP, UK; [11] Department of Statistics, University of Oxford, 1 South Parks Road, Oxford OX1 3TG, UK; [12] Cancer Research UK Genetic Epidemiology Unit, Strangeways Research Laboratory, Worts Causeway, Cambridge CB1 8RN, UK; [13] National Health Service Blood and Transplant, Sheffield Centre, Longley Lane, Sheffield S5 7JN, UK; [14] National Health Service Blood and Transplant, Brentwood Centre, Crescent Drive, Brentwood, CM15 8DP, UK; [15] The Welsh Blood Service, Ely Valley Road, Talbot Green, Pontyclun, CF72 9WB, UK; [16] The Scottish National Blood Transfusion Service, Ellen's Glen Road, Edinburgh, EH17 7QT, UK; [17] National Health Service Blood and Transplant, Southampton Centre, Coxford Road, Southampton, SO16 5AF, UK; [18] Avon Longitudinal Study of Parents and Children, University of Bristol, 24 Tyndall Avenue, Bristol, BS8 1TQ, UK; [19] Division of Community Health Services, St George's University of London, Cranmer Terrace, London SW17 0RE, UK; [20] Institute of Child Health, University College London, 30 Guilford St, London WC1N 1EH, UK; [21] University of Aberdeen, Institute of Medical Sciences, Foresterhill, Aberdeen, AB25 2ZD, UK; [22] Department of Psychiatry, Division of Neuroscience, Birmingham University, Birmingham, B15 2QZ, UK; [23] Department of Psychological Medicine, Henry Wellcome Building, School of Medicine, Cardiff University, Heath Park, Cardiff CF14 4XN, UK; [24] SGDP, The Institute of Psychiatry, King's College London, De Crespigny Park Denmark Hill London SE5 8AF, UK; [25] School of Neurology, Neurobiology and Psychiatry, Royal Victoria Infirmary, Queen Victoria Road, Newcastle upon Tyne, NE1 4LP, UK; [26] LIGHT and LIMM Research Institutes, Faculty of Medicine and Health, University of Leeds, Leeds, LS1 3EX, UK; [27] IBD Research Group, Addenbrooke's Hospital, University of Cambridge, Cambridge, CB2 2QQ, UK; [28] Gastrointestinal Unit, School of Molecular and Clinical Medicine, University of Edinburgh, Western General Hospital, Edinburgh EH4 2XU UK; [29] Department of Medical & Molecular Genetics, King's College London School of Medicine, 8th Floor Guy's Tower, Guy's Hospital, London, SE1 9RT, UK; [30] Institute for Digestive Diseases, University College London Hospitals Trust, London, NW1 2BU, UK; [31] Department of Gastroenterology, Guy's and St Thomas' NHS Foundation Trust, London, SE1 7EH, UK; [32] Department of Gastroenterology & Hepatology, University of Newcastle upon Tyne, Royal Victoria Infirmary, Newcastle upon

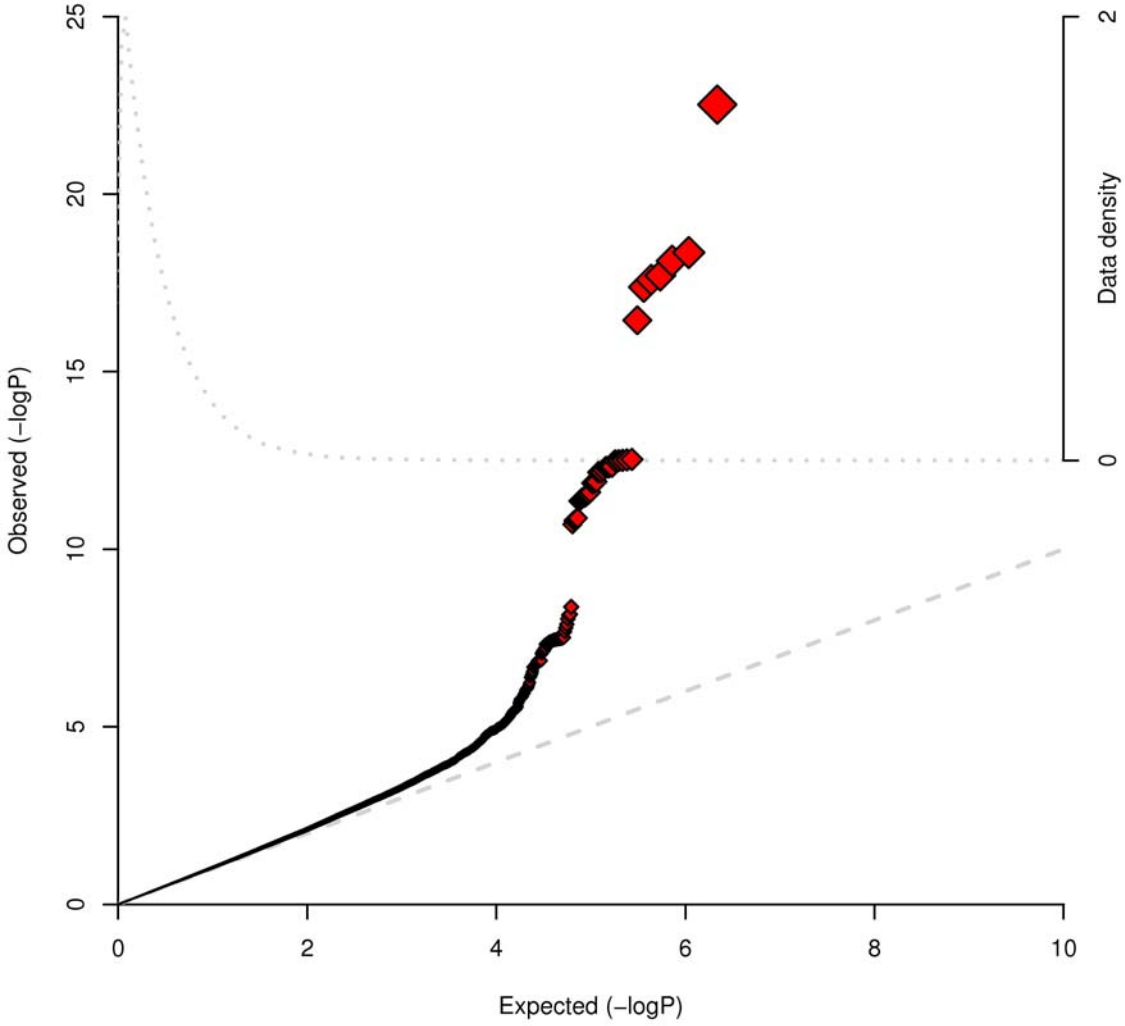
Tyne, NE1 4LP, UK; [33] Gastroenterology Unit, Radcliffe Infirmary, University of Oxford, Oxford, OX2 6HE, UK; [34] Medicine and Therapeutics, Aberdeen Royal Infirmary, Foresterhill, Aberdeen, Grampian AB9 2ZB, UK; [35] Clinical Pharmacology Unit and the Diabetes and Inflammation Laboratory, University of Cambridge, Addenbrookes Hospital, Hills Road, Cambridge CB2 2QQ, UK; [36] Centre National de Genotypage, 2, Rue Gaston Cremieux, Evry, Paris 91057.; [37] BHF Glasgow Cardiovascular Research Centre, University of Glasgow, 126 University Place, Glasgow, G12 8TA, UK; [38] Clinical Pharmacology and Barts and The London Genome Centre, William Harvey Research Institute, Barts and The London, Queen Mary's School of Medicine, Charterhouse Square, London EC1M 6BQ, UK; [39] Cardiovascular Medicine, University of Oxford, Wellcome Trust Centre for Human Genetics, Roosevelt Drive, Oxford OX3 7BN, UK; [40]arc Epidemiology Research Unit, University of Manchester, Stopford Building, Oxford Rd, Manchester, M13 9PT, UK; [41] Department of Paediatrics, University of Cambridge, Addenbrooke's Hospital, Cambridge, CB2 2QQ, UK; [42] Genetics of Complex Traits, Institute of Biomedical and Clinical Science, Peninsula Medical School, Magdalen Road, Exeter EX1 2LU UK; [43] Diabetes Genetics, Institute of Biomedical and Clinical Science, Peninsula Medical School, Barrack Road, Exeter EX2 5DU UK; [44] Centre for Diabetes and Metabolic Medicine, Barts and The London, Royal London Hospital, Whitechapel, London, E1 1BB UK; [45] Diabetes Research Group, School of Clinical Medical Sciences, Newcastle University, Framlington Place, Newcastle upon Tyne NE2 4HH, UK; [46] The MRC Centre for Causal Analyses in Translational Epidemiology, Bristol University, Canynge Hall, Whiteladies Rd, Bristol BS2 8PR, UK; [47] MRC Laboratories, Fajara, The Gambia; [48] Diamantina Institute for Cancer, Immunology and Metabolic Medicine, Princess Alexandra Hospital, University of Queensland, Woolloongabba, Qld 4102, Australia; [49] Botnar Research Centre, University of Oxford, Headington, Oxford OX3 7BN, UK; [50] Department of Medicine, Division of Medical Sciences, Institute of Biomedical Research, University of Birmingham, Edgbaston, Birmingham B15 2TT, UK; [51] Section of Cancer Genetics, Institute of Cancer Research, 15 Cotswold Road, Sutton, SM2 5NG, UK; [52] Cancer Genome Project, The Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, UK; [53] Department of Clinical Neurosciences, University of Cambridge, Addenbrooke's Hospital, Hills Road, Cambridge CB2 2QQ, UK; [54] PRESENT ADDRESS: Illumina Cambridge, Chesterford Research Park, Little Chesterford, Nr Saffron Walden, Essex, CB10 1XL, UK.

Supplementary Figure 1. Q-Q plots for the meta-analysis (z-score based on p values) of T2D across the DGI, FUSION, and WTCCC stage 1 studies (n=10,128). The p values for the corresponding z-scores are plotted (as -log10 values) as a function of p values from the expected (uniform) null distribution either before (a,c) or after (b,d) genomic control within each study on genotyped and imputed SNPs separately. For reference, the frequency of data across the expected p value distribution is plotted in the upper portion of each plot; note that the great majority of data have a -log10(p value) less than 2, as expected. Plots (a,b) retain the 11 "established" T2D loci, whereas plots (c,d) exclude SNPs surrounding them (n=1,981). Note that after genomic control, and after previously validated associations are removed, the observed distribution of p values matches expectation for the bulk of the data, but starts to subtly depart from the null at $p<10^{-3}$.

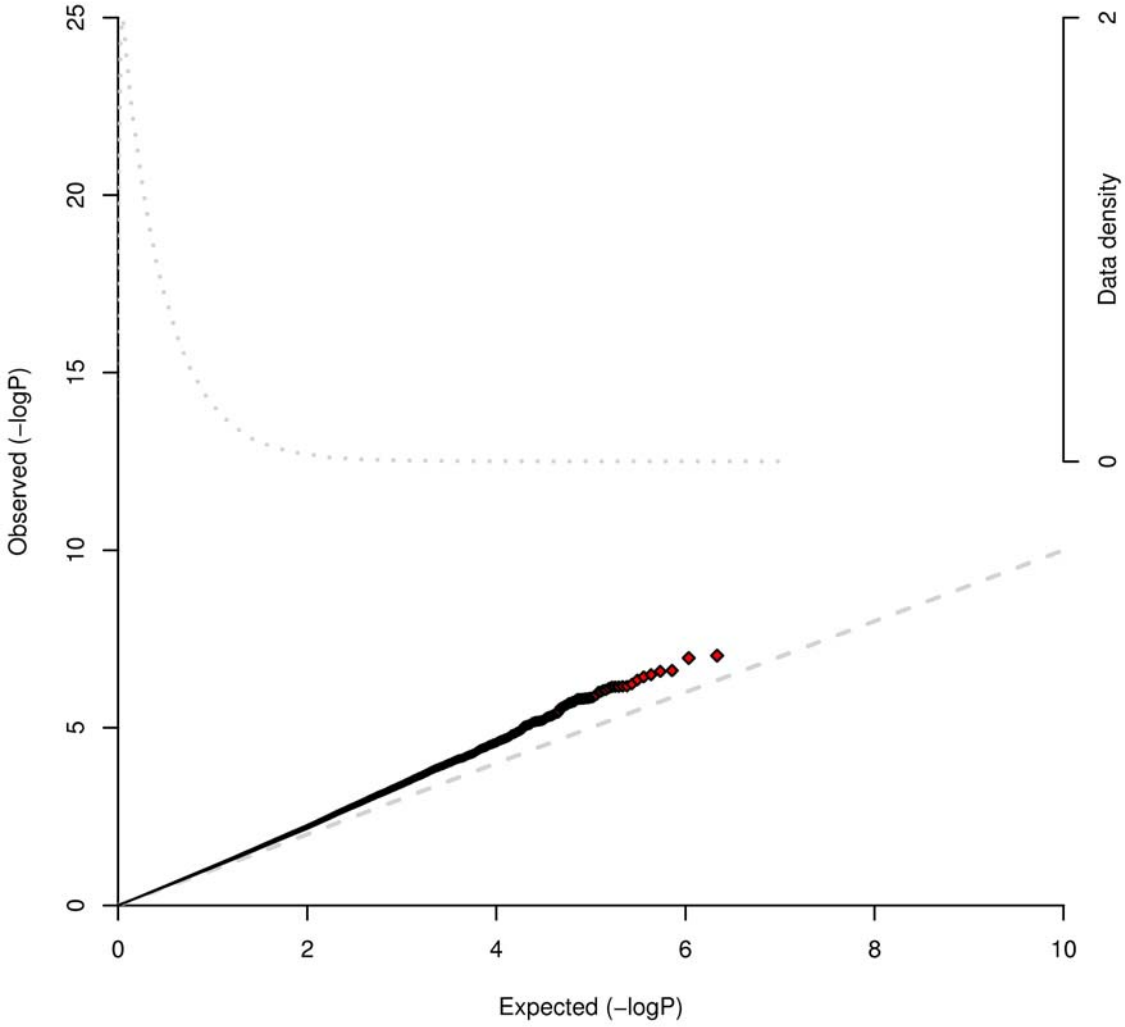
(a) Q-Q plot for the meta-analysis of stage 1 studies before genomic control correction, including signals representing established T2D loci.



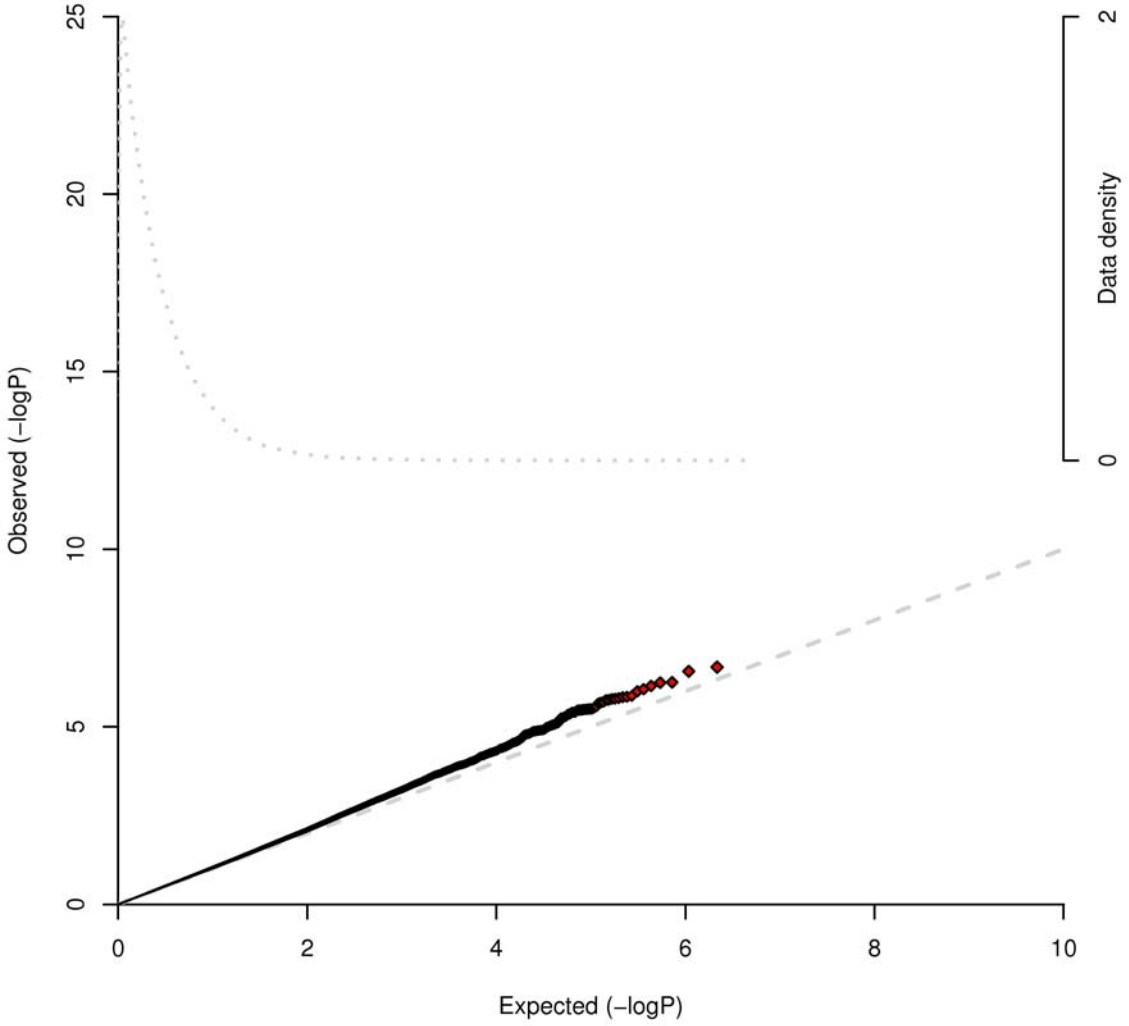
(b) Q-Q plot for the meta-analysis of stage 1 studies after genomic control correction, including signals representing established T2D loci.



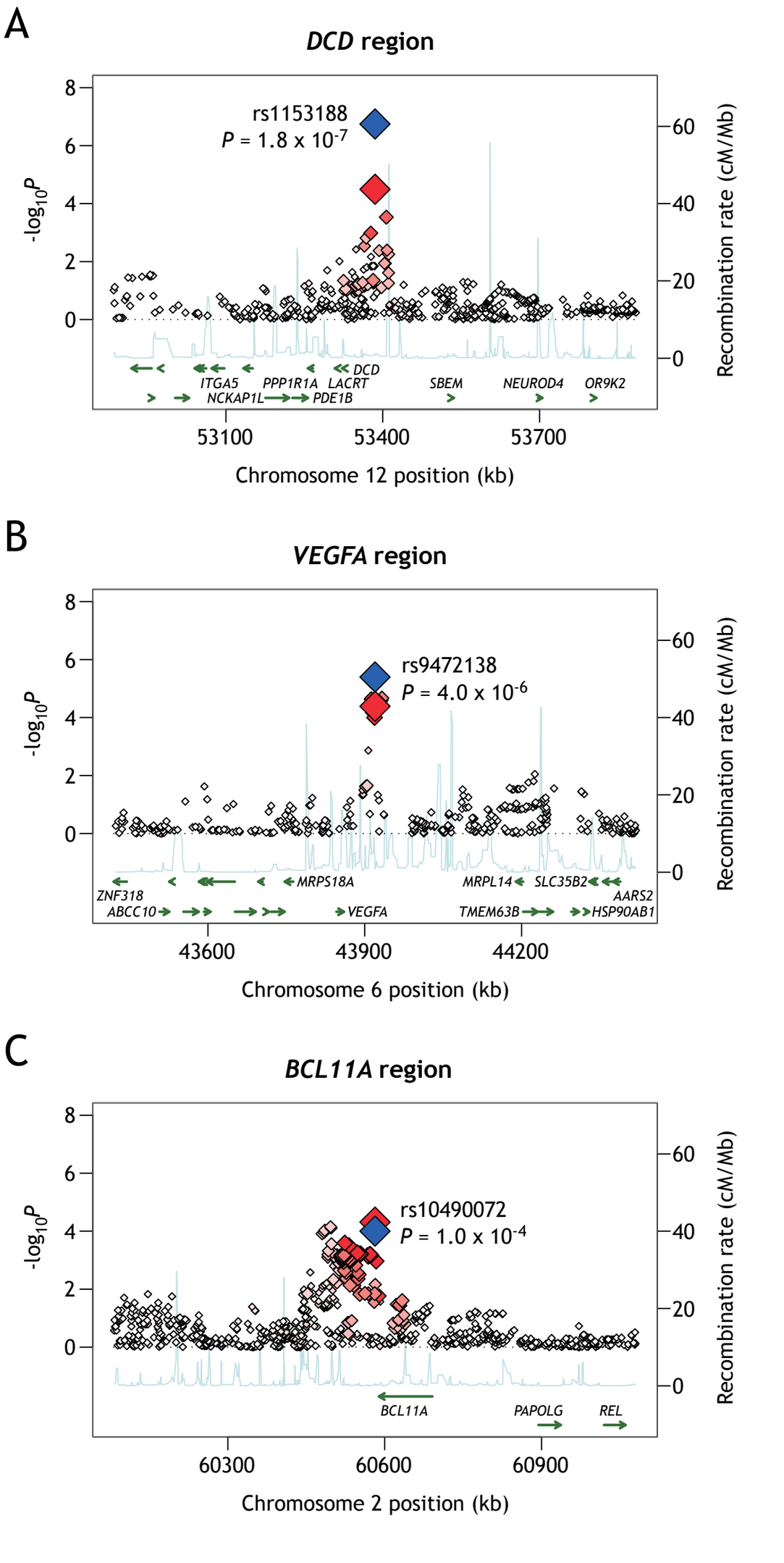
(c) Q-Q plot for the meta-analysis of stage 1 studies before genomic control correction, excluding signals representing established T2D loci.



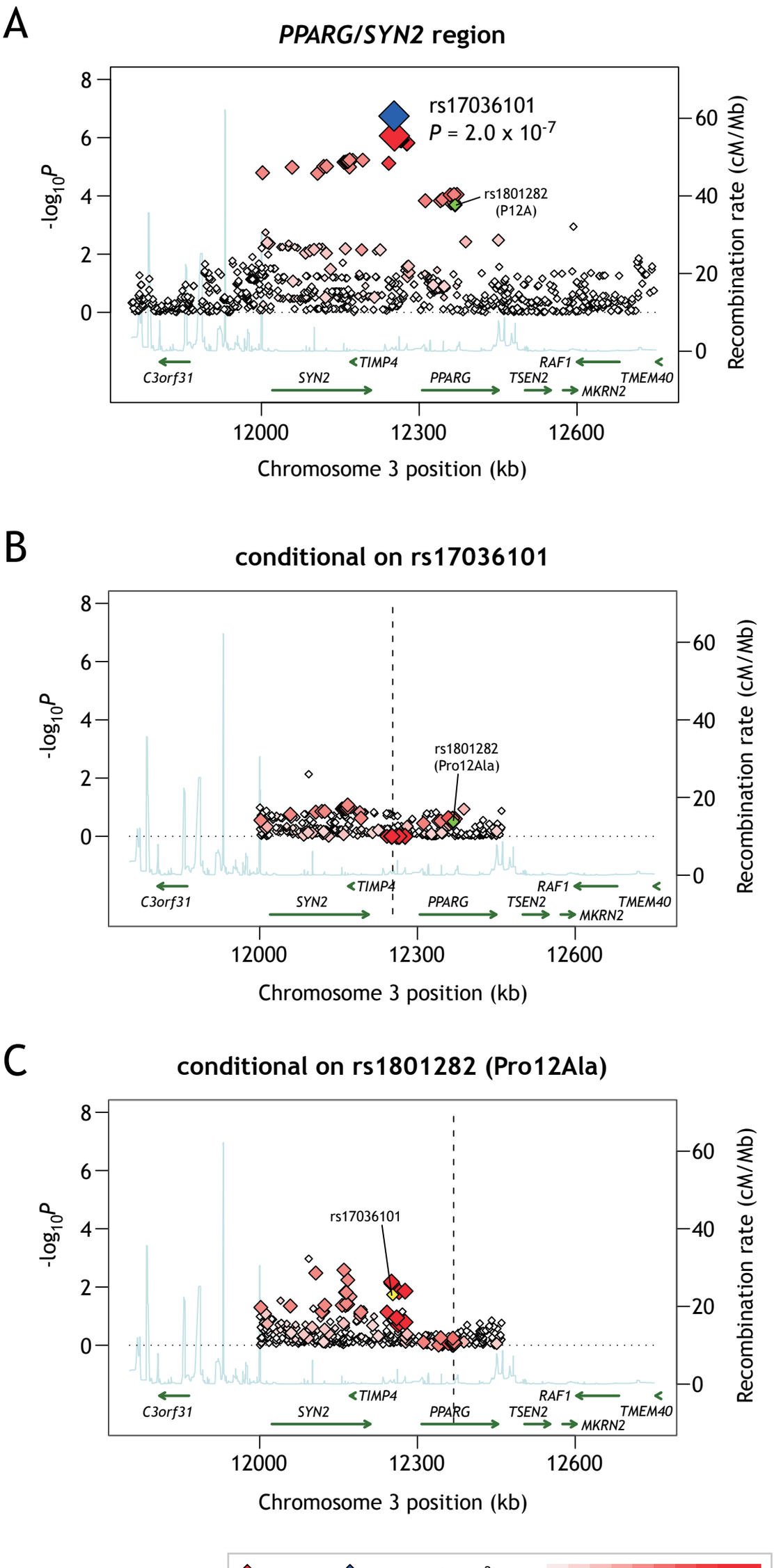
(d) Q-Q plot for the meta-analysis of stage 1 studies after genomic control correction, excluding signals representing established T2D loci.



Supplementary Figure 2. Regional plots of the (A) *DCD*, (B) *VEGFA*, and (C) *BCL11A* regions. Genotyped and imputed SNPs passing QC across all three stage 1 studies are plotted with their meta-analysis p values (as -log10 values) as a function of genomic position (with NCBI Build 35). In each panel, the SNP taken forward to stages 2 and 3 is represented by a blue diamond (meta-analysis p value across stages 1-3), and its initial p value in stage 1 data is denoted by a red diamond. Estimated recombination rates (taken from HapMap)[13] are plotted to reflect the local LD structure around the associated SNPs and their correlated proxies (according to a white to red scale from r^2=0 to r^2=1; based on pairwise r^2 values from HapMap CEU)[13]. Gene annotations were taken from the University of California-Santa Cruz genome browser.

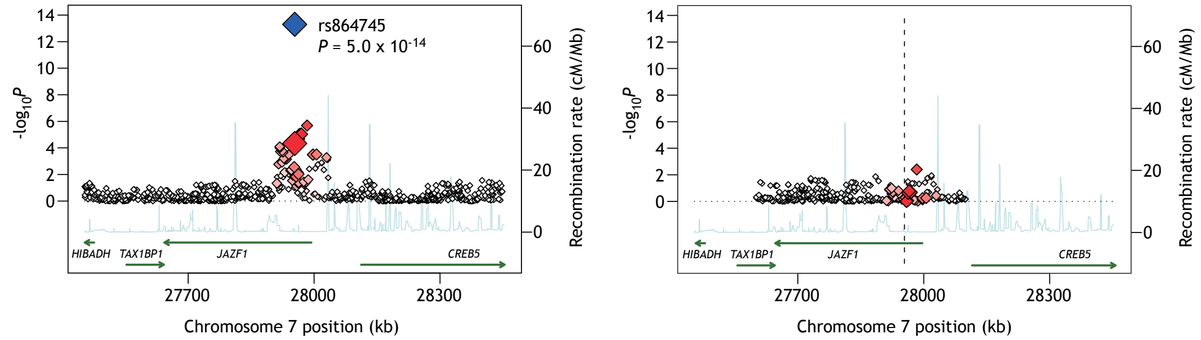

A
DCD region
rs1153188
P = 1.8 x 10-7
-log10P
Recombination rate (cM/Mb)
ITGA5
NCKAP1L
PPP1R1A
PDE1B
LACRT
DCD
SBEM
NEUROD4
OR9K2
53100
53400
53700
Chromosome 12 position (kb)
B
VEGFA region
rs9472138
P = 4.0 x 10-6
ZNF318
ABCC10
MRPS18A
VEGFA
MRPL14
TMEM63B
SLC35B2
AARS2
HSP90AB1
43600
43900
44200
Chromosome 6 position (kb)
C
BCL11A region
rs10490072
P = 1.0 x 10-4
BCL11A
PAPOLG
REL
60300
60600
60900
Chromosome 2 position (kb)
stage 1
stages 1-3
r2 scale
0.5
1

Supplementary Figure 3. Regional plot of *PPARG*. Genotyped and imputed SNPs passing QC across all three stage 1 studies are plotted with their meta-analysis p values (as -log10 values) as a function of genomic position (with NCBI Build 35). (A) Unconditional association analysis. SNP rs17036101 was taken forward to stages 2 and 3, and is represented by a blue diamond (meta-analysis p value across stages 1-3), and its initial p value in stage 1 data is denoted by a red diamond. (B) Conditional association analysis based on rs17036101 (with its position indicated by the vertical dashed line) in stage 1. The Pro12Ala variant is indicated by a green diamond. (C) Conditional association analysis based on rs1801282 (Pro12Ala, with its position indicated by the vertical dashed line) in stage 1. Estimated recombination rates (taken from HapMap)[13] are plotted to reflect the local LD structure around the associated SNPs and their correlated proxies (according to a white to red scale from r^2=0 to r^2=1; based on pairwise r^2 values from HapMap CEU)[13]. Gene annotations were taken from the University of California-Santa Cruz genome browser.


A
PPARG/SYN2 region
rs17036101
P = 2.0 x 10-7
rs1801282
(P12A)
-log10P
Recombination rate (cM/Mb)
C3orf31
SYN2
TIMP4
PPARG
TSEN2
RAF1
TMEM40
MKRN2
12000
12300
12600
Chromosome 3 position (kb)
B
conditional on rs17036101
rs1801282
(Pro12Ala)
C
conditional on rs1801282 (Pro12Ala)
rs17036101
stage 1
stages 1-3
r2 scale
0.5
1

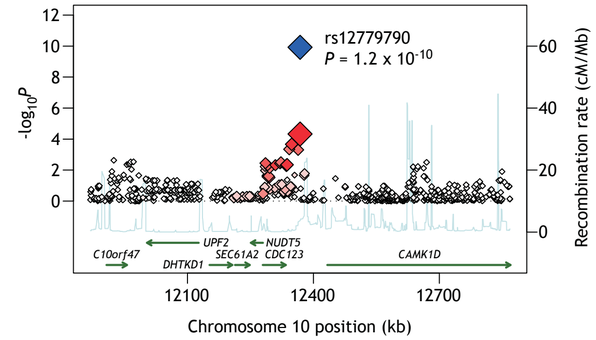
Supplementary Figure 4. Regional plots for the described associations. For each of the (A) *JAZF1*, (B) *CDC123*/*CAMK1D*, (C) *TSPAN8*/*LGR5*, (D) *THADA*, (E) *ADAMTS9*, (F) *NOTCH2*/*ADAM30*, (G) *DCD* and (H) *VEGFA* regions, genotyped and imputed SNPs passing QC across all three stage 1 studies are plotted with their meta-analysis p values (as -log10 values) as a function of genomic position (with NCBI Build 35). For each region, we plot the association results from the unconditional analysis (where the index SNP was taken forward to stages 2 and 3, and is represented by a blue diamond (meta-analysis p value across stages 1-3), and a red diamond (initial p value in stage 1), and the association results from the conditional analysis in stage 1 based on that index SNP (with its position indicated by the vertical dashed line). Estimated recombination rates (taken from HapMap)[13] are plotted to reflect the local LD structure around the associated SNPs and their correlated proxies (according to a white to red scale from r^2=0 to r^2=1; based on pairwise r^2 values from HapMap CEU)[13]. Gene annotations were taken from the University of California-Santa Cruz genome browser.

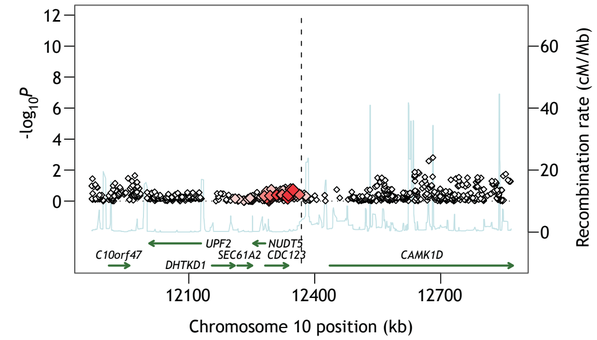
A


JAZF1 region
rs864745
P = 5.0 x 10-14
-log10P
Recombination rate (cM/Mb)
HIBADH TAX1BP1 JAZF1 CREB5
Chromosome 7 position (kb)

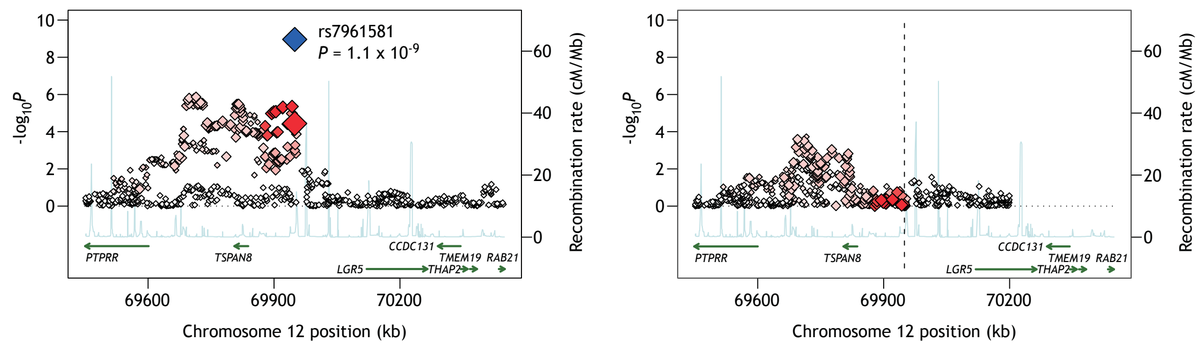
conditional on rs864745
-log10P
Recombination rate (cM/Mb)
HIBADH TAX1BP1 JAZF1 CREB5
Chromosome 7 position (kb)

B


CDC123/CAMK1D region
rs12779790
P = 1.2 x 10-10
-log10P
Recombination rate (cM/Mb)
C10orf47 UPF2 NUDT5 DHTKD1 SEC61A2 CDC123 CAMK1D
Chromosome 10 position (kb)

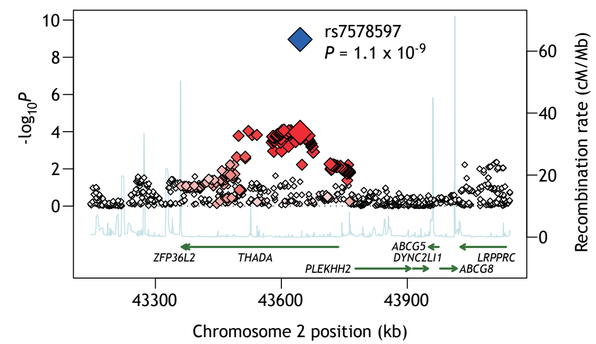

conditional on rs12779790
-log10P
Recombination rate (cM/Mb)
C10orf47 UPF2 NUDT5 DHTKD1 SEC61A2 CDC123 CAMK1D
Chromosome 10 position (kb)

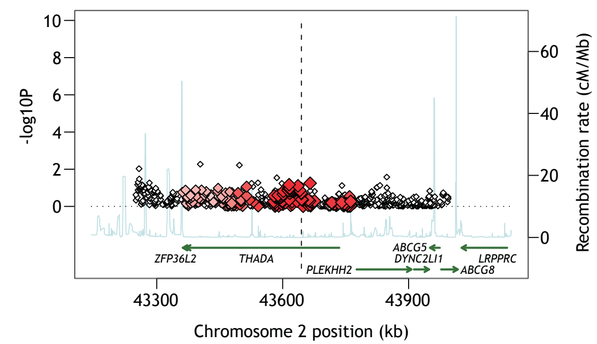
C


TSPAN8/LGR5 region
rs7961581
P = 1.1 x 10-9
-log10P
Recombination rate (cM/Mb)
PTPRR TSPAN8 CCDC131 TMEM19 RAB21 LGR5 THAP2
Chromosome 12 position (kb)

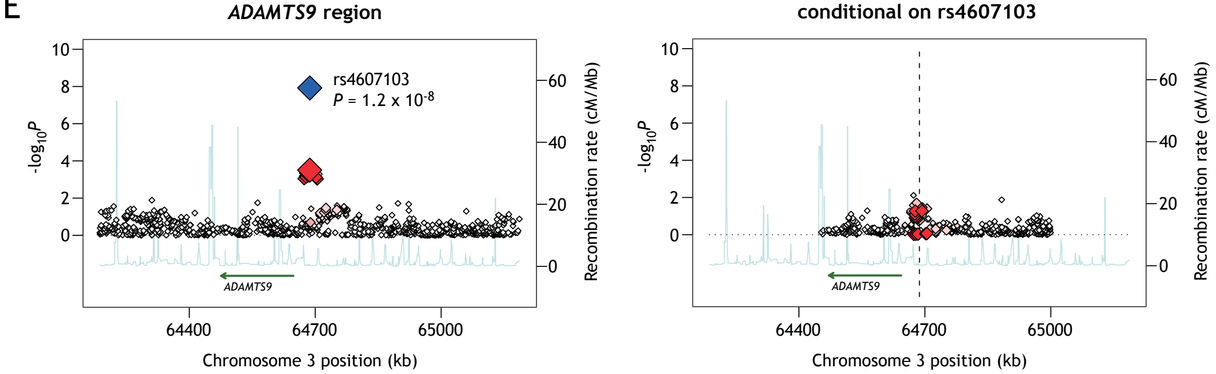
conditional on rs7961581
-log10P
Recombination rate (cM/Mb)
PTPRR TSPAN8 CCDC131 TMEM19 RAB21 LGR5 THAP2
Chromosome 12 position (kb)

D


THADA region
rs7578597
P = 1.1 x 10-9
-log10P
Recombination rate (cM/Mb)
ZFP36L2 THADA PLEKHH2 ABCG5 DYNC2LI1 LRPPRC ABCG8
Chromosome 2 position (kb)

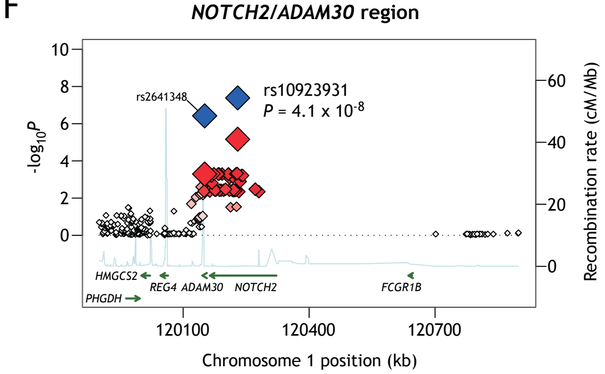

conditional on rs7578597
-log10P
Recombination rate (cM/Mb)
ZFP36L2 THADA PLEKHH2 ABCG5 DYNC2LI1 LRPPRC ABCG8
Chromosome 2 position (kb)

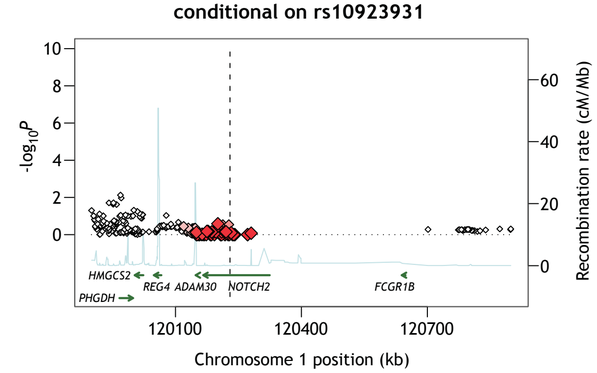
E


ADAMTS9 region
rs4607103
P = 1.2 x 10-8
-log10P
Recombination rate (cM/Mb)
ADAMTS9
Chromosome 3 position (kb)

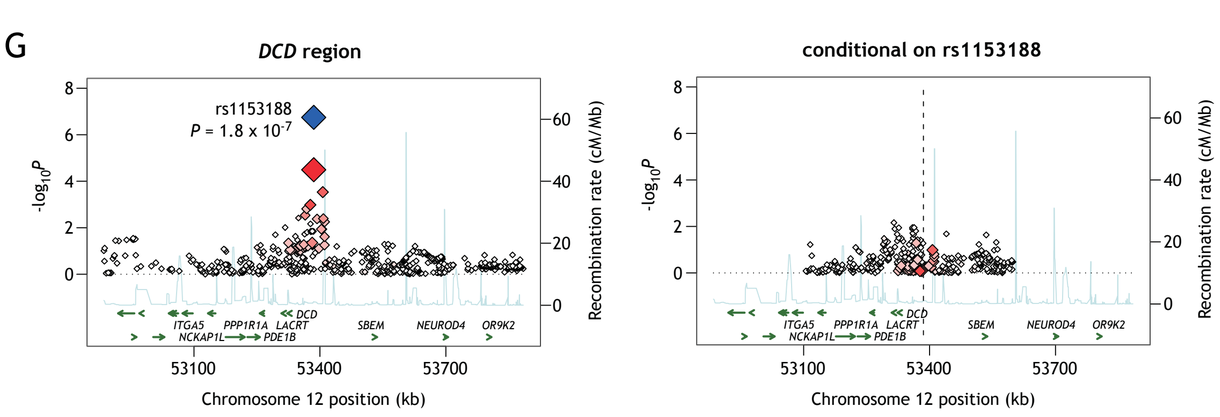
conditional on rs4607103
-log10P
Recombination rate (cM/Mb)
ADAMTS9
Chromosome 3 position (kb)

F


NOTCH2/ADAM30 region
rs2641348
rs10923931
P = 4.1 x 10-8
-log10P
Recombination rate (cM/Mb)
HMGCS2 REG4 ADAM30 NOTCH2 FCGR1B PHGDH
Chromosome 1 position (kb)

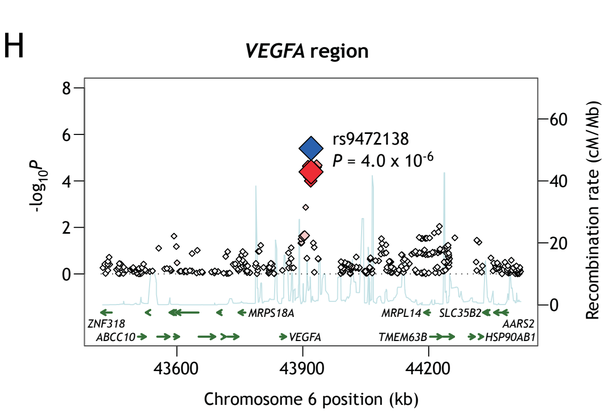

conditional on rs10923931
-log10P
Recombination rate (cM/Mb)
HMGCS2 REG4 ADAM30 NOTCH2 FCGR1B PHGDH
Chromosome 1 position (kb)

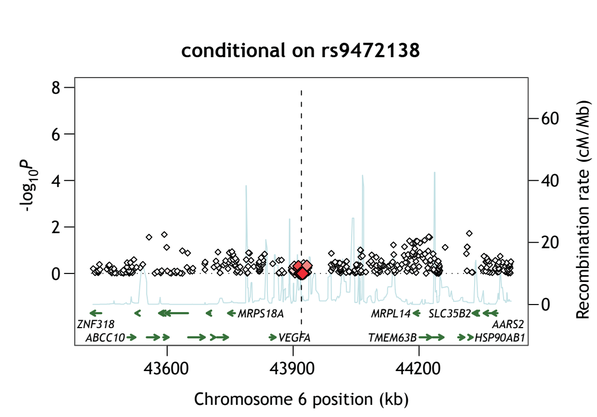
G

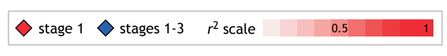

DCD region
rs1153188
P = 1.8 x 10-7
-log10P
Recombination rate (cM/Mb)
ITGA5 PPP1R1A LACRT DCD SBEM NEUROD4 OR9K2 NCKAP1L PDE1B
Chromosome 12 position (kb)

conditional on rs1153188
-log10P
Recombination rate (cM/Mb)
ITGA5 PPP1R1A LACRT DCD SBEM NEUROD4 OR9K2 NCKAP1L PDE1B
Chromosome 12 position (kb)

H


VEGFA region
rs9472138
P = 4.0 x 10-6
-log10P
Recombination rate (cM/Mb)
ZNF318 MRPS18A MRPL14 SLC35B2 AARS2 ABCC10 VEGFA TMEM63B HSP90AB1
Chromosome 6 position (kb)


conditional on rs9472138
-log10P
Recombination rate (cM/Mb)
ZNF318 MRPS18A MRPL14 SLC35B2 AARS2 ABCC10 VEGFA TMEM63B HSP90AB1
Chromosome 6 position (kb)


stage 1
stages 1-3
r2 scale
0.5
1

Supplementary Table 1a. Clinical characteristics of DGI T2D study samples.

Study sample	n (M/F)	Mean age (years)	Age at onset (years)	BMI (kg/m^2)	Fasting PG (mmol/l)	PG at 2h OGTT (mmol/l)
Stage 1						
Unrelated case-control						
Cases	529/493	65 ± 10	58 ± 10	28.1 ± 4.1	9.5 ± 3.1	15.3 ± 5.4
Controls	540/535	58 ± 10		27.6 ± 3.7	5.3 ± 0.5	5.4 ± 1.3
Discordant sibships						
Cases	213/229	62 ± 11	55 ± 10	29.5 ± 5.1	9.8 ± 3.4	15.0 ± 5.2
Controls	167/225	61 ± 11		26.4 ± 3.9	5.4 ± 0.5	6.1 ± 1.2
Stage 2						
Sweden case-control						
Cases	1,667/1,163	59 ± 12	na	29.6 ± 5.5	11.9 ± 4.3	na
Controls	1,340/2,210	57 ± 6		25.1 ± 3.6	4.8 ± 0.4	na
GCI US case-control						
Cases	644/582	63 ± 11	na	32.9 ± 6.9	9.8 ± 3.0	na
Controls	644/582	61 ± 10		27.4 ± 5.2	5.1 ± 0.9	na
GCI Poland case-control						
Cases	422/587	62 ± 10	na	29.6 ± 4.8	8.9 ± 4.0	na
Controls	422/587	59 ± 7		26.1 ± 3.6	4.8 ± 1.2	na

BMI: body mass index, PG: plasma glucose, OGTT: oral glucose tolerance test, na: not available

Supplementary Table 1b. Clinical characteristics of UK samples.

Study Sample	n	% male participants	Age of diagnosis (years) (mean, SD)	Age at study (years) (mean, SD)	Waist circumference (cm) (mean, SD)		Body mass index (kg/m^2) (geometric mean, SD range)
					Male	Female	
Stage 1							
WTCCC cases	1,924	58.1	50.3 (9.2)	58.6 (10.1)	106.2 (13.4)	102.5 (15.4)	30.7 (25.4, 37.1)
WTCCC controls	2,938	49.2	na	na	na	na	na
Stage 2							
RS1 cases	2,022	57.9	55.7 (8.9)	64.2 (9.4)	107.1 (13.1)	101.1 (13.7)	31.0 (25.9, 37.2)
RS2 cases	632	58.3	48.1 (9.1)	58.4 (9.0)	108.1 (13.6)	104.5 (14.6)	31.5 (26.1, 38.0)
RS3 cases	1,103	54.2	55.3 (9.1)	63.7 (9.9)	107.5 (13.3)	103.0 (15.3)	31.2 (25.6, 38.0)
RS1 controls	2,037	51.2	na	58.7 (12.0)	97.6 (10.6)	85.4 (12.3)	26.3 (22.4, 30.9)
RS2 controls	1,750	48.9	na	31.6 (5.7)	92.0 (11.0)	na	25.0 (21.1, 29.5)
RS3 controls	1,559	52.6	na	59.0 (11.8)	98.2 (10.9)	87.5 (12.9)	26.6 (22.7, 31.2)

WTCCC controls came from two sources. No data on age, waist circumference or BMI are available for the UK Blood Service controls. Control individuals from the 1958 Birth Cohort were last reviewed at age 41. Under the terms of access, waist circumference and BMI values from these controls are not available to WTCCC researchers. Only 46% (all male) of the RS2 control individuals had available waist circumference measures. na: not available

Supplementary Table 1c. Characteristics of FUSION stage 1 and stage 2 case and control samples.

Study Sample	n	n male	n female	Age of diagnosis (years) (median, IQR)	Age at study (years) (median, IQR)	Body mass index (kg/m^2) (median, IQR)	Fasting plasma glucose (mM/l) (median, IQR)
Stage 1							
Cases	1,161	653	508	53.0, 12.0	63.4, 11.2	29.8, 6.1	8.4, 3.9
Controls	1,174	574	600	na	64.0, 11.7	26.8, 5.0	5.4, 0.7
Stage 2							
Cases	1,215	724	491	56.0, 12.0	60.0, 11.5	30.2, 6.6	7.2, 2.1*
Controls	1,258	768	490	na	59.0, 10.6	26.4, 4.9	5.4, 0.6**

IQR = interquartile range; na: not applicable

*n=204 and **n=583 values converted from whole blood to plasma glucose equivalent using prediction equation from the European Diabetes Epidemiology Group (201), of which **n=262 fasted <8 hours.

Supplementary Table 1d. Overview of sample characteristics.

Study /stage	n cases[#]	n controls[#]	Ethnic origin[*]	Study design overview
WTCCC /UK stage 1	1,924	2,938	UK	T2D cases, population-based controls
DGI /stage 1	1,464	1,467	Sweden/ Finland	T2D cases, normoglycaemic controls matched for BMI, sex and geographic origin; discordant siblings
FUSION /stage 1	1,161	1,174	Finland	T2D cases, normal glucose tolerant controls
RS1 /UK stage 2	2,022	2,037	UK	T2D cases, population-based controls
RS2 /UK stage 2	632	1,750	UK	
RS3 /UK stage 2	1,103	1,559	UK	
Sweden /DGI stage 2	2,830	3,550	Sweden	T2D cases, normoglycaemic controls
GCI US /DGI stage 2	1,226	1,226	US European-ancestry	T2D cases and controls matched by age, sex and grandparental origin
GCI Poland /DGI stage 2	1,009	1,009	Poland	T2D cases and controls matched by age and sex
FUSION /stage 2	1,215	1,258	Finland	T2D cases, normal glucose tolerant controls
deCODE /stage 3	1,520 (1,422)	25,235 (3,455)	Iceland	T2D cases, population-based controls
KORA /stage 3	1,241	1,458	Germany	T2D cases, non-diabetic age- and sex-matched controls
Danish /stage 3	4,089	5,043	Denmark	T2D cases, normal glucose-tolerant controls
HUNT /stage 3	1,004	1,503	Norway	T2D cases, non-diabetic controls
NHS /stage 3	1,506	2,014	US European-ancestry	T2D cases, non-diabetic sex-matched controls
CCC /stage 3	547	533	UK	T2D cases, non-diabetic age- and sex-matched controls
EPIC /stage 3	388	774	UK	T2D cases, non-diabetic controls matched for age, sex, recruitment date and general practice
ADDITION/Ely /stage 3	892	1,610	UK	T2D cases, non-diabetic controls
Norfolk /stage 3	2,311	2,400	UK	T2D cases, non-diabetic controls
METSIM /stage 3	659	2,639	Finnish	T2D cases, normal glucose tolerant controls

[#]Sample sizes presented here are the maximum available for each study. For the deCODE stage 3 study, we used direct genotype data from the Icelandic GWA scan[5] where available. SNPs that had not been directly typed as part of this scan

were genotyped separately, in a subset of the GWA scan samples (numbers indicated in parentheses) (Supplementary Methods). *UK ethnic origin denotes UK individuals of European ancestry.

Supplementary Table 2. Calculations for power attained by stage 1 samples (4,549 cases and 5,579 controls) to detect an association at p=0.05 (nominal significance), $p=10^{-4}$ (used as a cut-off for prioritising SNPs), $p=10^{-6}$ (assuming 1 false positive and 1M hypothesis tests across the genome) and $p=5x10^{-8}$ (assuming 0.05 false positives and 1M hypothesis tests across the genome), under an additive model, disease variant minor allele frequencies and genotype relative risks (effect sizes), assuming causal variant typed (or SNP with $r^2=1$ to causal variant typed) and a disease prevalence of 5%.

MAF	GRR	α	% power
0.10	1.10	0.05	54
0.10	1.15	0.05	86
0.10	1.20	0.05	98
0.20	1.10	0.05	78
0.20	1.15	0.05	98
0.20	1.20	0.05	>99
0.30	1.10	0.05	88
0.30	1.15	0.05	>99
0.30	1.20	0.05	>99
0.10	1.10	10^{-4}	<5
0.10	1.15	10^{-4}	20
0.10	1.20	10^{-4}	55
0.20	1.10	10^{-4}	13
0.20	1.15	10^{-4}	56
0.20	1.20	10^{-4}	92
0.30	1.10	10^{-4}	22
0.30	1.15	10^{-4}	76
0.30	1.20	10^{-4}	98
0.10	1.10	10^{-6}	<5
0.10	1.15	10^{-6}	<5
0.10	1.20	10^{-6}	19
0.20	1.10	10^{-6}	<5
0.20	1.15	10^{-6}	20
0.20	1.20	10^{-6}	65
0.30	1.10	10^{-6}	<5
0.30	1.15	10^{-6}	38
0.30	1.20	10^{-6}	87
0.10	1.10	$5x10^{-8}$	<5
0.10	1.15	$5x10^{-8}$	<5
0.10	1.20	$5x10^{-8}$	8
0.20	1.10	$5x10^{-8}$	<5
0.20	1.15	$5x10^{-8}$	8
0.20	1.20	$5x10^{-8}$	44
0.30	1.10	$5x10^{-8}$	<5
0.30	1.15	$5x10^{-8}$	20
0.30	1.20	$5x10^{-8}$	71

Supplementary Table 3. Most significantly associated SNPs for known T2D susceptibility loci in the meta-analysis of DGI, FUSION, and WTCCC genome scans (including directly typed and imputed SNPs)

					Stage 1 (DGI)				Stage 1 (FUSION)				Stage 1 (WTCCC)				All stage 1 data combined			
SNP	Chr	Position NCBI35 (bp)	Risk allele /nonrisk allele	Nearest genes	OR	95%CI	IC*	*P*	OR	95%CI	r^2hat*	*P*	OR	95%CI	IC*	*P*	OR	95%CI	*P*	P_{het}
rs13071168	3	12,250,447	G/A	*SYN2/PPARG*	1.23	(0.97-1.54)	0.99	0.04	1.53	(1.23-1.90)	0.99	1.5E-04	1.31	(1.09-1.58)	0.99	2.7E-03	1.35	(1.20-1.52)	**8.2E-07**	0.36
rs1801282	3	12,368,125	C/G	*PPARG*	1.02	(0.87-1.19)	-	0.96	1.29	(1.10-1.51)	0.99	1.7E-03	1.24	(1.09-1.41)	-	1.3E-03	1.18	(1.09-1.29)	**2.0E-04**	0.07
rs4402960	3	186,994,389	T/G	*IGF2BP2*	1.14	(1.01-1.28)	-	5.2E-03	1.27	(1.12-1.44)	0.99	4.3E-04	1.15	(1.05-1.25)	-	4.2E-03	1.17	(1.10-1.25)	**7.5E-08**	0.37
rs4580722	4	6,422,314	A/C	*WFS1*	1.17	(1.03-1.33)	0.93	0.03	1.11	(0.98-1.25)	0.91	0.11	1.15	(1.05-1.25)	0.95	0.04	1.11	(1.04-1.18)	**1.0E-03**	0.61
rs6931514	6	20,811,931	G/A	*CDKAL1*	1.25	(1.09-1.43)	0.92	1.4E-03	1.22	(1.08-1.38)	0.99	1.7E-03	1.08	(1.00-1.18)	0.94	1.0E-07	1.25	(1.17-1.33)	**1.3E-11**	0.87
rs10282940	8	118,257,007	A/G	*SLC30A8*	1.11	(0.91-1.36)	-	0.24	1.33	(1.02-1.74)	0.96	0.03	1.27	(1.16-1.39)	-	0.11	1.15	(1.03-1.27)	**6.1E-03**	0.48
rs7020996	9	22,119,579	C/T	*CDKN2A/CDKN2B*	1.48	(1.22-1.79)	0.71	1.1E-04	1.27	(1.05-1.55)	0.83	0.02	1.12	(0.97-1.28)	0.79	2.1E-03	1.26	(1.15-1.38)	**1.8E-07**	0.14
rs5015480	10	94,455,539	C/T	*HHEX*	1.15	(1.03-1.28)	-	0.01	1.13	(1.01-1.27)	0.99	0.05	1.21	(1.11-1.32)	-	2.9E-05	1.17	(1.11-1.24)	**7.2E-08**	0.57
rs7903146	10	114,748,339	T/C	*TCF7L2*	1.32	(1.15-1.52)	0.96	1.0E-07	1.39	(1.20-1.61)	-	3.8E-05	1.39	(1.27-1.51)	0.98	2.3E-13	1.37	(1.28-1.47)	**3.0E-23**	0.84
rs5215	11	17,365,206	C/T	*KCNJ11*	1.14	(1.03-1.28)	-	4.0E-03	1.19	(1.06-1.34)	-	6.5E-03	1.15	(1.06-1.25)	-	3.4E-03	1.16	(1.09-1.23)	**4.1E-07**	0.87
rs8050136	16	52,373,776	A/C	*FTO*	1.07	(0.95-1.19)	-	0.38	1.03	(0.92-1.16)	-	0.34	1.26	(1.16-1.37)	-	7.5E-08	1.15	(1.09-1.22)	**6.9E-06**	6.37E-03
rs17705177	17	33,197,639	A/T	*TCF2*	1.12	(1.00-1.25)	-	0.06	1.07	(0.95-1.2)	0.99	0.26	1.10	(1.01-1.19)	-	0.03	1.10	(1.04-1.16)	**2.8E-03**	0.85
Total sample size						2,931				2,335				4,862				10,128		
Cases/controls						1,464/1,467				1,161/1,174				1,924/2,938				4,549/5,579		

Combined estimates of OR were calculated using a fixed effects, inverse variance meta-analysis; DGI sibs were not included in OR estimates

*Information content (IC) and r^2hat are imputation accuracy measures

We observe evidence for association (p<0.001) at eight of the ten established T2D loci (as well as at the *FTO* obesity locus).

This is unsurprising, as these same data supported discovery of many of these loci. Power to detect the remaining signals may be low due to small effect sizes and / or poor correlation structure around the association signals.

Supplementary Table 4. Comparison of uncorrected p values from imputation based association analysis with p values from association analysis using directly genotyped data for each stage 1 study

indexSNP	nearest gene(s)	chr	position	non-risk allele	risk allele	DGI imputed p value	DGI directly typed p value	FUSION imputed p value	FUSION directly typed p value	WTCCC imputed p value	WTCCC directly typed p value	imputed_WTCCC	imputed_FUSION	imputed_DGI
rs319598	*C5orf14,DCOHM*	5	134268134	T	C	2.59E-02	4.92E-02			1.99E-03	2.51E-02	yes	no	yes
rs17036101	*SYN2/PPARG*	3	12252845	A	G	3.22E-02	7.89E-02	1.49E-04	1.26E-04	2.86E-03	1.35E-02	yes	yes	yes
rs12779790	*CAMK1D*	10	12368016	A	G	6.33E-02	2.37E-01	8.53E-03	1.86E-02	5.66E-03	7.95E-03	yes	yes	yes
rs7578597	*THADA*	2	43644474	C	T	4.41E-02	1.56E-01			2.16E-02	1.12E-02	yes	no	yes
rs864745	*JAZF1*	7	27953796	C	T			1.75E-02	7.37E-02			no	yes	no
rs10490072	*BCL11A*	2	60581582	C	T	1.02E-02	6.20E-03	3.43E-02	8.92E-03			no	yes	yes
rs1153188	*DCD*	12	53385263	T	A			1.12E-01	1.39E-01			no	yes	no
rs4607103	*ADAMTS9*	3	64686944	T	C			4.93E-01	6.40E-01			no	yes	no
rs10923931	*NOTCH2*	1	120230001	G	T			9.01E-01	7.73E-01			no	yes	no
rs2641348	*ADAM30*	1	120149926	A	G	2.20E-06	2.37E-05					yes	no	yes
rs12137794	*DNAJC11*	1	6640210	T	C	1.30E-04	3.72E-04					yes	no	yes
rs431722	*NOTCH4*	6	32295700	C	T	2.93E-04	2.78E-05					yes	yes	yes
rs11072447	*MGC34741*	15	71953508	C	T	3.23E-04	1.60E-03					yes	yes	yes
rs1452075	*CADPS*	3	62456103	C	T	2.81E-03	1.02E-04					yes	no	yes
rs1534544	*ALK*	2	29508139	G	T	5.35E-03	2.09E-02					yes	no	yes
rs13088	*C10orf72*	10	49985899	A	G	9.34E-03	1.77E-02					yes	no	yes
rs1232597		20	10553631	T	G	1.48E-02	1.28E-02					yes	no	yes
rs2485597	*RYR2*	1	233599325	G	A	2.13E-02	3.93E-02					yes	yes	yes
rs8072774	*C17orf27*	17	75977649	G	A	2.68E-02	1.30E-02					yes	no	yes
rs6450472		5	57557248	T	C	3.69E-02	7.26E-03					yes	yes	yes
rs12658264		5	141744373	A	G	4.43E-02	9.27E-01					yes	yes	yes
rs9583036		13	105187386	G	T	5.42E-02	2.01E-02					no	yes	yes
rs8049156		16	25370548	A	C	6.00E-02	1.26E-01					yes	no	yes
rs7094128	*LRRC20*	10	71728540	C	T	8.35E-02	7.86E-02					yes	yes	yes
rs16896390	*ANKS1*	6	35090577	T	C	1.43E-01	3.69E-02					yes	yes	yes
rs703698		12	95492143	T	C	2.78E-01	2.40E-01					yes	yes	yes
rs16894945	*C6orf107*	6	34932085	C	A	2.91E-01	6.50E-02					yes	yes	yes
rs12332927		6	28063094	T	C	3.18E-01	3.92E-01					yes	yes	yes
rs1114702	*PKP2*	12	32853786	G	T	3.25E-01	1.97E-01					yes	yes	yes
rs2789686	*ANXA11*	10	81905116	C	T	4.28E-01	2.22E-01					yes	no	yes
rs4604170		5	104384351	C	T	6.36E-01	9.05E-01					yes	yes	yes
rs2741200	*TG,SLA*	8	134141015	T	C	6.37E-01	9.38E-01					yes	no	yes
rs6650596		17	13176800	A	T	8.47E-01	3.99E-01					yes	yes	yes
rs12205899		6	118806621	C	G	9.33E-01	5.09E-01					yes	yes	yes
rs11647813		16	12613037	G	C	9.71E-01	5.04E-01					yes	yes	yes
rs7610589		3	41349377	C	T	9.77E-01	8.60E-01					yes	yes	yes

Supplementary Table 6. Details of meta-analyses (p value-based GC corrected and uncorrected, fixed effects inverse variance, and random effects) of 11 SNPs across stages 1, 2 and 3.

indexSNP	nearest gene(s)	chr	position	non-risk allele	risk allele	n eff all stages	p meta all stages GC p value	p meta all stages uncorrected p value	p meta stage 1 uncorrected p value	odds ratio all stages (95% CI)	fixed effects all stages p value	p heterogeneity	% I^2 (95%CI)	random effects OR (95%CI)	random effects all stages p value
rs864745	*JAZF1*	7	27,953,796	C	T	59,617	5.0E-14	3.3E-14	9.2E-05	1.10 (1.07-1.13)	2.43E-14	0.703	0 (0 - 46.4)	1.10 (1.07-1.13)	2.43E-14
rs12779790	*CAMK1D/CDC23*	10	12,368,016	A	G	62,366	1.2E-10	9.2E-11	2.8E-04	1.11 (1.07-1.14)	7.81E-11	0.6669	0 (0 - 45.4)	1.11 (1.07-1.14)	7.81E-11
rs7961581	*TSPAN8/LGR5*	12	69,949,369	T	C	62,301	1.1E-09	7.5E-10	8.7E-06	1.09 (1.06-1.12)	1.63E-09	0.1963	23.2 (0 - 58.2)	1.09 (1.05-1.13)	3.22E-07
rs7578597	*THADA*	2	43,644,474	C	T	60,832	1.1E-09	8.6E-10	1.2E-04	1.15 (1.10-1.20)	1.82E-10	0.0079	53.2 (0 - 72.6)	1.16 (1.08-1.24)	2.58E-05
rs4607103	*ADAMTS9*	3	64,686,944	T	C	62,387	1.2E-08	8.9E-09	3.5E-04	1.09 (1.06-1.12)	2.35E-09	0.1695	25.2 (0 - 58.4)	1.09 (1.05-1.12)	1.45E-06
rs10923931	*NOTCH2*	1	120,230,001	G	T	58,667	4.1E-08	3.1E-08	6.9E-05	1.13 (1.08-1.17)	1.49E-09	0.0036	57.8 (8.7 - 75.3	1.14 (1.07-1.21)	7.09E-05
rs1153188	*DCD*	12	53,385,263	T	A	62,301	1.8E-07	1.3E-07	1.8E-05	1.08 (1.05-1.11)	4.00E-08	0.7875	0 (0 - 45.4)	1.08 (1.05-1.11)	4.00E-08
rs17036101	*SYN2/PPARG*	3	12,252,845	A	G	59,682	2.0E-07	1.5E-07	5.9E-06	1.15 (1.10-1.21)	3.95E-08	0.1895	24.5 (0 - 59.6)	1.15 (1.09-1.23)	5.41E-06
rs2641348	*ADAM30*	1	120,149,926	A	G	60,048	4.0E-07	3.2E-07	9.3E-04	1.10 (1.06-1.15)	2.53E-07	0.0818	35.9 (0 - 64.3)	1.10 (1.05-1.16)	1.15E-04
rs9472138	*VEGFA*	6	43,919,740	C	T	63,537	4.0E-06	3.1E-06	3.0E-05	1.06 (1.04-1.09)	4.48E-06	0.4291	2 (0 - 46.5)	1.06 (1.04-1.09)	6.53E-06
rs10490072	*BCL11A*	2	60,581,582	C	T	59,682	1.0E-04	5.4E-05	6.0E-06	1.05 (1.03-1.08)	9.22E-05	0.0035	56.8 (9.1 - 74.4)	1.05 (1.00-1.10)	3.13E-02

Supplementary Table 7. Conditional analysis of *SYN2*/*PPARG* SNPs rs17036101 and rs1801282 in stage 1 & 2 samples from UK, FUSION and DGI.

rs17036101 (near *SYN2*/*PPARG*)					
Sample	**n**	**odds ratio (95% CI)**	**p value**	**odds ratio conditional on rs1801282**	**p value conditional on rs1801282**
UK stage 1	4372	1.28 (1.05-1.55)	0.013	1.11 (0.87-1.42)	0.41
UK stage 2	7603	1.05 (0.93-1.19)	0.45	0.89 (0.74-1.07)	0.21
FUSION stage 1	2335	1.51 (1.22-1.87)	1.5E-04	1.41 (1.07-1.86)	0.02
FUSION stage 2	2473	1.06 (0.87-1.30)	0.56	1.02 (0.77-1.34)	0.89
DGI stage 1*	2504	1.16 (0.94-1.42)	0.17	1.29 (0.97-1.71)	0.08
DGI stage 2	10147	1.20 (1.07-1.33)	1.2E-03	1.16 (1.00-1.33)	0.04
Stage 1 & 2	24626	1.17 (1.10-1.25)	1.08E-06	1.11 (1.03-1.20)	0.022
rs1801282 (*PPARG* P12A)					
Sample	**n**	**odds ratio (95% CI)**	**p value**	**odds ratio conditional on rs17036101**	**p value conditional on rs17036101**
UK stage 1	4372	1.24 (1.09-1.41)	1.3E-03	1.17 (0.98-1.41)	0.08
UK stage 2	7603	1.16 (1.06-1.29)	2.5E-03	1.23 (1.07-1.40)	3.0E-03
FUSION stage 1	2335	1.30 (1.11-1.53)	1.1E-03	1.09 (0.88-2.34)	0.43
FUSION stage 2	2473	1.07 (0.92-1.25)	0.37	1.05 (0.85-1.29)	0.64
DGI stage 1*	2504	1.02 (0.86-1.20)	0.85	0.89 (0.71-1.11)	0.29
DGI stage 2	10147	1.11 (1.03-1.20)	8.5E-03	1.04 (0.94-1.15)	0.45
Stage 1 & 2	24626	1.14 (1.09-1.20)	3.19E-08	1.09 (1.02-1.16)	0.013

*OR calculations do not include the DGI sibship component

Supplementary Table 8. Association of 11 SNPs with BMI in cases and controls in stage 1 and stage 2 samples

					Controls									Cases								
indexSNP	**nearest gene(s)**	**chr**	**position**	**T2D risk allele**	**n_eff stages 1 and 2**	**p_meta**	**dir**	**n_eff_stage_1**	**p_stage_1**	**dir**	**n_eff_stage_2**	**p_stage_2**	**dir**	**n_eff stages 1 and 2**	**p_meta**	**dir**	**n_eff_stage_1**	**p_stage_1**	**dir**	**n_eff_stage_2**	**p_stage_2**	**dir**
rs864745	*JAZF1*	7	27,953,796	T	14241	0.834	+	4147	0.250	-	10094	0.324	+	13817	0.993	+	4493	0.778	+	9324	0.853	-
rs12779790	*CAMK1D*	10	12,368,016	G	15856	0.485	+	4102	0.070	+	11754	0.795	-	13421	0.057	-	4099	0.055	-	9322	0.314	-
rs7961581	*TSPAN8/LGR5*	12	69,949,369	C	14246	0.795	+	4173	0.582	+	10073	0.964	-	13864	0.535	-	4514	0.107	-	9350	0.716	+
rs7578597	*THADA*	2	43,644,474	T	15890	0.730	-	4163	0.917	+	11727	0.643	-	13487	0.144	-	4155	0.352	-	9332	0.257	-
rs4607103	*ADAMTS9*	3	64,686,944	C	15829	0.765	-	4147	0.546	-	11682	0.990	+	13804	0.164	-	4495	0.691	-	9309	0.156	-
rs10923931	*NOTCH2*	1	120,230,001	T	10814	0.032	+	4130	0.169	+	6684	0.099	+	10355	1.63E-03	-	4491	0.008	-	5864	0.061	-
rs1153188	*DCD*	12	53,385,263	A	14270	0.835	-	4174	0.939	-	10096	0.843	-	13852	0.796	+	4515	0.286	+	9337	0.670	-
rs17036101	*SYN2/PPARG*	3	12,252,845	G	15918	0.595	-	4135	0.693	+	11783	0.394	-	13494	0.708	-	4118	0.289	-	9376	0.799	+
rs2641348	*ADAM30*	1	120,149,926	G	14563	0.045	+	2767	0.021	+	11796	0.267	+	11974	0.003	-	2619	0.022	-	9355	0.033	-
rs9472138	*VEGFA*	6	43,919,740	T	14169	0.690	+	4172	0.698	-	9997	0.468	+	13811	0.433	-	4514	0.045	-	9297	0.658	+
rs10490072	*BCL11A*	2	60,581,582	T	15868	0.074	-	4143	0.384	-	11725	0.118	-	13756	0.020	-	4476	0.274	-	9280	0.038	-

Analysis was adjusted for age and gender in DGI and WTCCC stage 1, and for age, age^2, gender, birth province, and study indicator in FUSION.

n_eff denotes effective sample size

dir denotes direction of effect relative to the T2D risk allele

Supplementary Table 9. DGI and FUSION quantitative trait association results for 11 SNPs

							All			FUSION			DGI		
Trait	**SNP**	**nearest gene(s)**	**chr**	**pos**	**non risk allele**	**T2D risk allele**	**N**	**p_meta**	**dir**	**n**	**p value**	**dir**	**n**	**p value**	**dir**
Fasting glucose															
	rs864745	*JAZF1*	7	27953796	C	T	6543	0.10	-	1840	0.036	-	4703	0.517	-
	rs12779790	*CAMK1D/CDC23*	10	12368016	A	G	6535	0.26	+	1820	0.796	+	4715	0.239	+
	rs7961581	*TSPAN8/LGR5*	12	69949369	T	C	6586	0.014	+	1878	0.333	+	4708	0.021	+
	rs7578597	*THADA*	2	43644474	C	T	6599	0.89	+	1889	0.426	+	4710	0.728	-
	rs4607103	*ADAMTS9*	3	64686944	T	C	6588	0.099	+	1868	0.633	+	4720	0.099	+
	rs10923931	*NOTCH2*	1	120230001	G	T	6560	0.78	-	1846	0.725	+	4714	0.58	-
	rs1153188	*DCD*	12	53385263	T	A	6604	0.39	+	1905	0.764	+	4699	0.41	+
	rs17036101	*SYN2/PPARG*	3	12252845	A	G	6588	0.025	+	1882	0.152	+	4706	0.080	+
	rs2641348	*ADAM30*	1	120149926	A	G	6601	0.72	-	1889	0.490	+	4712	0.39	-
	rs9472138	*VEGFA*	6	43919740	C	T	6543	0.028	+	1905	0.244	+	4638	0.062	+
	rs10490072	*BCL11A*	2	60581582	C	T	6570	0.53	-	1864	0.618	+	4706	0.29	-
Fasting insulin															
	rs864745	*JAZF1*	7	27953796	C	T	5748	0.43	+	1209	0.887	-	4539	0.332	+
	rs12779790	*CAMK1D/CDC23*	10	12368016	A	G	5732	0.87	-	1181	0.423	+	4551	0.552	-
	rs7961581	*TSPAN8/LGR5*	12	69949369	T	C	5779	0.85	+	1234	0.347	-	4545	0.479	+
	rs7578597	*THADA*	2	43644474	C	T	5780	0.83	-	1234	0.726	+	4546	0.667	-
	rs4607103	*ADAMTS9*	3	64686944	T	C	5768	0.39	+	1212	0.178	-	4556	0.096	+
	rs10923931	*NOTCH2*	1	120230001	G	T	5740	0.64	+	1190	0.515	-	4550	0.390	+
	rs1153188	*DCD*	12	53385263	T	A	5769	0.60	+	1234	0.575	+	4535	0.763	+
	rs17036101	*SYN2/PPARG*	3	12252845	A	G	5756	0.41	+	1213	0.681	-	4543	0.252	+
	rs2641348	*ADAM30*	1	120149926	A	G	5782	0.65	+	1234	0.588	-	4548	0.431	+
	rs9472138	*VEGFA*	6	43919740	C	T	5710	0.43	+	1234	0.778	+	4476	0.462	+
	rs10490072	*BCL11A*	2	60581582	C	T	5755	0.17	-	1212	0.075	-	4543	0.542	-
HOMA-IR															
	rs864745	*JAZF1*	7	27953796	C	T	5751	0.45	+	1207	0.688	-	4544	0.294	+
	rs12779790	*CAMK1D/CDC23*	10	12368016	A	G	5735	0.98	+	1179	0.408	+	4556	0.697	-
	rs7961581	*TSPAN8/LGR5*	12	69949369	T	C	5782	0.43	+	1232	0.373	-	4550	0.174	+
	rs7578597	*THADA*	2	43644474	C	T	5783	0.66	-	1232	0.780	+	4551	0.526	-

	rs4607103	*ADAMTS9*	3	64686944	T	C	5771	0.14	+	1210	0.249	-	4561	0.025	+
	rs10923931	*NOTCH2*	1	120230001	G	T	5743	0.98	-	1188	0.527	-	4555	0.764	+
	rs1153188	*DCD*	12	53385263	T	A	5772	0.37	+	1232	0.535	+	4540	0.485	+
	rs17036101	*SYN2/PPARG*	3	12252845	A	G	5759	0.31	+	1211	0.829	-	4548	0.211	+
	rs2641348	*ADAM30*	1	120149926	A	G	5785	0.92	-	1232	0.640	-	4553	0.894	+
	rs9472138	*VEGFA*	6	43919740	C	T	5713	0.27	+	1232	0.619	+	4481	0.327	+
	rs10490072	*BCL11A*	2	60581582	C	T	5758	0.13	-	1210	0.070	-	4548	0.440	-
2 hour glucose															
	rs864745	*JAZF1*	7	27953796	C	T	3277	0.34	+	1840	0.484	+	1437	0.524	+
	rs12779790	*CAMK1D/CDC23*	10	12368016	A	G	3257	0.93	-	1820	0.838	+	1437	0.717	-
	rs7961581	*TSPAN8/LGR5*	12	69949369	T	C	3315	0.19	+	1878	0.609	+	1437	0.162	+
	rs7578597	*THADA*	2	43644474	C	T	3326	0.59	+	1889	0.404	+	1437	0.890	-
	rs4607103	*ADAMTS9*	3	64686944	T	C	3305	0.58	-	1868	0.298	-	1437	0.735	+
	rs10923931	*NOTCH2*	1	120230001	G	T	3283	0.86	+	1846	0.958	+	1437	0.833	+
	rs1153188	*DCD*	12	53385263	T	A	3342	0.10	+	1905	0.128	+	1437	0.435	+
	rs17036101	*SYN2/PPARG*	3	12252845	A	G	3319	0.12	+	1882	0.186	+	1437	0.405	+
	rs2641348	*ADAM30*	1	120149926	A	G	3326	0.94	-	1889	0.865	-	1437	0.932	+
	rs9472138	*VEGFA*	6	43919740	C	T	3342	0.57	+	1905	0.543	+	1437	0.861	+
	rs10490072	*BCL11A*	2	60581582	C	T	3301	0.11	-	1864	0.072	-	1437	0.697	-
2 hour insulin															
	rs864745	*JAZF1*	7	27953796	C	T	1598	0.08	+	567	0.150	+	1031	0.260	+
	rs12779790	*CAMK1D/CDC23*	10	12368016	A	G	1582	0.96	+	551	0.170	+	1031	0.348	-
	rs7961581	*TSPAN8/LGR5*	12	69949369	T	C	1612	0.10	+	581	0.645	+	1031	0.091	+
	rs7578597	*THADA*	2	43644474	C	T	1612	0.83	+	581	0.587	+	1031	0.886	-
	rs4607103	*ADAMTS9*	3	64686944	T	C	1600	0.34	-	569	0.938	+	1031	0.216	-
	rs10923931	*NOTCH2*	1	120230001	G	T	1592	0.88	-	561	0.681	-	1031	0.911	+
	rs1153188	*DCD*	12	53385263	T	A	1612	0.98	+	581	0.145	-	1031	0.260	+
	rs17036101	*SYN2/PPARG*	3	12252845	A	G	1600	0.78	+	569	0.503	-	1031	0.397	+
	rs2641348	*ADAM30*	1	120149926	A	G	1612	0.97	+	581	0.533	-	1031	0.608	+
	rs9472138	*VEGFA*	6	43919740	C	T	1612	0.36	+	581	0.867	-	1031	0.205	+
	rs10490072	*BCL11A*	2	60581582	C	T	1600	8.84E-03	-	569	0.038	-	1031	0.085	-
Total Cholesterol															
	rs864745	*JAZF1*	7	27953796	C	T	7195	0.61	+	3945	0.279	+	3250	0.673	-
	rs12779790	*CAMK1D/CDC23*	10	12368016	A	G	7173	0.96	-	3911	0.966	+	3262	0.903	-
	rs7961581	*TSPAN8/LGR5*	12	69949369	T	C	7260	0.14	+	4005	0.277	+	3255	0.308	+
	rs7578597	*THADA*	2	43644474	C	T	7279	0.12	+	4022	0.460	+	3257	0.135	+

	rs4607103	*ADAMTS9*	3	64686944	T	C	7255	0.95	+	3988	0.936	-	3267	0.849	+
	rs10923931	*NOTCH2*	1	120230001	G	T	7231	0.23	+	3970	0.085	+	3261	0.912	-
	rs1153188	*DCD*	12	53385263	T	A	7316	1.82E-03	-	4070	9.98E-04	-	3246	0.319	-
	rs17036101	*SYN2/PPARG*	3	12252845	A	G	7282	0.26	+	4029	0.015	+	3253	0.308	-
	rs2641348	*ADAM30*	1	120149926	A	G	7278	0.49	+	4019	0.215	+	3259	0.734	-
	rs9472138	*VEGFA*	6	43919740	C	T	7257	0.34	-	4072	0.337	-	3185	0.727	-
	rs10490072	*BCL11A*	2	60581582	C	T	7206	0.59	+	3953	0.323	+	3253	0.782	-
HDL cholesterol															
	rs864745	*JAZF1*	7	27953796	C	T	6541	0.43	+	3943	0.509	+	2598	0.656	+
	rs12779790	*CAMK1D/CDC23*	10	12368016	A	G	6507	0.92	+	3909	0.893	+	2598	0.997	-
	rs7961581	*TSPAN8/LGR5*	12	69949369	T	C	6601	0.89	-	4003	0.442	-	2598	0.462	+
	rs7578597	*THADA*	2	43644474	C	T	6618	0.16	+	4020	0.393	+	2598	0.246	+
	rs4607103	*ADAMTS9*	3	64686944	T	C	6584	0.22	-	3986	0.857	-	2598	0.086	-
	rs10923931	*NOTCH2*	1	120230001	G	T	6566	0.17	+	3968	0.141	+	2598	0.706	+
	rs1153188	*DCD*	12	53385263	T	A	6666	0.82	-	4068	0.934	-	2598	0.796	-
	rs17036101	*SYN2/PPARG*	3	12252845	A	G	6625	0.61	+	4027	0.765	+	2598	0.662	+
	rs2641348	*ADAM30*	1	120149926	A	G	6615	0.41	+	4017	0.278	+	2598	0.962	-
	rs9472138	*VEGFA*	6	43919740	C	T	6668	0.093	+	4070	0.081	+	2598	0.613	+
	rs10490072	*BCL11A*	2	60581582	C	T	6549	0.46	+	3951	0.407	+	2598	0.885	+
LDL cholesterol															
	rs864745	*JAZF1*	7	27953796	C	T	5972	0.058	+	3404	0.167	+	2568	0.194	+
	rs12779790	*CAMK1D/CDC23*	10	12368016	A	G	5943	0.84	+	3375	0.632	-	2568	0.801	+
	rs7961581	*TSPAN8/LGR5*	12	69949369	T	C	6036	0.66	+	3468	0.621	+	2568	0.921	+
	rs7578597	*THADA*	2	43644474	C	T	6048	0.19	+	3480	0.265	+	2568	0.465	+
	rs4607103	*ADAMTS9*	3	64686944	T	C	6017	0.41	-	3449	0.833	-	2568	0.312	-
	rs10923931	*NOTCH2*	1	120230001	G	T	5997	0.53	+	3429	0.385	+	2568	0.961	-
	rs1153188	*DCD*	12	53385263	T	A	6088	0.012	-	3520	2.56E-03	-	2568	0.751	-
	rs17036101	*SYN2/PPARG*	3	12252845	A	G	6045	0.16	+	3477	0.054	+	2568	0.920	-
	rs2641348	*ADAM30*	1	120149926	A	G	6049	0.48	+	3481	0.426	+	2568	0.884	+
	rs9472138	*VEGFA*	6	43919740	C	T	6089	0.057	-	3521	0.201	-	2568	0.153	-
	rs10490072	*BCL11A*	2	60581582	C	T	5992	0.55	+	3424	0.621	+	2568	0.732	+
Triglycerides															
	rs864745	*JAZF1*	7	27953796	C	T	9340	0.74	-	3470	0.516	-	5870	0.930	+
	rs12779790	*CAMK1D/CDC23*	10	12368016	A	G	9324	0.11	+	3442	0.935	+	5882	0.048	+
	rs7961581	*TSPAN8/LGR5*	12	69949369	T	C	9411	0.55	+	3536	0.141	+	5875	0.699	-
	rs7578597	*THADA*	2	43644474	C	T	9426	0.17	-	3549	0.624	-	5877	0.179	-

rs4607103	*ADAMTS9*	3	64686944	T	C	9403	0.15	+	3516	0.882	-	5887	0.055	+
rs10923931	*NOTCH2*	1	120230001	G	T	9378	0.68	+	3497	0.476	+	5881	0.970	-
rs1153188	*DCD*	12	53385263	T	A	9455	0.82	-	3589	0.112	-	5866	0.339	+
rs17036101	*SYN2/PPARG*	3	12252845	A	G	9418	0.54	+	3545	0.507	+	5873	0.801	+
rs2641348	*ADAM30*	1	120149926	A	G	9429	0.54	+	3550	0.228	+	5879	0.871	-
rs9472138	*VEGFA*	6	43919740	C	T	9395	0.19	-	3590	0.772	-	5805	0.150	-
rs10490072	*BCL11A*	2	60581582	C	T	9365	0.70	-	3492	0.221	+	5873	0.150	-
Systolic Blood Pressure														
rs864745	*JAZF1*	7	27953796	C	T	11753	0.39	+	5610	0.288	+	6143	0.868	+
rs12779790	*CAMK1D/CDC23*	10	12368016	A	G	11701	0.31	-	5546	0.762	-	6155	0.272	-
rs7961581	*TSPAN8/LGR5*	12	69949369	T	C	10715	0.39	+	4567	0.689	-	6148	0.137	+
rs7578597	*THADA*	2	43644474	C	T	11876	0.04	+	5726	2.77E-03	+	6150	0.994	+
rs4607103	*ADAMTS9*	3	64686944	T	C	11821	0.46	-	5661	0.671	-	6160	0.531	-
rs10923931	*NOTCH2*	1	120230001	G	T	11789	1.00	-	5635	0.374	+	6154	0.391	-
rs1153188	*DCD*	12	53385263	T	A	11918	0.30	+	5779	0.279	+	6139	0.685	+
rs17036101	*SYN2/PPARG*	3	12252845	A	G	11848	0.66	-	5702	0.812	+	6146	0.400	-
rs2641348	*ADAM30*	1	120149926	A	G	11873	0.70	-	5721	0.589	+	6152	0.290	-
rs9472138	*VEGFA*	6	43919740	C	T	11861	0.21	+	5783	0.609	-	6078	0.025	+
rs10490072	*BCL11A*	2	60581582	C	T	11760	0.80	+	5614	0.614	+	6146	0.891	-
Diastolic Blood Pressure														
rs864745	*JAZF1*	7	27953796	C	T	8989	0.46	+	4494	0.528	+	4495	0.685	+
rs12779790	*CAMK1D/CDC23*	10	12368016	A	G	8963	0.48	-	4456	0.980	-	4507	0.327	-
rs7961581	*TSPAN8/LGR5*	12	69949369	T	C	9067	0.49	+	4567	0.136	+	4500	0.597	-
rs7578597	*THADA*	2	43644474	C	T	9083	0.013	+	4581	0.065	+	4502	0.093	+
rs4607103	*ADAMTS9*	3	64686944	T	C	9053	0.74	-	4541	0.777	-	4512	0.855	-
rs10923931	*NOTCH2*	1	120230001	G	T	9028	0.93	+	4522	0.967	-	4506	0.873	+
rs1153188	*DCD*	12	53385263	T	A	9128	0.99	-	4637	0.706	+	4491	0.692	-
rs17036101	*SYN2/PPARG*	3	12252845	A	G	9082	0.96	+	4584	0.899	+	4498	0.955	-
rs2641348	*ADAM30*	1	120149926	A	G	9080	0.62	-	4576	0.569	-	4504	0.895	-
rs9472138	*VEGFA*	6	43919740	C	T	9068	0.44	+	4638	0.927	-	4430	0.233	+
rs10490072	*BCL11A*	2	60581582	C	T	8996	0.39	-	4498	0.812	-	4498	0.322	-

P values were combined using a weighted z score-based meta-analysis of FUSION stage1&2 and DGI stage 1&2 results as available; dir denotes direction of effect relative to the T2D risk allele
Fasting glucose, fasting insulin, HOMA-IR, 2 hr glucose and 2 hr insulin association results were combined for non diabetic individuals only

Supplementary Table 10. Case-control analysis adjusted for BMI, age and gender.

Meta-analysis of DGI, FUSION and UK Stage 1&2

						Age, gender adjusted		BMI, age, gender adjusted	
SNP	**nearest gene(s)**	**chr**	**pos**	**non risk allele**	**risk allele**	**OR (95% CI)**	**p value**	**OR (95% CI)**	**p value**
rs864745	*JAZF1*	7	27953796	C	T	1.09 (1.05-1.13)	2.88E-06	1.10 (1.06-1.14)	3.21E-06
rs12779790	*CAMK1D*	10	12368016	A	G	1.12 (1.07-1.17)	9.19E-07	1.14 (1.09-1.20)	4.70E-07
rs4607103	*ADAMTS9*	3	64686944	T	C	1.12 (1.08-1.17)	8.09E-08	1.13 (1.08-1.19)	3.37E-07
rs7578597	*THADA*	2	43644474	C	T	1.16 (1.09-1.24)	3.09E-06	1.18 (1.11-1.27)	1.88E-07
rs7961581	*TSPAN8/LGR5*	12	69949369	T	C	1.10 (1.05-1.15)	1.35E-05	1.10 (1.05-1.15)	1.38E-04
rs10923931*	*NOTCH2*	1	120230001	G	T	1.15 (1.08-1.22)	3.03E-05	1.16 (1.09-1.25)	7.83E-06
rs2641348	*ADAM30*	1	120149926	A	G	1.13 (1.07-1.19)	2.07E-05	1.15 (1.08-1.22)	1.76E-06
rs1153188	*DCD*	12	53385263	T	A	1.09 (1.04-1.13)	3.23E-05	1.08 (1.03-1.12)	1.09E-03
rs10490072	*BCL11A*	2	60581582	C	T	1.10 (1.06-1.14)	1.21E-06	1.12 (1.07-1.16)	9.54E-08
rs17036101	*SYN2/PPARG*	3	12252845	A	G	1.23 (1.14-1.33)	8.07E-08	1.21 (1.12-1.32)	7.50E-06
rs9472138	*VEGFA*	6	43919740	C	T	1.08 (1.03-1.12)	3.89E-04	1.08 (1.04-1.13)	2.58E-04

DGI Stage 1&2

						Age, gender adjusted		BMI, age, gender adjusted	
SNP	**nearest gene(s)**	**chr**	**pos**	**non risk allele**	**risk allele**	**OR (95% CI)**	**p value**	**OR (95% CI)**	**p value**
rs864745	*JAZF1*	7	27953796	C	T	1.12 (1.06-1.18)	1.63E-05	1.13 (1.07-1.20)	3.36E-05
rs12779790	*CAMK1D*	10	12368016	A	G	1.06 (1.00-1.14)	6.47E-02	1.08 (1.01-1.16)	3.29E-02
rs4607103	*ADAMTS9*	3	64686944	T	C	1.15 (1.09-1.23)	3.73E-06	1.17 (1.10-1.25)	3.32E-06
rs7578597	*THADA*	2	43644474	C	T	1.18 (1.08-1.29)	2.10E-04	1.23 (1.11-1.35)	5.52E-05
rs7961581	*TSPAN8/LGR5*	12	69949369	T	C	1.07 (1.01-1.13)	2.38E-02	1.07 (1.01-1.15)	3.13E-02
rs10923931*	*NOTCH2*	1	120230001	G	T	1.16 (1.07-1.26)	6.05E-04	1.18 (1.08-1.30)	3.14E-04
rs2641348	*ADAM30*	1	120149926	A	G	1.17 (1.08-1.28)	1.81E-04	1.20 (1.10-1.32)	9.42E-05
rs1153188	*DCD*	12	53385263	T	A	1.09 (1.03-1.15)	5.40E-03	1.07 (1.00-1.15)	3.64E-02
rs10490072	*BCL11A*	2	60581582	C	T	1.08 (1.02-1.15)	8.20E-03	1.10 (1.03-1.17)	2.76E-03
rs17036101	*SYN2/PPARG*	3	12252845	A	G	1.23 (1.11-1.36)	8.28E-05	1.21 (1.08-1.35)	9.51E-04
rs9472138	*VEGFA*	6	43919740	C	T	1.07 (1.01-1.14)	1.59E-02	1.07 (1.01-1.14)	2.78E-02

FUSION Stage 1&2

SNP	nearest gene(s)	chr	pos	non risk allele	risk allele	Age, gender adjusted OR (95% CI)	Age, gender adjusted p value	BMI, age, gender adjusted OR (95% CI)	BMI, age, gender adjusted p value
rs864745	*JAZF1*	7	27953796	C	T	1.08 (1.00-1.17)	5.71E-02	1.10 (1.00-1.20)	3.90E-02
rs12779790	*CAMK1D*	10	12368016	A	G	1.20 (1.08-1.32)	5.41E-04	1.21 (1.08-1.35)	7.75E-04
rs4607103	*ADAMTS9*	3	64686944	T	C	1.08 (0.98-1.18)	1.25E-01	1.09 (0.98-1.21)	1.13E-01
rs7578597	*THADA*	2	43644474	C	T	1.12 (0.92-1.35)	2.66E-01	1.14 (0.93-1.41)	2.15E-01
rs7961581	*TSPAN8/LGR5*	12	69949369	T	C	1.05 (0.97-1.14)	2.19E-01	1.04 (0.95-1.14)	3.71E-01
rs10923931*	*NOTCH2*	1	120230001	G	T	1.05 (0.93-1.18)	4.05E-01	1.07 (0.95-1.22)	2.70E-01
rs2641348	*ADAM30*	1	120149926	A	G	1.07 (0.95-1.20)	2.57E-01	1.09 (0.97-1.24)	1.58E-01
rs1153188	*DCD*	12	53385263	T	A	1.11 (1.01-1.21)	2.69E-02	1.09 (0.99-1.21)	6.94E-02
rs10490072	*BCL11A*	2	60581582	C	T	1.12 (1.02-1.23)	1.81E-02	1.15 (1.03-1.27)	9.10E-03
rs17036101	*SYN2/PPARG*	3	12252845	A	G	1.25 (1.08-1.44)	2.74E-03	1.23 (1.05-1.44)	9.56E-03
rs9472138	*VEGFA*	6	43919740	C	T	1.05 (0.97-1.15)	2.36E-01	1.06 (0.97-1.16)	2.24E-01

UK Stage 1 & 2

SNP	nearest gene(s)	chr	pos	non risk allele	risk allele	Age, gender adjusted OR (95% CI)	Age, gender adjusted p value	BMI, age, gender adjusted OR (95% CI)	BMI, age, gender adjusted p value
rs864745	*JAZF1*	7	27953796	C	T	1.06 (1.00-1.13)	4.72E-02	1.04 (0.97-1.12)	2.35E-01
rs12779790	*CAMK1D*	10	12368016	A	G	1.24 (1.12-1.38)	4.15E-05	1.27 (1.13-1.42)	7.68E-05
rs4607103	*ADAMTS9*	3	64686944	T	C	1.08 (0.98-1.19)	1.11E-01	1.07 (0.96-1.19)	2.36E-01
rs7578597	*THADA*	2	43644474	C	T	1.15 (1.04-1.28)	7.60E-03	1.21 (1.08-1.36)	1.28E-03
rs7961581	*TSPAN8/LGR5*	12	69949369	T	C	1.21 (1.11-1.33)	2.74E-05	1.22 (1.10-1.35)	1.46E-04
rs10923931*	*NOTCH2*	1	120230001	G	T	1.28 (1.07-1.53)	7.10E-03	1.39 (1.13-1.72)	1.77E-03
rs2641348	*ADAM30*	1	120149926	A	G	1.11 (1.00-1.24)	5.81E-02	1.16 (1.02-1.32)	1.93E-02
rs1153188	*DCD*	12	53385263	T	A	1.07 (1.00-1.15)	5.90E-02	1.08 (1.00-1.17)	5.57E-02
rs10490072	*BCL11A*	2	60581582	C	T	1.12 (1.05-1.21)	1.06E-03	1.17 (1.08-1.27)	1.05E-04
rs17036101	*SYN2/PPARG*	3	12252845	A	G	1.20 (1.01-1.43)	4.34E-02	1.18 (0.97-1.44)	1.03E-01
rs9472138	*VEGFA*	6	43919740	C	T	1.12 (1.03-1.21)	6.40E-03	1.14 (1.04-1.25)	4.07E-03

*rs2934381 used as a proxy for rs10923931 in UK analyses

Stage 1 DGI sample consisted of unrelated case-control component only (n=2097; 1022 cases, 1075 controls)

Stage 1 WTCCC sample consisted of T2D cases and controls, for whom BMI data were available (n=3335; 1913 cases, 1422 controls)