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

DIAbetes Genetics Replication And Meta-analysis (DIAGRAM) Consortium

SUPPLEMENTARY INFORMATION

Contents

SUPPLEMENTARY MATERIAL

Research participants

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

Overview of Study Design and Analysis Strategy

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

Genotyping, imputation and quality control

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

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

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

Statistical analyses

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

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

Comparison between HapMap and 1000G reference variant sets

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

Power calculations

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

Approximate conditional analysis with GCTA

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

Finemapping analyses using credible set mapping

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

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

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

T1D/T2D discrimination analysis

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

Expression quantitative trait loci (eQTL) analysis

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

Hierarchical clustering of T2D-related metabolic phenotypes

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

Functional annotation and enrichment analysis

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

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

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

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

Pathway analyses with DEPICT

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

Supplementary material and methods references

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

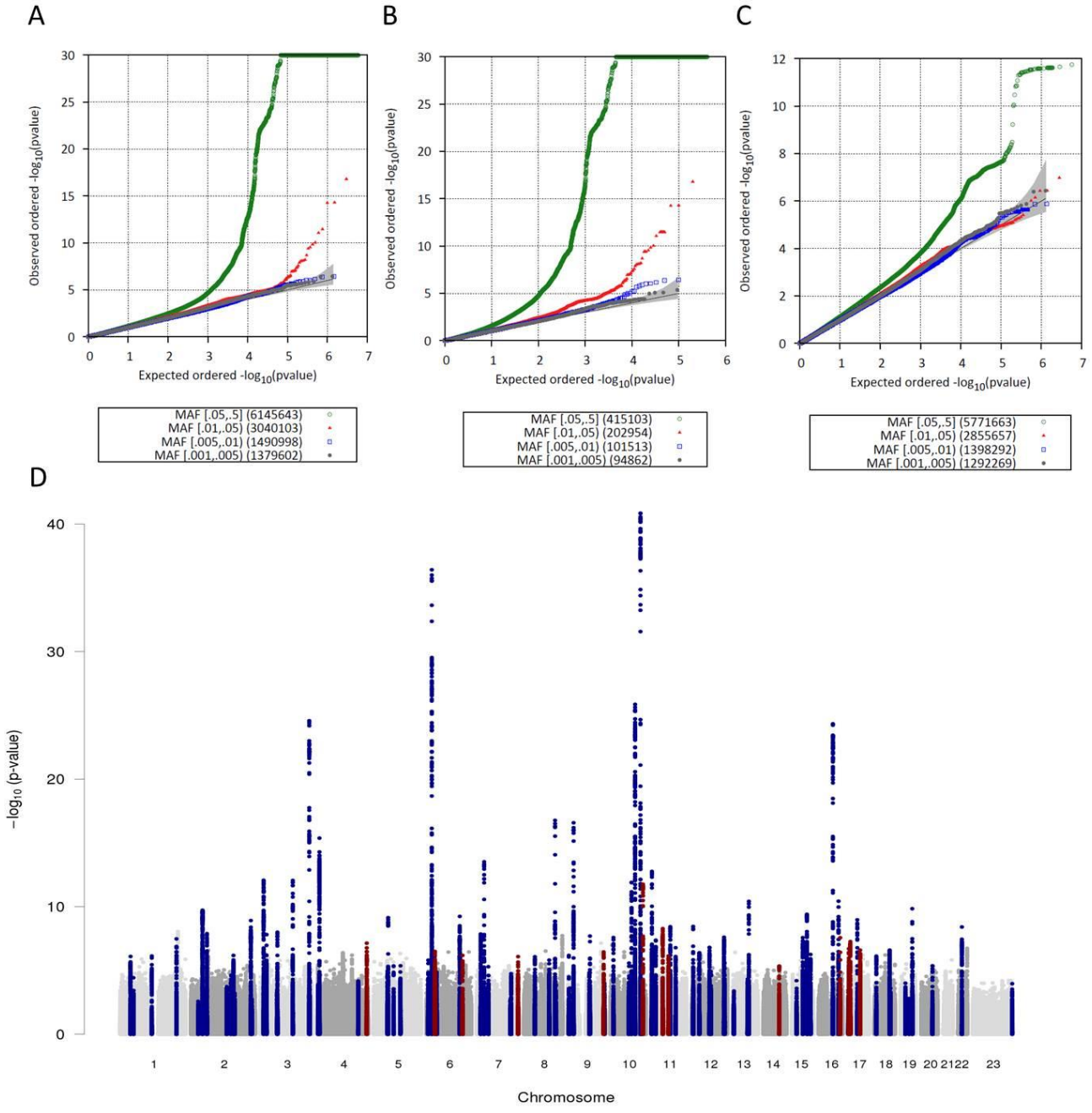
55. Dunham I, Kundaje A, Aldred SF, Collins PJ, Davis C a, Doyle F, et al. An integrated encyclopedia of DNA elements in the human genome. Nature. 2012;489(7414):57–74.

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

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

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

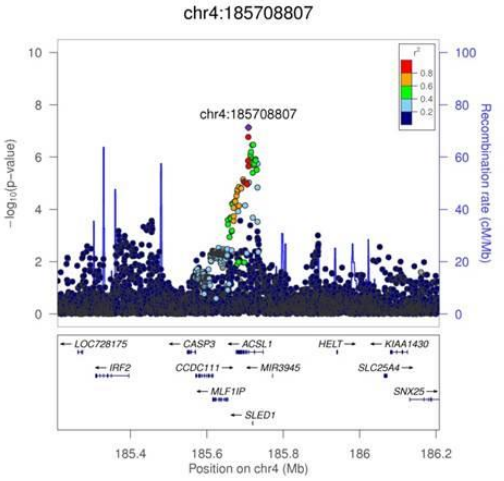
SUPPLEMENTARY FIGURES

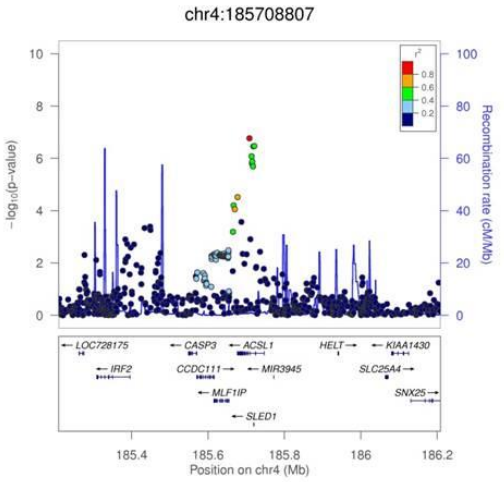


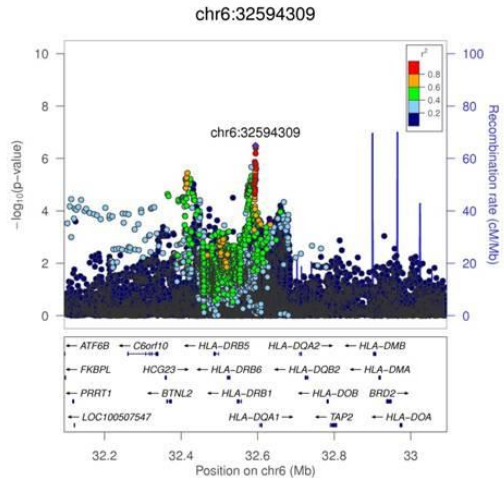
Supplementary Figure 1. QQ- and Manhattan plots of the discovery association meta-analysis results. A) QQ-plot of all the signals. B) QQ-plot of previously established signals. C) QQ-plot of novel signals. D) Manhattan plot. Signals of association reaching genome-wide significance for the first time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

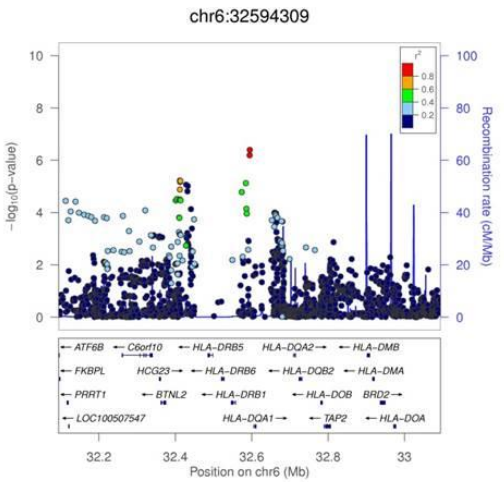
All 1000G SNVs

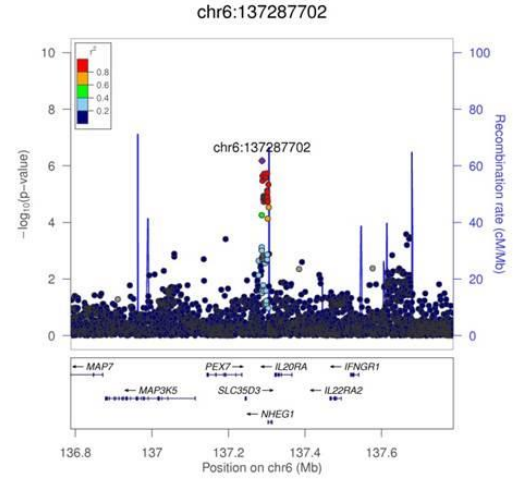
Restricted to HapMap SNVs

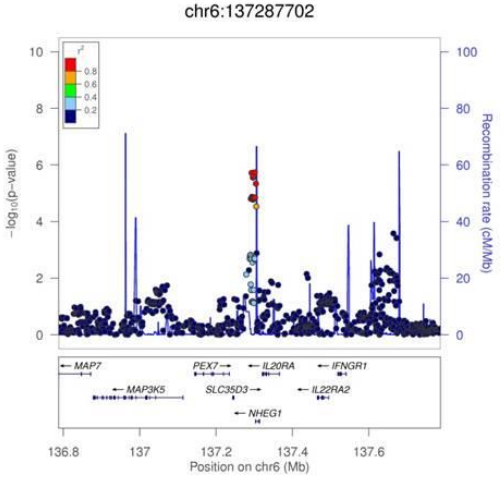


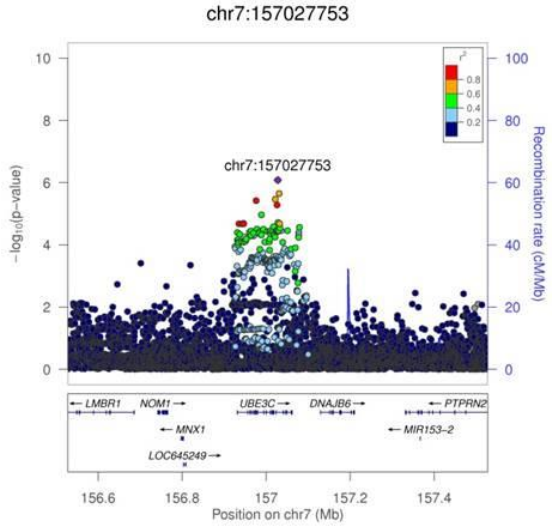


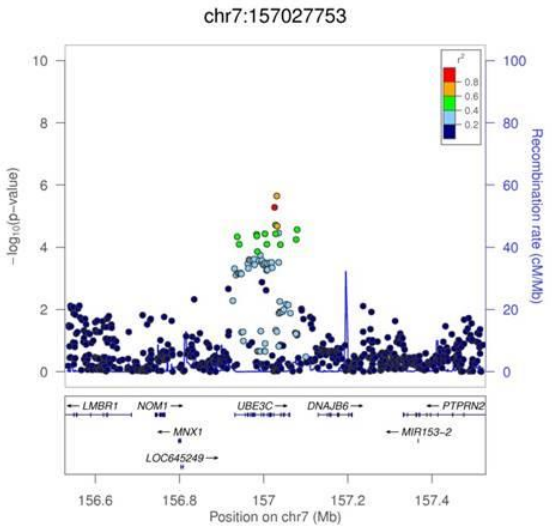


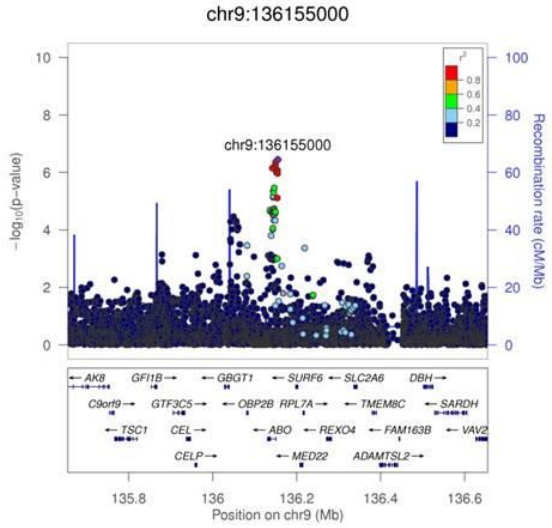


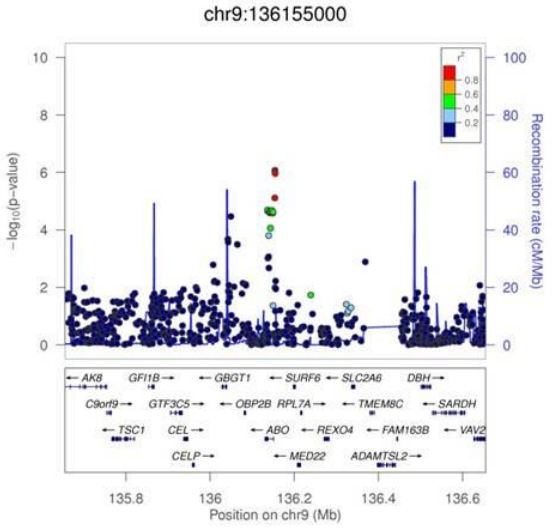


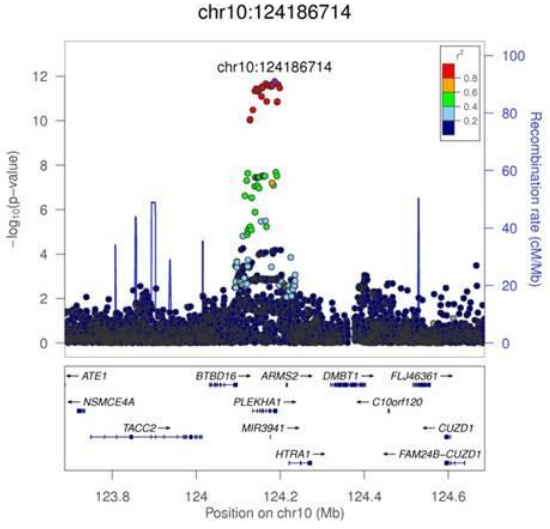


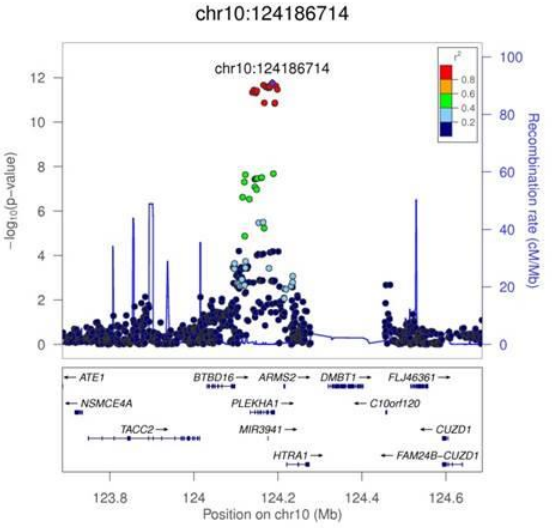

chr7:157027753
chr7:157027753
−log10(p−value)
Recombination rate (cM/Mb)
r²
0.8
0.6
0.4
0.2
LMBR1
NOM1
UBE3C
DNAJB6
PTPRN2
MNX1
MIR153−2
LOC645249
156.6
156.8
157
157.2
157.4
Position on chr7 (Mb)

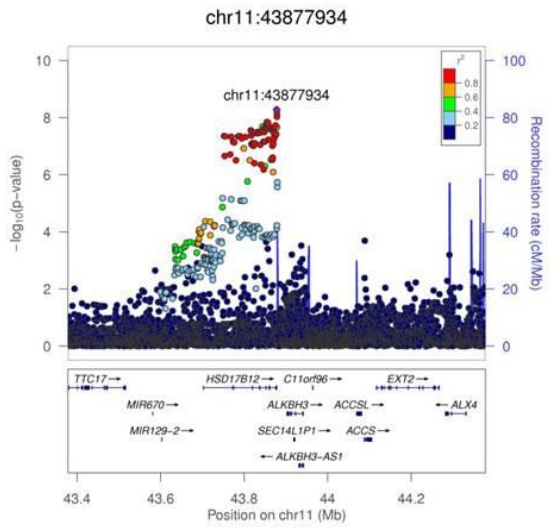

chr7:157027753
−log10(p−value)
Recombination rate (cM/Mb)
r²
0.8
0.6
0.4
0.2
LMBR1
NOM1
UBE3C
DNAJB6
PTPRN2
MNX1
MIR153−2
LOC645249
156.6
156.8
157
157.2
157.4
Position on chr7 (Mb)

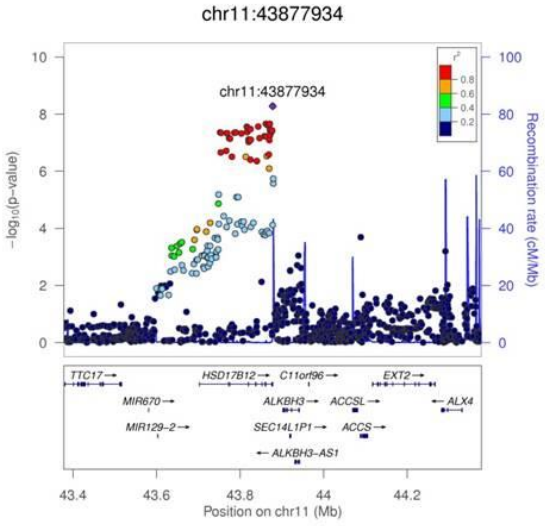

chr9:136155000
chr9:136155000
−log10(p−value)
Recombination rate (cM/Mb)
r²
0.8
0.6
0.4
0.2
AK8
GFI1B
GBGT1
SURF6
SLC2A6
DBH
C9orf9
GTF3C5
OBP2B
RPL7A
TMEM8C
SARDH
TSC1
CEL
ABO
REXO4
FAM163B
VAV2
CELP
MED22
ADAMTSL2
135.8
136
136.2
136.4
136.6
Position on chr9 (Mb)

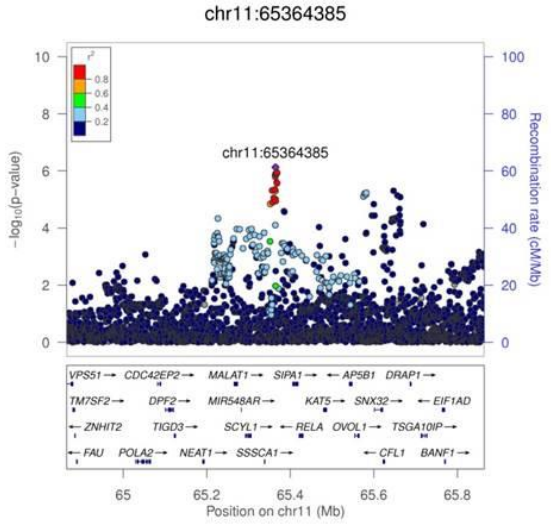

chr9:136155000
−log10(p−value)
Recombination rate (cM/Mb)
r²
0.8
0.6
0.4
0.2
AK8
GFI1B
GBGT1
SURF6
SLC2A6
DBH
C9orf9
GTF3C5
OBP2B
RPL7A
TMEM8C
SARDH
TSC1
CEL
ABO
REXO4
FAM163B
VAV2
CELP
MED22
ADAMTSL2
135.8
136
136.2
136.4
136.6
Position on chr9 (Mb)

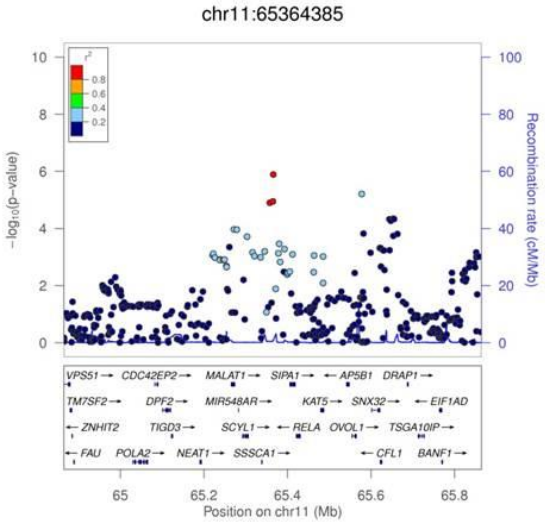

chr10:124186714
chr10:124186714
−log10(p−value)
Recombination rate (cM/Mb)
r²
0.8
0.6
0.4
0.2
ATE1
BTBD16
ARMS2
DMBT1
FLJ46361
NSMCE4A
PLEKHA1
C10orf120
TACC2
MIR3941
CUZD1
HTRA1
FAM24B−CUZD1
123.8
124
124.2
124.4
124.6
Position on chr10 (Mb)

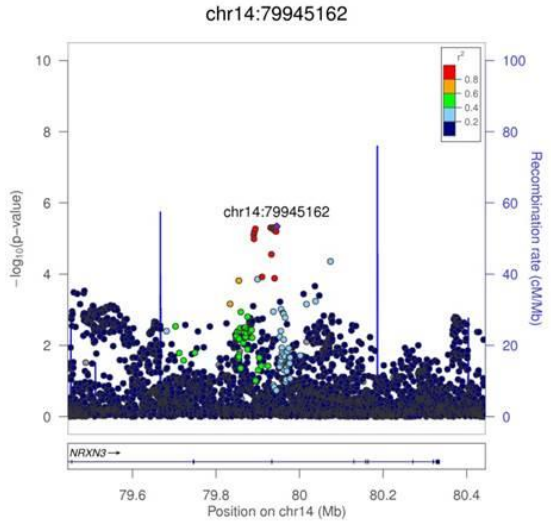

chr10:124186714
−log10(p−value)
Recombination rate (cM/Mb)
r²
0.8
0.6
0.4
0.2
ATE1
BTBD16
ARMS2
DMBT1
FLJ46361
NSMCE4A
PLEKHA1
C10orf120
TACC2
MIR3941
CUZD1
HTRA1
FAM24B−CUZD1
123.8
124
124.2
124.4
124.6
Position on chr10 (Mb)

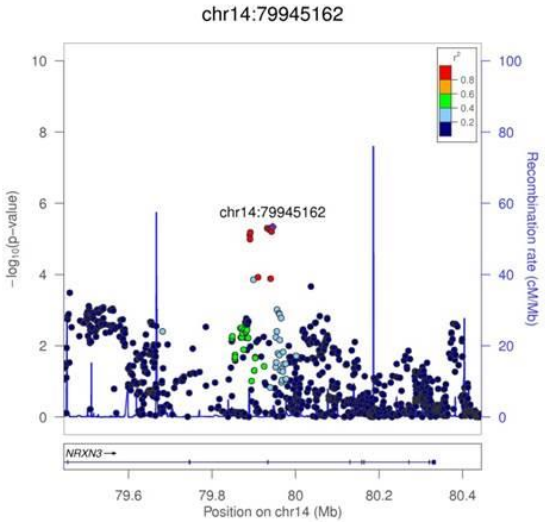

chr11:43877934
chr11:43877934
−log10(p−value)
Recombination rate (cM/Mb)
TTC17
HSD17B12
C11orf96
EXT2
MIR670
ALKBH3
ACCSL
ALX4
MIR129−2
SEC14L1P1
ACCS
ALKBH3−AS1
Position on chr11 (Mb)

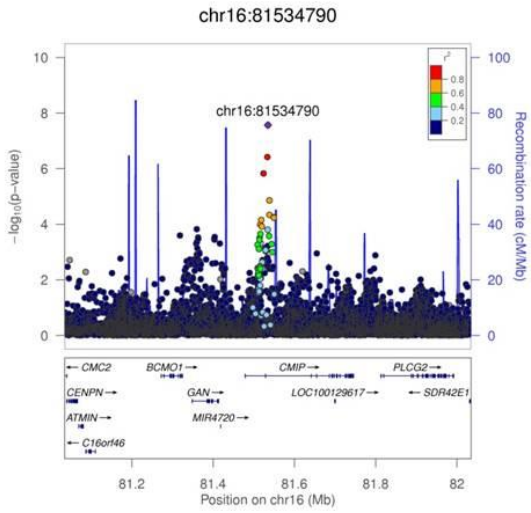

chr11:43877934
chr11:43877934
−log10(p−value)
Recombination rate (cM/Mb)
TTC17
HSD17B12
C11orf96
EXT2
MIR670
ALKBH3
ACCSL
ALX4
MIR129−2
SEC14L1P1
ACCS
ALKBH3−AS1
Position on chr11 (Mb)

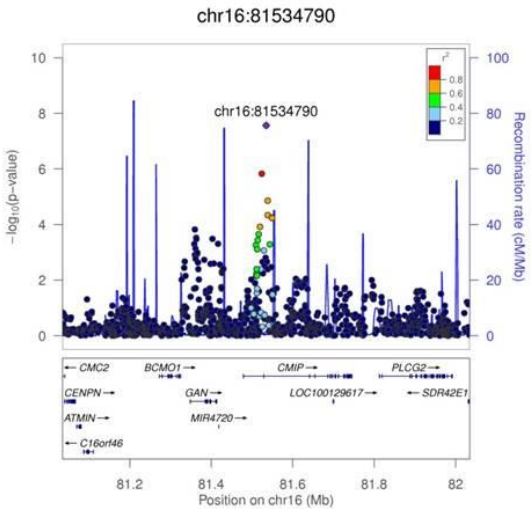

chr11:65364385
chr11:65364385
−log10(p−value)
Recombination rate (cM/Mb)
VPS51
CDC42EP2
MALAT1
SIPA1
AP5B1
DRAP1
TM7SF2
DPF2
MIR548AR
KAT5
SNX32
EIF1AD
ZNHIT2
TIGD3
SCYL1
RELA
OVOL1
TSGA10IP
FAU
POLA2
NEAT1
SSSCA1
CFL1
BANF1
Position on chr11 (Mb)

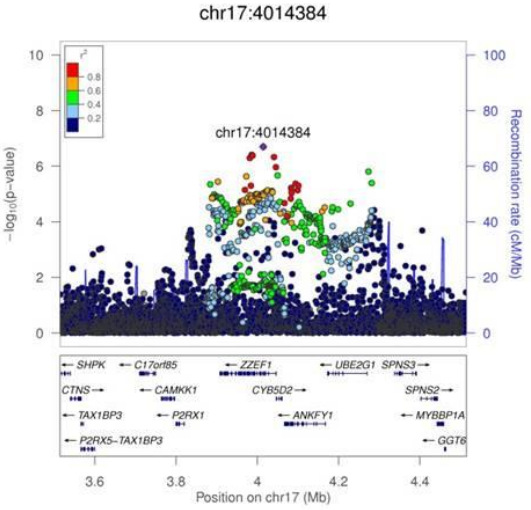

chr11:65364385
chr11:65364385
−log10(p−value)
Recombination rate (cM/Mb)
VPS51
CDC42EP2
MALAT1
SIPA1
AP5B1
DRAP1
TM7SF2
DPF2
MIR548AR
KAT5
SNX32
EIF1AD
ZNHIT2
TIGD3
SCYL1
RELA
OVOL1
TSGA10IP
FAU
POLA2
NEAT1
SSSCA1
CFL1
BANF1
Position on chr11 (Mb)

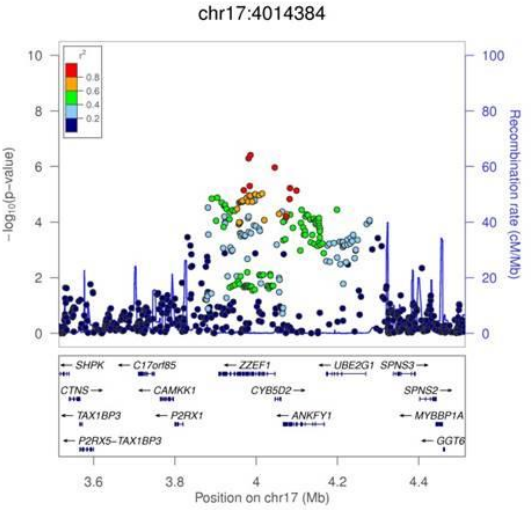

chr14:79945162
chr14:79945162
−log10(p−value)
Recombination rate (cM/Mb)
NRXN3
Position on chr14 (Mb)

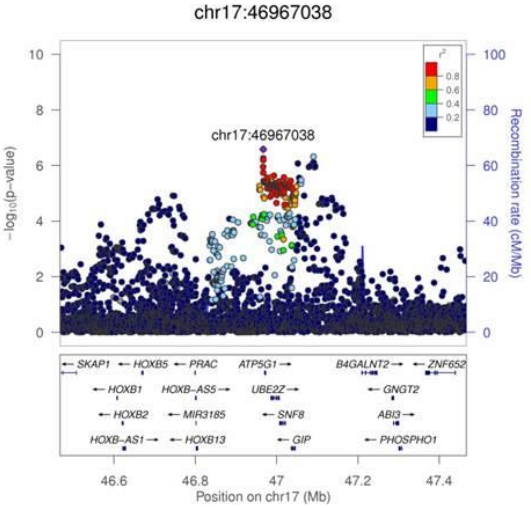

chr14:79945162
chr14:79945162
−log10(p−value)
Recombination rate (cM/Mb)
NRXN3
Position on chr14 (Mb)

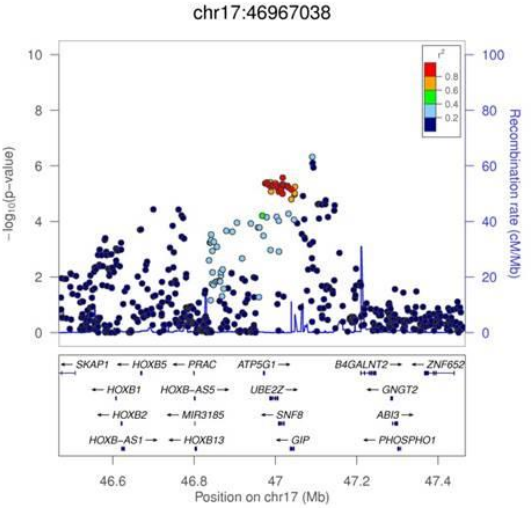

chr16:81534790
chr16:81534790
−log10(p−value)
Recombination rate (cM/Mb)
r²
0.8
0.6
0.4
0.2
CMC2
BCMO1
CMIP
PLCG2
CENPN
GAN
LOC100129617
SDR42E1
ATMIN
MIR4720
C16orf46
81.2
81.4
81.6
81.8
82
Position on chr16 (Mb)

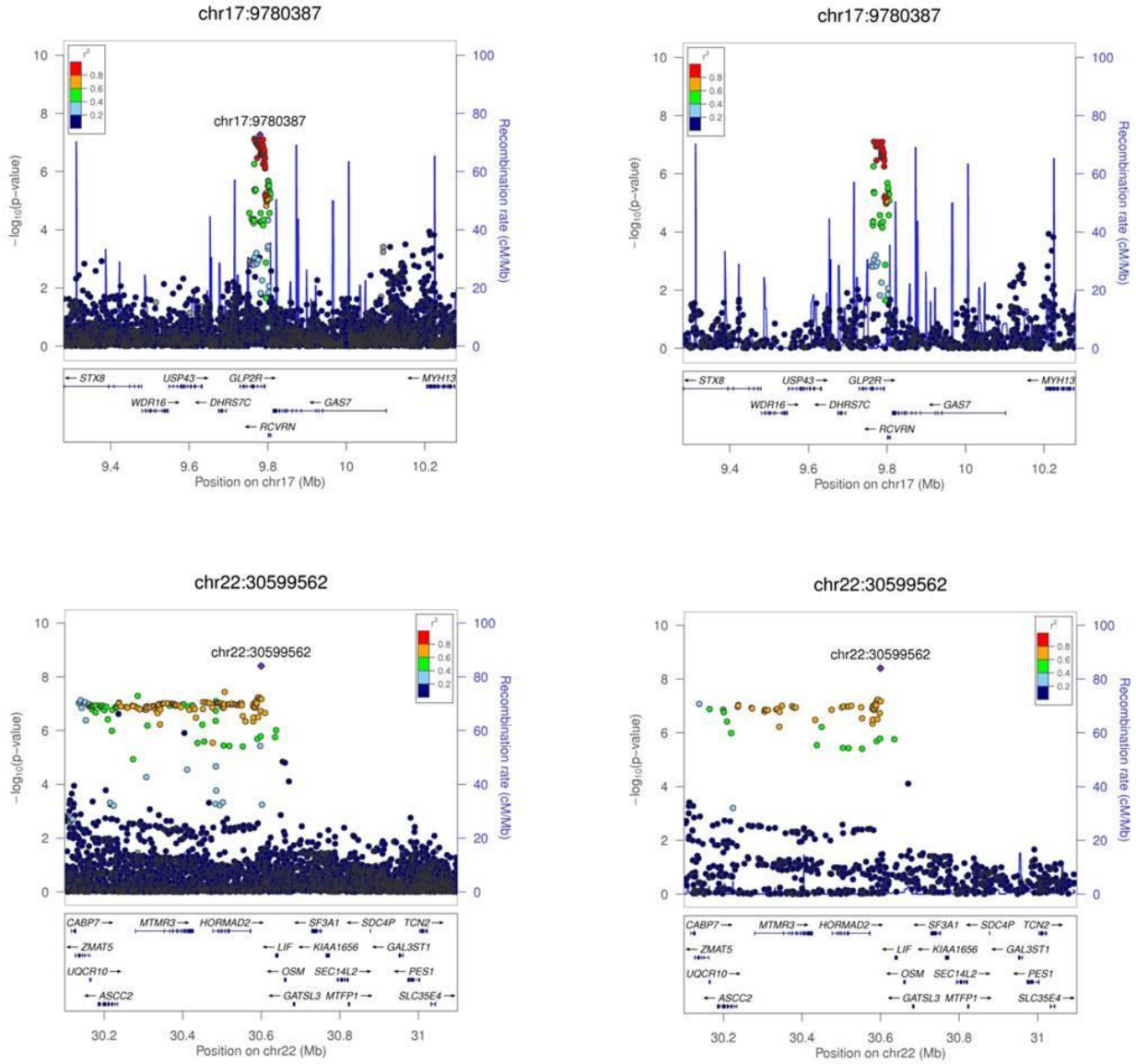

chr16:81534790
chr16:81534790
−log10(p−value)
Recombination rate (cM/Mb)
r²
0.8
0.6
0.4
0.2
CMC2
BCMO1
CMIP
PLCG2
CENPN
GAN
LOC100129617
SDR42E1
ATMIN
MIR4720
C16orf46
81.2
81.4
81.6
81.8
82
Position on chr16 (Mb)


chr17:4014384
chr17:4014384
−log10(p−value)
Recombination rate (cM/Mb)
r²
0.8
0.6
0.4
0.2
SHPK
C17orf85
ZZEF1
UBE2G1
SPNS3
CTNS
CAMKK1
CYB5D2
SPNS2
TAX1BP3
P2RX1
ANKFY1
MYBBP1A
P2RX5−TAX1BP3
GGT6
3.6
3.8
4
4.2
4.4
Position on chr17 (Mb)

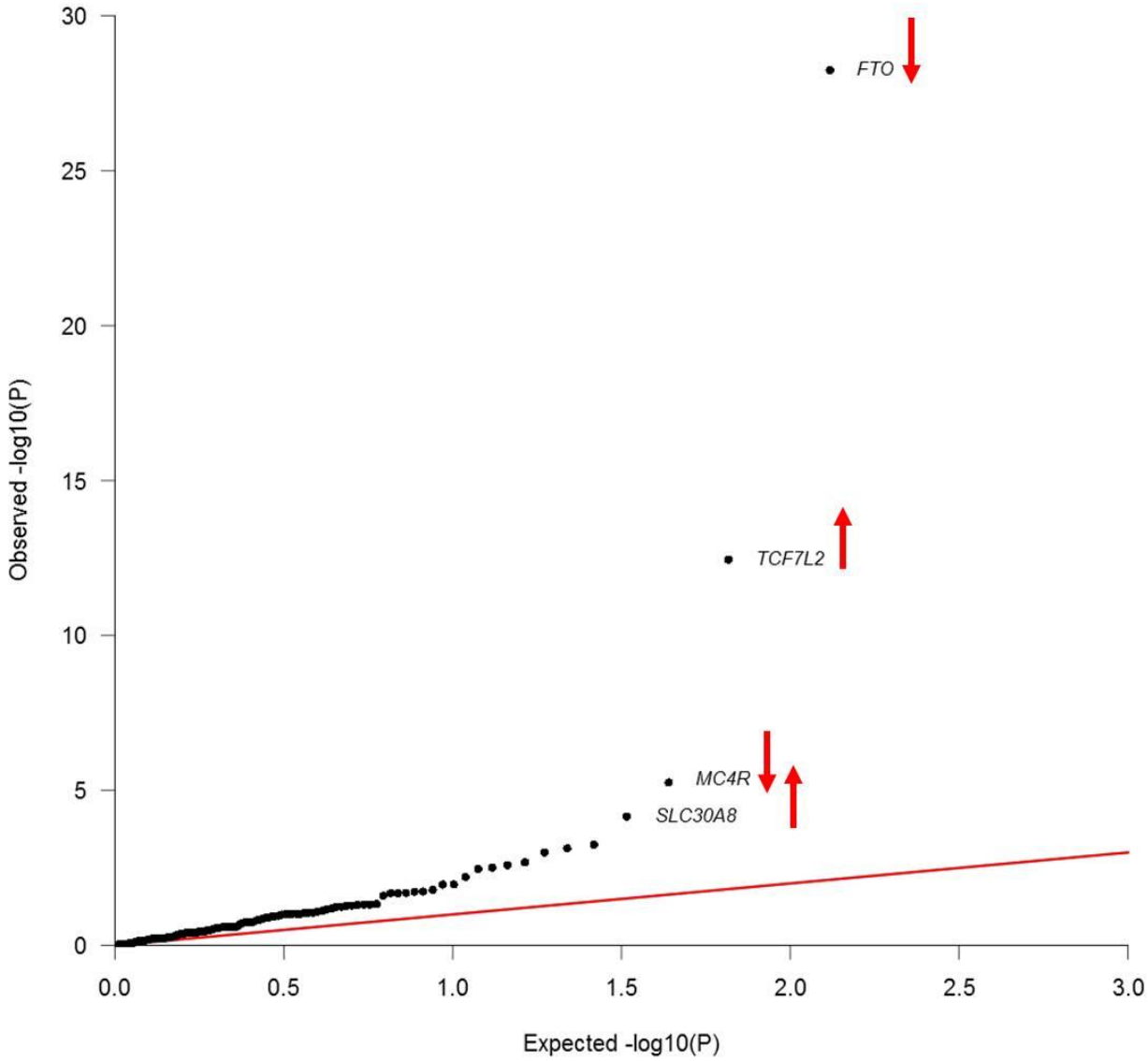

chr17:4014384
chr17:4014384
−log10(p−value)
Recombination rate (cM/Mb)
r²
0.8
0.6
0.4
0.2
SHPK
C17orf85
ZZEF1
UBE2G1
SPNS3
CTNS
CAMKK1
CYB5D2
SPNS2
TAX1BP3
P2RX1
ANKFY1
MYBBP1A
P2RX5−TAX1BP3
GGT6
3.6
3.8
4
4.2
4.4
Position on chr17 (Mb)


chr17:46967038
chr17:46967038
−log10(p−value)
Recombination rate (cM/Mb)
r²
0.8
0.6
0.4
0.2
SKAP1
HOXB5
PRAC
ATP5G1
B4GALNT2
ZNF652
HOXB1
HOXB−AS5
UBE2Z
GNGT2
HOXB2
MIR3185
SNF8
ABI3
HOXB−AS1
HOXB13
GIP
PHOSPHO1
46.6
46.8
47
47.2
47.4
Position on chr17 (Mb)


chr17:46967038
chr17:46967038
−log10(p−value)
Recombination rate (cM/Mb)
r²
0.8
0.6
0.4
0.2
SKAP1
HOXB5
PRAC
ATP5G1
B4GALNT2
ZNF652
HOXB1
HOXB−AS5
UBE2Z
GNGT2
HOXB2
MIR3185
SNF8
ABI3
HOXB−AS1
HOXB13
GIP
PHOSPHO1
46.6
46.8
47
47.2
47.4
Position on chr17 (Mb)



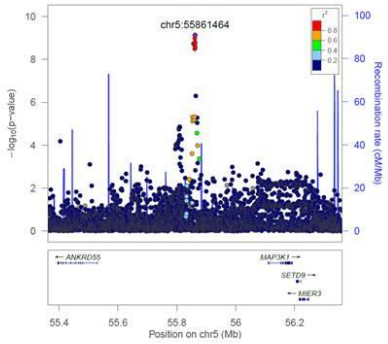
Supplementary Figure 2. Regional plots for the thirteen novel T2D loci. In the left panel, the plot is based using all 1000 Genomes March 2012 multi-ethnic SNV set, whereas in the right panel the plot is restricted to SNVs present in HapMap CEU reference set.



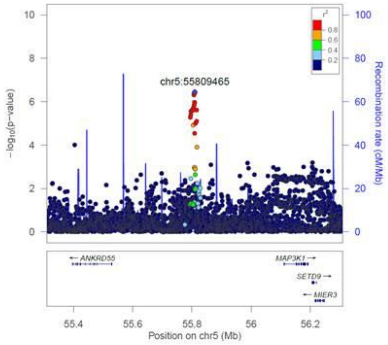
Supplementary Figure 3. QQ-plot of the expected vs. observed P-values for heterogeneity between BMI-adjusted and unadjusted association analysis models for established and novel T2D loci. The *FTO, TCF7L2, MC4R* and *SLC30A8* loci show large differences between models ($p_{heterogeneity}=5.70x10^{-29}$,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

1) *ANKRD55*

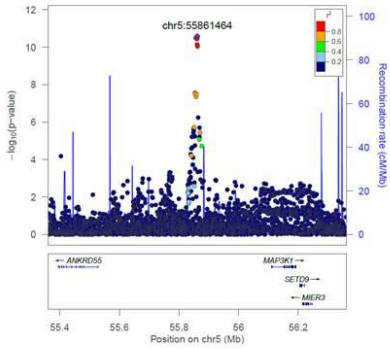
Lead SNV, unconditional



Distinct signal, conditional on lead SNV

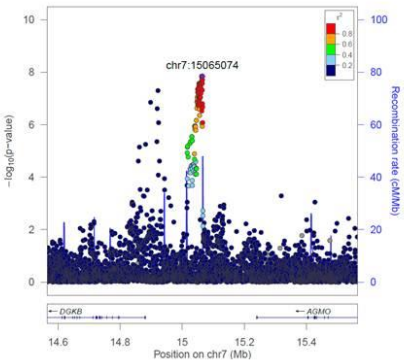


Lead SNV, conditional on distinct signal

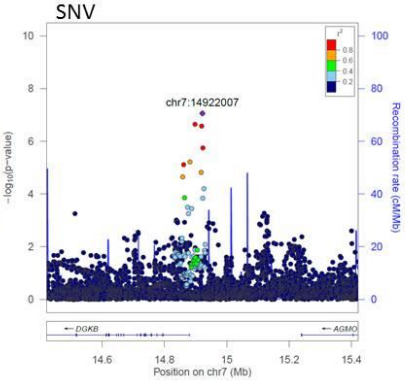


2) *DGKB*

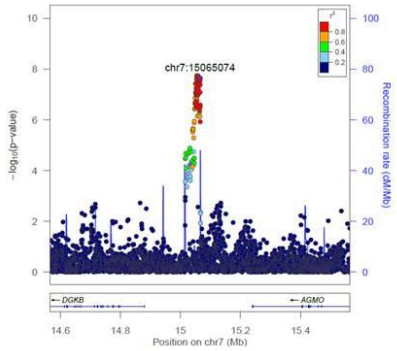
Lead SNV, unconditional



Distinct signal, conditional on lead SNV

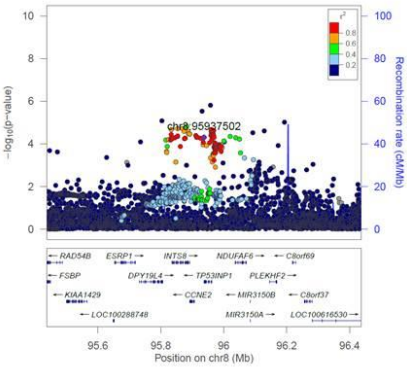


Lead SNV, conditional on distinct signal

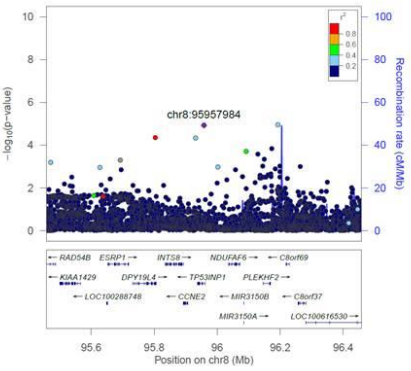


3) *TP53INP1*

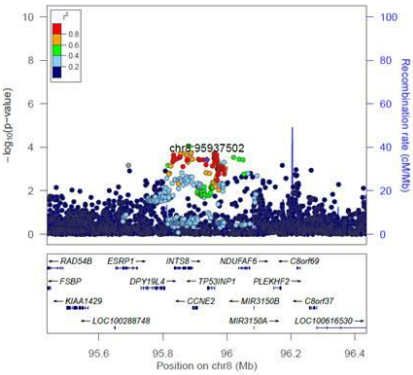
Lead SNV, unconditional



Distinct signal, conditional on lead SNV

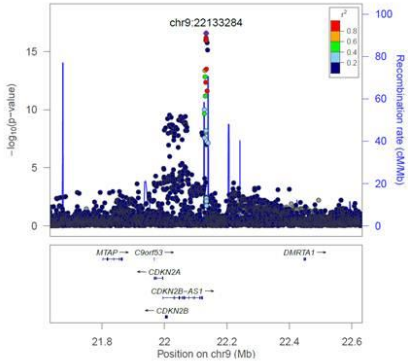


Lead SNV, conditional on distinct signal

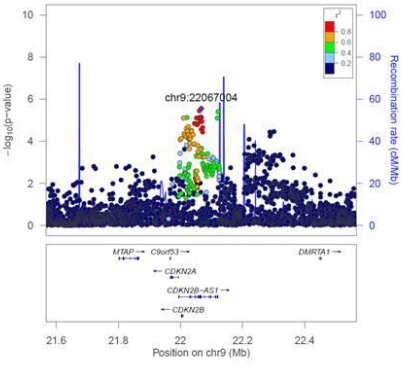


4) *CDKN2A/B*

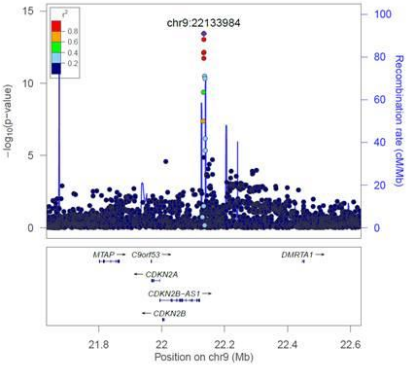
Lead SNV, unconditional



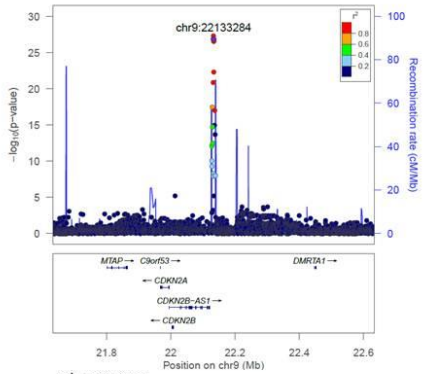
First distinct signal, conditional on lead SNV and other distinct signals



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

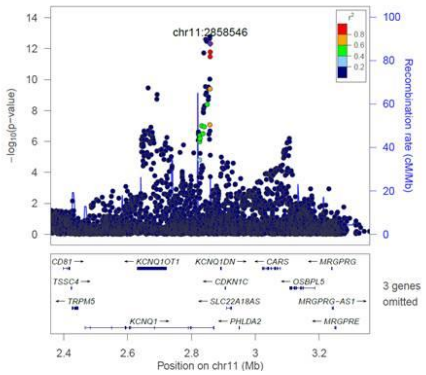


Lead SNV, conditional on distinct signals

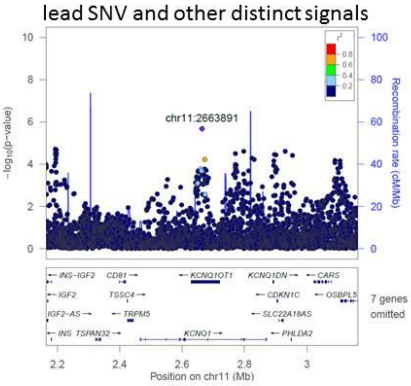


5) *KCNQ1*

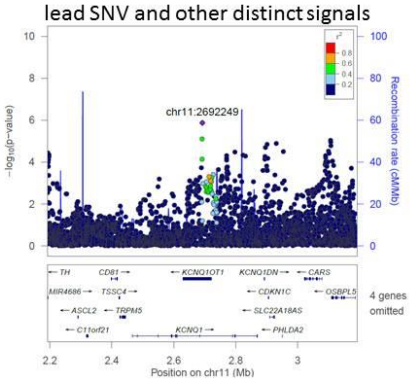
Lead SNV, unconditional



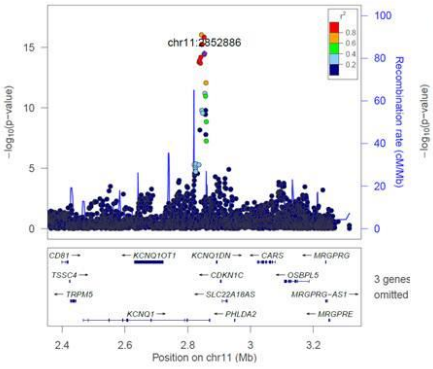
First distinct signal, conditional on lead SNV and other distinct signals



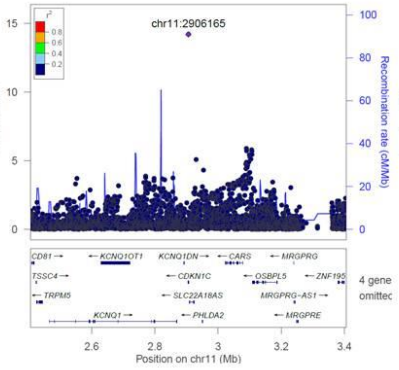
Second distinct signal, conditional on lead SNV and other distinct signals



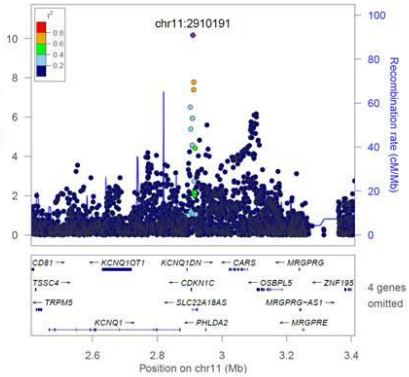
Third distinct signal, conditional on lead SNV and other distinct signals



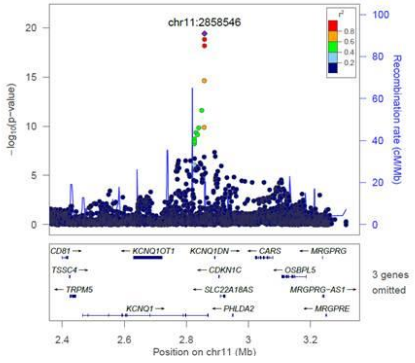
Fourth distinct signal, conditional on lead SNV and other distinct signals

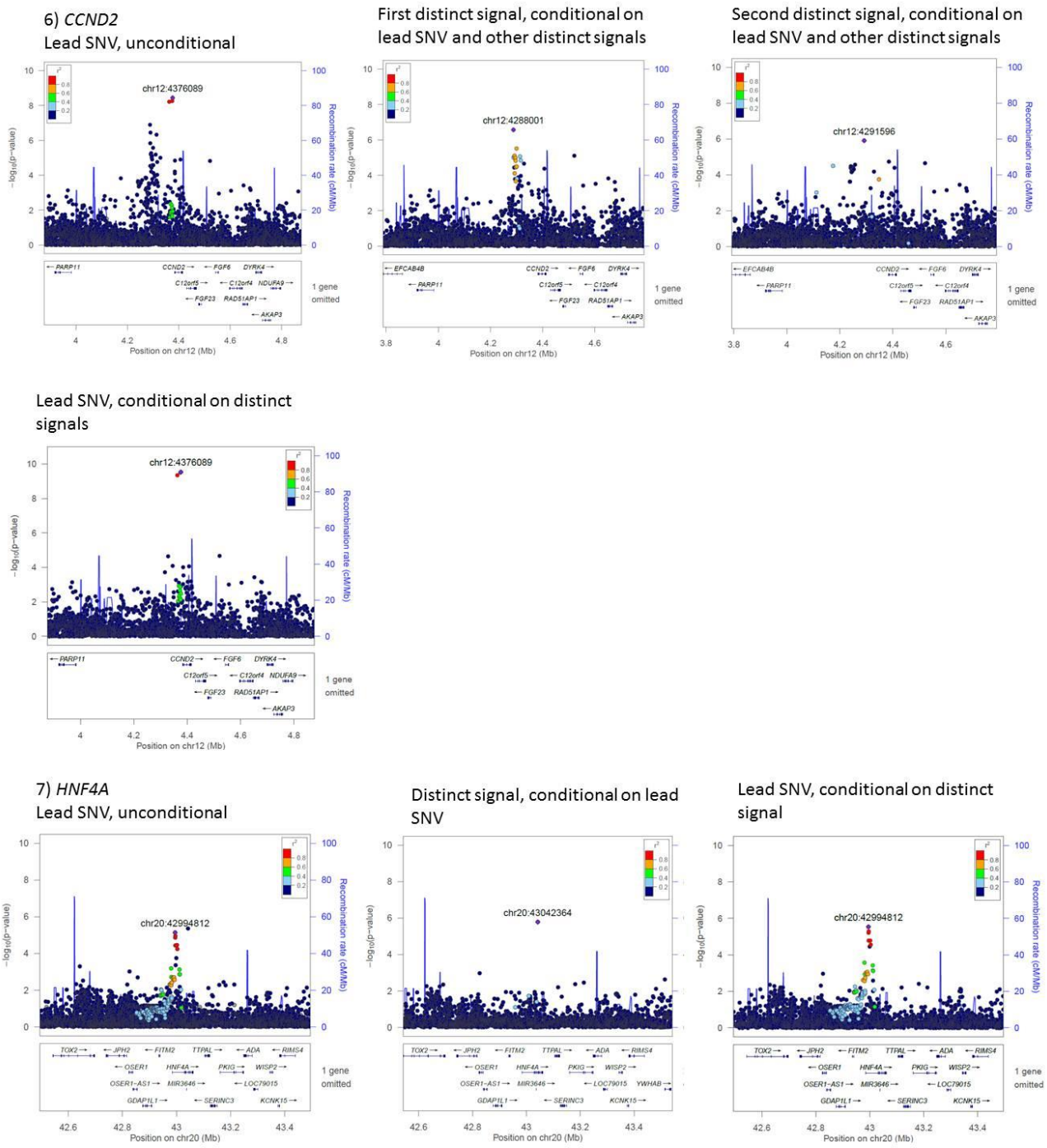


Fifth distinct signal, conditional on lead SNV and other distinct signals

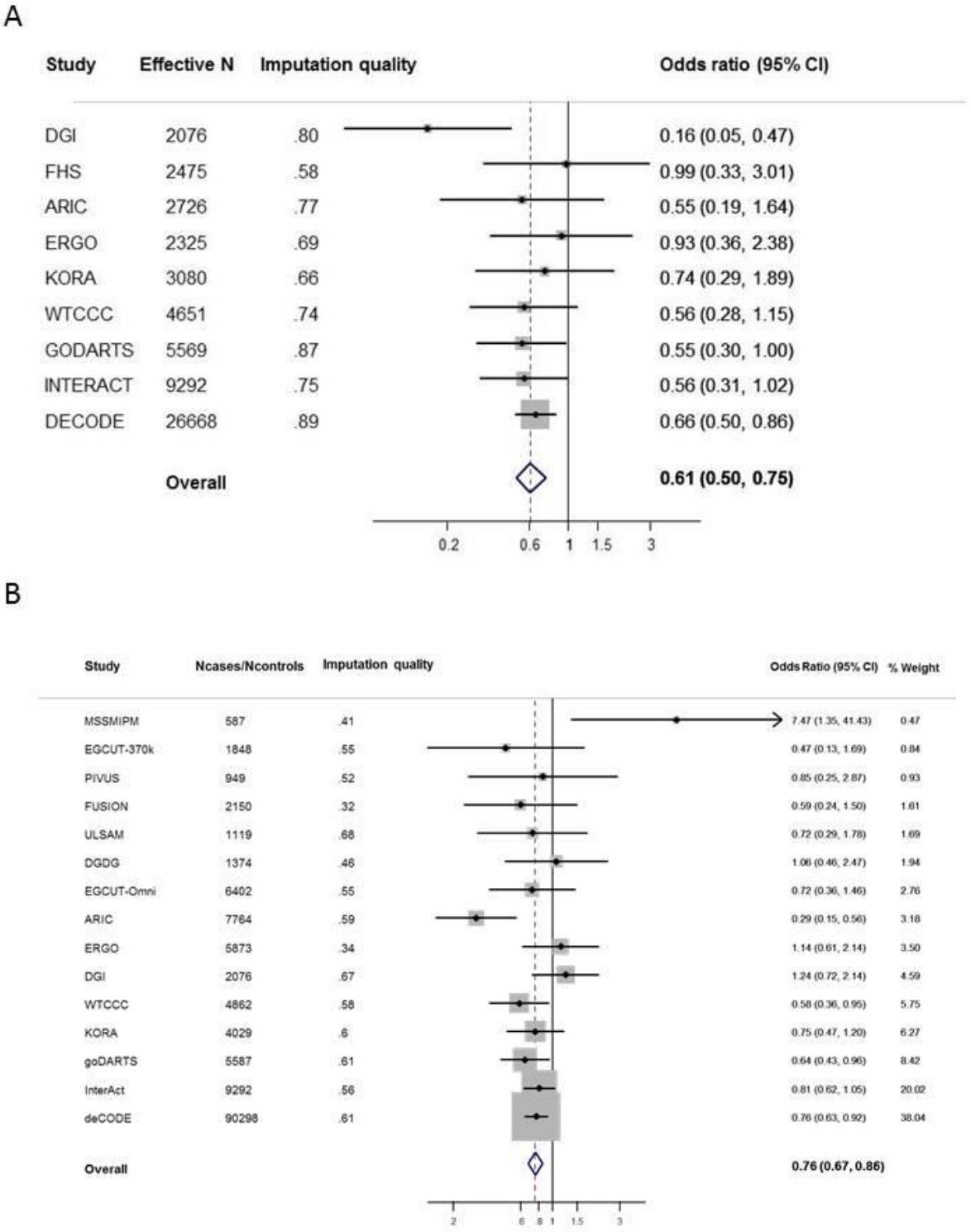


Lead SNV, conditional on distinct signals

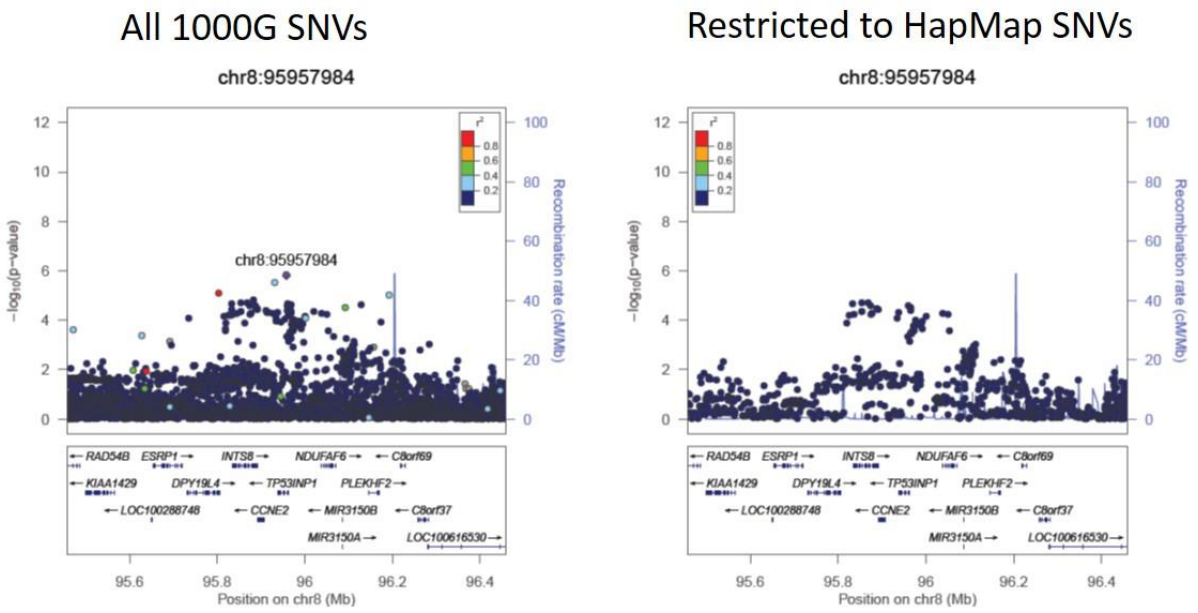




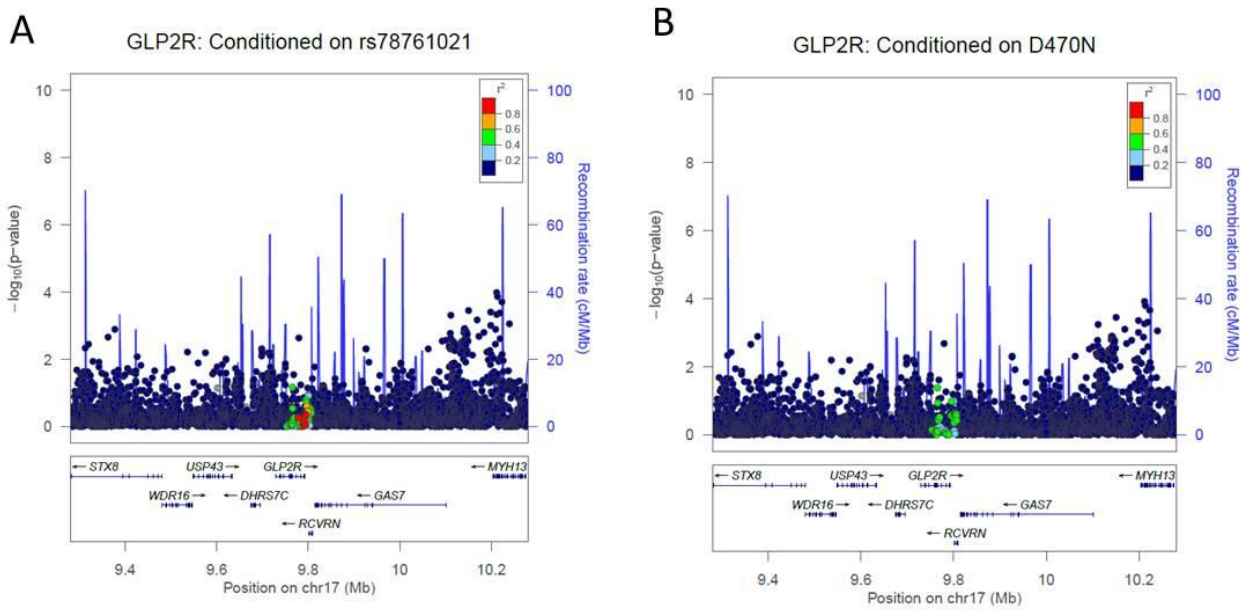
Supplementary Figure 4. Regional plots for T2D loci showing additional distinct signals ($p<10^{-5}$) in the approximate conditional analysis. First, unconditional analysis results are shown, followed by results conditioned on the lead SNV and other distinct signals. In the last plot for each locus the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



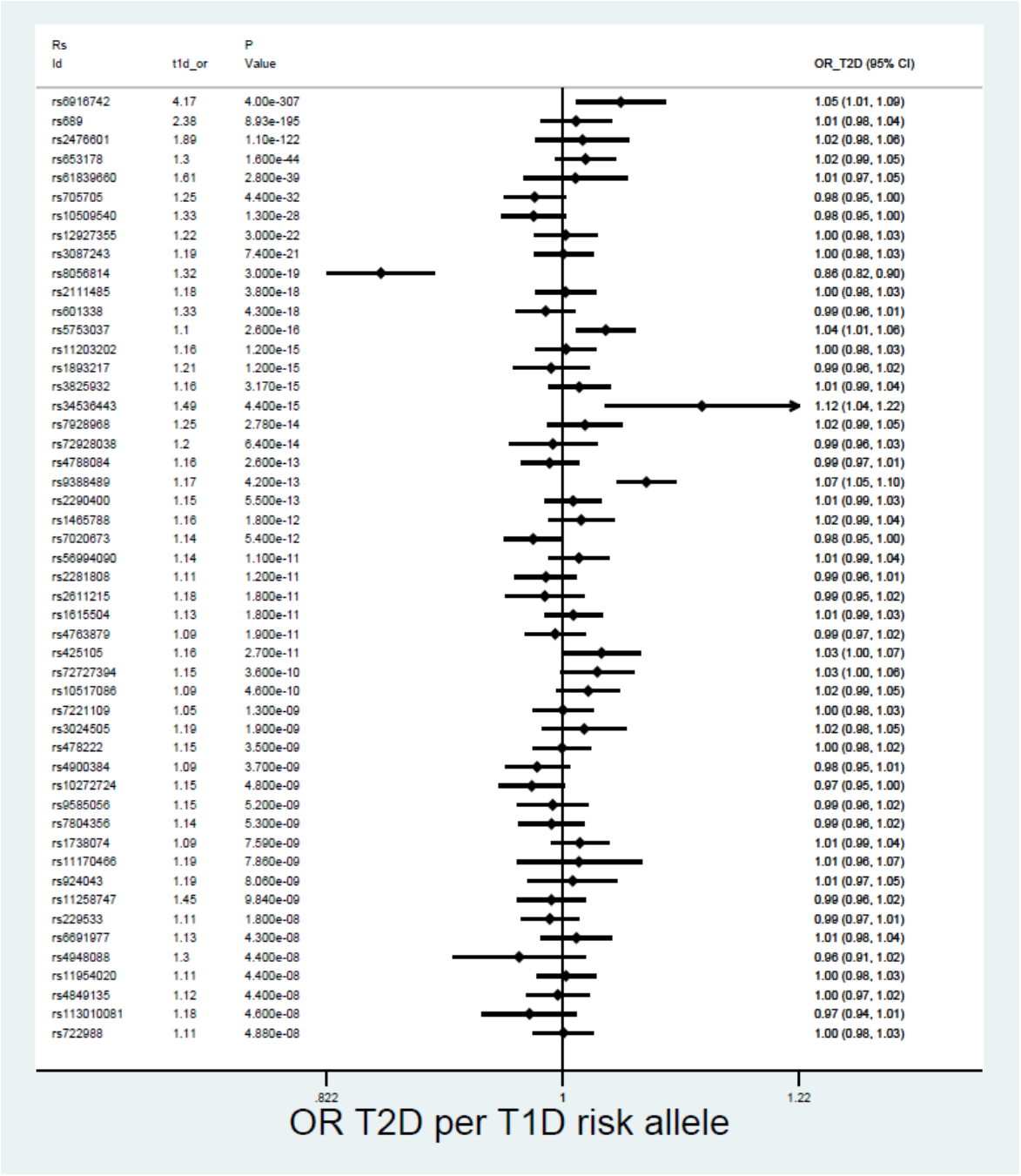
Supplementary Figure 5. Forest plots of the A) putative low frequency distinct signal (rs188827514) and B) previously established (Steinthorsdottir et al.) **low-frequency variant (rs76895963) at *CCND2* for their associations with T2D.** Odds ratios (OR) with their 95% confidence intervals (CI) are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



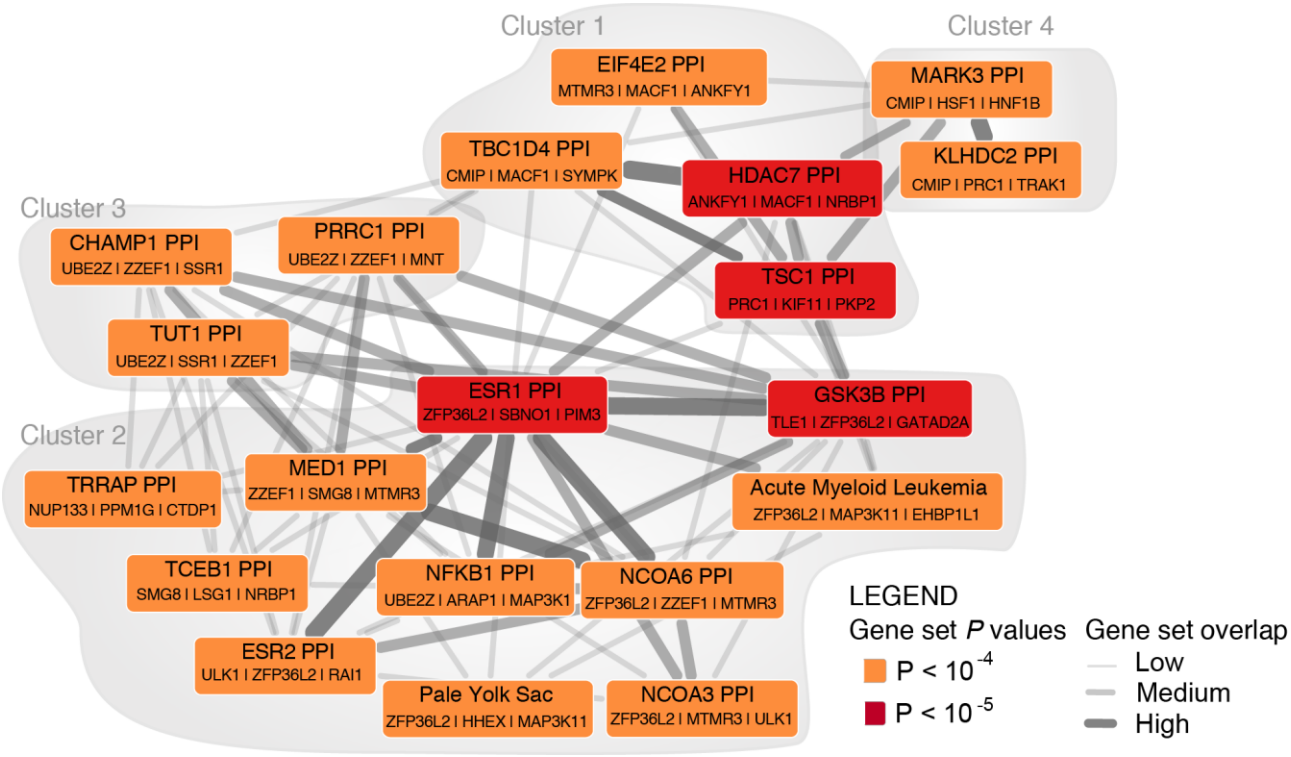
Supplementary Figure6. Regional architecture of *TP53INP1* locus. In the right panel the figure is plotted using all 1000 Genomes SNVs and highlights the new lead SNV (rs11786613) independent from the previous lead variant, signal visible in the left panel the plot is restricted to SNVs present in HapMap.



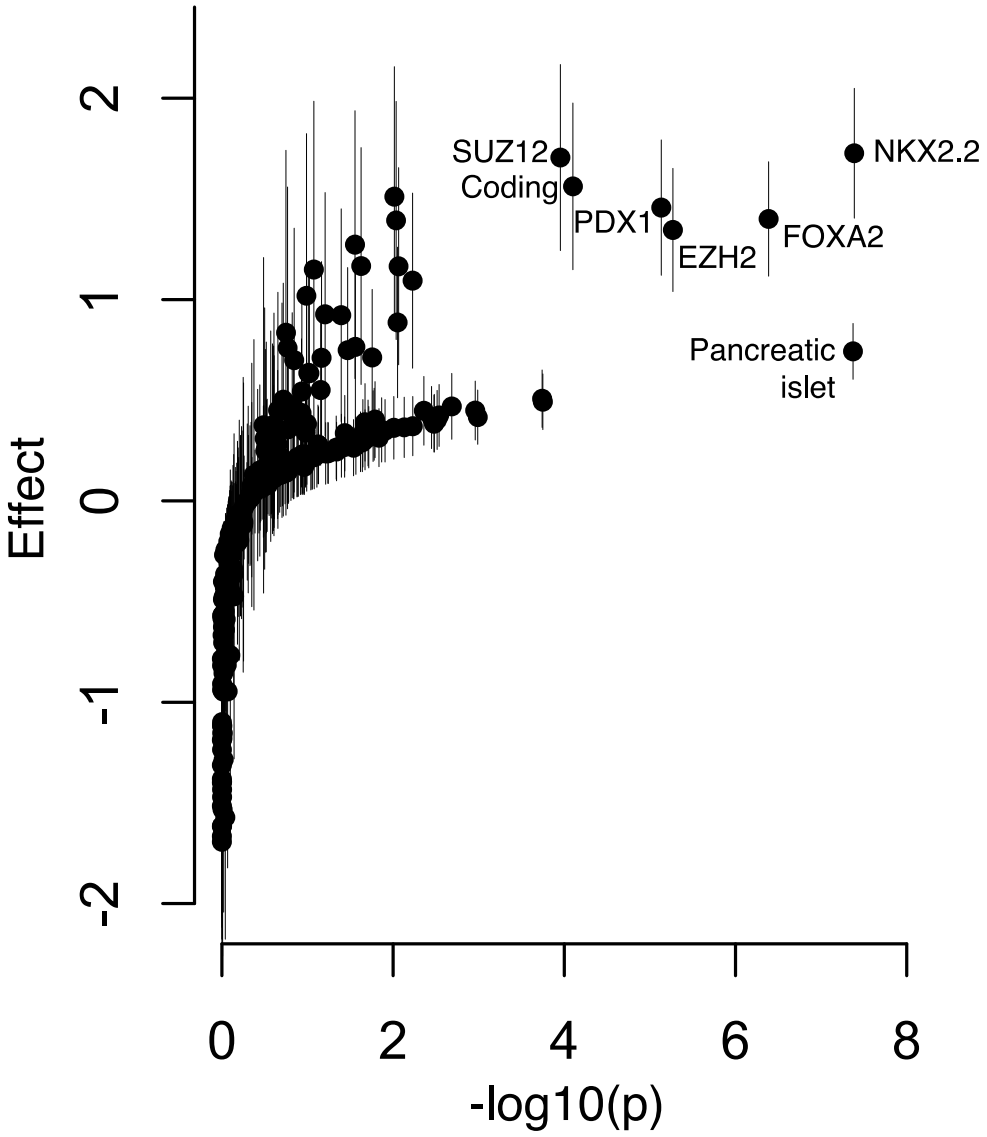
Supplementary Figure 7. Association of variation in *GLP2R* with T2D after approximate conditional analyses on either A) the lead SNV (rs78761021), or B) D470N.



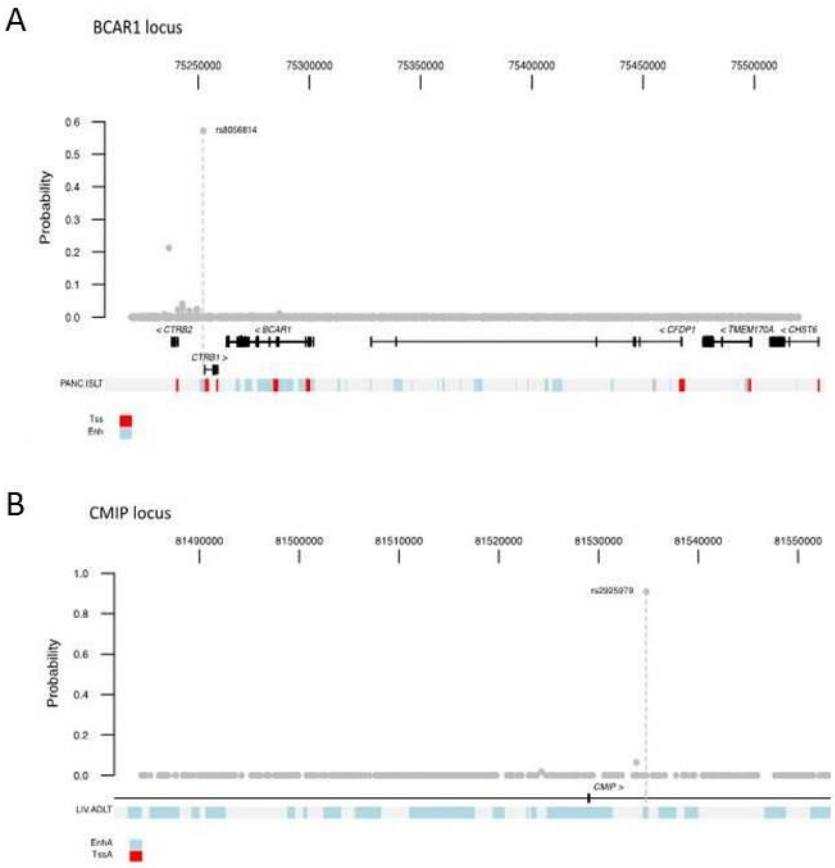
Supplementary Figure 8. Effects on T2D of 50 established T1D variants. All effects are aligned to T1D risk-raising allele. Loci are sorted from top to bottom by the magnitude of association with T1D.



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



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



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

BIOLOGY BOX

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

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

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

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

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

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

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

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

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

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

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

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

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

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

BIOLOGY BOX REFERENCES:

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

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

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