ESM Methods

Study population and trait measurement

The Atherosclerosis Risk in Communities Study (ARIC) is part of Causal Variants Across the Life Course (CALiCo) consortium of well characterised population-based studies in bi-racial populations (white and African American) with a central genotyping and resequencing core laboratory [1]. ARIC is a multi-center prospective investigation of atherosclerotic disease in a predominantly bi-racial population conducted in four U.S. communities [1]. ARIC recruited 15,792 individuals (of which 4,266 are African American) aged 45-64 years from four communities in Forsyth County, N.C., Jackson, M.S., Minneapolis, M.N., and Washington County, M.D. for a baseline examination beginning in 1987, with annual follow-up until 2012, and semi-annual follow-up after occurring by telephone to maintain contact and assess health status. The data used in this study were from the baseline visit. BMI was calculated from weight and height measurements. Current smoking was dichotomised as yes/no according to the question “Do you now smoke cigarettes?” All African American subjects in ARIC were genotyped on the Metabochip. A detailed study protocol is available on the ARIC study website (https://www2.cscc.unc.edu/aric/).

The Coronary Artery Risk Development in Young Adults (CARDIA) Study is part of the CALiCo consortium and is a study examining the development and determinants of clinical and subclinical cardiovascular disease and their risk factors [2]. This study recruited 5,115 white and black individuals between 18 and 30 years old (52% African American and 55% women) in Birmingham, AL; Chicago, IL; Minneapolis, MN; and Oakland, CA [3]. CARDIA cohort participants, born between 1955 and 1968, have been influenced substantially by the obesity epidemic at ages younger than participants in other established NHLBI cohorts. Data were collected on a variety of factors believed to be related to heart disease, including fasting glucose levels. Physical measurements (height and weight), as well as lifestyle factors such as such as smoking and other chemistries (e.g., insulin), were collected. Current smoking was dichotomised using the following questions at the baseline questionnaire: "have you ever smoked cigarettes regularly for at least 3 months? By regularly, we mean at least 5 cigarettes per week, almost every week?" Those who stated “no” were considered to be non-smokers, while those who stated “yes” were then asked whether they currently smoked cigarettes. Those who responded “yes” were considered to be current smokers and those who responded “no” were considered to be former smokers. All study participants provided written informed consent. Only the African American participants were genotyped on Metabochip.
The Hispanic Community Health Study / Study of Latinos (HCHS/SOL) is a population-based cohort study of 16,415 self-identified Hispanic/Latino individuals aged 18-74 years randomly selected from households in four U.S. field centers (Chicago, IL; Miami, FL; Bronx, NY; San Diego, CA) [4]. The cohort includes participants who self-identified as having Hispanic/Latino background, the largest groups being Central American ($n = 1,730$), Cuban ($n = 2,348$), Dominican ($n = 1,460$), Mexican ($n = 6,471$), Puerto-Rican ($n = 2,728$), and South American ($n = 1,068$). The baseline examination during 2008 and 2011 included a clinical visit with comprehensive biological, behavioral, and socio-demographic assessments. Smoking status was measured by self-report and categorised into three groups: current, former, and never smokers. Two questions were used: “Have you ever smoked at least 100 cigarettes in your entire life?” and “Do you now smoke daily, some days or not at all?” If participants had smoked at least 100 cigarettes in their entire life and reported smoking daily or some days, then they were considered current smokers; if participants had smoked at least 100 cigarettes in their entire life and did not report smoking daily or some days, then they were considered former smokers; and if participants had not smoked at least 100 cigarettes in their entire life, they were considered never smokers.

The Women’s Health Initiative (WHI) is a prospective study investigating post-menopausal women’s health [5]. A total of 161,808 women aged 50–79 years old were recruited from 40 U.S. clinical centers between 1993 and 1998. WHI consists of two parts: randomised clinical trials of hormone therapy, dietary modification, and calcium/Vitamin D supplementation, and an observational cohort study. Socio-demographic characteristics, lifestyle factors (e.g. smoking), medical history, medication use and physical measures of height and weight were collected at the baseline visit. Data on lifetime active and passive smoking were collected. Women were initially classified by active smoking status into current, former or never smokers (participants that had not smoked 100 cigarettes in their life). A subset of African American participants was genotyped through the WHI SNP Health Association Resource (SHARE). All African American, Hispanic/Latino, Asian, and Native American/American Indian individuals who provided informed consent to submit their genotype data to dbGaP were either directly genotyped on the Metabochip or had genome-wide data on the Affymetrix 6.0 array available to impute Metabochip SNPs.

The Multiethnic Cohort (MEC) is a population-based prospective cohort study of over 215,000 men and women in Hawaii and California aged 45-75 at baseline (1993-1996) and primarily of five ethnic/racial groups: African Americans, Native Hawaiians, Whites, Latinos, and Japanese.
MEC was funded by the National Cancer Institute in 1993 to examine lifestyle risk factors and genetic susceptibility to cancer. All eligible cohort members completed baseline and follow-up questionnaires. Each participant completed a mailed, epidemiologic self-administered questionnaire regarding demographic, dietary, and lifestyle traits. This questionnaire included history of daily cigarette smoking during the past two weeks, smoking duration, and a record of current medications. Subjects were selected for Metabochip genotyping based on availability of biomarker data or from a pool of controls for a study of type 2 diabetes.

The BioMe Biobank Program is an ongoing, prospective, hospital- and outpatient-based population research program operated by The Charles Bronfman Institute for Personalized Medicine (IPM) at Mount Sinai. BioMe is an Electronic Medical Record (EMR)-linked biobank that integrates research data and clinical care information for consented patients at The Mount Sinai Medical Center, which serves diverse local communities of upper Manhattan with broad health disparities. BioMe populations include 36% Hispanics, 25% AA, 30% EA, and 9% of other ancestry. Biobank operations are fully integrated in clinical care processes, including direct recruitment from clinical sites waiting areas and phlebotomy stations by dedicated Biobank recruiters independent of clinical care providers, prior to or following a clinician standard of care visit. Recruitment currently occurs at a broad spectrum of over 30 clinical care sites. Information on anthropometrics, demographics, and medication was derived from both participants' EMR and a medical history questionnaire given at baseline [7]. Smoking status was measured by self-reported based on the following two questions: “Have you ever smoked cigarettes?” and “If you have quit, what year did you quit?” Height and Weight data are from questionnaires submitted at the time of Biobank enrollment. For the current analyses, genotype and phenotype data was available on Hispanic/Latinos and individuals with African Ancestry. Samples were genotyped using the Illumnia HumanOmniExpress+ v1.1 and then imputed with Impute2 [8] to genotypes of the 1000 Genomes project [9].

IRB Statement All participants included in these analyses have given consent for genetic studies and data sharing.

Continuous BMI measurement In all CALiCo studies, and WHI, BMI was calculated from height and weight measured at time of study enrolment in a clinic setting. In WHI only, measurements collected 1 or 3 years after enrolment were substituted for 140 participants missing enrolment.
height and/or weight. In MEC and BioMe, self-reported height and weight were used to calculate baseline BMI.

*Smoking status measurement* Smoking status was harmonised across studies as current versus former/never. See PAGE cohort descriptions above for study-specific details on ascertainment.

*Racial/ethnic grouping* In all studies, self-reported racial/ethnic group was collected via epidemiological questionnaires at baseline.

*Glucose/insulin measurement* Analyses were performed for fasting glucose (mmol/l) and natural log transformed fasting insulin (pmol/l). Individuals were excluded from the analysis if they were on diabetes treatment (oral or insulin), had a fasting plasma glucose equal to or greater than 7 mmol/l (126mg/dl), or fasting status for less than 8 hours. Individuals with BMI<16.5 kg/m² and BMI>70 kg/m² were also excluded with the assumption that these extremes could be attributable to data coding errors, an underlying illness, or possibly to a familial syndrome. In BioMe, oral type 2 diabetes medications used for exclusion included Acetohexamide, Tolazamide, Chlorpropamide, Glipizide, Glyburide, Glimepiride, Repaglinide, Nateglinide, Metformin, Rosiglitazone, Pioglitazone, Troglitazone, Acarbose, Miglitol, Sitagliptin, and Exenatide.

**Genotyping and quality control**

*DNA Extraction* In the MEC and WHI, DNA was purified from buffy coat samples. A subset of MEC DNA samples were whole-genome amplified by Molecular Staging Inc. following their standard protocol. For CALiCo and BioMe studies, DNA was extracted from blood samples drawn at baseline.

*Genotyping* Study specific details are summarised in ESM Table 1. Genotyping was performed using the Metabochip, a custom Illumina iSelect genotyping array. The Metabochip includes nearly 200,000 SNP markers and was designed to cost-effectively analyse putative association signals identified through GWAS meta-analyses of many obesity-related metabolic and cardiovascular traits and to fine-map established loci. Metabochip SNPs were selected from the catalogs developed by the International HapMap and 1000 Genomes Projects. More than 122,000 SNPs were included to fine-map 257 GWAS loci of 23 traits. The boundaries around each GWAS index SNP were determined by identifying all SNPs with $r^2 \geq 0.5$ with the index SNP, and then expanding the initial boundaries by 0.02cM in either direction using the HapMap-based genetic map. SNPs were
excluded if (a) the Illumina design score was <0.5 or (b) there were SNPs within 15bp in both
directions of the SNP of interest with allele frequency of >0.02 among Europeans (CEU). SNPs
annotated as nonsynonymous, essential splice site, or stop codon were included regardless of allele
frequency, design score, or nearby SNPs in the primer.

Samples were genotyped at the Human Genetics Center of the University of Texas-Houston (ARIC
and CARDIA), the University of Southern California Epigenome Center (MEC), the Genomics Core
Facility of the Icahn Institute for Genomics and Multiscale Biology, Icahn School of Medicine at
Mount Sinai (BioMe) and the Translational Genomics Research Institute (WHI). Each center
genotyped the same 90 HapMap YRI (Yoruba in Ibadan, Nigeria) samples to facilitate cross-study
quality control (QC), as well as 2-3% study-specific blinded replicates to assess genotyping quality.
Genotypes were called separately for each study using GenomeStudio with the GenCall 2.0
algorithm. Samples were called using study-specific cluster definitions (based on samples with call
rate >95%, ARIC, CARDIA, MEC, WHI, BioMe) and kept in the analysis if call rate was >95%. We
excluded SNPs with GenTrain score <0.6 (ARIC, CARDIA, MEC, WHI), Proper info score ≥ 0.4
(BioMe), cluster separation score <0.4, call rate <0.95, and Hardy-Weinberg Equilibrium p value <
1×10^{-6}. We utilised the common 90 YRI samples and excluded any SNP that had more than 1
Mendelian error (in 30 YRI trios), any SNP that had more than two replication errors with
discordant calls when comparisons were made across studies in 90 YRI samples, and any SNP that
had more than three discordant calls for 90 YRI genotyped in PAGE versus the HapMap database.
SNPs were excluded from the meta-analyses if they were present in less than three studies.

**Replication and Fine-mapping of known glycaemic trait loci**

For replication of index SNPs from GWAS, we used a nominal significance level (p value = 0.05). For
transethnic signal fine-mapping we used the locus-specific p value (ranging from α = 1.41 × 10^{-5} to α
= 4.1 × 10^{-4}), which is 0.05 divided by the number of variants passing quality control at each locus.

To fine-map previously identified loci, we investigated the patterns of association at 15 known
fasting glucose and fasting insulin loci using meta-analysis results from 13,613, and 2,406 SNPs
genotyped on the Metabochip in up to 26,760, and 22,674 individuals for fasting glucose, and
fasting insulin, respectively. At loci that exhibited evidence of regional significance (0.05/# of SNPs
in region), we performed a series of sequential conditional analyses adding the most significant
lead SNP into the regression model as an additional covariate and testing all remaining regional SNPs not already in the model as a covariate for association. Sequential conditional analyses were performed adding in lead SNPs to the model until the strongest SNP association showed a conditional \( p \) value > than the regional significance level.

Regional plots for each locus were created showing data publicly available from the Scott et al European MetaboChip analysis and using both the European lead SNP and the new lead SNPs from the trans-ethnic meta-analysis or from one of the population specific analyses. If the lead SNP from the current analysis was not present in the publicly available European MetaboChip results a suitable proxy \( (r^2 > 0.8) \) was used. The plots were generated on the LocusZoom web-based plotting software using LD information from the 1000 Genomes Project in the most representative population European (hg19/Mar2012EUR data), AA (hg19/ Mar2012AFR data), H/L (hg19/Mar2012AMR data).

**Strategy for selecting novel associations**

The MetaboChip selected genotyping content for type 2 diabetes, 2 hour glucose, glycated hemoglobin, fasting glucose, fasting insulin, myocardial infarction and coronary artery disease, high-density lipoprotein/low-density lipoprotein /triglycerides/total cholesterol, BMI, waist to hip ratio, body fat percentage, height, waist circumference, diastolic/systolic blood pressure, QT interval, mean platelet volume, platelet count, and white blood cell count. Given that MetaboChip SNPs were included specifically due to prior evidence for association with these traits, the MetaboChip-wide analyses were defined as testing for pleiotropy with any of these cardio, metabolic traits. Novel trait associations were subsequently investigated in the GWAS catalog to identify which trait(s) the locus was previously reported for. MetaboChip-wide results were considered statistically significant if they reached a threshold of \( p \) value \( \leq 2.5 \times 10^{-7} \) \((0.05/196975)\), and were not in LD \((r^2 < 0.2)\) or within 500 kb of a reported index SNP. Secondary independent or population-specific associations in the 15 fine-mapping regions are reported as being significant if they met region specific significance of \(0.05/\#\) SNPs in the locus.

**Statistical analysis**

All analyses were adjusted for age, sex, smoking status (current versus former/never), BMI, and ancestry principal components (PCs) in each study. Some studies also adjusted for center when applicable.
Each study performed race/ethnic specific analyses to test the association between continuous fasting glucose or natural log-transformed fasting insulin levels with genotypes or imputed dosages assuming an additive mode of inheritance. For studies of unrelated individuals, we applied multiple linear regression including age, sex, center site (as applicable), smoking status (current versus former/never), continuous BMI, and ancestry PCs (number varied by study) as model covariates. Like previous studies, primary analyses adjusted for BMI because it is a major risk factor for type 2 diabetes and is correlated with glycaemic traits. For sensitivity analysis, all models were also run without BMI as a covariate. Adjustment for smoking was decided a priori given the racial/ethnic differences in smoking patterns in the US and incident type 2 diabetes, as well as the association of cotinine with glycaemic related traits in non-diabetics [10]. HCHS/SOL has approximately 2000 related individuals, and a complex sampling design was used for recruitment. Therefore, in this study we employed the W-PS method (http://dlin.web.unc.edu/software/SUGEN/) by Lin et al, which is a weighted version of generalised estimation equations, to account for unequal inclusion probabilities and family relationships [11].

We combined SNP effect estimates and their standard errors across studies for H/L, AA, NA/AI and ASN by inverse-variance weighted fixed effects meta-analysis using METAL [12]. Results from these ethnic/race-specific meta-analyses are presented. Quantile-quantile plots for the Metabochip-wide analysis of fasting glucose and fasting insulin are shown in ESM Figure 4. Consistency between study/race-ethnicity effect size was assessed using the $Q$ test (Chi-squared p value) and the $I^2$ metric, where low $I^2$ suggests little difference between study/race-ethnicity variability. A two-stage conventional fixed-effects analysis approach was chosen to enable comparison of effects between races/ethnicities and because this strategy has been shown to provide a well-controlled type 1 error [13, 14]. Furthermore, $I^2$ metrics indicated little heterogeneity between race/ethnicities for the significant trans-ethnic results. Results for all analyses were reported as betas and standard errors (SE).

**Functional Annotation**

We interrogated each of the fine-mapped loci to determine if the identified non-coding variants were positioned within regulatory regions such as enhancers, promoters, insulators and/or silencers, which can potentially modulate transcript levels and thereby explain the underlying biology of the glucose/insulin association. To identify variants that overlapped putative regulatory elements we aligned a custom track with a list of correlated variants ($r^2 \geq 0.2$ in 1000 Genome
Project AMR and AFR) at each locus with several relevant tracks. Bedfiles were obtained using http://raggr.usc.edu/. Given that our traits of interest were glycaemic loci, we particularly utilised DNAseI hypersensitivity, histone modification and transcription factor occupancy data assayed in tissues and cell types derived from pancreatic, pancreatic islet, brain or other metabolically relevant tissues highlighted by expression patterns of strong candidate genes. By integrating the signal from several epigenomics tracks in the aforementioned tissues/cell types, a genomic element was annotated to be a putative enhancer if it showed enrichment for signal in DNASe hypersensitivity and H3K4me1/H3k27ac marks; as a promoter if it showed enrichment for signal in DNAse hypersensitivity and H3K4me3; as an insulator if it showed enrichment for signal for DNAse hypersensitivity, CTCF binding; and as a silencer if it showed enrichment for signal for DNAse Hypersensitivity and H3K9me3. A summary of the annotation for variants overlapping regulatory elements including noteworthy features like occupancy by key transcription factors and potentially interesting nearest genes candidates are provided in ESM Table 19. The epigenomics datasets were sourced from publicly available projects including ENCODE, Roadmap Epigenomics, and FANTOM5. Furthermore, to identify the motifs disrupted by putative risk alleles, we utilised Haploreg (v5) and the JASPAR motif database. To query GTEx eQTL data we utilised Haploreg (v5). Splice site misregulation was predicted using SPANR [15].

In addition to the use of epigenomic data, we also included summary functional scores from in silico prediction algorithms including RegulomeDB, the Combined Annotation Dependent Depletion (CADD) score, a PHRED-like score indicating deleteriousness of variant versus all other substitutions in genome. Additionally, we utilized the UCSC Genome Browser to interrogate TargetScan miRNA Regulatory Sites from 3’ UTR regions.
References

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

2. Friedman GD, Cutter GR, Donahue RP, et al. (1988) CARDIA: study design, recruitment, and some characteristics of the examined subjects. J Clin Epidemiol 41:1105–1116.

3. Hughes GH, Cutter G, Donahue R, et al. (1987) Recruitment in the Coronary Artery Disease Risk Development in Young Adults (Cardia) Study. Control Clin Trials 8:68S–73S.

4. Sorlie PD, Avilés-Santa LM, Wassertheil-Smoller S, et al. (2010) Design and implementation of the Hispanic Community Health Study/Study of Latinos. Ann Epidemiol 20:629–641.

5. The Women’s Health Initiative Study Group. (1998) Design of the Women’s Health Initiative clinical trial and observational study. Control Clin Trials 19:61–109.

6. Kolonel LN, Henderson BE, Hankin JH, et al. (2000) A multiethnic cohort in Hawaii and Los Angeles: baseline characteristics. Am J Epidemiol 151:346–357.

7. Gottesman O, Kuivaniemi H, Tromp G, et al. (2013) The Electronic Medical Records and Genomics (eMERGE) Network: past, present, and future. Genet Med 15:761–771.

8. Howie BN, Donnelly P, Marchini J (2009) A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. PLoS Genet 5:e1000529. doi: 10.1371/journal.pgen.1000529

9. 1000 Genomes Project Consortium, Abecasis GR, Auton A, et al. (2012) An integrated map of genetic variation from 1,092 human genomes. Nature 491:56–65.

10. Liu R, Zheng Z, Du J, et al. (2016) Racial Disparity in the Associations of Cotinine with Insulin Secretion: Data from the National Health and Nutrition Examination Survey, 2007-2012. PLoS ONE 11:e0167260. doi: 10.1371/journal.pone.0167260

11. Lin D-Y, Tao R, Kalsbeek WD, et al. (2014) Genetic association analysis under complex survey sampling: the Hispanic Community Health Study/Study of Latinos. Am J Hum Genet 95:675–688.

12. Willer CJ, Li Y, Abecasis GR (2010) METAL: fast and efficient meta-analysis of genomewide association scans. Bioinformatics 26:2190–2191.

13. Hong J, Lunetta KL, Cupples LA, et al. (2016) Evaluation of a Two-Stage Approach in Trans-Ethnic Meta-Analysis in Genome-Wide Association Studies. Genet Epidemiol 40:284–292.

14. Franceschini N, Carty CL, Lu Y, et al. (2016) Variant Discovery and Fine Mapping of Genetic Loci Associated with Blood Pressure Traits in Hispanics and African Americans. PLoS ONE 11:e0164132. doi: 10.1371/journal.pone.0164132
15. Xiong HY, Alipanahi B, Lee LJ, et al. (2015) RNA splicing. The human splicing code reveals new insights into the genetic determinants of disease. Science 347:1254806.

16. Gao N, LeLay J, Vatamaniuk MZ, et al. (2008) Dynamic regulation of Pdx1 enhancers by Foxa1 and Foxa2 is essential for pancreas development. Genes Dev 22:3435–3448.

17. Sund NJ, Vatamaniuk MZ, Casey M, et al. (2001) Tissue-specific deletion of Foxa2 in pancreatic beta cells results in hyperinsulinemic hypoglycemia. Genes Dev 15:1706–1715.

18. Zou F, Chai HS, Younkin CS, et al. (2012) Brain expression genome-wide association study (eGWAS) identifies human disease-associated variants. PLoS Genet 8:e1002707.

19. GTEx Consortium (2015) Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. Science 348:648–660.

20. Lyssenko V, Nagorny CLF, Erdos MR, et al. (2009) Common variant in MTNR1B associated with increased risk of type 2 diabetes and impaired early insulin secretion. Nat Genet 41:82–88.
Comparison of effect estimates for index SNP associations for fasting glucose and insulin between trans-ethnic (TE) meta-analysis and published European (EUR) Metabochip. Results are shown for models adjusted for BMI. PAGE transethnic meta-analysis results from Metabochip are compared to the Scott et al study of European-descent MetaboChip results in up to 99,029 samples for fasting insulin and 118,881 samples for fasting glucose.
ESM Figure 2
Regional plots for fine-mapping loci comparing Metabochip results for MAGIC European population (top) to PAGE transethnic meta-analysis (bottom). Each SNP in the 8 fasting glucose loci and 1 fasting insulin locus with evidence of fine-mapping is plotted with association results, -log10(p value), on the y-axis, and chromosomal position on the x-axis. Fasting glucose results are shown for: (a,b) GCKR, (c,d) G6PC2, (e,f) ADCY5, (g,h) DGKB, (i,j) GCK, (k,l) GLIS3, (m,n) FADS2, (o,p) MTNR1B. Fasting insulin results are shown for (q,r) GCKR. Not all SNPs used in the transethnic meta-analysis were present in the available MAGIC data (http://www.magicinvestigators.org/downloads/) due to mapping.
issues [16]. SNPs not passing QC or outside of fine-mapping region were removed from the TE plots. The colour scale legend indicates linkage disequilibrium (r²) between each fine-mapping SNP and the GWAS index SNPs (purple diamonds), which was calculated using 1000 Genomes Populations (CEU for MAGIC and AMR or AFR for PAGE). The population chosen for LD colouring in the tranethnic meta-analysis was based on population-specific analysis results (choosing the population with the most significant underlying SNP associations, AMR for all except GCKR (b), and ADCY5 (f). Solid line corresponds to locus specific significance threshold.

ESM Figure 3
### Table of SNPs

| SNP | Chromosome | Position |
|-----|------------|----------|
| rs11688384 | chr2 | 169,735,000 |
| rs560867 | chr2 | 169,740,000 |
| rs573225 | chr2 | 169,745,000 |
| rs76372731 | chr2 | 169,750,000 |
| 4510:169735034:A:C | chr2 | 169,755,000 |
| rs116279278 | chr2 | 169,760,000 |
| rs15507862 | chr2 | 169,765,000 |
| rs557462 | chr2 | 169,770,000 |
| rs508506 | chr2 | 169,775,000 |
| rs34453330 | chr2 | 169,780,000 |
| rs569628 | chr2 | 169,785,000 |

### Diagram: UCSC Genes

- **Pancreas LmSig**: Uniformly Signal from Roadmap
- **Pancreatic_Islets H3K4me3 Uniformly Signal from Roadmap**
- **Layered H3K27Ac**: Mark (Often Found Near Active Regulatory Elements) on 7 cell lines from ENCODE
- **DNase Hypersensitivity Clusters in 125 cell types from ENCODE (v3)**
- **Transcription Factor ChiP-seq (161 factors) from ENCODE with Factorbook Motifs**
UCSC Genome Browser screen shots (http://genome.ucsc.edu, 24 August 2016) of the bioinformatic follow-up using Roadmap and ENCODE data are shown for eleven loci (a) PROX1, (b) G6PC2, (c) ADCY5, (d) DGKB, (e) GCK, (f) GLIS3, (g) FADS2, (h) GCKR, (i) IGF1, (j) MTNR1B, (k) SLC2A2. Methylation and acetylation modifications of the histone in metabolically relevant tissues (e.g. pancreatic islets) are shown. Regulatory elements within histone modifications are further demarked by open chromatin structure (DNaseI-seq), which is correlated with transcription factor binding. Finally, available ChIP-seq transcription factor binding sites are shown. Blue highlighting indicates a putative function element harboring either the lead SNP or SNPs in LD with the lead. Further details about eQTL, and specific transcription factors is given in ESM Table 7. In the PROX1 locus (a), rs10494973, overlaps an intronic pancreatic enhancer and a FOXA2 binding site, which is an essential activator of genes governing insulin secretion. In the G6PC2 locus (b), rs560887 is a known functional splicing variant and bioinformatic annotation supports this variant as the strongest functional candidate. 9 LD SNPs in the ADCY5 locus (c) overlap 4 pancreatic enhancers and were shown to be associated with ADCY5 and SEC22A expression in multiple tissues. In the DGKB locus (d), 7 SNPs in LD with the lead, rs62448618, overlap a strong intergenic pancreatic islet enhancer. In the 7p13 locus (e), 7 variants in LD with the lead are positioned in 4 pancreatic islet enhancer/promoter regions that were also an eQTL for GCK in thyroid tissue. In the GLIS3 locus (f), the lead SNP, rs10974438, is associated with expression of GLIS3 in brain tissue and is in LD with 3 SNPs positioned in two pancreatic islet enhancers. In the FADS3-11q12.2 region (g), the lead SNP, rs174547, was in LD with 6 variants spanning 3 enhancer regions. In the GCKR region (h), the lead SNP, rs1260326, is a missense variant in exon 15 of GCKR that is predicted by SPANR to disrupt splicing and is associated with expression of GCKR in brain tissue. Additionally, 13 variants in LD with the lead span across six predicted enhancers. In the IGF1 locus (i), rs10860845 tagged two putative functional variants positioned within two different pancreatic islet intronic enhancers. The lead SNP in the MTNR1B locus (j), rs10830963, falls in a brain and pancreatic islet/pancreas enhancers in the intronic region of MTNR1B, which also bind the transcription factors FOXA1 and FOXA2. In the SLC2A2 locus (k), SNPs in LD with the lead SNP, rs1604038 are associated with SLC2A2 and EIF5A2 expression in brain tissue, overlap 7 pancreatic islet/pancreas enhancers, and one LD SNP was a missense variant predicted by SPANR to misregulate splicing of SLC2A2.
Quantile-Quantile (Q-Q) plots for MetaboChip-wide fasting glucose and fasting insulin results. Expected and observed distributions of the \(-\log_{10}(p)\) values for MetaboChip SNP associations with: fasting glucose before (a) and after (b) BMI adjustment, and natural log transformed fasting insulin before (c) and after (d) BMI adjustment. Gray colouring of SNPs show results after removing loci known to be associated with type 2 diabetes and associated biomarkers (fasting glucose, fasting insulin, HbA1C, 2 hour glucose), anthropometric traits like BMI, hip/waist ratio, and waist circumference, as well as levels of high/low density lipoprotein and triglycerides.
ESM Figure 5

a
Fasting glucose and fasting insulin association $p$ values for each Metabochip variant from the transethnic meta-analysis in models with and without BMI. The $-\log_{10}$ of $p$ values for each SNP on the Metabochip is plotted against chromosomal positions. Fasting glucose model results with and without BMI covariate are shown in a) and b), respectively. Fasting insulin model results with BMI covariate are shown in c. Gray
and dark gray colouring of SNPs alternates by chromosome. Dark blue dots represent the SNPs in previously reported glycaemic trait loci (within 1 Mb of index SNP). Green dots correspond to novel (unobserved for these models). The blue solid line is the threshold for Metabochip-wide significance $\alpha = 2.5 \times 10^{-7}$; dashed blue line is threshold for genome-wide significance $\alpha = 5.0 \times 10^{-8}$. 