Multi-ancestry genetic study of type 2 diabetes highlights the power of diverse populations for discovery and translation

ARTICLES
https://doi.org/10.1038/s41588-022-01058-3
nature genetics

Multi-ancestry genetic study of type 2 diabetes highlights the power of diverse populations for discovery and translation

We assembled an ancestrally diverse collection of genome-wide association studies (GWAS) of type 2 diabetes (T2D) in 180,834 affected individuals and 1,159,055 controls (48.9% non-European descent) through the Diabetes Meta-Analysis of Trans-Ethnic association studies (DIAMANTE) Consortium. Multi-ancestry GWAS meta-analysis identified 237 loci attaining stringent genome-wide significance ($P < 5 \times 10^{-9}$), which were delineated to 338 distinct association signals. Fine-mapping of these signals was enhanced by the increased sample size and expanded population diversity of the multi-ancestry meta-analysis, which localized 54.4% of T2D associations to a single variant with >50% posterior probability. This improved fine-mapping enabled systematic assessment of candidate causal genes and molecular mechanisms through which T2D associations are mediated, laying the foundations for functional investigations. Multi-ancestry genetic risk scores enhanced transferability of T2D prediction across diverse populations. Our study provides a step toward more effective clinical translation of T2D GWAS to improve global health for all, irrespective of genetic background.

The global prevalence of T2D has quadrupled over the last 30 years[1], affecting approximately 392 million individuals in 2015 (ref. [2]). Despite this worldwide impact, the largest T2D GWAS have predominantly featured populations of European ancestry[3–6], compromising prospects for clinical translation. Failure to detect causal variants that contribute to disease risk outside European ancestry populations limits progress toward a full understanding of disease biology and constrains opportunities for development of therapeutics[7]. Implementation of personalized approaches to disease management depends on accurate prediction of individual risk, irrespective of ancestry. However, genetic risk scores (GRS) derived from European ancestry GWAS provide unreliable prediction when deployed in other population groups, in part reflecting differences in effect sizes, allele frequencies and patterns of linkage disequilibrium (LD)[8].

To address the impact of this population bias, recent T2D GWAS have included individuals of non-European ancestry[9–11]. The DIAMANTE Consortium was established to assemble T2D GWAS across diverse ancestry groups. Analyses of the European and East Asian ancestry components of the DIAMANTE study have previously been reported[6,10]. Here, we describe the results of our multi-ancestry meta-analysis, which expands on these published components to a total of 180,834 individuals with T2D and 1,159,055 controls, with 20.5% of the effective sample size ascertained from African, Hispanic and South Asian ancestry groups. With these data, we demonstrate the value of analyses conducted on diverse populations to understand how T2D-associated variants impact downstream molecular and biological processes underlying the disease and advance clinical translation of GWAS findings for all, irrespective of genetic background.

Results

Study overview. We accumulated association summary statistics from 122 GWAS for 180,834 individuals with T2D and 1,159,055 controls (effective sample size, 492,191) across five ancestry groups (Supplementary Tables 1–3). We use the term 'ancestry group' to refer to individuals with similar genetic background: European ancestry (51.1% of the total effective sample size); East Asian ancestry (28.4%); South Asian ancestry (8.3%); African ancestry, including recently admixed African American populations (6.6%); and Hispanic individuals with recent admixture of American, African and European ancestry (5.6%). Each ancestry-specific GWAS was imputed to reference panels from the 1000 Genomes Project[12,13], the Haplotype Reference Consortium[14] or population-specific whole-genome sequence data. Subsequent association analyses were adjusted for population structure and relatedness (Supplementary Table 4). We considered 19,829,461 biallelic autosomal single-nucleotide variants (SNVs) that overlapped reference panels with minor allele frequency >0.5% in at least one of the five ancestry groups (Extended Data Fig. 1 and Methods).

Robust discovery of multi-ancestry T2D associations. We aggregated association summary statistics via multi-ancestry meta-regression, implemented in MR-MEGA[15], which models allelic effect heterogeneity correlated with genetic ancestry. We included three axes of genetic variation as covariates that separated genome-wide associations from the five major ancestry groups (Extended Data Fig. 2 and Methods). We identified 277 loci associated with T2D at the conventional genome-wide significance threshold of $P < 5 \times 10^{-8}$ (Extended Data Fig. 3 and Supplementary Table 5). By accounting for ancestry-correlated allelic effect heterogeneity in the multi-ancestry meta-regression, we observed lower genomic control inflation ($\lambda_{GC} = 1.05$) than when using either fixed- or random-effects meta-analysis ($\lambda_{GC} = 1.25$ under both models) and stronger signals of association at lead SNVs at most loci (Extended Data Fig. 4). Of the 277 loci, 11 have not previously been reported in recently published T2D GWAS meta-analyses[6,10,11] that account for 78.6% of the total effective sample size of this multi-ancestry meta-regression (Extended Data Fig. 3 and Supplementary Note). Of the 100 and 193 loci attaining genome-wide significance ($P < 5 \times 10^{-8}$) in East Asian and European ancestry-specific meta-analyses, respectively, lead SNVs at 94 (94.0%) and 164 (85.0%) demonstrated stronger evidence for association (smaller P values) in the multi-ancestry meta-regression (Extended Data Fig. 5 and Supplementary Note). These results demonstrate the power of multi-ancestry meta-analyses for locus discovery afforded by

A full list of author and affiliations appears at the end of the paper.

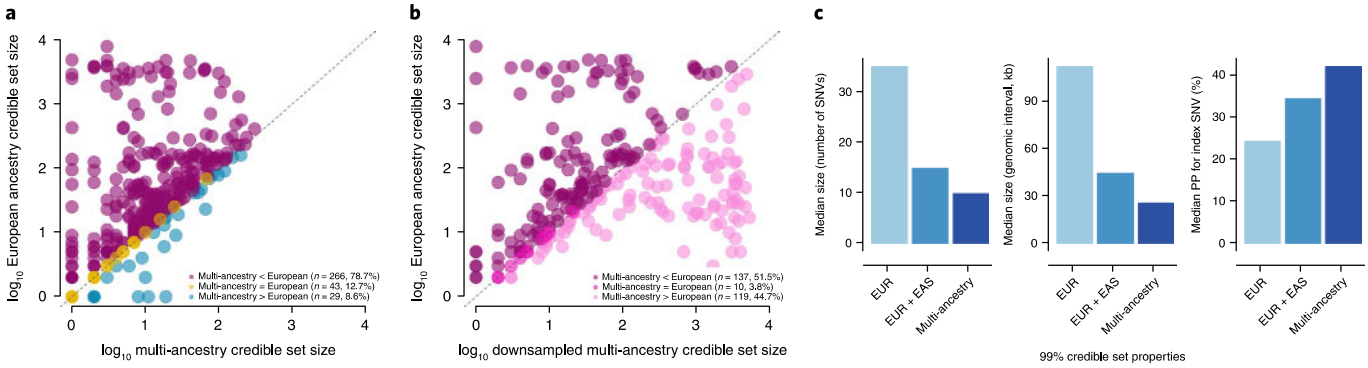


Fig. 1 | Comparison of fine-mapping resolution for distinct association signals for T2D obtained from ancestry-specific meta-analysis and multi-ancestry meta-regression. a, Each point corresponds to a distinct association signal, plotted according to the $\log_{10}$ credible set size in the multi-ancestry meta-regression on the *x* axis and the $\log_{10}$ credible set size in the European ancestry meta-analysis on the *y* axis. The 266 (78.7%) signals above the dashed $y = x$ line were more precisely fine-mapped in the multi-ancestry meta-regression. **b**, We 'downsampled' the multi-ancestry meta-regression to the effective sample size of the European ancestry-specific meta-analysis. Each point corresponds to one of the 266 signals that were more precisely fine-mapped in the multi-ancestry meta-regression. The 137 (51.5%) signals above the dashed $y = x$ line were more precisely fine-mapped in the 'downsampled' multi-ancestry meta-regression than in the equivalently sized European ancestry-specific meta-analysis. **c**, Properties of 99% credible sets of variants driving each distinct association signal in European (EUR) ancestry-specific meta-analysis, combined East Asian (EAS) and European ancestry meta-analysis and multi-ancestry meta-regression. The inclusion of the most under-represented ancestry groups (African, Hispanic and South Asian) in the multi-ancestry meta-regression reduced the median size of 99% credible sets and increased the median posterior probability (PP) ascribed to index SNVs.

increased sample size but also emphasize the importance of complementary ancestry-specific GWAS for identification of associations that are not shared across diverse populations.

The conventional genome-wide significance threshold does not allow for different patterns of LD across diverse populations in multi-ancestry meta-analysis. We therefore derived a multi-ancestry genome-wide significance threshold of $P < 5 \times 10^{-9}$ by estimating the effective number of independent SNVs across the five ancestry groups using haplotypes from the 1000 Genomes Project reference panel[13] (Methods). Of the 277 loci reported in this multi-ancestry meta-regression, 237 attained the more stringent significance threshold, which we considered for downstream analyses. Through approximate conditional analyses, conducted using ancestry-matched LD reference panels for each GWAS, we partitioned associations at the 237 loci into 338 distinct signals that were each represented by an index SNV at the same multi-ancestry genome-wide significance threshold (Methods, Supplementary Tables 6–8 and Supplementary Note). Allelic effect estimates for distinct association signals from approximate conditional analyses undertaken in admixed ancestry groups were robust to the choice of reference panel (Supplementary Note).

Allelic effect heterogeneity across ancestry groups. Allelic effect heterogeneity between ancestry groups can occur for several reasons, including differences in LD with causal variants or interactions with environment or polygenic background across diverse populations. An advantage of the multi-ancestry meta-regression model is that heterogeneity can be partitioned into two components. The first captures heterogeneity that is correlated with genetic ancestry (that is, it can be explained by the three axes of genetic variation). The second reflects residual heterogeneity due to differences in geographical location (for example, different environmental exposures) and study design (for example, different phenotype definition, case–control ascertainment or covariate adjustments between GWAS). We observed 136 (40.2%) distinct T2D associations with nominal evidence ($P_{HET} < 0.05$) of ancestry-correlated heterogeneity compared to 16.9 expected by chance (binomial test $P < 2.2 \times 10^{-16}$). By contrast, there was nominal evidence of residual heterogeneity at only 27 (8.0%) T2D-association signals (binomial test $P = 0.0037$), suggesting that differences in allelic effect size between GWAS are more likely due to factors related to genetic ancestry than to geography and/or study design (Supplementary Note).

Population diversity improves fine-mapping resolution. We sought to quantify the improvement in fine-mapping resolution offered by increased sample size and population diversity in the multi-ancestry meta-regression. For each of the 338 distinct signals, we first derived multi-ancestry and European ancestry-specific credible sets of variants that account for 99% of the posterior probability (π) of driving the T2D association under a uniform prior model of causality (Methods). Multi-ancestry meta-regression substantially reduced the median 99% credible set size from 35 variants (spanning 112 kb) to ten variants (spanning 26 kb) and increased the median posterior probability ascribed to the index SNV from 24.3% to 42.0%. The 99% credible sets for 266 (78.7%) distinct T2D associations were smaller in the multi-ancestry meta-regression than in the European ancestry-specific meta-analysis, while a further 26 (7.7%) signals were resolved to a single SNV in both (Fig. 1, Supplementary Table 9 and Supplementary Note). Causal variant localization was also more precise in the multi-ancestry meta-regression than in a meta-analysis of GWAS of European and East Asian ancestry, which together account for 79.5% of the total effective sample size, highlighting the important contribution of the most under-represented ancestry groups (African, Hispanic and South Asian) to fine-mapping resolution (Fig. 1 and Supplementary Note).

We next attempted to understand the relative contributions of population diversity and sample size to these improvements in fine-mapping resolution at the 266 distinct T2D associations that were more precisely localized after the multi-ancestry meta-regression. We downsampled studies contributing to the multi-ancestry meta-regression to approximate the effective sample size of the European ancestry-specific meta-analysis, while maintaining the distribution of population diversity (Methods and Supplementary Table 10). The associations were better resolved in the downsampled multi-ancestry meta-regression at 137 signals (51.5%), compared with 119 signals (44.7%) in the European ancestry-specific meta-analysis (Fig. 1 and Supplementary Table 11). These results highlight the value of diverse populations for causal variant localization in multi-ancestry meta-analysis, emphasizing the importance of increased sample size and differences in LD structure and

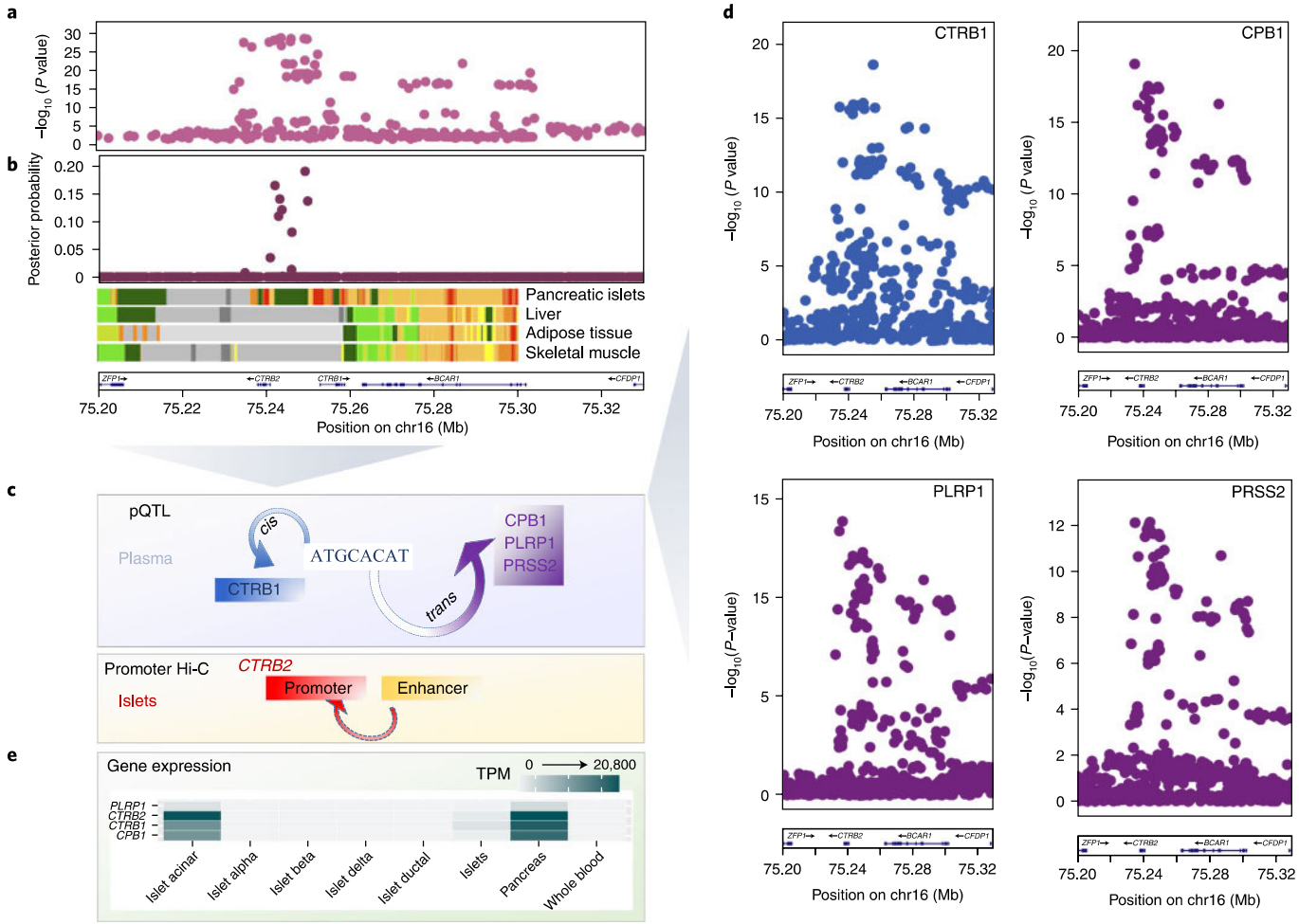


Fig. 2 | T2D-association signal at the *BCAR1* locus colocalizes with multiple circulating plasma pQTL. a, Signal plot for T2D association from multi-ancestry meta-regression of 180,834 affected individuals and 1,159,055 controls of diverse ancestry. Each point represents an SNV, plotted with its *P* value (on a $\log_{10}$ scale) as a function of genomic position (National Center for Biotechnology Information (NCBI) build 37). Gene annotations were taken from the University of California Santa Cruz genome browser. Recombination rates were estimated from the Phase II HapMap. Chr, chromosome. **b**, Fine-mapping of T2D-association signals from multi-ancestry meta-regression. Each point represents an SNV plotted with its posterior probability of driving T2D association as a function of genomic position (NCBI build 37). Chromatin states are presented for four diabetes-relevant tissues: active transcription start sites (TSS) (red), flanking active TSS (orange–red), strong transcription (green), weak transcription (dark green), genic enhancers (green–yellow), active enhancers (orange), weak enhancers (yellow), bivalent or poised TSS (Indian red), flanking bivalent TSS or enhancer (dark salmon), repressed polycomb (silver), weak repressed polycomb (gainsboro) and quiescent or low (white). **c**, Schematic presentation of the single *cis* and multiple *trans* effects mediated by the *BCAR1* locus on plasma proteins and the islet chromatin loop between islet enhancer and promoter elements near *CTRB2*. **d**, Signal plots for four circulating plasma proteins that colocalize with the T2D association in 3,301 European ancestry participants from the INTERVAL study. Each point represents an SNV, plotted with its *P* value (on a $\log_{10}$ scale) as a function of genomic position (NCBI build 37). **e**, Expression of genes (TPM, transcripts per million) encoding colocalized proteins in islets, the pancreas and whole blood.

allele-frequency distribution between ancestry groups that has also been reported for other complex human traits[16].

Multi-ancestry fine-mapping to single-variant resolution. Previous T2D GWAS have demonstrated improved localization of causal variants through integration of fine-mapping data with genomic annotation[6,17]. By mapping SNVs to three categories of functional and regulatory annotation, with an emphasis on diabetes-relevant tissues[18], we observed significant joint enrichment ($P < 0.00023$, Bonferroni correction for 220 annotations) for T2D associations mapping to protein-coding exons, transcription factor binding sites for NKX2.2, FOXA2, EZH and PDX1, and four chromatin states in pancreatic islets that mark active enhancers, active promoters and transcribed regions (Methods, Extended Data Fig. 6 and Supplementary Table 12). We used the enriched annotations to derive a prior model for causality and redefined 99% credible sets of variants for each distinct signal (Methods and Supplementary Table 13). Annotation-informed fine-mapping reduced the size of the 99% credible set, compared to the uniform prior, at 144 (42.6%) distinct association signals (Extended Data Fig. 7) and decreased the median from ten variants (spanning 26 kb) to eight variants (spanning 23 kb). For 184 (54.4%) signals, a single SNV accounted for >50% of the posterior probability of the T2D association (Supplementary Table 14). At 124 (36.7%) signals, >80% of the posterior probability could be attributed to a single SNV.

Missense variants implicate candidate causal genes. After annotation-informed multi-ancestry fine-mapping, 19 of the 184 SNVs accounting for >50% of the posterior probability of the T2D association were missense variants (Supplementary Table 15). Two of

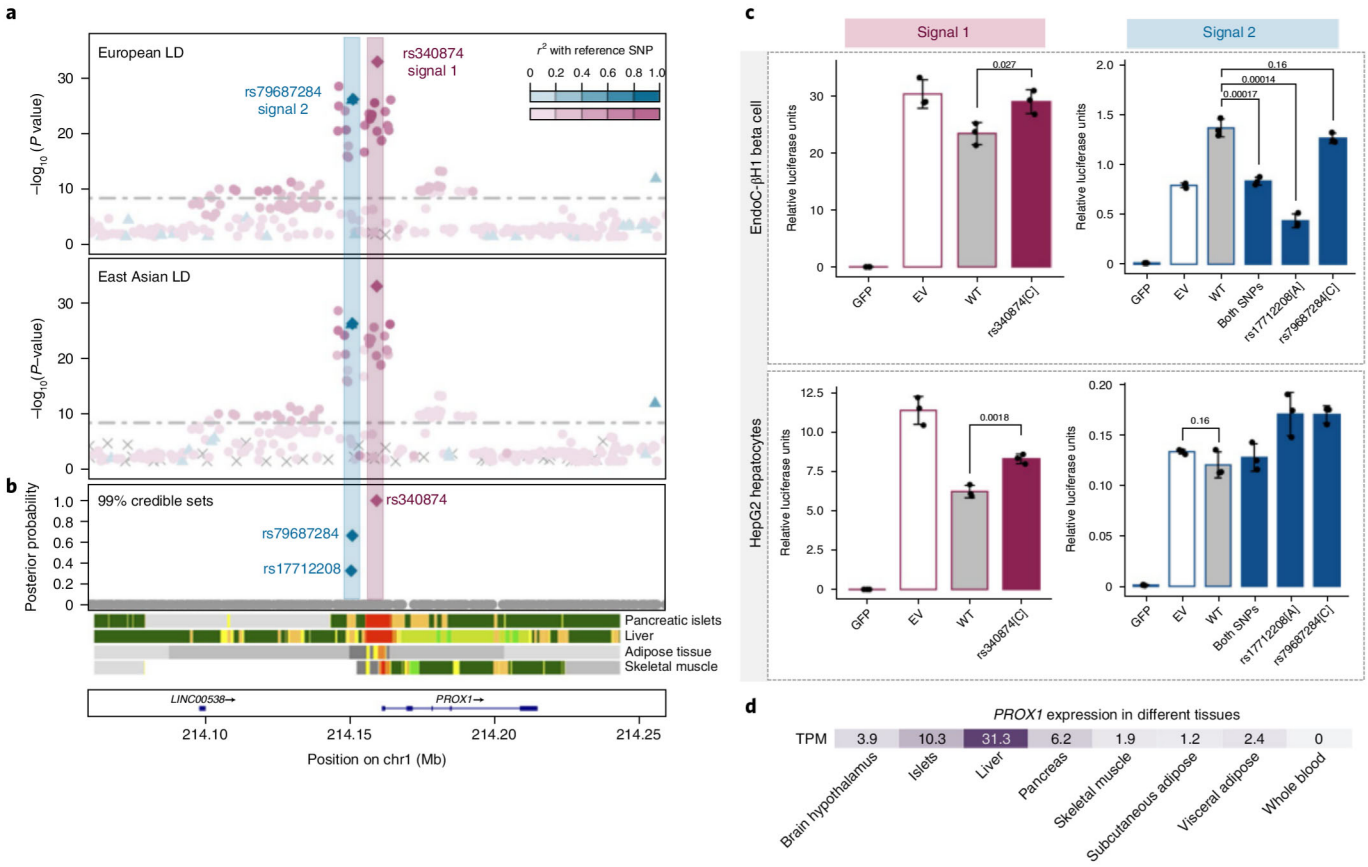


Fig. 3 | Defining causal molecular mechanisms at the *PROX1* locus. a, Signal plot for two distinct T2D associations from multi-ancestry meta-regression of 180,834 affected individuals and 1,159,055 controls of diverse ancestry. Each point represents an SNV, plotted with its *P* value (on a $-\log_{10}$ scale) as a function of genomic position (NCBI build 37). Index SNVs are represented by blue and purple diamonds. All other SNVs are colored according to the LD with the index SNVs in European and East Asian ancestry populations. Gene annotations were taken from the University of California Santa Cruz genome browser. **b**, Fine-mapping of T2D-association signals from multi-ancestry meta-regression. Each point represents an SNV plotted with its posterior probability of driving each distinct T2D association as a function of genomic position (NCBI build 37). The 99% credible sets for the two signals are highlighted by purple and blue diamonds. Chromatin states are presented for four diabetes-relevant tissues: active TSS (red), flanking active TSS (orange–red), strong transcription (green), weak transcription (dark green), genic enhancers (green–yellow), active enhancers (orange), weak enhancers (yellow), bivalent or poised TSS (Indian red), flanking bivalent TSS or enhancer (dark salmon), repressed polycomb (silver), weak repressed polycomb (gainsboro), quiescent or low (white). **c**, Transcriptional activity of the 99 credible set variants at the two T2D-association signals in human HepG2 hepatocytes and EndoC-βH1 beta cell models obtained from in vitro reporter assays. Biological replicates, $n=3$; technical replicates, $n=3$. WT, wild type (non-risk allele or haplotype); GFP, green fluorescent protein (negative control); EV, empty vector (baseline). Heights of bars represent means. Error bars represent s.e.m. Differences in luciferase activity between groups were tested using two-tailed two-sample *t*-tests, for which $P<0.05$ was considered statistically significant. **d**, Expression of *PROX1* across a range of diabetes-relevant tissues.

these implicate new candidate causal genes for the disease: *MYO5C* p.Glu1075Lys (rs3825801, $P=3.8\times10^{-11}$, $\pi=69.2\%$) at the *MYO5C* locus and *ACVR1C* p.Ile482Val (rs7594480, $P=4.0\times10^{-12}$, $\pi=95.2\%$) at the *CYTIP* locus. *ACVR1C* encodes ALK7, a transforming growth factor β receptor, overexpression of which induces growth inhibition and apoptosis of pancreatic beta cells[19]; *ACVR1C* p.Ile482Val has been previously associated with body fat distribution[20]. The multi-ancestry meta-regression also highlighted examples of previously reported associations that were better resolved by fine-mapping across diverse populations, including *SLC16A11*, *KCNJ11–ABCC8* and *ZFAND3–KCNK16–GLP1R* (Supplementary Note).

Multi-omics integration highlights candidate effector genes. We next sought to take advantage of the improved fine-mapping resolution offered by the multi-ancestry meta-regression to extend insights into candidate effector genes, tissue specificity and mechanisms through which regulatory variants at noncoding T2D-association signals impact disease risk. We integrated annotation-informed fine-mapping data with molecular quantitative trait loci (QTL) in *cis* for (1) circulating plasma proteins (pQTL)[21] and (2) gene expression (eQTL) in diverse tissues, including pancreatic islets, subcutaneous and visceral adipose tissue, liver, skeletal muscle and hypothalamus[22,23] (Methods). Bayesian colocalization[24] of each pair of distinct T2D associations and molecular QTL identified 97 candidate effector genes at 72 signals with posterior probability $\pi_{COLOC}>80\%$ (Supplementary Tables 16 and 17). The colocalizations reinforced evidence supporting several genes previously implicated in T2D through detailed experimental studies, including *ADCY5*, *STARD10*, *IRS1*, *KLF14*, *SIX3* and *TCF7L2* (refs. [25–29]). A single candidate effector gene was implicated at 49 T2D-association signals, of which ten colocalized with eQTL across multiple tissues: *CEP68*, *ITGB6*, *RBM6*, *PCGF3*, *JAZF1*, *ANK1*, *ABO*, *ARHGAP19*, *PLEKHA1* and *AP3S2*. By contrast, we observed that *cis* eQTL at 44 signals were specific to one tissue (24 to pancreatic islets, 11 to subcutaneous adipose tissue, five to skeletal muscle, two to visceral adipose tissue and one each to liver and hypothalamus),

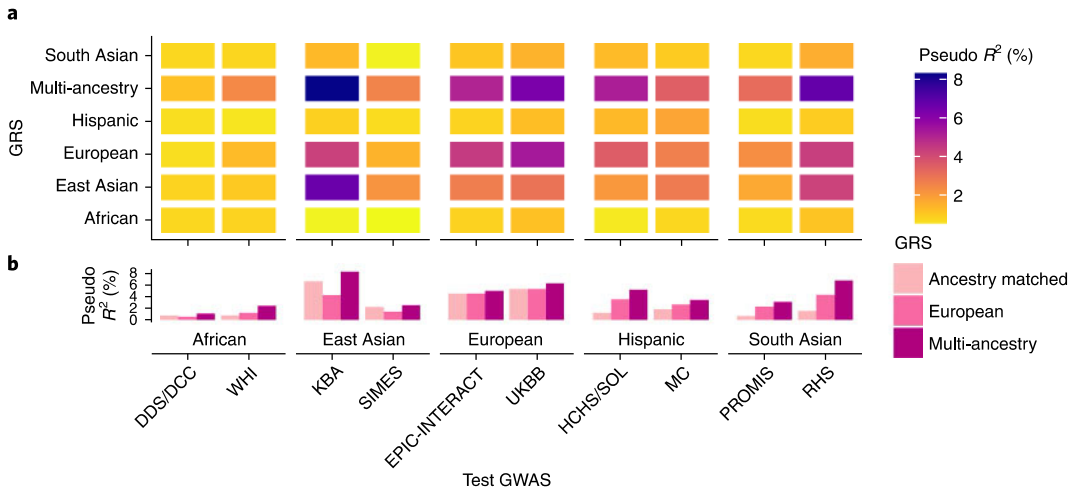


Fig. 4 | Transferability of multi-ancestry and ancestry-specific GRS into GWAS across diverse population groups. Each GRS was constructed using lead SNVs attaining genome-wide significance ($P < 5 \times 10^{-9}$ for multi-ancestry GRS and $P < 5 \times 10^{-8}$ for ancestry-specific GRS). For the multi-ancestry GRS, population-specific allelic effects on T2D were estimated from the meta-regression to generate different GRS weights for each test GWAS. Test GWAS acronyms are defined in Supplementary Table 1. For each ancestry-specific GRS, weights were generated from allelic effect estimates obtained from the fixed-effects meta-analysis. **a**, The trait variance explained (pseudo R^2) by each GRS was assessed in two test GWAS from each ancestry group. **b**, The multi-ancestry GRS out-performed ancestry-specific GRS into all test GWAS, reflecting the shared genetic contribution to T2D across diverse populations, despite differing allele frequencies and LD patterns.

emphasizing the importance of conducting colocalization analyses across multiple tissues. Genome-wide promoter-focused chromatin confirmation capture data (pcHi-C) from pancreatic islets, subcutaneous adipose tissue and liver (equivalent data are not available from hypothalamus or visceral adipose tissue)[30–32] provided complementary support for several candidate effector genes (Supplementary Table 18 and Supplementary Note). These results demonstrate how the increased fine-mapping resolution afforded by our multi-ancestry meta-analysis can be integrated with diverse molecular data resources to reveal putative mechanisms underlying T2D susceptibility.

At the *BCAR1* locus, multi-ancestry fine-mapping resolved the T2D-association signal to a 99% credible set of nine variants. These variants overlap a chromatin-accessible single-nucleus assay for transposase-accessible chromatin using sequencing (snATAC-seq) peak in human pancreatic acinar cells[33] and an enhancer element in human pancreatic islets that interacts with an active promoter upstream of *CTRB2* (encoding the pancreatic exocrine enzyme chymotrypsin B2)[31]. The observations in bulk pancreatic islets are likely to have arisen due to exocrine (acinar cell) contamination, as single-cell data do not support the expression of *CTRB2* in endocrine cells (Fig. 2). The T2D-association signal also colocalized with a *cis* pQTL for circulating plasma levels of chymotrypsin B1 (CTRB1; $\pi_{COLOC} = 98.6\%$). Interestingly, by extending our colocalization analyses at this locus to *trans* pQTL, we found that variants driving the T2D-association signal also regulate levels of three other pancreatic secretory enzymes produced by acinar cells and involved in the digestion of ingested fats and proteins: carboxypeptidase B1 (CPB1; $\pi_{COLOC} = 98.8\%$), pancreatic lipase-related protein 1 (PLRP1; $\pi_{COLOC} = 97.6\%$) and serine protease 2 (PRSS2; $\pi_{COLOC} = 98.3\%$). These observations are consistent with an effect of T2D-associated variants at this locus on gene and protein expression in the exocrine pancreas, with consequences for pancreatic endocrine function. This is in line with a recent study[34] reporting rare mutations in the gene encoding another protein produced by the exocrine pancreas, chymotrypsin-like elastase family member 2A, which were found to influence levels of digestive enzymes and glucagon (secreted from alpha cells in pancreatic islets). In sum, these complementary findings add to a growing body of evidence linking defects in the exocrine pancreas and T2D pathogenesis[35,36].

At the *PROX1* locus, multi-ancestry fine-mapping localized the two distinct association signals to only three variants (Fig. 3 and Extended Data Fig. 8). The index SNV at the first signal (rs340874, $P = 1.1 \times 10^{-18}$, $\pi > 99.9\%$) overlaps the *PROX1* promoter in both human liver and pancreatic islets[18,29]. At the second signal, the two credible set variants map to the same enhancer active in islets and liver (rs79687284, $P = 6.9 \times 10^{-19}$, $\pi = 66.7\%$; rs17712208, $P = 1.4 \times 10^{-18}$, $\pi = 33.3\%$). Recent studies have demonstrated that the T2D-risk allele at rs17712208 (but not rs79687284) results in significant repression of enhancer activity in mouse MIN6 (ref. [33]) and human EndoC-βH1 beta cell models[37]. Furthermore, this enhancer interacts with the *PROX1* promoter in islets[31] but not in liver[32]. Motivated by these observations, we sought to determine whether these distinct signals impact T2D risk (via *PROX1*) in a tissue-specific manner by assessing transcriptional activity of the credible set variants (rs340874, rs79687284 and rs17712208) in human HepG2 hepatocyte and EndoC-βH1 beta cell models using in vitro reporter assays (Methods and Fig. 3). At the first signal, we demonstrated significant differences in luciferase activity between alleles at rs340874 in both hepatocytes (33% increase for risk allele, $P = 0.0018$) and beta cells (24% increase for risk allele, $P = 0.027$). However, at the second signal, a significant difference in luciferase activity between alleles was observed only for rs17712208 in islets (68% decrease for risk allele, $P = 0.00014$). Interestingly, there was evidence that the risk allele at rs79687284 could attenuate the effect, as the combined effect of both risk alleles in the credible set was less severe. In HepG2 cells, both risk alleles increased transcription relative to wild type, although the difference for each variant alone or combined was not statistically significant. In sum, these results suggest that likely causal variants at these distinct association signals exert their impact on T2D through the same effector gene, *PROX1*, but act in different tissue-specific manners.

Transferability of T2D GRS across diverse populations. GRS derived from European ancestry GWAS have limited transferability

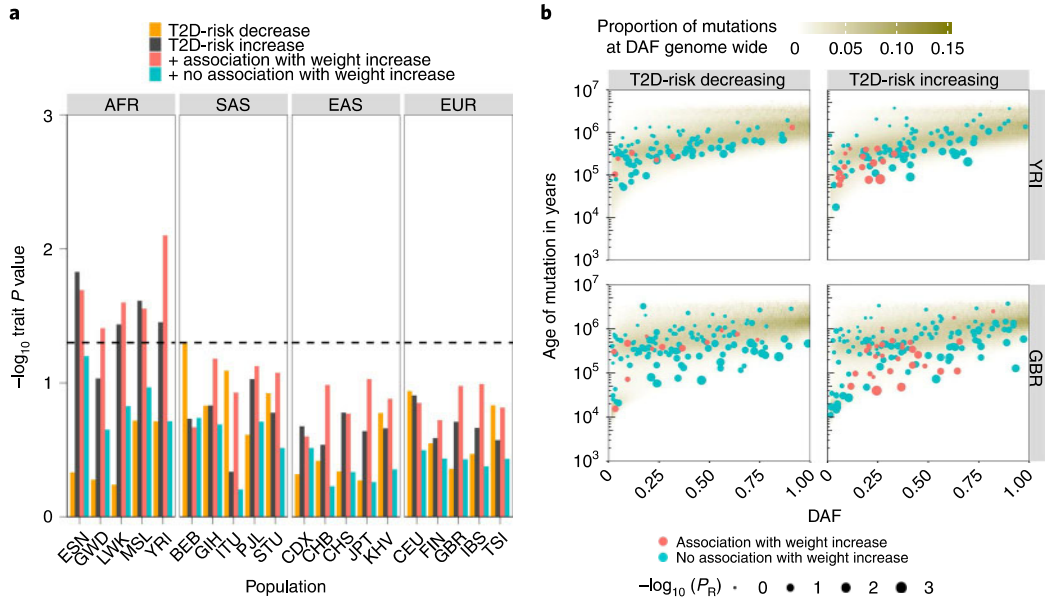


Fig. 5 | Positive selection acting on T2D index SNVs. a, Evidence of selection from Relate. Increased T2D risk is restricted to African (AFR) ancestry populations and is explained by those SNVs that are associated with increased weight. No evidence of selection was observed in South Asian (SAS) ancestry populations, East Asian (EAS) ancestry populations or European (EUR) ancestry populations. **b**, T2D-risk alleles that are associated with increased weight are particularly young for their derived allele frequency (DAF). P_R, *P* value for selection evidence. Population abbreviations (sample sizes): ESN (98), Esan in Nigeria; GWD (112), Gambian in Western Divisions of the Gambia; LWK (98), Luhya in Webuye, Kenya; MSL (84), Mende in Sierra Leone; YRI (107), Yoruba in Ibadan, Nigeria; BEB (85), Bengali in Bangladesh; GIH (102), Gujarati Indian from Houston, Texas; ITU (101), Indian Telegu from the UK; PJL (95), Punjabi from Lahore, Pakistan; STU (101), Sri Lankan Tamil from the UK; CDX (92), Chinese Dai in Xishuangbanna, China; CHB (102), Han Chinese in Beijing, China; CHS (104), Southern Han Chinese; JPT (103), Japanese in Tokyo, Japan; KHV (98), Kinh in Ho Chi Min City, Vietnam; CEU (98), Utah residents with northern and western European ancestry; FIN (98), Finnish in Finland; GBR (90), British in England and Scotland; IBS (106), Iberian population in Spain; TSI (106), Toscani in Italy.

into other population groups in part because of ancestry-correlated differences in the frequency and effect of risk alleles[38]. We took advantage of the population diversity in the DIAMANTE study to compare the prediction performance of multi-ancestry and ancestry-specific T2D GRS constructed using lead SNVs at loci attaining genome-wide significance. We selected two studies per ancestry group as test GWAS into which the prediction performance of the GRS was assessed using trait variance explained (pseudo R^2) and odds ratio (OR) per risk score unit. We repeated the multi-ancestry meta-regression and ancestry-specific meta-analyses after excluding the test GWAS and defined lead SNVs at loci attaining genome-wide significance ($P < 5 \times 10^{-9}$ for multi-ancestry GRS and $P < 5 \times 10^{-8}$ for ancestry-specific GRS). For each ancestry-specific GRS, we used allelic effect estimates for each lead SNV as weights, irrespective of the population in which the test GWAS was undertaken. However, for the multi-ancestry GRS, we derived weights for each lead SNV that were specific to each test GWAS population by allowing for ancestry-correlated heterogeneity in allelic effects (Methods).

As expected, ancestry-specific GRS performed best in test GWAS from their respective ancestry group (Fig. 4 and Supplementary Table 19). However, for the ancestry groups with the smallest effective sample size (African, Hispanic and South Asian), the predictive power of the ancestry-specific GRS was weak (pseudo $R^2 < 1\%$) because the number of lead SNVs attaining genome-wide significance was small. For test GWAS from these under-represented ancestry groups, the European ancestry-specific GRS out-performed the ancestry-matched GRS because (1) more lead SNVs attained genome-wide significance in the European ancestry meta-analysis; and (2) the T2D-association signals represented by these lead SNVs are mostly shared across ancestry groups despite differing allele frequencies and LD patterns. Notwithstanding these observations, the greatest predictive power for test GWAS from all ancestry groups was achieved by the multi-ancestry GRS weighted with population-specific allelic effect estimates.

We then tested the power of the multi-ancestry GRS to predict T2D status in 129,230 individuals of Finnish ancestry from FinnGen, a population-based biobank from Finland (Methods). Because FinnGen was not part of the DIAMANTE study, we used association summary statistics from the complete meta-regression to derive Finnish-specific allelic effect estimates to use as weights in the multi-ancestry GRS (Extended Data Fig. 9 and Supplementary Table 20). Individuals in the top decile of the GRS were at 5.3-fold increased risk of T2D compared to those in the bottom decile. Inclusion of the multi-ancestry GRS with Finnish-specific weights to a predictive model including age, sex and body mass index (BMI) increased the area under the receiver operating characteristic curve (AUROC) from 81.8% to 83.5%. We note that modest increases in AUROC attributable to the GRS over lifestyle and/or clinical factors in cross-sectional studies can mask impactful improvements in clinical performance, particularly among those individuals at the extremes of the GRS distribution who may have especially high lifetime disease risk and/or be prone to earlier disease onset[39]. In FinnGen, age impacted the power of a predictive model including the T2D GRS, sex and BMI: the AUROC decreased from 86.9% in individuals under 50 years old to 73.1% in those over 80 years old (Supplementary Table 21). Each unit of the weighted GRS was associated with earlier age of T2D diagnosis by 1.24 years ($P = 7.1 \times 10^{-57}$), indicating that those with a higher genetic burden are more likely to be affected earlier in life.

Positive selection of T2D-risk alleles. Previous investigations[40] have concluded that historical positive selection has not had the major impact on T2D envisaged by the thrifty genotype hypothesis[41].

We sought to re-evaluate the evidence for positive selection of T2D-risk alleles across our expanded collection of distinct multi-ancestry association signals. We fitted demographic histories to haplotypes for each population in the 1000 Genomes Project reference panel[13] using Relate[42]. We quantified the evidence for selection for each T2D index SNV by assessing the extent to which the mutation has more descendants than other lineages that were present when it arose (Methods). This approach is well powered to detect positive selection acting on polygenic traits over a period of a few thousand to a few tens of thousands of years. We detected evidence of selection ($P < 0.05$) in four of the five African ancestry populations in the 1000 Genomes Project reference panel (but not other ancestry groups) toward increased T2D risk (Fig. 5). Given that T2D itself is likely to have been an advantageous phenotype only via pleiotropic variants acting through beneficial traits, we tested for association of index SNVs at distinct T2D signals with phenotypes available in the UK Biobank[43] (Methods and Extended Data Fig. 10). We found that T2D-risk alleles that were also associated with increased weight (and other obesity-related traits) generally displayed more recent origin when compared to the genome-wide mutation age distribution at the same derived allele frequency ($P < 0.05$ in all African ancestry populations), consistent with positive selection (Extended Data Fig. 10). Excluding these weight-related SNVs removed the selection signature observed in African ancestry populations. These observations are consistent with positive selection of T2D-risk alleles that has been driven by the promotion of energy storage and use appropriate to the local environment[44]. Outside Africa, our analysis yields no evidence for selection of T2D-risk alleles. This suggests the absence of a selective advantage outside Africa or, alternatively, that the selective advantage is old and now masked in the relatively more strongly bottlenecked groups outside Africa. Further work is needed to characterize the specific pathways responsible for this adaptation and its finer-scale geographic impact.

Discussion

In consideration of the global burden of T2D, the DIAMANTE Consortium assembled the most ancestrally diverse collection of GWAS of the disease to date. We implemented a powerful meta-regression approach[15] to enable aggregation of GWAS summary statistics across diverse populations that allows for heterogeneity in allelic effects on disease risk that is correlated with ancestry. By representing the ancestry of each study as multidimensional and continuous axes of genetic variation, the meta-regression model is not restricted to broad continental ancestry categories and can allow for finer-scale differences between GWAS within ancestry groups[45]. Our study demonstrated the advantages of applying this approach to ancestrally diverse GWAS in DIAMANTE with regard to (1) discovery of association signals that are shared across populations through increased sample size and by reducing the genomic control inflation due to residual stratification, (2) defining the extent of heterogeneity in allelic effects at distinct association signals, (3) allowing for LD-driven heterogeneity to enable fine-mapping of causal variants and (4) deriving population-specific weights that substantially improve the transferability of multi-ancestry GRS over ancestry-specific GRS. Our analyses considered SNVs present in the 1000 Genomes Project[13] and Haplotype Reference Consortium[14] reference panels used for imputation, which potentially excludes low-frequency population-specific variants, but provides a uniform 'backbone' of variants for fine-mapping association signals that are shared across multiple population groups. The contribution of population-specific variants that do not overlap reference panels is more fully assessed in complementary ancestry-specific meta-analyses, such as those in European and East Asian components of DIAMANTE[6,10]. Further development of fine-mapping methods is required to localize such population-specific causal variants in multi-ancestry meta-analysis[46].

Our study has extended knowledge of T2D genetics over previous efforts that include GWAS that have contributed to our multi-ancestry meta-analysis[6,10,11], demonstrating the opportunities to deliver new biological insights and identify new target genes and mechanisms through which genetic variation impacts on disease risk. Annotation-informed multi-ancestry fine-mapping resolved 54.4% of distinct T2D-association signals to a single variant with >50% posterior probability. Through integration of these fine-mapping data with molecular QTL resources, we identified a total of 117 candidate causal genes at T2D loci, of which 40 were not reported in complementary analyses undertaken in previous efforts (Supplementary Note). Formal Bayesian colocalization analyses across diverse tissues highlighted complex cell type-specific mechanisms through which regulatory variants at noncoding T2D-association signals impact disease risk, exemplified by the *BCAR1* and *PROX1* loci, and lay the foundations for future functional investigations. Our study demonstrates the advantages of a GRS derived from multi-ancestry meta-regression for T2D prediction across five major ancestry groups. Finally, we built on our expanded collection of distinct multi-ancestry association signals to demonstrate evidence of positive selection of T2D-risk alleles in African populations that may have been driven by the promotion of energy storage and use through adaptation to the local environment.

Multi-ancestry meta-analysis maximizes power to detect association signals that are shared across ancestry groups. However, by modeling heterogeneity in allelic effects across ancestries, our meta-regression approach can also allow for association signals that are driven by ancestry-specific causal variants, although power will be limited by the sample size available in that ancestry group. Ancestry-specific variants tend to have lower frequency, with the result that discovery of T2D associations that are unique to African, Hispanic or South Asian ancestry groups in our study will have been limited to those with relatively large effects. To address this limitation, it remains essential that the human genetics research community continues to bolster GWAS collections in under-represented populations that often suffer the greatest burden of disease and to further expand diversity in imputation reference panels, as exemplified by the Trans-Omics for Precision Medicine (TOPMed) Program[47]. Increasing diversity in genetic research will ultimately provide a more comprehensive and refined view of the genetic contribution to complex human traits, powering understanding of the molecular and biological processes underlying common diseases, and offering the most promising opportunities for clinical translation of GWAS findings to improve global public health.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41588-022-01058-3.

Received: 1 March 2021; Accepted: 23 March 2022;
Published online: 12 May 2022

References

1. NCD Risk Factor Collaboration. Worldwide trends in diabetes since 1980: a pooled analysis of 751 population-based studies with 4.4 million participants. *Lancet* **387**, 1513–1530 (2016).
2. GBD 2015 Disease and Injury Incidence and Prevalence Collaborators. Global, regional, and national incidence, prevalence, and years lived with disability for 310 diseases and injuries, 1990–2015: a systematic analysis for the Global Burden of Disease Study 2015. *Lancet* **388**, 1545–1602 (2016).
3. Voight, B. F. et al. Twelve type 2 diabetes susceptibility loci identified through large-scale association analysis. *Nat. Genet.* **42**, 579–589 (2010).
4. Morris, A. P. et al. Large-scale association analysis provides insights into the genetic architecture and pathophysiology of type 2 diabetes. *Nat. Genet.* **44**, 981–990 (2012).

5. Scott, R. A. et al. An expanded genome-wide association study of type 2 diabetes in Europeans. *Diabetes* **66**, 2888–2902 (2017).
6. Mahajan, A. et al. Fine-mapping type 2 diabetes loci to single-variant resolution using high-density imputation and islet-specific epigenome maps. *Nat. Genet.* **50**, 1505–1513 (2018).
7. Moltke, I. et al. A common Greenlandic *TBC1D4* variant confers muscle insulin resistance and type 2 diabetes. *Nature* **512**, 190–193 (2014).
8. Martin, A. R. et al. Human demographic history impacts genetic risk prediction across diverse populations. *Am. J. Hum. Genet.* **100**, 635–649 (2017).
9. Suzuki, K. et al. Identification of 28 new susceptibility loci for type 2 diabetes in the Japanese population. *Nat. Genet.* **51**, 379–386 (2019).
10. Spracklen, C. N. et al. Identification of type 2 diabetes loci in 433,540 East Asian individuals. *Nature* **582**, 240–245 (2020).
11. Vujkovic, M. et al. Discovery of 318 new risk loci for type 2 diabetes and related vascular outcomes among 1.4 million participants in a multi-ancestry meta-analysis. *Nat. Genet.* **52**, 680–691 (2020).
12. The 1000 Genomes Project Consortium. An integrated map of genetic variation from 1,092 human genomes. *Nature* **491**, 56–65 (2012).
13. The 1000 Genomes Project Consortium. A global reference for human genetic variation. *Nature* **526**, 68–74 (2015).
14. McCarthy, S. et al. A reference panel of 64,976 haplotypes for genotype imputation. *Nat. Genet.* **48**, 1279–1283 (2016).
15. Mägi, R. et al. Trans-ethnic meta-regression of genome-wide association studies accounting for ancestry increases power for discovery and improves fine-mapping resolution. *Hum. Mol. Genet.* **26**, 3639–3650 (2017).
16. Chen, M.-H. et al. Trans-ethnic and ancestry-specific blood-cell genetics in 746,667 individuals from 5 global populations. *Cell* **182**, 1198–1213 (2020).
17. Mahajan, A. et al. Refining the accuracy of validated target identification through coding variant fine-mapping in type 2 diabetes. *Nat. Genet.* **50**, 559–571 (2018).
18. Varshney, A. et al. Genetic regulatory signatures underlying islet gene expression and type 2 diabetes. *Proc. Natl Acad. Sci. USA* **114**, 2301–2306 (2017).
19. Zhao, F. et al. Nodal induces apoptosis through activation of the ALK7 signaling pathway in pancreatic INS-1 β-cells. *Am. J. Physiol. Endocrinol. Metab.* **303**, E132–E143 (2012).
20. Emdin, C. A. et al. DNA sequence variation in *ACVR1C* encoding the activin receptor-like kinase 7 influences body fat distribution and protects against type 2 diabetes. *Diabetes* **68**, 226–234 (2019).
21. Sun, B. B. et al. Genomic atlas of the human plasma proteome. *Nature* **558**, 73–79 (2018).
22. GTEx Consortium. Genetic effects on gene expression across human tissues. *Nature* **550**, 204–213 (2017).
23. Vinuela, A. et al. Genetic variant effects on gene expression in human pancreatic islets and their implications for T2D. *Nat. Commun.* **11**, 4912 (2020).
24. Giambartolomei, C. et al. Bayesian test for colocalization between pairs of genetic association studies using summary statistics. *PLoS Genet.* **10**, e1004383 (2014).
25. van de Bunt, M. et al. Transcript expression data from human islets links regulatory signals from genome-wide association studies for type 2 diabetes and glycemic traits to their downstream effectors. *PLoS Genet.* **11**, e1005694 (2015).
26. Roman, T. S. et al. A type 2 diabetes-associated functional regulatory variant in a pancreatic islet enhancer at the *ADCY5* locus. *Diabetes* **66**, 2521–2530 (2017).
27. Carrat, G. R. et al. Decreased *STARD10* expression is associated with defective insulin secretion in humans and mice. *Am. J. Hum. Genet.* **100**, 238–256 (2017).
28. Small, K. S. et al. Regulatory variants at *KLF14* influence type 2 diabetes risk via a female-specific effect on adipocyte size and body composition. *Nat. Genet.* **50**, 572–580 (2018).
29. Thurner, M. et al. Integration of human pancreatic islet genomic data refines regulatory mechanisms at type 2 diabetes susceptibility loci. *eLife* **7**, e31977 (2018).
30. Pan, D. Z. et al. Integration of human adipocyte chromosomal interactions with adipose gene expression prioritizes obesity-related genes from GWAS. *Nat. Commun.* **9**, 1512 (2018).
31. Miguel-Escalada, I. et al. Human pancreatic islet three-dimensional chromatin architecture provides insights into the genetics of type 2 diabetes. *Nat. Genet.* **51**, 1137–1148 (2019).
32. Chesi, A. et al. Genome-scale Capture C promoter interactions implicate effector genes at GWAS loci for bone mineral density. *Nat. Commun.* **10**, 1260 (2019).
33. Chiou, J. et al. Single-cell chromatin accessibility reveals pancreatic islet cell type- and state-specific regulatory programs of diabetes risk. *Nat. Genet.* **53**, 455–466 (2021).
34. Esteghamat, F. et al. *CELA2A* mutations predispose to early-onset atherosclerosis and metabolic syndrome and affect plasma insulin and platelet activation. *Nat. Genet.* **51**, 1233–1243 (2019).
35. Ng, N. H. J. et al. Tissue-specific alteration of metabolic pathways influences glycemic regulation. Preprint at *bioRxiv* https://doi.org/10.1101/790618 (2019).
36. Gloyn, A. L. Exocrine or endocrine? A circulating pancreatic elastase that regulates glucose homeostasis. *Nat. Metab.* **1**, 853–855 (2019).
37. Wesolowska-Andersen, A. et al. Deep learning models predict regulatory variants in pancreatic islets and refine type 2 diabetes association signals. *eLife* **9**, e51503 (2020).
38. Martin, A. R. et al. Clinical use of current polygenic risk scores may exacerbate health disparities. *Nat. Genet.* **51**, 584–591 (2019).
39. Mars, N. et al. Polygenic and clinical risk scores and their impact on age at onset and prediction of cardiometabolic diseases and common cancers. *Nat. Med.* **26**, 549–557 (2020).
40. Ayub, Q. et al. Revisiting the thrifty gene hypothesis via 65 loci associated with susceptibility to type 2 diabetes. *Am. J. Hum. Genet.* **94**, 176–185 (2014).
41. Neel, J. V. Diabetes mellitus: a "thrifty" genotype rendered detrimental by "progress"?. *Am. J. Hum. Genet.* **14**, 353–362 (1962).
42. Speidel, L., Forest, M., Shi, S. & Myers, S. R. A method for genome-wide genealogy estimation for thousands of samples. *Nat. Genet.* **51**, 1321–1329 (2019).
43. Bycroft, C. et al. The UK Biobank resource with deep phenotyping and genomic data. *Nature* **562**, 203–209 (2018).
44. Chen, R. et al. Type 2 diabetes risk alleles demonstrate extreme directional differentiation among human populations, compared to other diseases. *PLoS Genet.* **8**, e1002621 (2012).
45. Lewis, A. C. F. et al. Getting genetic ancestry right for science and society. *Science* **376**, 250–252 (2022).
46. Kanai, M. et al. Insights from complex trait fine-mapping across diverse populations. Preprint at *medRxiv* https://doi.org/10.1101/2021.09.03.21262975 (2021).
47. Taliun, D. et al. Sequencing of 53,831 diverse genomes from the NHLBI TOPMed Program. *Nature* **590**, 290–299 (2021).

Anubha Mahajan[1,2,277] ✉, Cassandra N. Spracklen[3,4,291], Weihua Zhang[5,6,291], Maggie C. Y. Ng[7,8,9,291], Lauren E. Petty[7,291], Hidetoshi Kitajima[2,10,11,12,291], Grace Z. Yu[1,2,291], Sina Rüeger[13,291], Leo Speidel[14,15,291], Young Jin Kim[16], Momoko Horikoshi[17], Josep M. Mercader[18,19,20], Daniel Taliun[21], Sanghoon Moon[16,2], Soo-Heon Kwak[22,2], Neil R. Robertson[1,2], Nigel W. Rayner[1,2,23,24], Marie Loh[5,25,26], Bong-Jo Kim[16], Joshua Chiou[27,278], Irene Miguel-Escalada[28,29], Pietro della Briotta Parolo[13], Kuang Lin[30], Fiona Bragg[30,31], Michael H. Preuss[32], Fumihiko Takeuchi[33], Jana Nano[34], Xiuqing Guo[35], Amel Lamri[36,37], Masahiro Nakatochi[38], Robert A. Scott[39], Jung-Jin Lee[40], Alicia Huerta-Chagoya[41,279], Mariaelisa Graff[42], Jin-Fang Chai[43], Esteban J. Parra[44], Jie Yao[35], Lawrence F. Bielak[45], Yasuharu Tabara[46], Yang Hai[35], Valgerdur Steinthorsdottir[47], James P. Cook[48], Mart Kals[49],

Niels Grarup[50], Ellen M. Schmidt[21], Ian Pan[51], Tamar Sofer[52,53,54], Matthias Wuttke[55], Chloe Sarnowski[56,280], Christian Gieger[57,58,59], Darryl Nousome[60], Stella Trompet[61,62], Jirong Long[63], Meng Sun[2], Lin Tong[64], Wei-Min Chen[65], Meraj Ahmad[66], Raymond Noordam[62], Victor J. Y. Lim[43], Claudia H. T. Tam[67,68], Yoonjung Yoonie Joo[69,70,281], Chien-Hsiun Chen[71], Laura M. Raffield[3], Cécile Lecoeur[72,73], Bram Peter Prins[23], Aude Nicolas[74], Lisa R. Yanek[75], Guanjie Chen[76], Richard A. Jensen[77], Salman Tajuddin[78], Edmond K. Kabagambe[63,282], Ping An[79], Anny H. Xiang[80], Hyeok Sun Choi[81], Brian E. Cade[20,53], Jingyi Tan[35], Jack Flanagan[17,48], Fernando Abaitua[2,283], Linda S. Adair[82], Adebowale Adeyemo[76,2], Carlos A. Aguilar-Salinas[83], Masato Akiyama[84,85], Sonia S. Anand[36,37,86], Alain Bertoni[87], Zheng Bian[88], Jette Bork-Jensen[50], Ivan Brandslund[89,90], Jennifer A. Brody[77], Chad M. Brummett[91], Thomas A. Buchanan[92], Mickaël Canouil[72,73], Juliana C. N. Chan[67,68,93,94], Li-Ching Chang[71], Miao-Li Chee[95], Ji Chen[96,284], Shyh-Huei Chen[97], Yuan-Tsong Chen[71], Zhengming Chen[30,31], Lee-Ming Chuang[98,99], Mary Cushman[100], Swapan K. Das[101], H. Janaka de Silva[102], George Dedoussis[103], Latchezar Dimitrov[8], Ayo P. Doumatey[76], Shufa Du[82,104], Qing Duan[3], Kai-Uwe Eckardt[105,106], Leslie S. Emery[107], Daniel S. Evans[108], Michele K. Evans[78], Krista Fischer[49], James S. Floyd[77], Ian Ford[109], Myriam Fornage[110], Oscar H. Franco[34], Timothy M. Frayling[111], Barry I. Freedman[112], Christian Fuchsberger[21,113], Pauline Genter[114], Hertzel C. Gerstein[36,37,86], Vilmantas Giedraitis[115], Clicerio González-Villalpando[116], Maria Elena González-Villalpando[116], Mark O. Goodarzi[117], Penny Gordon-Larsen[82,104], David Gorkin[118], Myron Gross[119], Yu Guo[88], Sophie Hackinger[23], Sohee Han[16], Andrew T. Hattersley[120], Christian Herder[57,121,122], Annie-Green Howard[104,123], Willa Hsueh[124], Mengna Huang[51,125], Wei Huang[126], Yi-Jen Hung[127,128], Mi Yeong Hwang[16], Chii-Min Hwu[129,130], Sahoko Ichihara[131], Mohammad Arfan Ikram[34], Martin Ingelsson[115], Md Tariqul Islam[132], Masato Isono[33], Hye-Mi Jang[16], Farzana Jasmine[64], Guozhi Jiang[67,68], Jost B. Jonas[133], Marit E. Jørgensen[134,135], Torben Jørgensen[136,137,138], Yoichiro Kamatani[84,139], Fouad R. Kandeel[140], Anuradhani Kasturiratne[141], Tomohiro Katsuya[142,143], Varinderpal Kaur[19], Takahisa Kawaguchi[46], Jacob M. Keaton[8,63,285], Abel N. Kho[144,145], Chiea-Chuen Khor[146], Muhammad G. Kibriya[64], Duk-Hwan Kim[147], Katsuhiko Kohara[148,286], Jennifer Kriebel[57,58,59], Florian Kronenberg[149], Johanna Kuusisto[150], Kristi Läll[49,151], Leslie A. Lange[152], Myung-Shik Lee[153,154], Nanette R. Lee[155], Aaron Leong[19,156,157], Liming Li[158], Yun Li[3], Ruifang Li-Gao[159], Symen Ligthart[34], Cecilia M. Lindgren[2,160,161], Allan Linneberg[136,162], Ching-Ti Liu[56], Jianjun Liu[146,163], Adam E. Locke[164,165,287], Tin Louie[107], Jian'an Luan[39], Andrea O. Luk[67,68], Xi Luo[166], Jun Lv[158], Valeriya Lyssenko[167,168], Vasiliki Mamakou[169], K. Radha Mani[66,292], Thomas Meitinger[170,171,172], Andres Metspalu[49], Andrew D. Morris[173], Girish N. Nadkarni[32,174,175], Jerry L. Nadler[176], Michael A. Nalls[74,177,178], Uma Nayak[65], Suraj S. Nongmaithem[66], Ioanna Ntalla[179], Yukinori Okada[180,181,182], Lorena Orozco[183], Sanjay R. Patel[184], Mark A. Pereira[185], Annette Peters[57,58,172], Fraser J. Pirie[186], Bianca Porneala[157], Gauri Prasad[187,188], Sebastian Preissl[118], Laura J. Rasmussen-Torvik[189], Alexander P. Reiner[190], Michael Roden[57,121,122], Rebecca Rohde[42], Kathryn Roll[35], Charumathi Sabanayagam[95,191,192], Maike Sander[193,194,195], Kevin Sandow[35], Naveed Sattar[196], Sebastian Schönherr[149], Claudia Schurmann[32,174,197], Mohammad Shahriar[64,288], Jinxiu Shi[126], Dong Mun Shin[16], Daniel Shriner[76], Jennifer A. Smith[45,198], Wing Yee So[67,93,2], Alena Stančáková[150], Adrienne M. Stilp[107], Konstantin Strauch[2,199,200,201], Ken Suzuki[17,84,180,202], Atsushi Takahashi[84,203], Kent D. Taylor[35], Barbara Thorand[57,58], Gudmar Thorleifsson[47], Unnur Thorsteinsdottir[47,204], Brian Tomlinson[67,205], Jason M. Torres[2,289], Fuu-Jen Tsai[206], Jaakko Tuomilehto[207,208,209,210], Teresa Tusie-Luna[211,212], Miriam S. Udler[18,19,156],

Adan Valladares-Salgado[213], Rob M. van Dam[43,163], Jan B. van Klinken[214,215,216], Rohit Varma[217], Marijana Vujkovic[218], Niels Wacher-Rodarte[219], Eleanor Wheeler[39], Eric A. Whitsel[42,220], Ananda R. Wickremasinghe[141], Ko Willems van Dijk[214,215,221], Daniel R. Witte[222,223], Chittaranjan S. Yajnik[224], Ken Yamamoto[225], Toshimasa Yamauchi[202], Loïc Yengo[226], Kyungheon Yoon[16], Canqing Yu[158], Jian-Min Yuan[227,228], Salim Yusuf[36,37,86], Liang Zhang[95], Wei Zheng[63], FinnGen*, eMERGE Consortium*, Leslie J. Raffel[229], Michiya Igase[230], Eli Ipp[114], Susan Redline[20,53,231], Yoon Shin Cho[81], Lars Lind[232], Michael A. Province[79], Craig L. Hanis[233], Patricia A. Peyser[45], Erik Ingelsson[234,235], Alan B. Zonderman[78], Bruce M. Psaty[77,236,237], Ya-Xing Wang[238], Charles N. Rotimi[76], Diane M. Becker[75], Fumihiko Matsuda[46], Yongmei Liu[87,239], Eleftheria Zeggini[23,24,240], Mitsuhiro Yokota[241], Stephen S. Rich[242], Charles Kooperberg[190], James S. Pankow[185], James C. Engert[243,244], Yii-Der Ida Chen[35], Philippe Froguel[72,73,245], James G. Wilson[246], Wayne H. H. Sheu[128,130,247], Sharon L. R. Kardia[45], Jer-Yuarn Wu[71], M. Geoffrey Hayes[69,248,249], Ronald C. W. Ma[67,68,93,94], Tien-Yin Wong[95,191,192], Leif Groop[13,167], Dennis O. Mook-Kanamori[159], Giriraj R. Chandak[66], Francis S. Collins[250], Dwaipayan Bharadwaj[187,251], Guillaume Paré[37,252], Michèle M. Sale[65,293], Habibul Ahsan[64], Ayesha A. Motala[186], Xiao-Ou Shu[63], Kyong-Soo Park[22,253,254], J. Wouter Jukema[61,255], Miguel Cruz[213], Roberta McKean-Cowdin[60], Harald Grallert[57,58,59], Ching-Yu Cheng[95,191,192], Erwin P. Bottinger[32,174,197], Abbas Dehghan[5,34,256], E-Shyong Tai[43,163,257], Josée Dupuis[56], Norihiro Kato[33], Markku Laakso[150], Anna Köttgen[55], Woon-Puay Koh[258,259], Colin N. A. Palmer[260], Simin Liu[51,125,261], Goncalo Abecasis[21], Jaspal S. Kooner[6,256,262,263], Ruth J. F. Loos[32,50,264], Kari E. North[42], Christopher A. Haiman[60], Jose C. Florez[18,19,156], Danish Saleheen[40,265,266], Torben Hansen[50], Oluf Pedersen[50], Reedik Mägi[49], Claudia Langenberg[39,267], Nicholas J. Wareham[39], Shiro Maeda[17,268,269], Takashi Kadowaki[202,290], Juyoung Lee[16], Iona Y. Millwood[30,31], Robin G. Walters[30,31], Kari Stefansson[47,204], Simon R. Myers[2,270], Jorge Ferrer[28,29,271], Kyle J. Gaulton[193,194], James B. Meigs[18,156,157], Karen L. Mohlke[3], Anna L. Gloyn[1,2,272,273,294], Donald W. Bowden[8,9,274,294], Jennifer E. Below[7,294], John C. Chambers[5,6,25,256,262,294], Xueling Sim[43,294], Michael Boehnke[21,294], Jerome I. Rotter[35,294], Mark I. McCarthy[1,2,272,277,294] ✉ and Andrew P. Morris[2,48,49,275,276,294] ✉

[1]Oxford Centre for Diabetes, Endocrinology and Metabolism, Radcliffe Department of Medicine, University of Oxford, Oxford, UK. [2]Wellcome Centre for Human Genetics, Nuffield Department of Medicine, University of Oxford, Oxford, UK. [3]Department of Genetics, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA. [4]Department of Epidemiology and Biostatistics, University of Massachusetts Amherst, Amherst, MA, USA. [5]Department of Epidemiology and Biostatistics, Imperial College London, London, UK. [6]Department of Cardiology, Ealing Hospital, London North West Healthcare NHS Trust, London, UK. [7]Vanderbilt Genetics Institute, Division of Genetic Medicine, Vanderbilt University Medical Center, Nashville, TN, USA. [8]Center for Genomics and Personalized Medicine Research, Wake Forest School of Medicine, Winston-Salem, NC, USA. [9]Department of Biochemistry, Wake Forest School of Medicine, Winston-Salem, NC, USA. [10]The Advanced Research Center for Innovations in Next-Generation Medicine (INGEM), Tohoku University, Sendai, Japan. [11]Department of Integrative Genomics, Tohoku Medical Megabank Organization, Tohoku University, Sendai, Japan. [12]Cancer Center, Tohoku University Hospital, Tohoku University, Sendai, Japan. [13]Institute for Molecular Medicine Finland (FIMM), University of Helsinki, Helsinki, Finland. [14]Genetics Institute, University College London, London, UK. [15]Francis Crick Institute, London, UK. [16]Division of Genome Science, Department of Precision Medicine, National Institute of Health, Cheongju-si, Republic of Korea. [17]Laboratory for Genomics of Diabetes and Metabolism, RIKEN Center for Integrative Medical Sciences, Yokohama, Japan. [18]Programs in Metabolism and Medical & Population Genetics, Broad Institute of Harvard and MIT, Cambridge, MA, USA. [19]Diabetes Unit and Center for Genomic Medicine, Massachusetts General Hospital, Boston, MA, USA. [20]Harvard Medical School, Boston, MA, USA. [21]Department of Biostatistics and Center for Statistical Genetics, University of Michigan, Ann Arbor, MI, USA. [22]Department of Internal Medicine, Seoul National University Hospital, Seoul, South Korea. [23]Department of Human Genetics, Wellcome Sanger Institute, Hinxton, UK. [24]Institute of Translational Genomics, Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg, Germany. [25]Lee Kong Chian School of Medicine, Nanyang Technological University, Singapore, Singapore. [26]Translational Laboratory in Genetic Medicine (TLGM), Agency for Science, Technology and Research (A*STAR) and National University of Singapore (NUS), Singapore, Singapore. [27]Biomedical Sciences Graduate Studies Program, University of California San Diego, La Jolla, CA, USA. [28]Regulatory Genomics and Diabetes, Centre for Genomic Regulation, the Barcelona Institute of Science and Technology, Barcelona, Spain. [29]Centro de Investigación Biomédica en Red Diabetes y Enfermedades Metabólicas asociadas (CIBERDEM), Madrid, Spain. [30]Nuffield Department of Population Health, University of Oxford, Oxford, UK. [31]Medical Research Council Population Health Research Unit, University of Oxford, Oxford, UK. [32]The Charles Bronfman Institute for Personalized Medicine, Icahn School of Medicine at Mount Sinai, New York, NY, USA. [33]Department of Gene Diagnostics and Therapeutics, Research Institute, National Center for Global Health and Medicine, Tokyo, Japan. [34]Department of Epidemiology, Erasmus University Medical Center, Rotterdam, the Netherlands. [35]The Institute for Translational Genomics and Population

Sciences, Department of Pediatrics, the Lundquist Institute for Biomedical Innovation (formerly Los Angeles Biomedical Research Institute) at Harbor–UCLA Medical Center, Torrance, CA, USA. [36]Department of Medicine, McMaster University, Hamilton, Ontario, Canada. [37]Population Health Research Institute, Hamilton Health Sciences and McMaster University, Hamilton, Ontario, Canada. [38]Public Health Informatics Unit, Department of Integrated Health Sciences, Nagoya University Graduate School of Medicine, Nagoya, Japan. [39]MRC Epidemiology Unit, Institute of Metabolic Science, University of Cambridge, Cambridge, UK. [40]Division of Translational Medicine and Human Genetics, University of Pennsylvania, Philadelphia, PA, USA. [41]Consejo Nacional de Ciencia y Tecnología (CONACYT), Instituto Nacional de Ciencias Médicas y Nutrición Salvador Zubirán, Mexico City, Mexico. [42]Department of Epidemiology, Gillings School of Global Public Health, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA. [43]Saw Swee Hock School of Public Health, National University of Singapore and National University Health System, Singapore, Singapore. [44]Department of Anthropology, University of Toronto at Mississauga, Mississauga, Ontario, Canada. [45]Department of Epidemiology, School of Public Health, University of Michigan, Ann Arbor, MI, USA. [46]Center for Genomic Medicine, Kyoto University Graduate School of Medicine, Kyoto, Japan. [47]deCODE Genetics, Amgen Inc., Reykjavik, Iceland. [48]Department of Health Data Science, University of Liverpool, Liverpool, UK. [49]Estonian Genome Centre, Institute of Genomics, University of Tartu, Tartu, Estonia. [50]Novo Nordisk Foundation Center for Basic Metabolic Research, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark. [51]Department of Epidemiology, Brown University School of Public Health, Providence, RI, USA. [52]Department of Biostatistics, Harvard University, Boston, MA, USA. [53]Division of Sleep and Circadian Disorders, Brigham and Women's Hospital, Boston, MA, USA. [54]Department of Medicine, Harvard University, Boston, MA, USA. [55]Institute of Genetic Epidemiology, Department of Data Driven Medicine, Faculty of Medicine and Medical Center, University of Freiburg, Freiburg, Germany. [56]Department of Biostatistics, Boston University School of Public Health, Boston, MA, USA. [57]German Center for Diabetes Research (DZD), Neuherberg, Germany. [58]Institute of Epidemiology, Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg, Germany. [59]Research Unit of Molecular Epidemiology, Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg, Germany. [60]Department of Population and Public Health Sciences, Keck School of Medicine of USC, Los Angeles, CA, USA. [61]Department of Cardiology, Leiden University Medical Center, Leiden, the Netherlands. [62]Section of Gerontology and Geriatrics, Department of Internal Medicine, Leiden University Medical Center, Leiden, the Netherlands. [63]Division of Epidemiology, Department of Medicine, Institute for Medicine and Public Health, Vanderbilt Genetics Institute, Vanderbilt University Medical Center, Nashville, TN, USA. [64]Institute for Population and Precision Health, the University of Chicago, Chicago, IL, USA. [65]Department of Public Health Sciences and Center for Public Health Genomics, University of Virginia School of Medicine, Charlottesville, VA, USA. [66]Genomic Research on Complex Diseases (GRC-Group), CSIR–Centre for Cellular and Molecular Biology (CSIR-CCMB), Hyderabad, India. [67]Department of Medicine and Therapeutics, the Chinese University of Hong Kong, Hong Kong, China. [68]Chinese University of Hong Kong–Shanghai Jiao Tong University Joint Research Centre in Diabetes Genomics and Precision Medicine, the Chinese University of Hong Kong, Hong Kong, China. [69]Division of Endocrinology, Metabolism, and Molecular Medicine, Department of Medicine, Northwestern University Feinberg School of Medicine, Chicago, IL, USA. [70]Department of Health and Biomedical Informatics, Northwestern University Feinberg School of Medicine, Chicago, IL, USA. [71]Institute of Biomedical Sciences, Academia Sinica, Taipei, Taiwan. [72]Inserm U1283, CNRS UMR 8199, European Genomic Institute for Diabetes, Institut Pasteur de Lille, Lille, France. [73]University of Lille, Lille University Hospital, Lille, France. [74]Laboratory of Neurogenetics, National Institute on Aging, National Institutes of Health, Bethesda, MD, USA. [75]Department of Medicine, Johns Hopkins University School of Medicine, Baltimore, MD, USA. [76]Center for Research on Genomics and Global Health, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD, USA. [77]Cardiovascular Health Research Unit, Department of Medicine, University of Washington, Seattle, WA, USA. [78]Laboratory of Epidemiology and Population Sciences, National Institute on Aging, National Institutes of Health, Baltimore, MD, USA. [79]Division of Statistical Genomics, Washington University School of Medicine, St Louis, MO, USA. [80]Department of Research and Evaluation, Division of Biostatistics Research, Kaiser Permanente of Southern California, Pasadena, CA, USA. [81]Department of Biomedical Science, Hallym University, Chuncheon, South Korea. [82]Department of Nutrition, Gillings School of Global Public Health, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA. [83]Unidad de Investigación en Enfermedades Metabólicas and Departamento de Endocrinología y Metabolismo, Instituto Nacional de Ciencias Médicas y Nutrición Salvador Zubirán, Mexico City, Mexico. [84]Laboratory for Statistical and Translational Genetics, RIKEN Center for Integrative Medical Sciences, Yokohama, Japan. [85]Department of Ocular Pathology and Imaging Science, Graduate School of Medical Sciences, Kyushu University, Fukuoka, Japan. [86]Department of Health Research Methods, Evidence, and Impact, McMaster University, Hamilton, Ontario, Canada. [87]Department of Epidemiology and Prevention, Division of Public Health Sciences, Wake Forest School of Medicine, Winston-Salem, NC, USA. [88]Chinese Academy of Medical Sciences, Beijing, China. [89]Institute of Regional Health Research, University of Southern Denmark, Odense, Denmark. [90]Department of Clinical Biochemistry, Vejle Hospital, Vejle, Denmark. [91]Department of Anesthesiology, University of Michigan Medical School, Ann Arbor, MI, USA. [92]Department of Medicine, Division of Endocrinology and Diabetes, Keck School of Medicine of USC, Los Angeles, CA, USA. [93]Hong Kong Institute of Diabetes and Obesity, the Chinese University of Hong Kong, Hong Kong, China. [94]Li Ka Shing Institute of Health Sciences, the Chinese University of Hong Kong, Hong Kong, China. [95]Singapore Eye Research Institute, Singapore National Eye Centre, Singapore, Singapore. [96]Wellcome Sanger Institute, Hinxton, UK. [97]Department of Biostatistics and Data Science, Wake Forest School of Medicine, Winston-Salem, NC, USA. [98]Division of Endocrinology and Metabolism, Department of Internal Medicine, National Taiwan University Hospital, Taipei, Taiwan. [99]Institute of Epidemiology and Preventive Medicine, National Taiwan University, Taipei, Taiwan. [100]Department of Medicine, University of Vermont, Colchester, VT, USA. [101]Section on Endocrinology and Metabolism, Department of Internal Medicine, Wake Forest School of Medicine, Winston-Salem, NC, USA. [102]Department of Medicine, Faculty of Medicine, University of Kelaniya, Ragama, Sri Lanka. [103]Department of Nutrition and Dietetics, Harokopio University of Athens, Athens, Greece. [104]Carolina Population Center, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA. [105]Department of Nephrology and Medical Intensive Care Medicine, Charité Universitätsmedizin Berlin, Berlin, Germany. [106]Department of Nephrology and Hypertension, Friedrich-Alexander-Universität Erlangen-Nürnberg, Erlangen, Germany. [107]Department of Biostatistics, University of Washington, Seattle, WA, USA. [108]California Pacific Medical Center Research Institute, San Francisco, CA, USA. [109]Robertson Centre for Biostatistics, University of Glasgow, Glasgow, UK. [110]Institute of Molecular Medicine, University of Texas Health Science Center at Houston, Houston, TX, USA. [111]Genetics of Complex Traits, University of Exeter Medical School, University of Exeter, Exeter, UK. [112]Department of Internal Medicine, Wake Forest School of Medicine, Winston-Salem, NC, USA. [113]Institute for Biomedicine, Eurac Research, Affiliated Institute of the University of Lübeck, Bolzano, Italy. [114]Department of Medicine, Division of Endocrinology and Metabolism, Lundquist Research Institute at Harbor–UCLA Medical Center, Torrance, CA, USA. [115]Department of Public Health and Caring Sciences, Uppsala University, Uppsala, Sweden. [116]Centro de Estudios en Diabetes, Unidad de Investigacion en Diabetes y Riesgo Cardiovascular, Centro de Investigacion en Salud Poblacional, Instituto Nacional de Salud Publica, Mexico City, Mexico. [117]Department of Medicine, Division of Endocrinology, Diabetes and Metabolism, Cedars-Sinai Medical Center, Los Angeles, CA, USA. [118]Center for Epigenomics, University of California San Diego, La Jolla, CA, USA. [119]Department of Laboratory Medicine and Pathology, University of Minnesota, Minneapolis, MN, USA. [120]University of Exeter Medical School, University of Exeter, Exeter, UK. [121]Institute for Clinical Diabetology, German Diabetes Center, Leibniz Center for Diabetes Research at Heinrich Heine University Düsseldorf, Düsseldorf, Germany. [122]Department of Endocrinology and Diabetology, Medical Faculty and University Hospital Düsseldorf, Heinrich Heine University Düsseldorf, Düsseldorf, Germany. [123]Department of Biostatistics, Gillings School of Global Public Health, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA. [124]Department of Internal Medicine, Diabetes and Metabolism Research Center, the Ohio State University Wexner Medical Center, Columbus, OH, USA. [125]Center for Global Cardiometabolic Health, Brown University, Providence, RI, USA. [126]Shanghai-MOST Key Laboratory of Health and Disease Genomics, Chinese National Human Genome Center at Shanghai (CHGC) and Shanghai

Institute for Biomedical and Pharmaceutical Technologies (SIBPT), Shanghai, China. [127]Division of Endocrine and Metabolism, Tri-Service General Hospital Songshan Branch, Taipei, Taiwan. [128]School of Medicine, National Defense Medical Center, Taipei, Taiwan. [129]Section of Endocrinology and Metabolism, Department of Medicine, Taipei Veterans General Hospital, Taipei, Taiwan. [130]School of Medicine, National Yang Ming Chiao Tung University, Taipei, Taiwan. [131]Department of Environmental and Preventive Medicine, Jichi Medical University School of Medicine, Shimotsuke, Japan. [132]University of Chicago Research Bangladesh, Dhaka, Bangladesh. [133]Institute of Molecular and Clinical Ophthalmology Basel, Basel, Switzerland. [134]Steno Diabetes Center Copenhagen, Gentofte, Denmark. [135]National Institute of Public Health, Southern Denmark University, Copenhagen, Denmark. [136]Center for Clinical Research and Prevention, Bispebjerg and Frederiksberg Hospital, Frederiksberg, Denmark. [137]Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark. [138]Faculty of Medicine, Aalborg University, Aalborg, Denmark. [139]Laboratory of Complex Trait Genomics, Department of Computational Biology and Medical Sciences, Graduate School of Frontier Sciences, the University of Tokyo, Tokyo, Japan. [140]Department of Clinical Diabetes, Endocrinology & Metabolism, Department of Translational Research and Cellular Therapeutics, City of Hope, Duarte, CA, USA. [141]Department of Public Health, Faculty of Medicine, University of Kelaniya, Ragama, Sri Lanka. [142]Department of Clinical Gene Therapy, Osaka University Graduate School of Medicine, Osaka, Japan. [143]Department of Geriatric and General Medicine, Graduate School of Medicine, Osaka University, Osaka, Japan. [144]Division of General Internal Medicine and Geriatrics, Department of Medicine, Northwestern University Feinberg School of Medicine, Chicago, IL, USA. [145]Center for Health Information Partnerships, Institute for Public Health and Medicine, Northwestern University Feinberg School of Medicine, Chicago, IL, USA. [146]Genome Institute of Singapore, Agency for Science, Technology and Research, Singapore, Singapore. [147]Department of Molecular Cell Biology, Sungkyunkwan University School of Medicine, Suwon, South Korea. [148]Department of Regional Resource Management, Ehime University Faculty of Collaborative Regional Innovation, Ehime, Japan. [149]Institute of Genetic Epidemiology, Department of Genetics and Pharmacology, Medical University of Innsbruck, Innsbruck, Austria. [150]Institute of Clinical Medicine, Internal Medicine, University of Eastern Finland and Kuopio University Hospital, Kuopio, Finland. [151]Institute of Mathematics and Statistics, University of Tartu, Tartu, Estonia. [152]Department of Medicine, University of Colorado Denver, Anschutz Medical Campus, Aurora, CO, USA. [153]Severance Biomedical Science Institute and Department of Internal Medicine, Yonsei University College of Medicine, Seoul, South Korea. [154]Department of Medicine, Samsung Medical Center, Sungkyunkwan University School of Medicine, Seoul, South Korea. [155]USC—Office of Population Studies Foundation, Inc., University of San Carlos, Cebu City, Philippines. [156]Department of Medicine, Harvard Medical School, Boston, MA, USA. [157]Division of General Internal Medicine, Massachusetts General Hospital, Boston, MA, USA. [158]Department of Epidemiology and Biostatistics, Peking University Health Science Centre, Peking University, Beijing, China. [159]Department of Clinical Epidemiology, Leiden University Medical Center, Leiden, the Netherlands. [160]Program in Medical & Population Genetics, Broad Institute, Cambridge, MA, USA. [161]Big Data Institute, Li Ka Shing Centre for Health Information and Discovery, University of Oxford, Oxford, UK. [162]Department of Clinical Medicine, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark. [163]Department of Medicine, Yong Loo Lin School of Medicine, National University of Singapore and National University Health System, Singapore, Singapore. [164]McDonnell Genome Institute, Washington University School of Medicine, St Louis, MO, USA. [165]Department of Medicine, Division of Genomics and Bioinformatics, Washington University School of Medicine, St Louis, MO, USA. [166]Department of Biostatistics and Data Science, University of Texas Health Science Center at Houston, Houston, TX, USA. [167]Department of Clinical Sciences, Diabetes and Endocrinology, Lund University Diabetes Centre, Malmö, Sweden. [168]Department of Clinical Science, Center for Diabetes Research, University of Bergen, Bergen, Norway. [169]Dromokaiteio Psychiatric Hospital, National and Kapodistrian University of Athens, Athens, Greece. [170]Institute of Human Genetics, Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg, Germany. [171]Institute of Human Genetics, Technical University of Munich, Munich, Germany. [172]German Centre for Cardiovascular Research (DZHK), Partner Site Munich Heart Alliance, Munich, Germany. [173]The Usher Institute to the Population Health Sciences and Informatics, University of Edinburgh, Edinburgh, UK. [174]Digital Health Center, Digital Engineering Faculty of Hasso Plattner Institue and University Potsdam, Potsdam, Germany. [175]The Division of Data Driven and Digital Medicine (D3M), Department of Medicine, Icahn School of Medicine at Mount Sinai, New York, NY, USA. [176]Department of Medicine and Pharmacology, New York Medical College, Valhalla, NY, USA. [177]Data Tecnica International LLC, Glen Echo, MD, USA. [178]Center for Alzheimer's and Related Dementias, National Institutes of Health, Baltimore, MD, USA. [179]William Harvey Research Institute, Barts and the London School of Medicine and Dentistry, Queen Mary University of London, London, UK. [180]Department of Statistical Genetics, Osaka University Graduate School of Medicine, Osaka, Japan. [181]Laboratory of Statistical Immunology, Immunology Frontier Research Center (WPI-IFReC), Osaka University, Osaka, Japan. [182]Laboratory for Systems Genetics, RIKEN Center for Integrative Medical Sciences, Yokohama, Japan. [183]Instituto Nacional de Medicina Genómica, Mexico City, Mexico. [184]Division of Pulmonary, Allergy, and Critical Care Medicine, Department of Medicine, University of Pittsburgh, Pittsburgh, PA, USA. [185]Division of Epidemiology and Community Health, School of Public Health, University of Minnesota, Minneapolis, MN, USA. [186]Department of Diabetes and Endocrinology, Nelson R. Mandela School of Medicine, College of Health Sciences, University of KwaZulu-Natal, Durban, South Africa. [187]Academy of Scientific and Innovative Research, CSIR—Human Resource Development Centre Campus, Ghaziabad, India. [188]Genomics and Molecular Medicine Unit, CSIR—Institute of Genomics and Integrative Biology, New Delhi, India. [189]Department of Preventive Medicine, Northwestern University Feinberg School of Medicine, Chicago, IL, USA. [190]Fred Hutchinson Cancer Research Center, Seattle, WA, USA. [191]Ophthalmology and Visual Sciences Academic Clinical Program (Eye ACP), Duke-NUS Medical School, Singapore, Singapore. [192]Department of Ophthalmology, Yong Loo Lin School of Medicine, National University of Singapore and National University Health System, Singapore, Singapore. [193]Department of Pediatrics, Pediatric Diabetes Research Center, University of California San Diego, La Jolla, CA, USA. [194]Institute for Genomic Medicine, University of California San Diego, La Jolla, CA, USA. [195]Department of Cellular and Molecular Medicine, University of California San Diego, La Jolla, CA, USA. [196]Institute of Cardiovascular and Medical Sciences, University of Glasgow, Glasgow, UK. [197]Hasso Plattner Institute for Digital Health at Mount Sinai, Icahn School of Medicine at Mount Sinai, New York, NY, USA. [198]Survey Research Center, Institute for Social Research, University of Michigan, Ann Arbor, MI, USA. [199]Institute of Genetic Epidemiology, Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg, Germany. [200]Chair of Genetic Epidemiology, IBE, Faculty of Medicine, LMU Munich, Munich, Germany. [201]Institute of Medical Biostatistics, Epidemiology and Informatics (IMBEI), University Medical Center, Johannes Gutenberg University, Mainz, Germany. [202]Department of Diabetes and Metabolic Diseases, Graduate School of Medicine, the University of Tokyo, Tokyo, Japan. [203]Department of Genomic Medicine, National Cerebral and Cardiovascular Center, Osaka, Japan. [204]Faculty of Medicine, University of Reykjavik, Reykjavik, Iceland. [205]Faculty of Medicine, Macau University of Science and Technology, Macau, China. [206]Department of Medical Genetics and Medical Research, China Medical University Hospital, Taichung, Taiwan. [207]Department of Health, Finnish Institute for Health and Welfare, Helsinki, Finland. [208]National School of Public Health, Madrid, Spain. [209]Department of Neuroscience and Preventive Medicine, Danube University Krems, Krems, Austria. [210]Diabetes Research Group, King Abdulaziz University, Jeddah, Saudi Arabia. [211]Unidad de Biología Molecular y Medicina Genómica, Instituto Nacional de Ciencias Médicas y Nutrición Salvador Zubirán, Mexico City, Mexico. [212]Departamento de Medicina Genómica y Toxiología, Ambiental Instituto de Investigaciones Biomédicas, UNAM, Mexico City, Mexico. [213]Unidad de Investigacion Medica en Bioquimica, Hospital de Especialidades, Centro Medico Nacional Siglo XXI, IMSS, Mexico City, Mexico. [214]Einthoven Laboratory for Experimental Vascular Medicine, Leiden University Medical Center, Leiden, the Netherlands. [215]Department of Human Genetics, Leiden University Medical Center, Leiden, the Netherlands. [216]Department of Clinical Chemistry, Laboratory of Genetic Metabolic Disease, Amsterdam University Medical Center, Amsterdam, the Netherlands. [217]Southern California Eye Institute, CHA Hollywood Presbyterian Medical Center, Los Angeles, CA, USA. [218]Department of Medicine, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA, USA. [219]Unidad de Investigación Médica en Epidemiologia Clinica, Hospital de Especialidades, Centro Medico Nacional Siglo XXI, IMSS, Mexico City, Mexico. [220]Department of Medicine, School of

Medicine, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA. [221]Department of Internal Medicine, Division of Endocrinology, Leiden University Medical Center, Leiden, the Netherlands. [222]Department of Public Health, Aarhus University, Aarhus, Denmark. [223]Danish Diabetes Academy, Odense, Denmark. [224]Diabetology Research Centre, King Edward Memorial Hospital and Research Centre, Pune, India. [225]Department of Medical Biochemistry, Kurume University School of Medicine, Kurume, Japan. [226]Institute for Molecular Bioscience, University of Queensland, Brisbane, Queensland, Australia. [227]Division of Cancer Control and Population Sciences, UPMC Hillman Cancer Center, University of Pittsburgh, Pittsburgh, PA, USA. [228]Department of Epidemiology, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, PA, USA. [229]Department of Pediatrics, Division of Genetic and Genomic Medicine, UCI Irvine School of Medicine, Irvine, CA, USA. [230]Department of Anti-Aging Medicine, Ehime University Graduate School of Medicine, Ehime, Japan. [231]Division of Pulmonary, Critical Care, and Sleep Medicine, Beth Israel Deaconess Medical Center, Boston, MA, USA. [232]Department of Medical Sciences, Uppsala University, Uppsala, Sweden. [233]Human Genetics Center, University of Texas Health Science Center at Houston, Houston, TX, USA. [234]Department of Medicine, Division of Cardiovascular Medicine, Stanford University School of Medicine, Stanford, CA, USA. [235]Department of Medical Sciences, Molecular Epidemiology and Science for Life Laboratory, Uppsala University, Uppsala, Sweden. [236]Department of Epidemiology, University of Washington, Seattle, WA, USA. [237]Department of Health Services, University of Washington, Seattle, WA, USA. [238]Beijing Institute of Ophthalmology, Ophthalmology and Visual Sciences Key Laboratory, Beijing Tongren Hospital, Capital Medical University, Beijing, China. [239]Department of Medicine, Division of Cardiology, Duke University School of Medicine, Durham, NC, USA. [240]Technical University of Munich (TUM) and Klinikum Rechts der Isar, TUM School of Medicine, Munich, Germany. [241]Kurume University School of Medicine, Kurume, Japan. [242]Center for Public Health Genomics, University of Virginia School of Medicine, Charlottesville, VA, USA. [243]Department of Medicine, McGill University, Montreal, Quebec, Canada. [244]Department of Human Genetics, McGill University, Montreal, Quebec, Canada. [245]Department of Genomics of Common Disease, School of Public Health, Imperial College London, London, UK. [246]Department of Physiology and Biophysics, University of Mississippi Medical Center, Jackson, MS, USA. [247]Division of Endocrinology and Metabolism, Department of Medicine, Taichung Veterans General Hospital, Taichung, Taiwan. [248]Center for Genetic Medicine, Northwestern University Feinberg School of Medicine, Chicago, IL, USA. [249]Department of Anthropology, Northwestern University, Evanston, IL, USA. [250]Center for Precision Health Research, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD, USA. [251]Systems Genomics Laboratory, School of Biotechnology, Jawaharlal Nehru University, New Delhi, India. [252]Department of Pathology and Molecular Medicine, McMaster University, Hamilton, Ontario, Canada. [253]Department of Internal Medicine, Seoul National University College of Medicine, Seoul, South Korea. [254]Department of Molecular Medicine and Biopharmaceutical Sciences, Graduate School of Convergence Science and Technology, Seoul National University, Seoul, South Korea. [255]Netherlands Heart Institute, Utrecht, the Netherlands. [256]MRC-PHE Centre for Environment and Health, Imperial College London, London, UK. [257]Duke–NUS Medical School, Singapore, Singapore. [258]Singapore Institute for Clinical Sciences, Agency for Science Technology and Research (A*STAR), Singapore, Singapore. [259]Healthy Longevity Translational Research Programme, Yong Loo Lin School of Medicine, National University of Singapore, Singapore, Singapore. [260]Pat Macpherson Centre for Pharmacogenetics and Pharmacogenomics, University of Dundee, Dundee, UK. [261]Department of Medicine, Brown University Alpert School of Medicine, Providence, RI, USA. [262]Imperial College Healthcare NHS Trust, Imperial College London, London, UK. [263]National Heart and Lung Institute, Imperial College London, London, UK. [264]The Mindich Child Health and Development Institute, Ichan School of Medicine at Mount Sinai, New York, NY, USA. [265]Department of Biostatistics and Epidemiology, University of Pennsylvania, Philadelphia, PA, USA. [266]Center for Non-Communicable Diseases, Karachi, Pakistan. [267]Computational Medicine, Berlin Institute of Health at Charité Universitätsmedizin, Berlin, Germany. [268]Department of Advanced Genomic and Laboratory Medicine, Graduate School of Medicine, University of the Ryukyus, Okinawa, Japan. [269]Division of Clinical Laboratory and Blood Transfusion, University of the Ryukyus Hospital, Okinawa, Japan. [270]Department of Statistics, University of Oxford, Oxford, UK. [271]Section of Genetics and Genomics, Department of Metabolism, Digestion and Reproduction, Imperial College London, London, UK. [272]Oxford NIHR Biomedical Research Centre, Churchill Hospital, Oxford University Hospitals NHS Foundation Trust, Oxford, UK. [273]Division of Endocrinology, Department of Pediatrics, Stanford School of Medicine, Stanford University, Stanford, CA, USA. [274]Center for Diabetes Research, Wake Forest School of Medicine, Winston-Salem, NC, USA. [275]Centre for Genetics and Genomics Versus Arthritis, Centre for Musculoskeletal Research, Division of Musculoskeletal and Dermatological Sciences, University of Manchester, Manchester, UK. [276]NIHR Manchester Biomedical Research Centre, Manchester University NHS Foundation Trust, Manchester, UK. [277]Present address: Genentech, South San Francisco, CA, USA. [278]Present address: Internal Medicine Research Unit, Pfizer Worldwide Research, Cambridge, MA, USA. [279]Present address: Departamento de Medicina Genómica y Toxicologí, a AmbientalInstituto de Investigaciones Biomédicas, UNAM, Ciudad de Mexico, Mexico, Mexico. [280]Present address: Department of Epidemiology, Human Genetics, and Environmental Sciences, the University of Texas Health Science Center at Houston, School of Public Health, Houston, TX, USA. [281]Present address: Institute of Data Science, Korea University, Seoul, South Korea. [282]Present address: Division of Academics, Ochsner Health, New Orleans, LA, USA. [283]Present address: Vertex Pharmaceuticals Ltd, Oxford, UK. [284]Present address: Exeter Centre of Excellence in Diabetes (ExCEeD), Exeter Medical School, University of Exeter, Exeter, UK. [285]Present address: Center for Precision Health Research, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD, USA. [286]Present address: Ibusuki Kozenkai Hospital, Ibusuki, Japan. [287]Present address: Regeneron Genetics Center, Tarrytown, NY, USA. [288]Present address: Institute for Population and Precision Health (IPPH), Biological Sciences Division, the University of Chicago, Chicago, IL, USA. [289]Present address: Clinical Trial Service Unit and Epidemiological Studies Unit, Nuffield Department of Population Health, University of Oxford, Oxford, UK. [290]Present address: Toranomon Hospital, Tokyo, Japan. [291]These authors contributed equally: Cassandra N. Spracklen, Weihua Zhang, Maggie C. Y. Ng, Lauren E. Petty, Hidetoshi Kitajima, Grace Z. Yu, Sina Rüeger, Leo Speidel. [292]Deceased: K. Radha Mani. [293]Deceased: Michèle M. Sale. [294]These authors jointly supervised this work: Anna L. Gloyn, Donald W. Bowden, Jennifer E. Below, John C. Chambers, Xueling Sim, Michael Boehnke, Jerome I. Rotter, Mark I. McCarthy, Andrew P. Morris. *Lists of authors and their affiliations appear at the end of the paper.
✉e-mail: mahajan.anubha@gene.com; mccarthy.mark@gene.com; andrew.morris-5@manchester.ac.uk

FinnGen

Sina Rüeger[14] and Pietro della Briotta Parolo[14]

A full list of members and their affiliations appears in the Supplementary Information.

eMERGE Consortium

Yoonjung Yoonie Joo[68,69], Abel N. Kho[144,145] and M. Geoffrey Hayes[68,247,248]

A full list of members and their affiliations appears in the Supplementary Information.

Methods

Ethics statement. All human research was approved by the relevant Institutional Review Boards and conducted according to the Declaration of Helsinki. All participants provided written informed consent. Study-level ethical statements are provided in the Supplementary Note.

Study-level analyses. Individuals were assayed with a range of GWAS genotyping arrays, with sample and SNV quality control undertaken within each study (Supplementary Tables 2 and 4). Most GWAS were undertaken with individuals from one ancestry group (Supplementary Table 1), where population outliers were excluded using self-reported and genetic ancestry. For the remaining multi-ancestry GWAS (Supplementary Table 1), individuals were first assigned to an ancestry group using both self-reported and genetic ancestry, and analyses were then undertaken separately within each ancestry group. For each ancestry-specific GWAS, samples were pre-phased and imputed up to reference panels from the 1000 Genomes Project (phase 1, March 2012 release; phase 3, October 2014 release)[12,13], the Haplotype Reference Consortium[14] or population-specific whole-genome sequencing[48–50] (Supplementary Table 4). SNVs with poor imputation quality and/or minor allele count <5 were excluded from downstream association analyses (Supplementary Table 4). Association with T2D was evaluated in a regression framework under an additive model in the dosage of the minor allele, with adjustment for age and sex (when appropriate) and additional study-specific covariates (Supplementary Table 4). Analyses accounted for structure (population stratification and/or familial relationships) by (1) excluding related samples and adjusting for principal components derived from a genetic relatedness matrix as additional covariates in the regression model or (2) incorporating a random effect for the genetic relatedness matrix in a mixed model (Supplementary Table 4). Allelic effects and corresponding standard errors that were estimated from a linear (mixed) model were converted to the log odds scale[51]. Study-level association summary statistics (*P* values and standard error of allelic log ORs) were corrected for residual structure, not accounted for in the regression analysis, by means of genomic control[52] if the inflation factor was >1 (Supplementary Table 4).

Multi-ancestry meta-analyses. To account for the different reference panels used for imputation, we considered autosomal biallelic SNVs that overlap the 1000 Genomes Project reference panel (phase 3, October 2014 release)[13] and the Haplotype Reference Consortium reference panel[14]. We considered only those SNVs with minor allele frequency >0.5% in haplotypes in at least one of the five ancestry groups (Supplementary Table 22) in the 1000 Genomes Project (phase 3, October 2014 release)[13]. We excluded SNVs that differed in allele frequency by >20% when comparing reference panels in the same subsets of samples.

The most powerful methods for discovery of new loci through multi-ancestry meta-analysis allow for potential allelic effect heterogeneity between ancestry groups that cannot be accommodated in a fixed-effects model[53]. Random-effects meta-analysis allows for 'unstructured' heterogeneity but cannot allow for the expectation that GWAS from the same ancestry group are likely to have more similar allelic effects than those from different ancestry groups. Some of these limitations could be addressed with a two-stage hierarchical model (within and then between ancestry). However, we preferred a meta-regression approach, implemented in MR-MEGA[15], which models allelic effect heterogeneity that is correlated with genetic ancestry by including axes of genetic variation as covariates to capture ancestral diversity between GWAS. We constructed a distance matrix of differences in mean effect allele frequency between each pair of GWAS across a subset of 386,563 SNVs reported in all studies. We implemented multidimensional scaling of the distance matrix to obtain three principal components that defined axes of genetic variation to separate GWAS from the five ancestry groups (Extended Data Fig. 2).

For each SNV, we modeled allelic log ORs across GWAS in a linear regression framework, weighted by the inverse of the variance of the effect estimates, incorporating the three axes of genetic variation as covariates. We tested for (1) association with T2D allowing for allelic effect heterogeneity between GWAS that is correlated with ancestry, (2) heterogeneity in allelic effects on T2D between GWAS that is correlated with ancestry and (3) residual allelic effect heterogeneity between GWAS due to unmeasured confounders. We corrected the meta-regression association *P* values for inflation due to residual structure between GWAS using genomic control adjustment (allowing for four degrees of freedom): $\lambda_{TA} = 1.052$. We included SNVs reported in ≥50% of the total effective sample size ($N_{TA} \geq 246{,}095$) in downstream analyses.

We also aggregated association summary statistics across GWAS via fixed-effects meta-analysis using METAL[54] and random-effects (RE2 model) meta-analysis using METASOFT[55]. Both meta-analyses were based on inverse-variance weighting of allelic log ORs to obtain effect-size estimates. We corrected standard errors for inflation due to residual structure between GWAS by genomic control adjustment: $\lambda_{TA}^{FE} = 1.253$ and $\lambda_{TA}^{RE} = 1.253$. We assessed evidence for heterogeneity in allelic effects between GWAS by Cochran's *Q* statistic.

Defining T2D loci. We initially selected lead SNVs attaining genome-wide significant evidence of association ($P < 5 \times 10^{-8}$) in the multi-ancestry meta-regression that were separated by at least 500 kb. Loci were first defined by the flanking genomic interval mapping 500 kb upstream and downstream of lead SNVs. Next, when lead SNVs were separated by less than 1 Mb, the corresponding loci were aggregated as a single locus. The lead SNV for each locus was then selected as the SNV with minimum association *P* value.

Genome-wide significance threshold. We considered haplotypes from the 1000 Genomes Project reference panel (phase 3, October 2014 release)[13]. We extracted autosomal biallelic SNVs that overlapped between reference panels used in study-level analyses. We estimated the effective number of independent SNVs across ancestry groups using LD pruning in PLINK[56] to be 9,966,662 at $r^2 > 0.5$ (ref. [57]). We therefore chose a multi-ancestry genome-wide significance threshold by Bonferroni correction for the effective number of SNVs as $P < 5 \times 10^{-9}$. Exemplar power calculations are provided in the Supplementary Note.

Dissection of distinct multi-ancestry association signals. We used iterative approximate conditioning, implemented in GCTA[58], making use of forward selection and backward elimination, to identify index SNVs at multi-ancestry genome-wide significance ($P < 5 \times 10^{-9}$). We used haplotypes from the 1000 Genomes Project reference panel (phase 3, October 2014 release)[13] that were specific to each ancestry group (Supplementary Table 22) as a reference for LD between SNVs across loci in the approximate conditional analysis. Details of the iterative approximate conditioning are provided in the Supplementary Note.

Ancestry-specific meta-analyses. We aggregated association summary statistics across GWAS via fixed-effects meta-analysis using METAL[54] based on inverse-variance weighting of allelic log ORs to obtain effect-size estimates. Details are provided in the Supplementary Note.

Fine-mapping resolution. Within each locus, we approximated the Bayes factor[59] Λ_{ij} in favor of T2D association of the *j*th SNV at the *i*th distinct association signal using summary statistics from (1) the multi-ancestry meta-regression, (2) the European ancestry-specific meta-analysis and (3) the combined East Asian and European ancestry meta-analysis. For loci with a single association signal, association summary statistics were obtained from unconditional analysis. For loci with multiple distinct association signals, association summary statistics were obtained from approximate conditional analyses. Details of the derivation of approximate Bayes factors are provided in the Supplementary Note. The posterior probability for the *j*th SNV at the *i*th distinct signal was then given by $\pi_{ij} \propto \Lambda_{ij}$. We derived a 99% credible set[60] for the *i*th distinct association signal by (1) ranking all SNVs according to their posterior probability π_{ij} and (2) including ranked SNVs until their cumulative posterior probability attains or exceeds 0.99.

Downsampled multi-ancestry meta-regression. We selected GWAS contributing to the multi-ancestry meta-regression to approximate the effective sample size of the European ancestry-specific meta-analysis and maintain the distribution of effective sample size across ancestry groups (Supplementary Table 10). The selected GWAS are summarized in the Supplementary Note. We conducted a 'downsampled' multi-ancestry meta-regression implemented in MR-MEGA[15] for the selected studies. For each SNV, we modeled allelic log ORs across GWAS in a linear regression framework, weighted by the inverse of the variance of the effect estimates, incorporating the same three axes of genetic variation as covariates (Extended Data Fig. 2). We corrected the meta-regression association *P* values for inflation due to residual structure between the selected GWAS using genomic control adjustment (allowing for four degrees of freedom): $\lambda_{TA*} = 1.012$. For each distinct association signal identified in the complete multi-ancestry meta-regression, we derived a 99% credible set[60] using association summary statistics from the downsampled multi-ancestry meta-regression. Details of the fine-mapping procedure are provided in the Supplementary Note.

Enrichment of T2D-association signals in genomic annotations. We mapped each SNV across T2D loci to three categories of functional and regulatory annotations: (1) genic regions, as defined by the GENCODE Project[61], including protein-coding exons, and 3′ and 5′ UTRs as different annotations; (2) chromatin immunoprecipitation followed by sequencing (ChIP–seq) binding sites for 165 transcription factors (161 proteins from the ENCODE Project[62] and four additional factors assayed in primary pancreatic islets[63]); and (3) 13 unique and recurrent chromatin states, including promoter, enhancer, transcribed and repressed regions in four T2D-relevant tissues[18] (pancreatic islets, the liver, adipose tissue and skeletal muscle). This resulted in a total of 220 genomic annotations for downstream enrichment analyses. We used fGWAS[64] to identify a joint model of enriched annotations across distinct T2D-association signals from the multi-ancestry meta-regression. Details are provided in the Supplementary Note.

Annotation-informed fine-mapping. Within each locus, for each distinct signal, we recalibrated the posterior probability of driving the T2D association for each SNV under an annotation-informed prior derived from the joint model of enriched annotations identified by fGWAS. Specifically, for the *j*th SNV at the *i*th distinct signal, the posterior probability $\pi_{ij} \propto \gamma_j \Lambda_{ij}$, where Λ_{ij} is the Bayes factor in

favor of T2D association. In this expression, the relative annotation-informed prior for the SNV is given by

$$\gamma_j = \exp\left(\sum_k \hat{\beta}_k z_{jk}\right),$$

where the summation is over the enriched annotations, $\hat{\beta}_k$ is the estimated log fold enrichment of the kth annotation from the final joint model, and z_{jk} is an indicator variable taking the value 1 if the jth SNV maps to the kth annotation and 0 otherwise. We derived a 99% credible set[60] for the ith distinct association signal by (1) ranking all SNVs according to their posterior probability π_{ij} and (2) including ranked SNVs until their cumulative posterior probability attains or exceeds 0.99.

Dissection of molecular QTL in diverse tissues. We accessed association summary statistics for molecular QTL in diverse tissues from three published resources: (1) 3,622 circulating plasma proteins in 3,301 healthy blood donors of European ancestry from the INTERVAL study[21] (2) pancreatic islet expression in 420 individuals of European ancestry from the InsPIRE Consortium[23] and (3) multi-tissue expression in 620 donors from the GTEx Project (release version 7)[22], including subcutaneous adipose tissue (328 samples), visceral adipose tissue (273 samples), brain hypothalamus (108 samples), liver (134 samples) and skeletal muscle (421 samples). We defined *cis* molecular QTL as mapping within 1 Mb of the TSS of the gene. Recognizing that molecular QTL may also be driven by multiple causal variants, we dissected signals for each significant *cis* and *trans* pQTL ($P < 1.5 \times 10^{-11}$) and for each significant *cis* eQTL (FDR Q value < 5%) via approximate conditional analyses implemented in GCTA[58]. We used a genotype reference panel of 6,000 unrelated individuals of white British origin, randomly selected from the UK Biobank[43], to model LD between SNVs. We excluded SNVs from the reference panel with poor imputation quality (info < 0.4) and/or significant deviation from Hardy–Weinberg equilibrium ($P < 10^{-6}$). We first identified index SNVs for each distinct molecular QTL signal using the '–cojo-slct' option: $P < 1.5 \times 10^{-11}$ for *cis* and *trans* pQTL and $P < 5 \times 10^{-8}$ for *cis* eQTL. For each molecular QTL with multiple index SNVs, we dissected each distinct signal using GCTA, removing each index SNV, and adjusting for the remainder using the '–cojo-cond' option.

Colocalization of T2D associations and molecular QTL. For each distinct T2D-association signal, we used COLOC version 3.1 (ref. [24]) to assess the evidence for colocalization with (1) each distinct *cis* and *trans* pQTL signal and (2) each distinct *cis* eQTL signal across tissues. COLOC assumes that at most one variant is causal for each distinct T2D association and each distinct molecular QTL, which is reasonable after deconvolution of signals via approximate conditional analyses. Under this assumption, there are five hypotheses: association with neither T2D nor the molecular QTL (H_0); association only with T2D (H_1) or the molecular QTL (H_2); or association with both T2D and the molecular QTL, driven either by two different causal variants (H_3) or by the same causal variant (H_4). We assumed the default prior probabilities of (1) 10^{-4} that a variant is causal only for T2D or only for the molecular QTL and (2) 10^{-6} that a variant is causal for both T2D and the molecular QTL. To take account of our annotation-informed prior model of causality, we then replaced the Bayes factor in favor of T2D association, Λ_{ij}, for the jth SNV at the ith distinct signal by $\pi_{ij}\Psi_i$, where $\Psi_i = \sum \Lambda_{ij}$ is the total Bayes factor for the signal. For the molecular QTL, approximate Bayes factors in favor of association for each variant were derived using Wakefield's method[65]. Under this model, COLOC then estimates the posterior probability of colocalization of the T2D association and molecular QTL (that is, hypothesis H_4, denoted as π_{COLOC}).

Plasmid transfection and luciferase reporter assay. We experimentally validated 99% credible set variants for distinct T2D-association signals at the *PROX1* locus using a luciferase reporter assay. Briefly, human EndoC-βH1 cells[66] and human liver cells were grown at 50–60% confluence in 24-well plates and were transfected (2×10^5 EndoC-βH1 cells per well and 5×10^4 HepG2 cells per well) with 500 ng of empty pGL3-Promoter vector (Promega) or pGL3-Promoter-PROX_insert with FuGENE HD (Roche Applied Science) using a FuGENE:DNA ratio of 6:1 according to the manufacturer's instructions. Details are provided in the Supplementary Note and at https://www.promega.co.uk/products/luciferase-assays/genetic-reporter-vectors-and-cell-lines/pgl3-luciferase-reporter-vectors/?catNum=E1751. Luciferase activities were measured 48 h after transfection using the Dual-Luciferase Reporter Assay kit (Promega) according to the manufacturer's instructions in half-volume 96-well format on an EnSpire Multimode Plate Reader (PerkinElmer). Firefly luciferase activity was normalized to the Renilla luciferase activity obtained by cotransfection of 10 ng of the pGL4.74[hRluc/TK] Renilla luciferase vector (Promega). All experiments were performed in triplicate on three different passages of each cell type. Differences in luciferase activity between groups were tested using two-tailed two-sample t-tests, and $P < 0.05$ was considered statistically significant.

Transferability of GRS across ancestry groups. We selected two studies per ancestry group as test GWAS, prioritizing those with larger effective sample sizes and greater genetic diversity (Supplementary Note). We repeated the multi-ancestry meta-regression after excluding the ten test GWAS, incorporating the same three axes of genetic variation as covariates to account for ancestry. The association P values from this 'reduced' meta-regression were then corrected for inflation due to residual structure between GWAS by means of genomic control adjustment (allowing for four degrees of freedom): $\lambda_{\mathrm{TA}} = 1.037$. SNVs reported in ≥50% of the total effective sample size of the 'reduced' meta-regression ($N_{\mathrm{TA}} \geq 179{,}074$) were included in downstream analyses. We identified loci attaining genome-wide significant evidence of association ($P < 5 \times 10^{-9}$) in the 'reduced' meta-regression, and the lead SNV for each locus was selected as the variant with the minimum association P value. For each test GWAS, we next estimated population-specific 'predicted' allelic effects for each lead SNV to be used as weights in the GRS. We also repeated each of the ancestry-specific fixed-effects meta-analyses after excluding the ten test GWAS and identified lead SNVs attaining genome-wide significant evidence of association ($P < 5 \times 10^{-8}$). For each test GWAS, we estimated the OR per unit of the population-specific multi-ancestry GRS and each ancestry-specific weighted GRS and the corresponding percentage of T2D variance explained (pseudo R^2). Details are provided in the Supplementary Note.

Predictive power of GRS in FinnGen. Individuals from FinnGen were genotyped with Illumina and Affymetrix arrays and were imputed up to the Finnish population-specific reference panel (SISu version 3). We excluded individuals due to non-Finnish ancestry, relatedness or missing age and/or sex. We derived Finnish-specific 'predicted' allelic effect estimates for each lead SNV from the multi-ancestry meta-regression to be used as weights in calculating the centered GRS for each individual. We excluded lead SNVs from the GRS that were not reported in FinnGen. We excluded individuals with missing T2D status or BMI from subsequent analyses, resulting in a total of 18,111 affected individuals and 111,119 unaffected individuals. We calculated the variance in T2D status explained (pseudo R^2) and the AUROC (calculated with a tenfold cross-validation) for models including BMI and/or GRS. We also conducted age-stratified analyses and tested for association of the GRS with age of T2D diagnosis. Details are provided in the Supplementary Note.

Selection analyses. We used Relate[42] to reconstruct genealogies for haplotypes from the 1000 Genomes Project reference panel (phase 3, October 2014 release)[13] separately for each population after excluding African American and admixed American populations in whom high levels of admixture are likely to confound selection evidence. We then used P values calculated for selection evidence for any variant that segregated in the population and passed quality-control filters[42], which quantify the extent to which the mutation has more descendants than other lineages that were present when it arose. We tested for evidence of selection for index SNVs for distinct T2D-association signals, which were partitioned into two groups, risk and protective, according to the direction of the allelic effect when aligned to the derived allele. We also tested for selection on a range of traits available in the UK Biobank[43] at the subset of index SNVs for which the derived allele increased risk of T2D. Details are provided in the Supplementary Note.

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

Association summary statistics from the multi-ancestry meta-analysis and annotation-informed fine-mapping are available through the AMP T2D Knowledge Portal (http://www.type2diabetesgenetics.org/) and the DIAGRAM Consortium data download website (http://diagram-consortium.org/downloads.html). Source data are provided with this paper.

References

48. Jónsson, H. et al. Whole genome characterization of sequence diversity of 15,220 Icelanders. *Sci. Data* **4**, 170115 (2017).
49. Mitt, M. et al. Improved imputation accuracy of rare and low-frequency variants using population-specific high-coverage WGS-based imputation reference panel. *Eur. J. Hum. Genet.* **25**, 869–876 (2017).
50. Moon, S. et al. The Korea Biobank Array: design and identification of coding variants associated with blood biochemical traits. *Sci. Rep.* **9**, 1382 (2019).
51. Cook, J. P., Mahajan, A. & Morris, A. P. Guidance for the utility of linear models in meta-analysis of genetic association studies of binary phenotypes. *Eur. J. Hum. Genet.* **25**, 240–245 (2016).
52. Devlin, B. & Roeder, K. Genomic control for association studies. *Biometrics* **55**, 997–1004 (1999).
53. Gurdasani, D., Barroso, I., Zeggini, E. & Sandhu, M. S. Genomics of disease risk in globally diverse populations. *Nat. Rev. Genet.* **20**, 520–535 (2019).
54. Willer, C. J., Li, Y. & Abecasis, G. R. METAL: fast and efficient meta-analysis of genome-wide association scans. *Bioinformatics* **26**, 2190–2191 (2010).
55. Han, B. & Eskin, E. Random-effects model aimed at discovering associations in meta-analysis of genome-wide association studies. *Am. J. Hum. Genet.* **88**, 586–598 (2011).

56. Purcell, S. et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).
57. Sobota, R. S. et al. Addressing population-specific multiple testing burdens in genetic association studies. *Ann. Hum. Genet.* **79**, 136–147 (2015).
58. Yang, J. et al. Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. *Nat. Genet.* **44**, 369–375 (2012).
59. Kass, R. E. & Raftery, A. E. Bayes factors. *J. Am. Stat. Assoc.* **90**, 773–795 (1995).
60. Maller, J. B. et al. Bayesian refinement of association signals for 14 loci in 3 common diseases. *Nat. Genet.* **44**, 1294–1301 (2012).
61. Harrow, J. et al. GENCODE: the reference human genome annotation for the ENCODE Project. *Genome Res.* **22**, 1760–1774 (2012).
62. ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74 (2012).
63. Pasquali, L. et al. Pancreatic islet enhancer clusters enriched in type 2 diabetes risk-associated variants. *Nat. Genet.* **46**, 136–143 (2014).
64. Pickrell, J. Joint analysis of functional genomic data and genome-wide association studies of 18 human traits. *Am. J. Hum. Genet.* **94**, 559–573 (2014).
65. Wakefield, J. A. Bayesian measure of the probability of false discovery in genetic epidemiology studies. *Am. J. Hum. Genet.* **81**, 208–227 (2007).
66. Ravassard, P. et al. A genetically engineered human pancreatic β cell line exhibiting glucose-inducible insulin secretion. *J. Clin. Invest.* **121**, 3589–3597 (2011).

Acknowledgements

A complete list of acknowledgements and funding appears in the Supplementary Note. This research was funded in part by the Wellcome Trust (grant numbers 064890, 072960, 083948, 084723, 085475, 086113, 088158, 090367, 090532, 095101, 098017, 098051, 098381, 098395, 101033, 101630, 104085, 106130, 200186, 200837, 202922, 203141, 206194, 212259, 212284, 212946 and 220457). For the purpose of open access, the authors have applied a CC-BY public copyright licence to any author accepted manuscript version arising from this submission.

Author contributions

DIAMANTE Consortium coordination, A. Mahajan, M.I.M., A.P.M.; manuscript preparation, A. Mahajan, C.N.S., W. Zhang, M.C.Y.N., L.E.P., H.K., G.Z.Y., S. Rüeger, L.S., A.L.G., M.B., J.I.R., M.I.M., A.P.M.; coordination of ancestry-specific GWAS collections, A. Mahajan, C.N.S., W. Zhang, M.C.Y.N., L.E.P., D.W.B., J.E.B., J.C.C., X.S., M.B.; central analysis group, A. Mahajan, C.N.S., W. Zhang, M.C.Y.N., L.E.P., H.K., Y.J.K., M. Horikoshi, J.M.M., D.T., S. Moon, S.-H.K., N.R.R., N.W.R., M. Loh, B.-J.K., J. Flanagan, J.B.M., K.L.M., J.E.B., J.C.C., X.S., M.B., J.I.R., M.I.M., A.P.M.; *PROX1* functional analyses, G.Z.Y., F.A., J.M.T., A.L.G.; GRS analyses in FinnGen, S. Rüeger, P.d.B.P.; selection analyses, L.S., S.R.M.; single-cell chromatin accessibility data, J. Chiou, D.G., S.P., M. Sander, K.J.G.; islet promoter Hi-C data generation, I.M.-E., J. Ferrer; study-level primary analyses, A. Mahajan, C.N.S., W. Zhang, M.C.Y.N., L.E.P., Y.J.K., M. Horikoshi, J.M.M., D.T., S. Moon, S.-H.K., K. Lin, F.B., M.H.P., F.T., J.N., X.G., A. Lamri, M.N., R.A.S., J.-J.L., A.H.-C., M. Graff, J.-F.C., E.J.P., J.Y., L.F.B., Y.T., Y.H., V.S., J.P.C., M.K., N.G., E.M.S., I.P., T.S., M.W., C. Sarnowski, C.G., D.N., S. Trompet, J. Long, M. Sun, L.T., W.-M.C., M. Ahmad, R.N., V.J.Y.L., C.H.T.T., Y.Y.J., C.-H.C., L.M.R., C. Lecoeur, B.P.P., A.N., L.R.Y., G.C., R.A.J., S. Tajuddin, E.K.K., P.A., A.H.X., H.S.C., B.E.C., J. Tan, X.S., A.P.M.; study-level phenotyping, genotyping and additional analyses, L.S.A., A.A., C.A.A.-S., M. Akiyama, S.S.A., A.B., Z.B., J.B.-J., I.B., J.A.B., C.M.B., T.A.B., M. Canouil, J.C.N.C., L.-C.C., M.-L.C., J. Chen, S.-H.C., Y.-T.C., Z.C., L.-M.C., M. Cushman, S.K.D., H.J.d.S., G.D., L.D., A.P.D., S.D., Q.D., K.-U.E., L.S.E., D.S.E., M.K.E., K.F., J.S.F., I.F., M.F., O.H.F., T.M.F., B.I.F., C.F., P.G., H.C.G., V.G., C.G.-V., M.E.G.-V., M.O.G., P.G.-L., M. Gross, Y.G., S. Hackinger, S. Han, A.T.H., C.H., A.-G.H., W. Hsueh, M. Huang, W. Huang, Y.-J.H., M.Y.H., C.-M.H., S.I., M.A.I., M. Ingelsson, M.T.I., M. Isono, H.-M.J., F.J., G.J., J.B.J., M.E.J., T.J., Y.K., F.R.K., A. Kasturiratne, T. Katsuya, V.K., T. Kawaguchi, J.M.K., A.N.K., C.-C.K., M.G.K., K.K., J. Kriebel, F.K., J. Kuusisto, K. Läll, L.A.L., M.-S.L., N.R.L., A. Leong, L. Li, Y. Li, R.L.-G., S. Ligthart, C.M.L., A. Linneberg, C.-T.L., J. Liu, A.E.L., T.L., J. Luan, A.O.L., X.L., J. Lv, V.L., V.M., K.R.M., T.M., A. Metspalu, A.D.M., G.N.N., J.L.N., M.A.N., U.N., S.S.N., I.N., Y.O., L.O., S.R.P., M.A. Pereira, A.P., F.J.P., B.P., G. Prasad, L.J.R.-T., A.P.R., M.R., R.R., K.R., C. Sabanayagam, K. Sandow, N.S., S.S., C. Schurmann, M. Shahriar, J.S., D.M.S., D. Shriner, J.A.S., W.Y.S., A.S., A.M.S., K. Strauch, K. Suzuki, A.T., K.D.T., B. Thorand, G.T., U.T., B. Tomlinson, F.-J.T., J. Tuomilehto, T.T.-L., M.S.U., A.V.-S., R.M.v.D., J.B.v.K., R.V., M.V., N.W.-R., E.W., E.A.W., A.R.W., K.W.v.D., D.R.W., C.S.Y., K. Yamamoto, T.Y., L.Y., K. Yoon, C.Y., J.-M.Y., S.Y., L.Z., W. Zheng; study-level principal investigator, L.J.R., M. Igase, E. Ipp, S. Redline, Y.S.C., L. Lind, M.A. Province, C.L.H., P.A.P., E. Ingelsson, A.B.Z., B.M.P., Y.-X.W., C.N.R., D.M.B., F.M., Y. Liu, E.Z., M.Y., S.S.R., C.K., J.S.P., J.C.E., Y.-D.I.C., P.F., J.G.W., W.H.H.S., S.L.R.K., J.-Y.W., M.G.H., R.C.W.M., T.-Y.W., L.G., D.O.M.-K., G.R.C., F.S.C., D.B., G. Paré, M.M.S., H.A., A.A.M., X.-O.S., K.-S.P., J.W.J., M. Cruz, R.M.-C., H.G., C.-Y.C., E.P.B., A.D., E.-S.T., J.D., N.K., M. Laakso, A. Köttgen, W.-P.K., C.N.A.P., S. Liu, G.A., J.S.K., R.J.F.L., K.E.N., C.A.H., J.C.F., D. Saleheen, T.H., O.P., R.M., C. Langenberg, N.J.W., S. Maeda, T. Kadowaki, J. Lee, I.Y.M., R.G.W., K. Stefansson, J.B.M., K.L.M., D.W.B., J.C.C., M.B., J.I.R., M.I.M., A.P.M.

Competing interests

A. Mahajan is now an employee of Genentech and a holder of Roche stock. R.A.S. is now an employee of GlaxoSmithKline. V.S. is an employee of deCODE Genetics–Amgen. L.S.E. is now an employee of Bristol Myers Squibb. J.S.F. has consulted for Shionogi. T.M.F. has consulted for Sanofi and Boerhinger Ingelheim and received funding from GSK. H.C.G. holds the McMaster–Sanofi Population Health Institute Chair in Diabetes Research and Care; reports research grants from Eli Lilly, AstraZeneca, Merck, Novo Nordisk and Sanofi; reports honoraria for speaking from AstraZeneca, Boehringer Ingelheim, Eli Lilly, Novo Nordisk, DKSH, Zuellig, Roche and Sanofi; and reports consulting fees from Abbott, AstraZeneca, Boehringer Ingelheim, Eli Lilly, Merck, Novo Nordisk, Pfizer, Sanofi, Kowa and Hanmi. M. Ingelsson is a paid consultant for BioArctic. R.L.-G. is a part-time consultant for Metabolon. A.E.L. is now an employee of the Regeneron Genetics Center and holds shares in Regeneron Pharmaceuticals. M.A.N. currently serves on the scientific advisory board for Clover Therapeutics and is an advisor to Neuron23. S.R.P. has received grant funding from Bayer Pharmaceuticals, Philips Respironics and Respicardia. N.S. has consulted for or been on speaker bureaus for Abbott, Amgen, AstraZeneca, Boehringer Ingelheim, Eli Lilly, Hanmi, Novartis, Novo Nordisk, Sanofi and Pfizer and has received grant funding from AstraZeneca, Boehringer Ingelheim, Novartis and Roche Diagnostics. A.M.S. receives funding from Seven Bridges Genomics to develop tools for the NHLBI BioData Catalyst consortium. G.T. is an employee of deCODE Genetics–Amgen. U.T. is an employee of deCODE Genetics–Amgen. E. Ingelsson is now an employee of GlaxoSmithKline. B.M.P. serves on the steering committee of the Yale Open Data Access Project funded by Johnson & Johnson. R.C.W.M. reports research funding from AstraZeneca, Bayer, Novo Nordisk, Pfizer, Tricida and Sanofi and has consulted for or received speakers fees from AstraZeneca, Bayer and Boehringer Ingelheim, all of which have been donated to the Chinese University of Hong Kong to support diabetes research. D.O.M.-K. is a part-time clinical research consultant for Metabolon. S. Liu reports consulting payments and honoraria or promises of the same for scientific presentations or reviews at numerous venues, including but not limited to Barilla, by-Health, Ausa Pharmed, the Fred Hutchinson Cancer Center, Harvard University, the University of Buffalo, Guangdong General Hospital and the Academy of Medical Sciences; is a consulting member for Novo Nordisk; is a member of the data safety and monitoring board for a trial of pulmonary hypertension in patients with diabetes at Massachusetts General Hospital; receives royalties from UpToDate; and receives an honorarium from the American Society for Nutrition for his duties as an associate editor. K. Stefansson is an employee of deCODE Genetics–Amgen. K.J.G. consults for Genentech and holds stock in Vertex Pharmaceuticals. A.L.G.'s spouse is an employee of Genentech and holds stock options in Roche. M.I.M. has served on advisory panels for Pfizer, Novo Nordisk and Zoe Global; has received honoraria from Merck, Pfizer, Novo Nordisk and Eli Lilly and research funding from AbbVie, AstraZeneca, Boehringer Ingelheim, Eli Lilly, Janssen, Merck, Novo Nordisk, Pfizer, Roche, Sanofi Aventis, Servier and Takeda; is now an employee of Genentech and a holder of Roche stock. The remaining authors declare no competing interests. The views expressed in this article are those of the authors and do not necessarily represent those of the NHS, the NIHR or the UK Department of Health; the National Heart, Lung, and Blood Institute, the National Institutes of Health or the US Department of Health and Human Services.

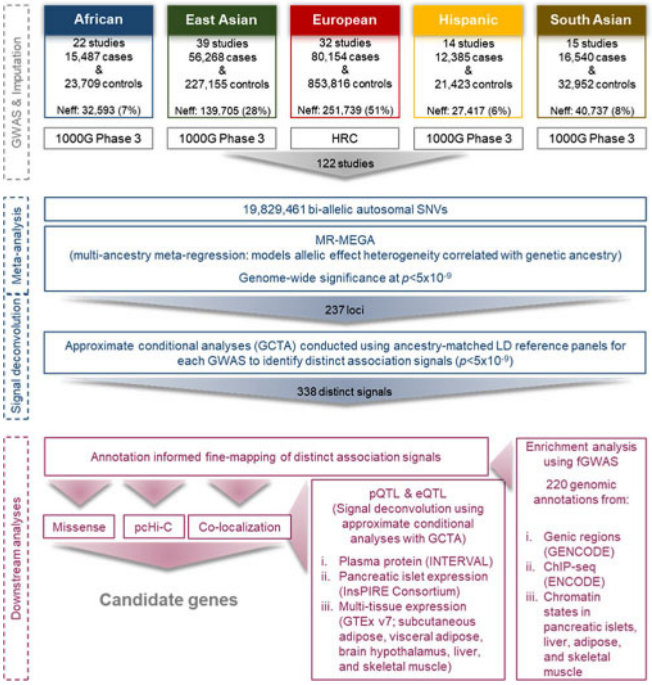
Additional information

Extended data is available for this paper at https://doi.org/10.1038/s41588-022-01058-3.

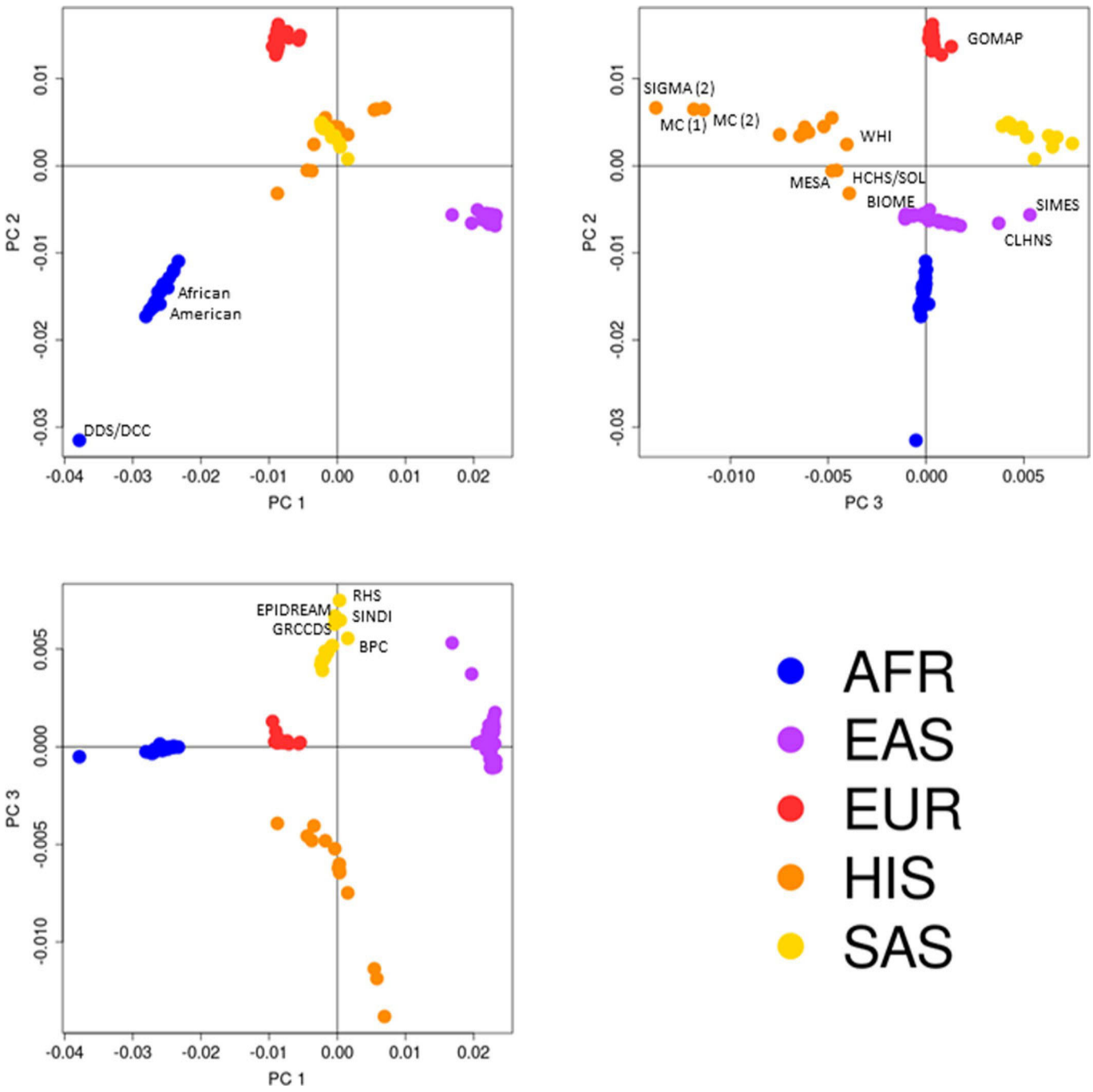
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41588-022-01058-3.

Correspondence and requests for materials should be addressed to Anubha Mahajan, Mark I. McCarthy or Andrew P. Morris.

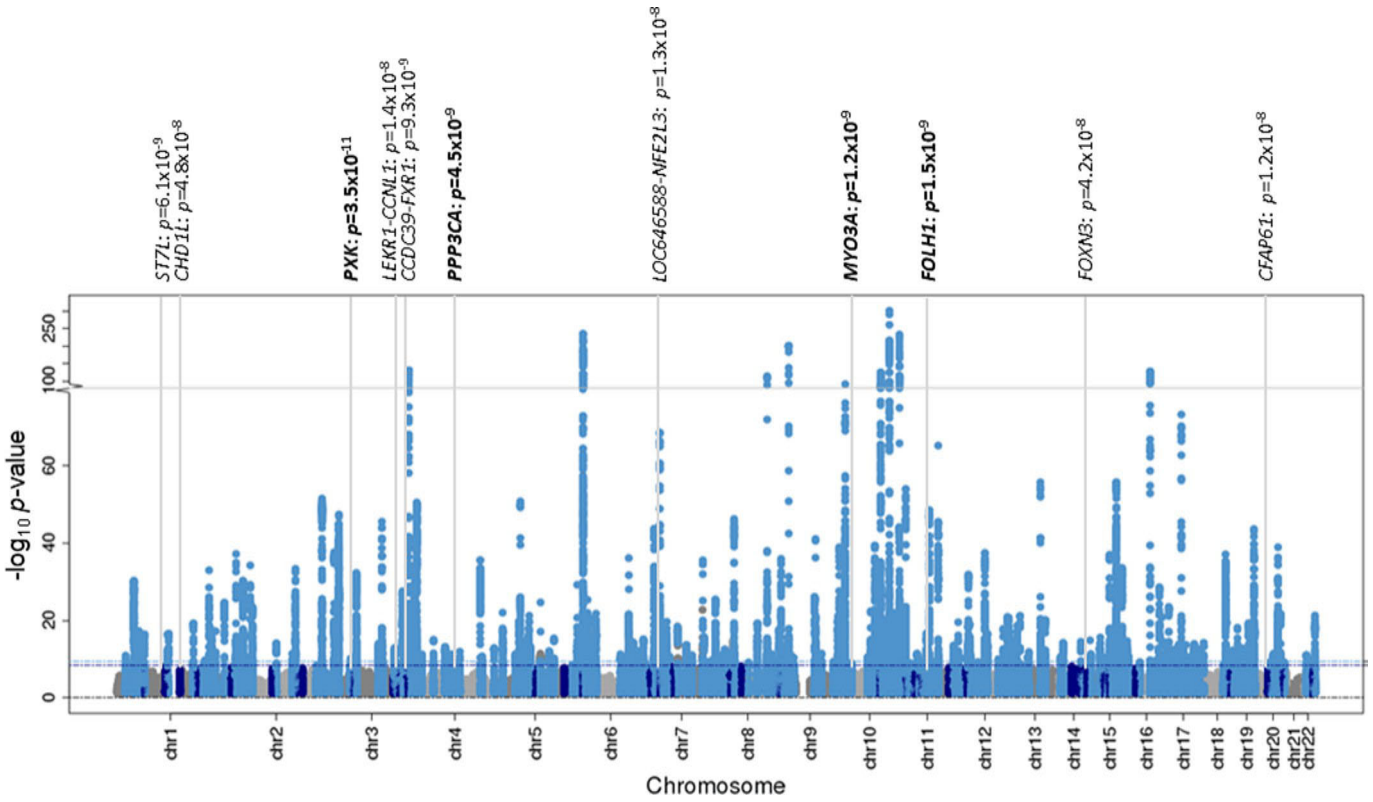
Peer review information *Nature Genetics* thanks Constantin Polychronakos and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.



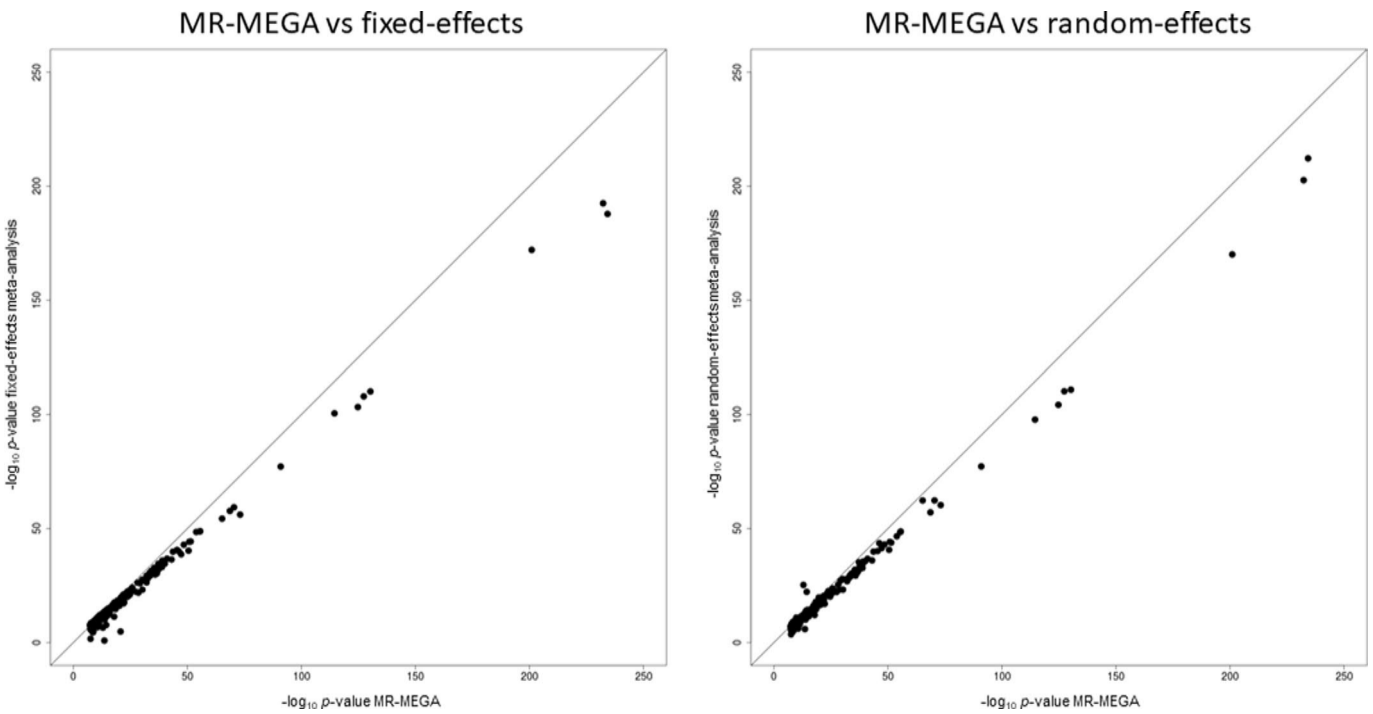
Extended Data Fig. 1 | Study overview. Summary of data resources and downstream analyses to identify candidate causal genes at T2D susceptibility loci.



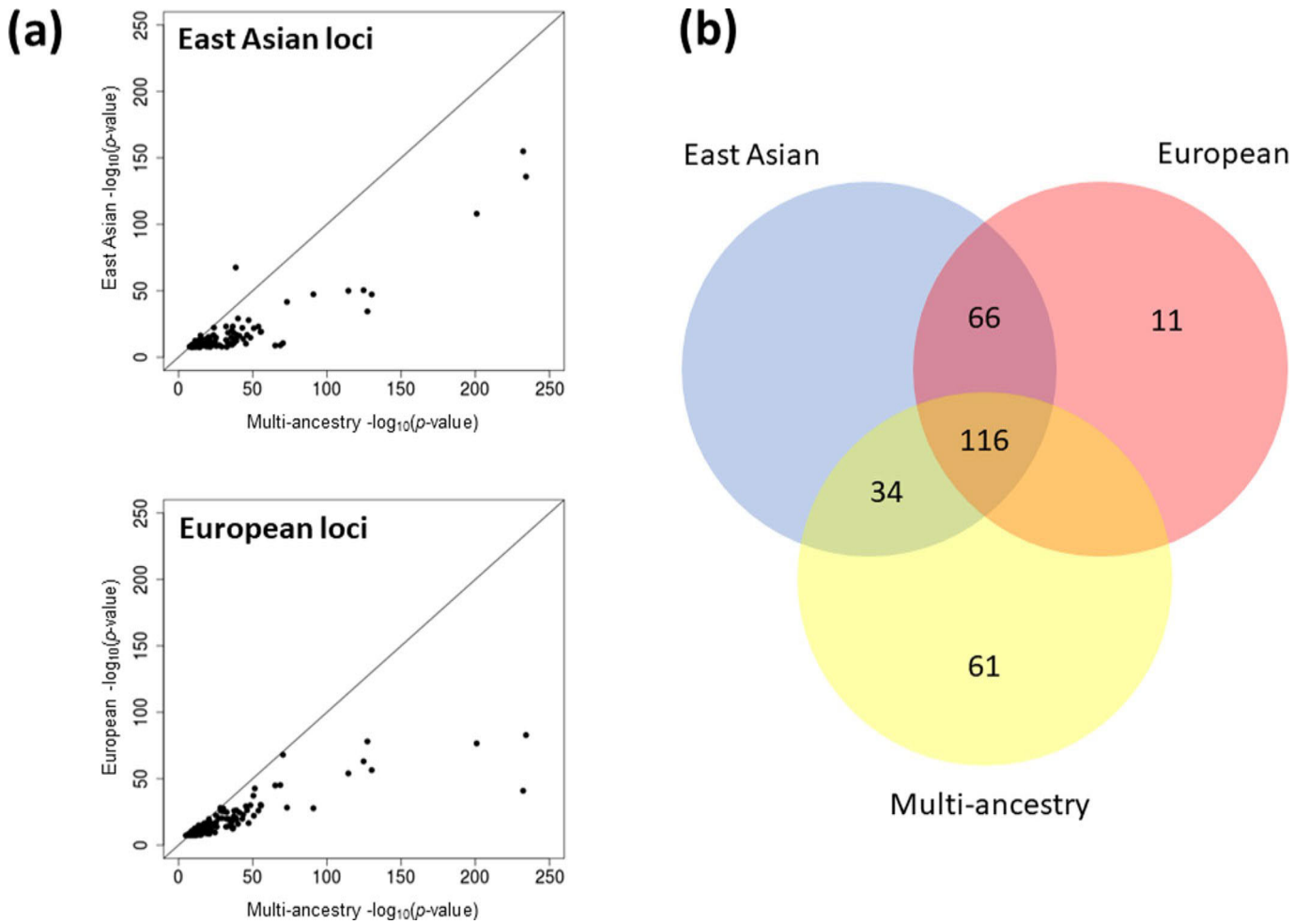
Extended Data Fig. 2 | Axes of genetic variation separating GWAS of T2D across diverse populations. The first three axes of genetic variation (PC 1, PC 2 and PC 3) from multi-dimensional scaling of the Euclidean distance matrix between populations are sufficient to separate five ancestry groups: African (AFR), East Asian (EAS), European (EUR), Hispanic (HIS) and South Asian (SAS). GWAS acronyms are defined in Supplementary Table 1. The second axis of genetic variation (PC 2) separates African American and continental African GWAS. The third axis of genetic variation (PC 3) reveals finer-scale differences between GWAS within ancestry groups: Hispanic studies with a greater proportion of American ancestry (SIGMA (2), MC (1) and MC (2)) or African ancestry (WHI, MESA, HCHS/SOL and BIOME); East Asian studies of Chinese, Japanese and Korean ancestry from those of Malay and Filipino ancestry (SIMES and CLHNS); South Asian studies of Sri Lankan, Bangladeshi and South Indian ancestry (RHS, EPIDREAM, SINDI, GRCCDS and BPC) from those of North Indian and Pakistani ancestry; and Northern European ancestry studies from the study of Greek ancestry from Southern Europe (GOMAP). GWAS were aligned to ancestry groups based on self-report at the study level.



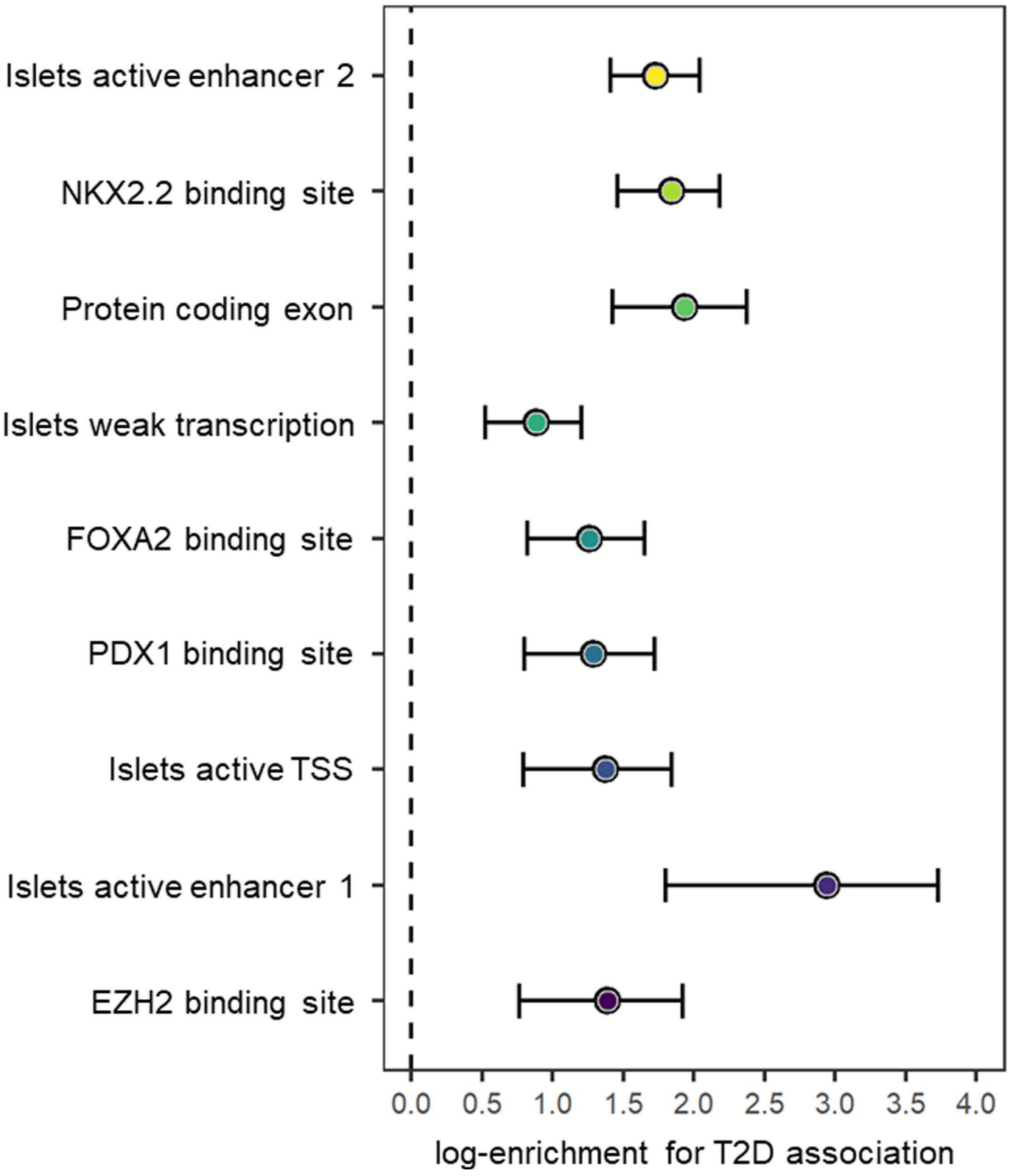
Extended Data Fig. 3 | Manhattan plot of genome-wide T2D association from multi-ancestry meta-regression (MR-MEGA) of up to 180,834 cases and 1,159,055 controls. Each point represents an SNV passing quality control in the multi-ancestry meta-regression, plotted with their association *P*-value (on a -$\log_{10}$ scale, truncated at 300) as a function of genomic position (NCBI build 37). Association signals attaining genome-wide significance are highlighted in pale blue ($P < 5 \times 10^{-9}$) and dark blue ($P < 5 \times 10^{-8}$). The names of novel loci names are highlighted with their association *P*-value from the multi-ancestry meta-regression.



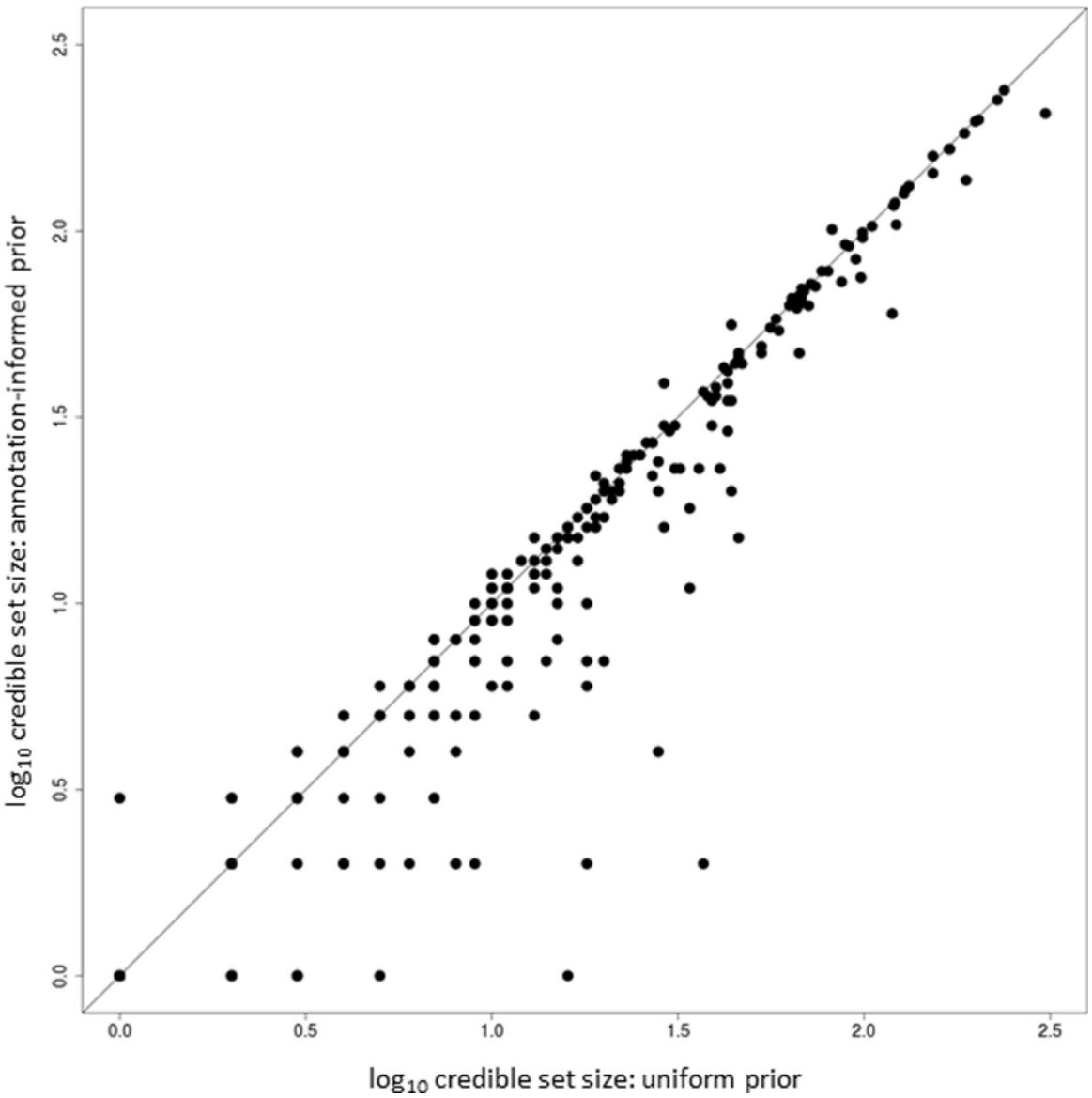
Extended Data Fig. 4 | Comparison of association *P*-values at lead SNVs at T2D loci between multi-ancestry meta-regression (MR-MEGA), fixed-effects meta-analysis and random-effects (RE2) meta-analysis of up to 180,834 cases and 1,159,055 controls. Each point corresponds to an SNV, plotted according to *P*-values (on a -$\log_{10}$ scale) from MR-MEGA on the *x*-axis and fixed- or random-effects meta-analysis on the *y*-axis. SNVs below the *y* = *x* line demonstrate stronger association with MR-MEGA. The lead SNV at the *TCF7L2* locus has been removed to improve clarity of presentation.



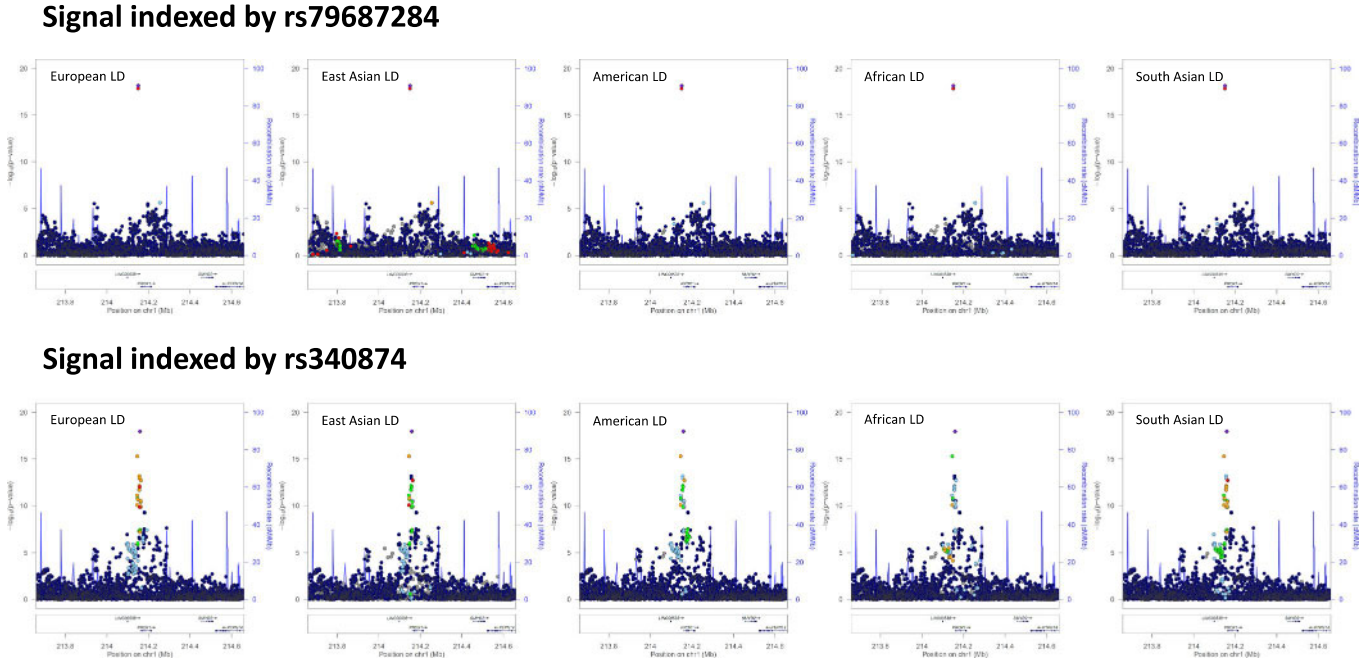
Extended Data Fig. 5 | Comparison of loci identified at genome-wide significance ($P < 5 \times 10^{-8}$) in multi-ancestry meta-regression (180,834 cases and 1,159,055 controls), and East Asian and European ancestry-specific meta-analyses (56,268 cases and 227,155 controls, and 80,154 cases and 853,816 controls, respectively). a, Association *P*-values at loci identified in East Asian and European ancestry-specific meta-analyses. Each point corresponds to a locus, plotted according to the *P*-value (on a $-\log_{10}$ scale) for the lead SNP in the multi-ancestry meta-regression on the *x*-axis and the lead SNP in the ancestry-specific meta-analysis on the *y*-axis. The *TCF7L2* locus has been removed to improve clarity of presentation. Loci plotted below the $y = x$ line show stronger evidence for association in the multi-ancestry meta-regression. **b**, Overlap of loci identified in multi-ancestry meta-regression and ancestry-specific meta-analyses.



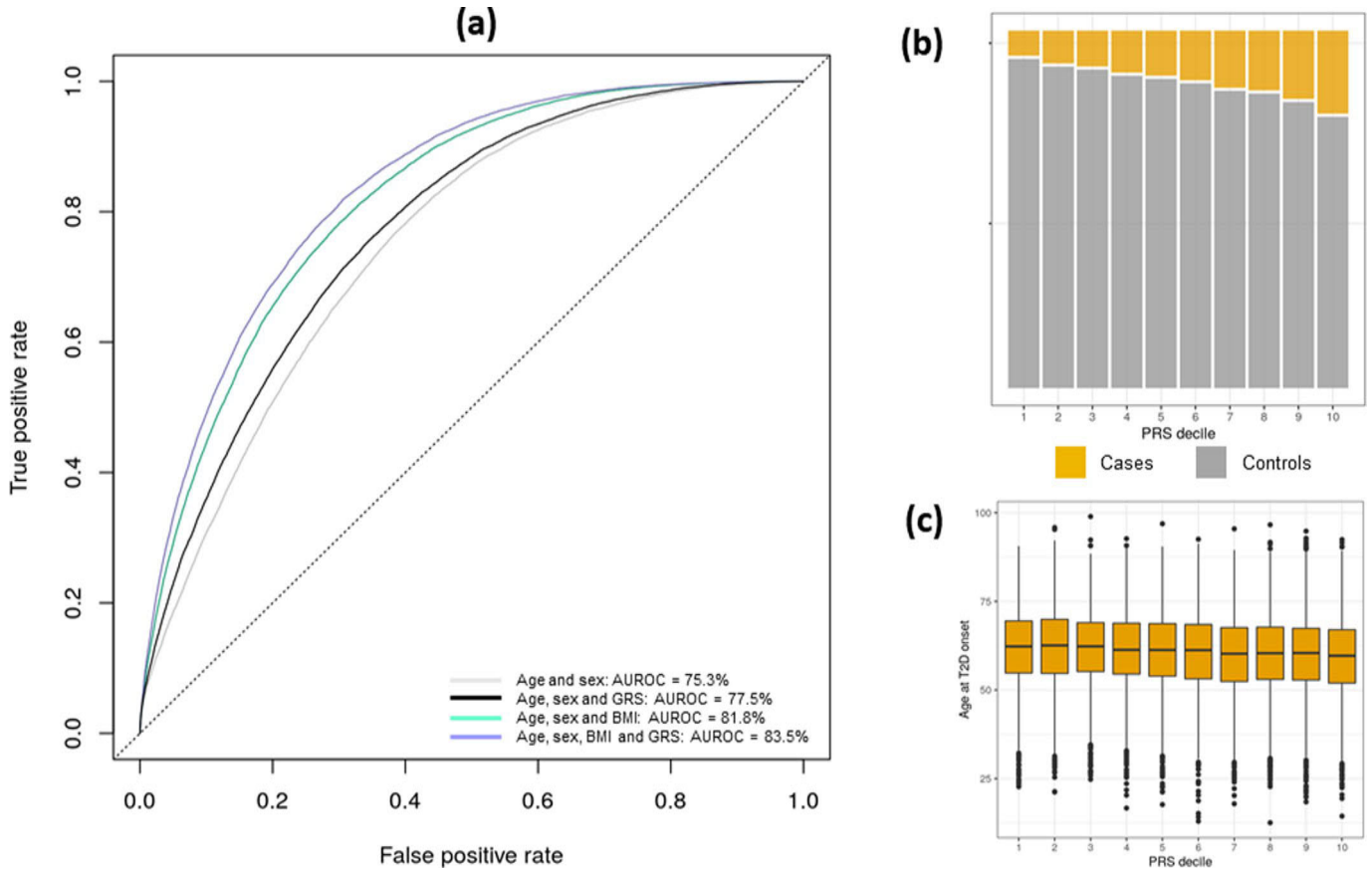
Extended Data Fig. 6 | Summary statistics from joint fGWAS model of enriched functional and regulatory annotations across distinct T2D association signals from multi-ancestry meta-regression (MR-MEGA) of up to 180,834 cases and 1,159,055 controls. Each point corresponds to an annotation, plotted for the log-enrichment for T2D association on the *x*-axis, with bars representing the corresponding 95% confidence interval (CI).



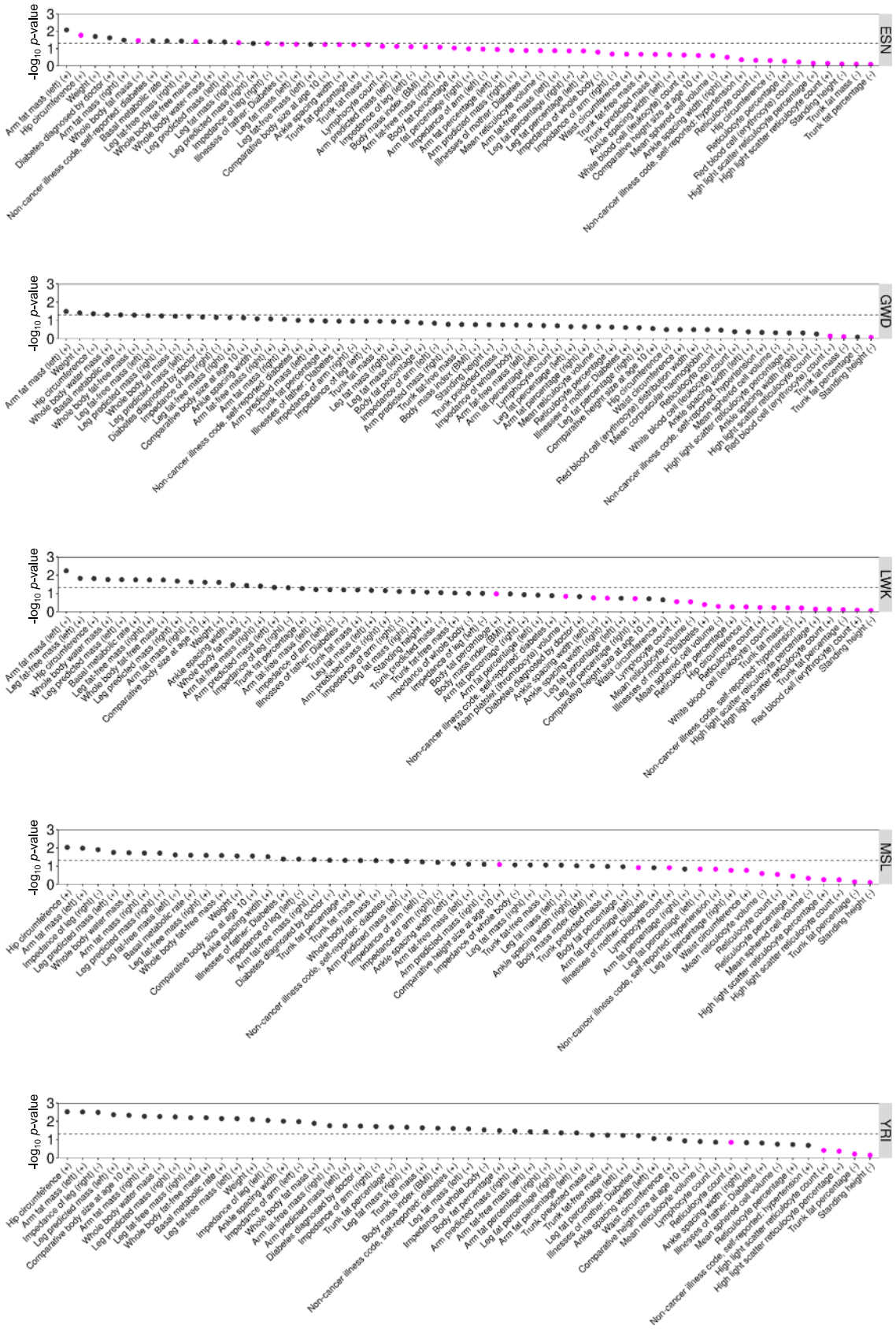
Extended Data Fig. 7 | Comparison of number of SNVs in 99% credible set for distinct association signals for T2D obtained from the multi-ancestry meta-regression of 180,834 cases and 1,159,055 controls under uniform and annotation-informed prior models of causality. Each point corresponds to a distinct association signal, plotted according to the $\log_{10}$ credible set size under the uniform prior on the *x*-axis and the $\log_{10}$ credible set size under the annotation-informed prior on the *y*-axis. The 144 (42.6%) signals below the $y = x$ line were more precisely fine-mapped under the annotation-informed prior.



Extended Data Fig. 8 | Differences in LD structure between ancestry groups at the *PROX1* locus for distinct association signals from multi-ancestry meta-regression (MR-MEGA) of up to 180,840 cases and 1,159,185 controls. Each point represents an SNV passing quality control in the multi-ancestry meta-regression (after conditional analysis), plotted with their association *P*-value (on a $\log_{10}$ scale) as a function of genomic position (NCBI build 37). The index SNV is represented by the purple symbol. The color coding of all other SNVs indicates LD with the index variant in the ancestry-matched reference haplotypes from the 1000 Genomes Project panel: red, $r^2 \geq 0.8$; gold, $0.6 \leq r^2 < 0.8$; green, $0.4 \leq r^2 < 0.6$; cyan, $0.2 \leq r^2 < 0.4$; blue, $r^2 < 0.2$; grey, r^2 unknown. Recombination rates are estimated from Phase II HapMap and gene annotations are taken from the University of California Santa Cruz genome browser.



Extended Data Fig. 9 | Power of multi-ancestry GRS to predict T2D status in 129,230 individuals of Finnish ancestry from FinnGen. a, Age under receiver operating characteristic curve (AUROC) after adding BMI and GRS to a baseline model adjusting for age and sex. **b**, Prevalence of T2D across GRS deciles. **c**, Boxplot of the distribution of age at T2D diagnosis across GRS deciles: box defines upper quartile, median and lower quartile, bars define maximum and minimum values within 1.5 x interquartile range of the upper and lower quartiles, other points are outliers.



Extended Data Fig. 10 | Evidence for selection from Relate in African ancestry populations of subsets of T2D risk variants (effect aligned to derived allele) that are associated with other traits available in the UK Biobank. Nominal evidence for selection ($P < 0.05$) is indicated by the dashed line. The color of each point indicates the evidence for selection of subsets of T2D risk variants that are not associated with the other trait: $P < 0.05$ (pink) and $P \geq 0.05$ (black). Population abbreviations: ESN, Esan in Nigeria; GWD, Gambian in Western Divisions in the Gambia; LWK, Luhya in Webuye, Kenya; MSL, Mende in Sierra Leone; YRI, Yoruba in Ibadan, Nigeria.

nature research

Corresponding author(s): Andrew Morris

Last updated by author(s): Nov 29, 2021

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☒	☐	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☐	☒	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	No software was used.
Data analysis	MR-MEGA v0.2, METAL v2011-03-25, METASOFT v2.0.0, PLINK v1.9, GCTA v1.26.0, fGWAS v0.3.6, coloc v3.1, R v3.4.2 (gtx package), Relate v1.0

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

Association summary statistics from the trans-ancestry meta-analysis and annotation informed fine-mapping will be made available through the AMP-T2D Knowledge Portal (http://www.type2diabetesgenetics.org/) and the DIAGRAM Consortium repository (http://diagram-consortium.org/downloads.html).

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	GWAS meta-analysis. We combined the largest sample size of type 2 diabetes cases and (population) controls that was available to the DIAMANTE Consortium. At our trans-ancestry genome-wide significance threshold (p<5x10-9), under an additive genetic model, we had ≥80% power to detect association of SNVs with MAF ≥5% and OR ≥1.045 or MAF ≥0.5% and OR ≥1.145. Luciferase reporter assays. The sample size was set as n=3; which means the vector transfection was performed three time (using different passage numbers) of each cell type.
Data exclusions	GWAS meta-analysis. Within each contributing study, individuals were excluded on the basis of well-established individual and variant quality control (QC) procedures to remove poor quality genotypes, samples and SNVs. These QC procedures are described in Supplementary Table 3 for each study. Luciferase reporter assays. There were no data exclusions.
Replication	GWAS meta-analysis. We did not conduct replication since we had already brought together all study data available to us via meta-analysis. All reported association signals were checked to confirm that effects were not driven by false positives in single studies. Luciferase reporter assays. Assays were performed with three biological replicates by using three different passage numbers of cells of each cell type. Within each assay, three technical replicates were included for each condition.
Randomization	GWAS meta-analysis. Randomization was not performed. Within each study, covariates were adjusted for to account for potential confounding. Covariate adjustments are reported in Supplementary Table 3. Luciferase reporter assays. Randomization was not performed.
Blinding	GWAS meta-analysis. Group allocation was not relevant to this study, so blinding was not necessary. Luciferase reporter assays. Blinding was not needed because the construction of each vector was designed before performing the assays.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☐	☒ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☐	☒ Human research participants
☒	☐ Clinical data
☒	☐ Dual use research of concern

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Eukaryotic cell lines

Policy information about cell lines

Cell line source(s)	Two cell lines were used for the Luciferase reporter assays. The EndoC-BH1 cell line, which is a commerically available genetically engineered from a human Beta cell line (https://www.jci.org/articles/view/58447) purchased from Human Cell Design (https://www.humancelldesign.com/). The HepG2 cell line was generated from human liver tissue and was purchased from ATCC (https://www.atcc.org/products/hb-8065).
Authentication	The EndoC-BH1 cell line was authenticated at a transcriptomic level (European Nucleotide Archive [ENA;] http://

Authentication	www.ebi.ac.uk/ena) under accession number PRJEB15283) and extensively characterized (Hastoy et al Scientific Reports 8, 16994; 2018). The HepG2 cell line (BH-8065) purchased from ATCC was authenticated by ATCC through the accessioning process.
Mycoplasma contamination	Both the EndoC-BH1 and HepG2 cell lines tested negative for mycoplasma contamination.
Commonly misidentified lines (See ICLAC register)	No misidentified cell line was used in the Luciferase reporter assays.

Human research participants

Policy information about studies involving human research participants

Population characteristics	Characteristics are presented for each contributing study in Supplementary Table 2.
Recruitment	Ascertainment of type 2 diabetes cases and controls for each contributing study are presented in Supplementary Table 1.
Ethics oversight	All human research was approved within each contributing study by the relevant institutional review boards and conducted according to the Declaration of Helsinki. All participants provided written informed consent. Ethics statements from each contributing study are provided in the Supplementary Note.

Note that full information on the approval of the study protocol must also be provided in the manuscript.