Article

Genetic drivers of heterogeneity in type 2 diabetes pathophysiology

Type 2 diabetes (T2D) is a heterogeneous disease that develops through diverse pathophysiological processes[1,2] and molecular mechanisms that are often specific to cell type[3,4]. Here, to characterize the genetic contribution to these processes across ancestry groups, we aggregate genome-wide association study data from 2,535,601 individuals (39.7% not of European ancestry), including 428,452 cases of T2D. We identify 1,289 independent association signals at genome-wide significance ($P < 5 \times 10^{-8}$) that map to 611 loci, of which 145 loci are, to our knowledge, previously unreported. We define eight non-overlapping clusters of T2D signals that are characterized by distinct profiles of cardiometabolic trait associations. These clusters are differentially enriched for cell-type-specific regions of open chromatin, including pancreatic islets, adipocytes, endothelial cells and enteroendocrine cells. We build cluster-specific partitioned polygenic scores[5] in a further 279,552 individuals of diverse ancestry, including 30,288 cases of T2D, and test their association with T2D-related vascular outcomes. Cluster-specific partitioned polygenic scores are associated with coronary artery disease, peripheral artery disease and end-stage diabetic nephropathy across ancestry groups, highlighting the importance of obesity-related processes in the development of vascular outcomes. Our findings show the value of integrating multi-ancestry genome-wide association study data with single-cell epigenomics to disentangle the aetiological heterogeneity that drives the development and progression of T2D. This might offer a route to optimize global access to genetically informed diabetes care.

Diabetes mellitus is a huge public-health burden, with an estimated prevalence of 537 million adults worldwide in 2021, of whom more than 90% are affected by T2D[6]. The biological processes through which T2D develops are diverse and include impaired insulin secretion and insulin resistance. This aetiological heterogeneity leads to substantial variability in patient phenotypes, including age of disease onset, manifestation of disease complications and response to management strategies[1,2]. Although environment and lifestyle are well-established risk factors for T2D, heritability has been estimated to be 69% amongst individuals of 35–60 years of age[7]. Previous genome-wide association studies (GWASs) of T2D have identified more than 500 risk loci[8,9], which showed variable patterns of association with clinical features mediated by effector genes acting through distinct molecular mechanisms that are often cell-type specific[3,4]. Through the newly established Type 2 Diabetes Global Genomics Initiative, we present findings from a very large meta-analysis of T2D GWAS data, comprising more than 2.5 million individuals of diverse ancestry—an increase of nearly threefold in the effective sample size compared with previous efforts[8,9]. We take advantage of the power afforded by this increased sample size and combine the GWAS data with emerging single-cell functional genomics data derived from disease-relevant tissues to uncover the aetiological heterogeneity of T2D. Furthermore, we construct partitioned polygenic scores (PSs)[5] across multiple ancestry groups, and assess their association with T2D-related macrovascular outcomes and progression to microvascular complications.

Study overview

We assembled GWAS data, including 428,452 cases of T2D and 2,107,149 controls (Supplementary Fig. 1 and Supplementary Tables 1 and 2). We organized these GWASs into six subsets of genetically similar studies, which we refer to as 'ancestry groups' (Extended Data Fig. 1). Specifically, we considered: a European ancestry group (EUR, 60.3% of the effective sample size); an East Asian ancestry group (EAS, 19.8%); an admixed African American group with ancestry predominantly from West Africa and Europe (AFA, 10.5%); an admixed Hispanic group with ancestry predominantly from the Americas, West Africa and Europe (HIS, 5.9%); a South Asian ancestry group (SAS, 3.3%); and a South African ancestry group (SAF, 0.2%). Association analyses accounted for study-level population structure and relatedness, and adjusted for age and sex, where appropriate, and additional study-specific covariates (Supplementary Table 3 and Methods).

Discovery of T2D loci

We aggregated association summary statistics across GWASs through multi-ancestry meta-regression, implemented in MR-MEGA (ref. 10), which allows for allelic effect heterogeneity that is correlated with ancestry. We included three axes of genetic variation as covariates in the meta-regression model that separated GWASs from different ancestry groups (Extended Data Fig. 1 and Methods), which resulted in

A list of authors and their affiliations appears at the end of the paper.
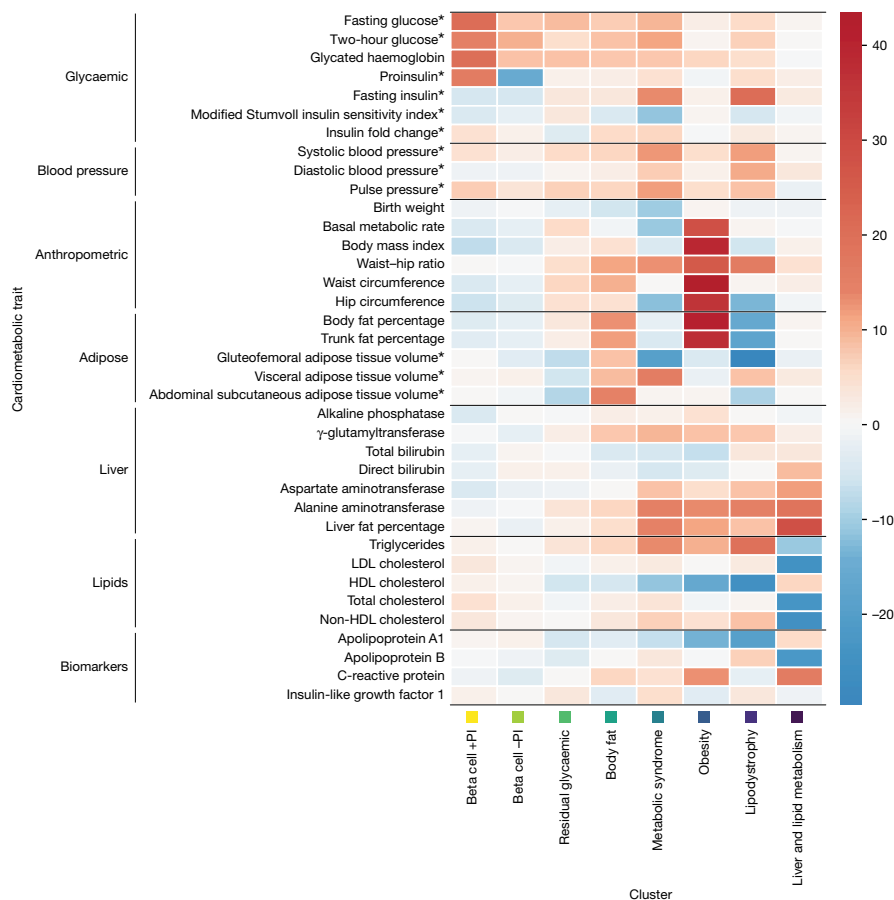
Fig. 1 | Heat map of associations of 37 cardiometabolic phenotypes with 8 mechanistic clusters of index SNVs for T2D association signals. Each column corresponds to a cluster. Each row corresponds to a cardiometabolic phenotype. The 'temperature' of each cell represents the z-score (aligned to the T2D risk allele) of association of the phenotype with index SNVs assigned to the cluster. *Phenotype is adjusted for body mass index.

lower genomic control inflation than did a fixed-effects meta-analysis ($\lambda_{GC} = 1.120$ and $\lambda_{GC} = 1.396$, respectively).

The DIAMANTE Consortium previously advocated the use of a multi-ancestry genome-wide significance threshold ($P < 5 \times 10^{-9}$) to define loci, which takes account of the weaker linkage disequilibrium (LD) between single-nucleotide variants (SNVs) expected after multi-ancestry meta-analysis[9]. To gain insight into true positive signals meeting conventional genome-wide significance ($P < 5 \times 10^{-8}$) that would be overlooked at this more stringent threshold, we considered loci reported by the DIAMANTE Consortium, which contributed 39.5% of the effective sample size of the current study. Of 39 loci with association signals meeting $5 \times 10^{-9} \le P < 5 \times 10^{-8}$ in the DIAMANTE Consortium analysis, 36 (92.3%) attained multi-ancestry genome-wide significance with the larger sample size available to us in the current study (Supplementary Text). We therefore focused our downstream analyses on SNVs that met the conventional genome-wide significance threshold.

We identified a total of 1,289 distinct T2D association signals ($P < 5 \times 10^{-8}$) that were represented by independent ($r^2 < 0.05$) index SNVs (Supplementary Fig. 2, Supplementary Table 4 and Methods). The 1,289 association signals mapped to 611 loci, of which 145 (23.7%) loci have not to our knowledge been previously reported in GWASs of T2D. At association signals that mapped to loci not previously reported for T2D, index SNVs were predominantly common (minor allele frequency (MAF) higher than 5% in at least one ancestry group) with odds ratios (ORs) lower than 1.05 (Supplementary Fig. 3).

Mechanistic clusters of T2D index SNVs

To understand the genetic contribution to phenotypic heterogeneity in T2D, we classified the 1,289 index SNVs according to their profile of associations (aligned to the T2D risk allele) with 37 cardiometabolic phenotypes. These included glycaemic traits, anthropometric measures, body fat and adipose tissue volume, blood pressure, levels of circulating plasma lipids, and biomarkers of liver function and lipid metabolism[11–19] (Supplementary Table 5). We applied an unsupervised 'hard clustering' approach with imputation of missing phenotype associations, which identified eight non-overlapping but exhaustive subsets of index SNVs with similar cardiometabolic profiles (Fig. 1, Table 1, Extended Data Fig. 2, Supplementary Fig. 4, Supplementary Tables 6 and 7 and Methods).

We observed that the cardiometabolic features and loci of five of our identified clusters overlapped with those reported in previous efforts[3,4,20,21], representing beta-cell dysfunction with a positive or negative association with proinsulin (PI), and insulin resistance mediated through obesity, lipodystrophy, and liver and lipid metabolism (Supplementary Table 8). T2D risk alleles at index SNVs in the two beta-cell-dysfunction clusters are associated with increased fasting glucose, two-hour glucose and glycated haemoglobin, and with decreased fasting insulin. Index SNVs in both clusters are also associated with PI, but with opposite directions of effect for the T2D risk allele. The clusters reflecting mechanisms of insulin resistance mediated through obesity, lipodystrophy, and liver and lipid metabolism include index SNVs that are associated with anthropometric measures and levels of

Table 1 | Cardiometabolic profile, example loci and physiological effect of index SNVs at T2D association signals allocated to eight mechanistic clusters

Mechanistic cluster	Cardiometabolic profile	Number of T2D associations	Example loci	Physiological effect	
				Insulin secretion	Insulin sensitivity
Beta cell +PI	+FG*, +2hG*, +HbA1c, +PI*	91	TCF7L2, KCNQ1, CDKAL1, CDKN2A–CDKN2B, SLC30A8	–	+
Beta cell –PI	+FG*, +2hG*, +HbA1c, –PI*	89	CDC123–CAMK1D, HNF1B, KCNJ11–ABCC8, HNF4A, HNF1A	–	+
Residual glycaemic	+FG*, +HbA1c	389	GCC1–PAX4–LEP, ANKRD55, GCKR, UBE2E2	–	–
Body fat	+Body fat, +ASAT*	273	ZMIZ1, HMGA2, CTBP1	+	–
Metabolic syndrome	+FG*, +FI*, +WHR, +VAT*, –GFAT*, +TG, –HDL, +BP	166	IGF2BP2, CCND2, HHEX–IDE, JAZF1, GPSM1	+	–
Obesity	+BMI, +WHR, +body fat, +BMR, +TG, –HDL	233	FTO, MC4R, MACF1, TMEM18	+	–
Lipodystrophy	+FI*, +WHR, –body fat, –GFAT*, +TG, –HDL, +BP	45	IRS1, GRB14–COBLL1, PPARG	+	–
Liver and lipid metabolism	–LDL, –TC, +liver fat, +liver biomarkers	3	TOMM40–APOE–GIPR, TM6SF2, PNPLA3	–	–

+/–: T2D risk alleles associated with increased or decreased phenotype values.

ASAT, abdominal subcutaneous adipose tissue volume; BMI, body mass index; BMR, basal metabolic rate; BP, blood pressure; FG, fasting glucose; FI, fasting insulin; GFAT, gluteofemoral adipose tissue volume; HbA1c, glycated haemoglobin; HDL, high-density lipoprotein cholesterol; LDL, low-density lipoprotein cholesterol; PI, proinsulin; TC, total cholesterol; TG, triglycerides; VAT, visceral adipose tissue volume; WHR, waist–hip ratio; 2hG, two-hour glucose.

*Adjusted for BMI.

circulating plasma lipids. T2D risk alleles at index SNVs in the obesity cluster are associated with increased body mass index (BMI), waist–hip ratio (WHR), body fat percentage and basal metabolic rate, and with decreased high-density lipoprotein (HDL) cholesterol. The lipodystrophy cluster comprises index SNVs for which T2D risk alleles are associated with increased fasting insulin, WHR, blood pressure and triglycerides, and with decreased body fat percentage, gluteofemoral adipose tissue (GFAT) volume and HDL cholesterol. T2D risk alleles at index SNVs assigned to the liver and lipid metabolism cluster are associated with increased liver fat and liver-related biomarkers, and with decreased low-density lipoprotein (LDL) cholesterol and total cholesterol.

By increasing the number of index SNVs in the clustering by nearly fourfold relative to previous efforts, we provide a more granular view of the biological processes through which T2D associations affect disease, and highlight three previously unreported clusters of signals with cardiometabolic profiles that are representative of metabolic syndrome, body fat and residual glycaemic effects. We observed significantly weaker allelic effects on T2D in these three clusters than in those previously reported (mean OR of 1.028 versus 1.033, $P = 2.2 \times 10^{-7}$), but there was no noticeable difference in disparity around the centroid between clusters (Extended Data Fig. 3, Supplementary Table 9 and Supplementary Fig. 5). T2D risk alleles at index SNVs assigned to the metabolic syndrome cluster are associated with increased fasting glucose, WHR, triglycerides and blood pressure, and with decreased HDL cholesterol, which together are used to define metabolic syndrome. T2D risk alleles in this cluster are also associated with increased fasting insulin, with accumulations of unhealthy fat depots (increased visceral adipose tissue (VAT) volume and liver fat) and with decreased GFAT volume. Previous investigations have shown that individuals with metabolic syndrome are at increased risk of T2D[22], although Mendelian randomization studies indicate that a causal effect is driven by increased waist circumference and increased fasting glucose[23]. T2D risk alleles at index SNVs assigned to the body fat cluster are associated with increased abdominal subcutaneous adipose tissue volume, VAT volume and body fat percentage. Although the body fat cluster profile of associations with cardiometabolic phenotypes shares these features in common with obesity-mediated insulin resistance, index SNVs in the body fat cluster are not strongly associated with BMI, lipid

levels or basal metabolic rate. Previous investigations have highlighted that body fat percentage is predictive of abnormal blood glucose in individuals with a healthy BMI[24]. Finally, T2D risk alleles at index SNVs assigned to the residual glycaemic cluster are most strongly associated with increased fasting glucose and glycated haemoglobin, but, unlike the two beta-cell-dysfunction clusters, are not associated with PI or decreased fasting insulin.

Clustering provides a framework to better understand the diverse physiological processes through which T2D develops and the shared biological pathways that drive genetic correlations with other insulin-resistance-related disorders, including gestational diabetes mellitus (GDM) and polycystic ovary syndrome (PCOS). T2D risk alleles at index SNVs showed a gradient of effects on insulin-related endophenotypes across clusters (Supplementary Text, Extended Data Fig. 4 and Supplementary Tables 10 and 11), representing a cline from insulin production and processing in the two beta-cell-dysfunction clusters through to insulin resistance that was most extreme in the lipodystrophy cluster. Index SNVs in the beta cell +PI cluster showed the strongest associations with GDM, whereas those in the obesity cluster were most strongly associated with PCOS (Supplementary Text, Extended Data Fig. 5 and Supplementary Table 12).

Regulatory processes underlying clusters

To gain insight into tissue-specific regulatory processes underpinning mechanistic clusters, we integrated T2D association signals with assay for transposase-accessible chromatin using sequencing (ATAC-seq) peaks from single-cell atlases of chromatin accessibility (CATLAS and DESCARTES) for 222 cell types derived from 30 human adult and 15 human fetal tissues[25,26] and an additional 106 cell types from the human brain[27] (Fig. 2, Supplementary Tables 13 and 14 and Methods).

We observed significant enrichment for regions of open chromatin in fetal islets and adult neuroendocrine cells in pancreatic islets (alpha, beta, gamma and delta) in the beta cell +PI, beta cell –PI and residual glycaemic clusters. In addition, the residual glycaemic cluster was enriched in fetal and adult pancreatic ductal cells, whereas the beta cell –PI cluster was enriched in adult enterochromaffin cells—a type of enteroendocrine cell that has an essential role in regulating intestinal motility and secretion in the gastrointestinal tract[28]. Enterochromaffin

Article
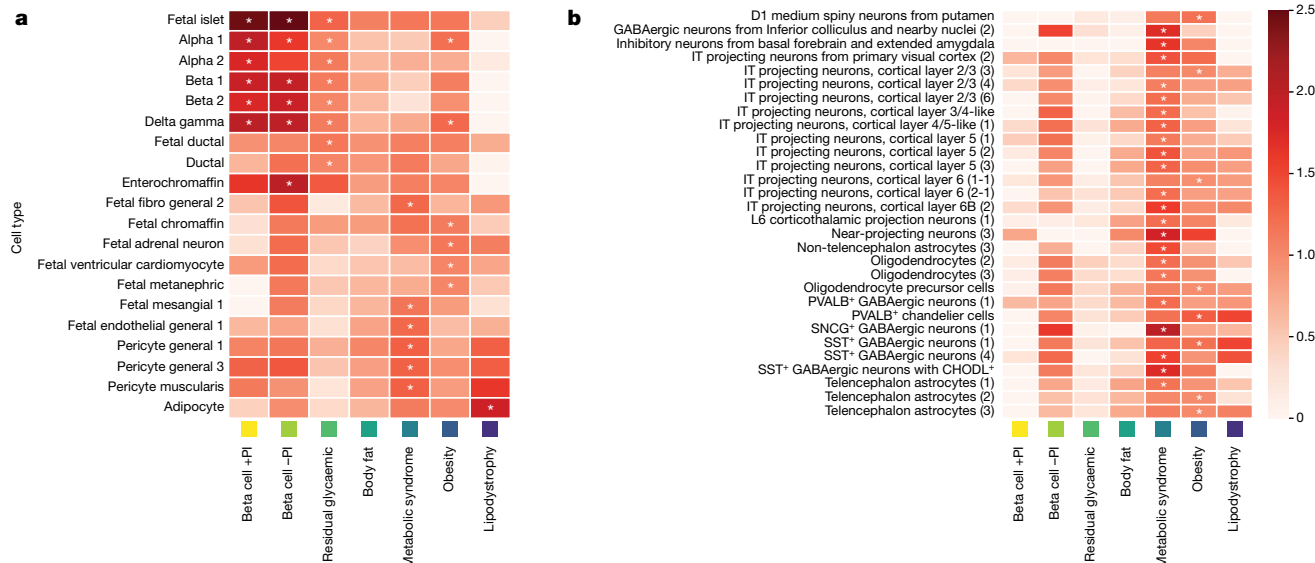
Fig. 2 | Heat map of cluster-specific enrichments of T2D associations for cell-type-specific regions of open chromatin derived from single-cell ATAC-seq peaks in adult and fetal tissue. a, Cell types (222 types) from 30 human adult tissues and 15 human fetal tissues. **b**, Cell types (106 types) from the human brain. In each panel, columns represent mechanistic clusters. Each row represents a cell type that was significantly enriched (Bonferroni correction for the number of cell types) for T2D associations in at least one cluster (indicated by an asterisk). The 'temperature' of each cell defines the magnitude of the log fold enrichment. The liver and lipid metabolism cluster is not presented because it includes only three T2D association signals and the model parameter estimates were unstable.

cells are a major target for glucagon-like peptide 1 (GLP-1) and highly express the GLP-1 receptor, agonists of which are widely used as medications for T2D[29] (Supplementary Text).

The obesity cluster was also significantly enriched for regions of open chromatin in adult pancreatic islets, although not as strongly as were the beta-cell-dysfunction clusters. Enrichment was observed only for alpha, gamma and delta cells, suggesting that there are alternative pathways through which islets affect the development of T2D, other than through the secretion of insulin from beta cells. The obesity cluster was further enriched in fetal adrenal gland cells (chromaffin cells and adrenal neurons), fetal heart cells (ventricular cardiomyocytes) and fetal kidney cells (metanephric cells). Previous studies have reported an enrichment of BMI loci or heritability for epigenomic annotations in pancreatic islets and adrenal gland[30,31], consistent with our findings. In the human brain, the obesity cluster was significantly enriched for regions of open chromatin in cell types including intratelencephalic (IT) projecting neurons, somatostatin-positive (SST+) GABAergic inhibitory neurons and D1 medium spiny neurons. SST+ GABAergic neurons exist in the hypothalamus and regulate food intake[32]. D1 medium spiny neurons are a type of GABAergic neuron in the human striatum that expresses D1-type dopamine receptors; these neurons have been implicated in food motivation and the development of diet-induced obesity in mice[33].

The remaining four clusters (lipodystrophy; metabolic syndrome; body fat; and liver and lipid metabolism) were not significantly enriched for regions of open chromatin in pancreatic islets. The lipodystrophy cluster was enriched only in adult adipocytes, which confirms previous reports in bulk adipose tissue[4,20]. Consistent with these results, association signals for WHR, triglycerides and HDL cholesterol, which are strongly affected by index SNVs in the lipodystrophy cluster, have been shown to be enriched in candidate *cis*-regulatory elements in adipocytes[26]. The metabolic syndrome cluster was enriched in cells that reside in the walls of blood vessels (adult pericytes and fetal endothelial cells), fetal kidney cells (mesangial cells) and fetal fibroblasts. Association signals for systolic and diastolic blood pressure, a key component of metabolic syndrome, have been shown to be enriched in candidate *cis*-regulatory elements in these cell types[26]. Endothelial dysfunction is not only a consequence of insulin resistance, but also impairs insulin

signalling to further reduce insulin sensitivity, thereby providing a pathophysiological mechanism that links the metabolic and cardiovascular components of metabolic syndrome[34]. In human brain, the metabolic syndrome cluster was significantly enriched for regions of open chromatin in cell types including IT projecting neurons and SST+ GABAergic inhibitory neurons. IT projecting neurons are a type of glutamatergic excitatory pyramidal neuron in the cerebral cortex, and metabolic syndrome was previously associated with pyramidal neurons and GABAergic neurons in cell-type specificity analyses in a GWAS that examined genetic factors in metabolic syndrome[35]. We observed no significant enrichments in the body fat cluster or in the liver and lipid metabolism cluster.

Ancestry-correlated heterogeneity

Previous multi-ancestry GWASs have shown widespread heterogeneity in allelic effects at T2D association signals across ancestry groups[9,36]. We took advantage of the meta-regression model to partition heterogeneity into an ancestry-correlated component explained by three axes of genetic variation, and a residual component reflecting differences in environmental exposures (that are not correlated with ancestry) and/or study design (Supplementary Table 15). We observed 127 (9.9%) independent T2D association signals with significant evidence for ancestry-correlated heterogeneity ($P_{HET} < 3.9 \times 10^{-5}$, Bonferroni correction for 1,289 signals). We would expect less than one signal to meet this threshold of significance, highlighting that ancestry-correlated heterogeneity is strongly enriched at T2D associations (one-sided binomial test $P < 2.2 \times 10^{-16}$). By contrast, we observed significant evidence of residual heterogeneity at only four (0.3%) association signals (one-sided binomial test $P = 0.031$). These results therefore suggest that differences in allelic effects at index SNVs are more strongly correlated with genetic ancestry than other factors that vary between GWASs.

We next sought to better understand the impact of genetic diversity on differences in allelic effects between GWASs at the 127 association signals with significant evidence of ancestry-correlated heterogeneity (Methods). For 118 (92.9%) signals, allelic effect sizes were most strongly associated with the first two axes of genetic variation, which
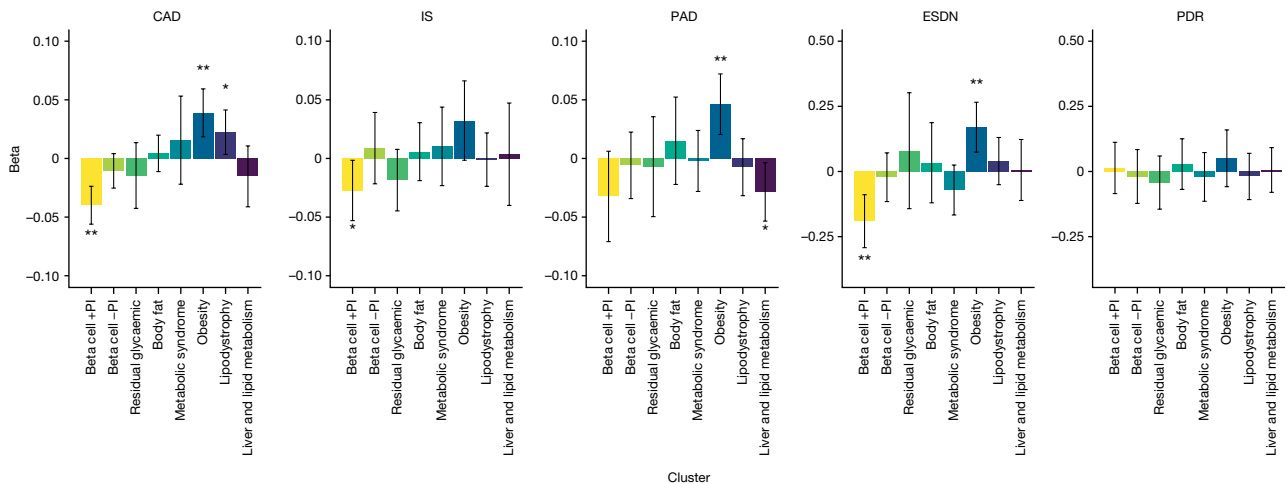
Fig. 3 | Associations of cluster-specific components of the partitioned PS with five T2D-related vascular outcomes in up to 279,552 individuals from multiple ancestry groups. Summaries of the associations of each cluster-specific component of the partitioned PS with CAD, ischaemic stroke (IS), peripheral artery disease (PAD), ESDN and proliferative diabetic retinopathy (PDR). The height of each bar corresponds to the log-odds ratio (beta) per standard deviation of the PS, and the grey bar shows the 95% confidence interval. Analyses of T2D-related macrovascular complications (CAD, PAD and IS) were undertaken in all individuals, with adjustment for T2D status. Analyses of microvascular complications were undertaken in individuals with T2D only. *$P < 0.05$, nominal association; **$P < 0.0063$, Bonferroni correction for eight clusters. Exact P values are provided in Supplementary Table 21.

reflect differences between AFA/EUR and EAS GWASs (AFA–EAS axis), and between AFA/EAS and EUR GWASs (AFA–EUR axis), respectively (Supplementary Text, Extended Data Figs. 1 and 6 and Supplementary Table 16).

We observed significant differences in mean z-scores for association between clusters for both the AFA–EAS axis ($P = 4.1 \times 10^{-6}$) and the AFA–EUR axis ($P = 1.5 \times 10^{-6}$). Index SNVs in the two beta-cell-dysfunction clusters were most positively associated with the AFR–EAS axis, indicating allelic effects on T2D that were greater in EAS GWASs than in AFA and EUR GWASs (Extended Data Fig. 7 and Supplementary Table 17). By contrast, index SNVs in the lipodystrophy and obesity clusters were most positively associated with the AFA–EUR axis, indicating allelic effects on T2D that were greater in EUR GWASs than in EAS and AFA GWASs. These results indicate that ancestry-correlated heterogeneity varies between mechanistic clusters, with allelic effects greatest for EAS GWASs at association signals assigned to clusters acting through beta-cell dysfunction and greatest for EUR GWASs at those assigned to clusters operating through insulin resistance.

Ancestry-correlated heterogeneity in allelic effects between GWASs is not driven by differences in allele frequency between ancestry groups, but can occur because of interaction between index SNVs and environmental and lifestyle factors, if not accounted for in the association analysis[37]. We observed substantial variation in the distribution of study-level mean BMI in T2D cases and controls across ancestry groups (Supplementary Fig. 6). Such variation could affect ancestry-correlated heterogeneity because, when cases and controls are selected from the extremes of the BMI distribution, the magnitude of allelic effect estimates at T2D signals acting through beta-cell dysfunction can be inflated[38]. We therefore extended the MR-MEGA meta-regression model to allow for allelic effect heterogeneity at index SNVs due to mean BMI in T2D cases and controls, in addition to axes of genetic variation (Methods).

After adjustment for study-level mean BMI in cases of T2D and in controls, only 24 association signals retained significant evidence of ancestry-correlated heterogeneity ($P < 3.9 \times 10^{-5}$), compared with 127 signals without adjustment (Supplementary Text and Supplementary Table 18). After adjustment for BMI, significant differences in mean z-scores for association between clusters for the AFA–EUR axis were maintained ($P = 3.2 \times 10^{-5}$ versus $P = 1.5 \times 10^{-6}$ without

adjustment), whereas those for the AFA–EAS axis were not ($P = 0.18$ versus $P = 4.1 \times 10^{-6}$ without adjustment). Furthermore, after adjustment for BMI, the two beta-cell-dysfunction clusters were no longer strongly positively associated with the AFA–EAS axis (Extended Data Fig. 7 and Supplementary Table 19). Together, these results suggest that heterogeneity in allelic effects between EAS GWASs and EUR/AFA GWASs, which occur most often at association signals assigned to the beta-cell-dysfunction clusters, can be mostly accounted for by differences in the distributions of mean BMI in T2D cases and in controls between these ancestry groups.

Associations of partitioned PS with outcomes

The major complications in individuals with T2D are macrovascular outcomes including coronary artery disease (CAD), ischaemic stroke and peripheral artery disease, and microvascular outcomes, including end-stage diabetic nephropathy (ESDN) and proliferative diabetic retinopathy. We tested for association of a cluster-specific partitioned PS with these vascular outcomes in up to 279,552 individuals (including 30,288 cases of T2D) across five ancestry groups (AFA, EAS, EUR, HIS and SAS) from the All of Us Research Program, Biobank Japan and the Genes & Health study (Methods). These individuals were not included in the multi-ancestry meta-analysis, thus avoiding potential inflated type I error rates owing to overlap between the discovery and the testing datasets. To maximize sample size, we tested macrovascular outcomes in all individuals, adjusted for T2D status, and microvascular complications only in individuals with T2D (Methods and Supplementary Table 20). To assess the additional information afforded by the partitioned PS over an overall T2D PS, agnostic to cluster membership, we tested for association of each cluster-specific component of the partitioned PS after adjustment for the overall PS. Figure 3 provides an overview of the associations of each cluster-specific component of the partitioned PS with the five vascular outcomes across ancestry groups.

We observed a significant association ($P < 0.0063$, Bonferroni correction for eight clusters) of two components of the partitioned PS with CAD: a negative association with the beta cell +PI cluster (OR = 0.96 per standard deviation of the PS, $P = 1.3 \times 10^{-6}$) and a positive association with the obesity cluster (OR = 1.04, $P = 0.00019$). There was no evidence of heterogeneity in the effects of these two clusters on CAD

Article

across ancestry groups (Supplementary Fig. 7 and Supplementary Table 21). Notably, after adjustment for a CAD PS derived from a previously published multi-ancestry meta-analysis of CAD GWASs[39], the positive CAD association with both components of the partitioned PS remained significant (Extended Data Fig. 8 and Supplementary Table 22): beta cell +PI cluster (OR = 0.96, $P = 4.4 \times 10^{-5}$) and obesity cluster (OR = 1.04, $P = 0.00065$). We also observed a significant positive association of the obesity cluster from the partitioned PS with peripheral artery disease (OR = 1.05, $P = 0.00045$), with no evidence of heterogeneity in effects across ancestry groups (Supplementary Fig. 8 and Supplementary Table 21). Across all three macrovascular outcomes, there was a general trend of negative association with the beta cell +PI cluster and positive association with the obesity cluster, although no cluster-specific components of the partitioned PS attained significance for ischaemic stroke (Supplementary Fig. 9 and Supplementary Table 21). There was no strong association of the overall T2D PS with CAD ($P = 0.17$), ischaemic stroke ($P = 0.022$) or peripheral artery disease ($P = 0.77$) after meta-analysis across ancestry groups. Together, these results highlight the advantages of the partitioned PS over an overall T2D PS for detecting associations with macrovascular outcomes, and provide insight into the biological processes that lead to their development.

We observed significant associations of two components of the partitioned PS with ESDN: a negative association with the beta cell +PI cluster (OR = 0.83, $P = 0.00024$) and a positive association with the obesity cluster (OR = 1.19, $P = 0.00050$). There was no evidence of heterogeneity in the effects of these two clusters across ancestry groups, (Supplementary Fig. 10 and Supplementary Table 21), and the overall PS was not strongly associated with ESDN ($P = 0.048$). By contrast, none of the cluster-specific components of the partitioned PS were associated with proliferative diabetic retinopathy. However, there was a strong positive association of the overall PS with this microvascular outcome (OR = 1.32, $P = 1.1 \times 10^{-9}$), with no evidence of heterogeneity in effects across ancestry groups (Supplementary Fig. 11 and Supplementary Table 21). Together, these results suggest that ESDN is associated with obesity and beta-cell dysfunction with opposite directions of effect, and confirm previous reports that proliferative diabetic retinopathy is driven by hyperglycaemia[40] and therefore strongly associated with the overall burden of T2D risk variants.

Finally, we tested for associations of the cluster-specific components of the partitioned PS and the overall T2D PS with age of onset of T2D (Extended Data Fig. 9 and Methods). The overall PS was strongly associated with an earlier age of onset (1.15 years per standard deviation of the PS, $P = 5.1 \times 10^{-8}$), although the effects were highly heterogeneous across ancestry groups (Supplementary Fig. 12 and Supplementary Table 23). However, even after adjustment for the overall PS, the obesity cluster was significantly associated with an earlier age of onset (0.38 years, $P = 1.4 \times 10^{-7}$), with no evidence of heterogeneity across ancestry groups. These findings highlight the importance of obesity-related processes for the onset of T2D, in addition to the development of vascular complications.

Associations with vascular outcomes in clinical trials

To gain insight into the associations of the obesity and beta cell +PI clusters with a broader range of vascular outcomes, we assessed the performance of the partitioned PS (after adjustment for the overall PS) in prospective GWASs in up to 29,827 EUR individuals with T2D from six clinical trials from the Thrombolysis in Myocardial Infarction (TIMI) Study Group (Methods and Supplementary Table 24). We observed the strongest associations of cluster-specific components of the partitioned PS with risk of hospitalization for heart failure: positive with the obesity cluster (hazard ratio (HR) = 1.15 per standard deviation of the PS, $P = 4.8 \times 10^{-6}$) and negative with the beta cell +PI cluster (HR = 0.90, $P = 0.00092$). Amongst macrovascular outcomes, the

beta cell +PI cluster was also negatively associated with cardiovascular death (HR = 0.90, $P = 0.0020$), major cardiovascular events (HR = 0.94, $P = 0.0050$) and myocardial infarction (HR = 0.94, $P = 0.027$). For microvascular outcomes, the two clusters showed associations with opposite directions of effect for albuminuria: obesity cluster (HR = 1.06, $P = 0.012$) and beta cell +PI cluster (HR = 0.95, $P = 0.047$). Across all outcomes, there was a general trend of positive association with the obesity cluster and negative association with the beta cell +PI cluster (Extended Data Fig. 10), consistent with the associations observed from our analyses of retrospective GWASs across ancestry groups.

Discussion

To better understand the aetiological heterogeneity of T2D across diverse populations, we assembled a large collection of T2D GWASs for six ancestry groups through the Type 2 Diabetes Global Genomics Initiative. By increasing the effective sample size by almost threefold compared with previous efforts, we identified a total of 611 loci attaining the conventional threshold of genome-wide significance ($P < 5 \times 10^{-8}$), 145 (23.7%) of which have not to our knowledge been previously reported. This conventional threshold is equivalent to a Bonferroni correction for the effective number of independent SNVs in EUR reference data[41]. Using empirical data from the 1000 Genomes Project, the DIAMANTE Consortium and others have advocated more stringent thresholds for multi-ancestry meta-analysis because the structure of LD is broken down across ancestry groups and the effective number of independent SNVs is increased[9,42]. In fact, our analyses suggest that loci meeting conventional genome-wide significance are unlikely to be false positive association signals, but instead are driven by index SNVs that have modest effects that require larger sample sizes to meet more stringent thresholds. We therefore recommend the use of this conventional threshold but advocate careful review of reported signals to ensure that associations are not driven by single studies or poorly imputed variants to protect against false positives.

Multi-ancestry meta-regression maximizes power to detect associations that are shared across ancestry groups by allowing for heterogeneity in allelic effects at index SNVs. MR-MEGA is not restricted to broad continental ancestry labels that can be used to reinforce the concept of fundamental genetic differences between groups[43], but instead represents ancestry as continuous axes of genetic variation, which better reflect the continuum of human genetic diversity and demographic history[44]. Still, it is important to emphasize that our meta-analysis does not fully capture global genetic diversity, in particular underrepresented populations across Africa, South and Central America, the Middle East and Oceania. For example, 98.2% of the total effective sample size of individuals with the highest proportion of ancestry from Africa are African Americans. The ancestry of these individuals represents a cline of admixture that is predominantly from West Africa and is therefore not representative of other regions in Africa, where the level of genetic variation is equivalent to the differences observed between other continental groups[43]. Bolstering GWAS collections in these underrepresented populations remains an urgent priority for the human genetics research community and highlights the need for careful interpretation of results that does not generalize findings across ancestry groups that are sensitive to biased representation.

Within the landscape of the genetic architecture of T2D, we identified eight clusters of index SNVs with distinct profiles of associations with 37 cardiometabolic phenotypes, which defined pathophysiology-relevant groupings. The addition of previously unreported T2D signals identified through the multi-ancestry meta-analysis helped define three clusters that were not detected in previous clustering efforts[3,4,20,21], with cardiometabolic profiles that are consistent with residual glycaemic effects, accumulations of body fat and metabolic syndrome. These previous efforts have implemented 'soft clustering' approaches, such as Bayesian non-negative matrix factorization, that generate weights

for cluster membership for each index SNV[4]. The assignment of index SNVs to clusters is then determined given a threshold weight for cluster membership, allowing for the possibility that a T2D association signal affects disease through multiple pathophysiological pathways. However, depending on the threshold for cluster membership, some index SNVs will be unassigned. Bayesian non-negative matrix factorization also considers positive and negative associations with the same phenotype as independent variables, and most clustering methods cannot directly accommodate missing phenotype associations. To address these potential limitations, we implemented methodology that jointly conducts k-means clustering of index SNVs with powerful iterative multiple imputation of missing phenotype associations. In this 'hard clustering' approach, each index SNV is assigned to exactly one cluster. This has the potential disadvantage, therefore, that index SNVs with outlying or intermediate profiles of trait associations are 'forced' into a cluster that does not fit well. However, the previously unreported clusters that we identified in our hard clustering were not noticeably more disparate than the clusters reported previously, suggesting that we have not introduced substantial noise by forcing all SNVs into exactly one cluster. Ultimately, the choice of clustering approach may depend on the objectives of any downstream investigations.

Our analyses highlighted a significant excess of T2D association signals with ancestry-correlated heterogeneity, which is driven mainly by differences in allelic effects between AFA, EAS and EUR GWASs. The two beta-cell-dysfunction clusters are most strongly associated with the AFA–EAS axis, in which effects are typically larger in EAS GWASs than in those for other ancestry groups. These two clusters are also most strongly associated with reduced insulin secretion and lower insulin resistance. By contrast, the lipodystrophy and obesity clusters, which are characterized by reduced insulin sensitivity and higher insulin resistance, are most strongly associated with the AFA–EUR axis, in which effects are typically larger in EUR than in other ancestry groups. These observations are consistent with studies reporting differences in the pathogenesis of T2D between ancestry groups, whereby T2D is initiated mainly through increased insulin resistance in EUR individuals, but is characterized by reduced insulin secretion with lower insulin resistance in EAS individuals[45,46]. We have shown that most signals with ancestry-correlated heterogeneity can be explained by differences in the distribution of BMI in T2D cases and controls between ancestry groups. Furthermore, after adjustment for study-level mean BMI, we observe no difference in allelic effects between clusters along the AFA–EAS axis. This is consistent with previous studies that reported that body composition is the main determinant of variation in T2D pathogenesis between EAS and EUR individuals, because insulin sensitivity and beta-cell response are similar in the two ancestry groups after accounting for differences in BMI[45,47].

We reveal—across multiple ancestry groups—significant associations of vascular outcomes with cluster-specific components of the partitioned PS after adjustment for the overall PS, which suggests that disease trajectories are associated with genetic burden in certain biological pathways that are consistent across diverse populations. Although the effect sizes of the cluster-specific components of the partitioned PS were small, they motivate future work to strengthen these effects through the identification of further T2D associations in larger sample sizes. Through integration with single-cell chromatin accessibility data across diverse cell types, they also enhance understanding of key biological processes driving heterogeneity in the clinical features of T2D phenotypes. For example, the obesity-cluster-specific component of the PS was positively associated with CAD and ESDN, and included index SNVs that were enriched for regions of open chromatin in fetal ventricular cardiomyocytes, fetal adrenal neuron, adult chromaffin cells in the adrenal gland and fetal metanephric cells. These findings are in line with the reported enrichments of CAD association signals for transcriptomic and epigenomic annotations in bulk tissues including the aorta and arteries, the heart and the adrenal gland[39,48,49], and of

renal function association signals in kidney-tissue-specific regulatory annotations[50]. Together, these findings provide a clear link to shared biological mechanisms that drive the development of T2D and other vascular diseases.

In conclusion, our findings show the value of integrating multi-ancestry GWASs of T2D and cardiometabolic traits with single-cell epigenomics across diverse tissues to disentangle the aetiological heterogeneity driving the development and progression of T2D across population groups. Improved understanding of the varied pathophysiological processes that link T2D to vascular outcomes could offer a route to genetically informed diabetes care and global opportunities for the clinical translation of findings from T2D GWASs.

Online content

Any methods, additional references, Nature Portfolio 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/s41586-024-07019-6.

1. McCarthy, M. I. Painting a new picture of personalised medicine for diabetes. *Diabetologia* **60**, 793–799 (2017).
2. Pearson, E. R. Type 2 diabetes: a multifaceted disease. *Diabetologia* **62**, 1107–1112 (2019).
3. 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).
4. Udler, M. S. et al. Type 2 diabetes genetic loci informed by multi-trait associations point to disease mechanisms and subtypes: a soft clustering analysis. *PLoS Med.* **15**, e1002654 (2018).
5. Udler, M. S., McCarthy, M. I., Florez, J. C. & Mahajan, A. Genetic risk scores for diabetes diagnosis and precision medicine. *Endocr. Rev.* **40**, 1500–1520 (2019).
6. International Diabetes Federation. *IDF Diabetes Atlas* 10th edn https://diabetesatlas.org/ (IDF, 2021).
7. Almgren, P. et al. Heritability and familiality of type 2 diabetes and related quantitative traits in the Botnia Study. *Diabetologia* **54**, 2811–2819 (2011).
8. 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).
9. Mahajan, A. et al. Multi-ancestry genetic study of type 2 diabetes highlights the power of diverse populations for discovery and translation. *Nat. Genet.* **54**, 560–572 (2022).
10. 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).
11. Warrington, N. M. et al. Maternal and fetal genetic effects on birth weight and their relevance to cardio-metabolic risk factors. *Nat. Genet.* **51**, 804–814 (2019).
12. Chen, J. et al. The trans-ancestral genomic architecture of glycemic traits. *Nat. Genet.* **53**, 840–860 (2021).
13. Liu, Y. et al. Genetic architecture of 11 organ traits derived from abdominal MRI using deep learning. *eLife* **10**, e65554 (2021).
14. Sulc, J. et al. Composite trait Mendelian randomization reveals distinct metabolic and lifestyle consequences of differences in body shape. *Commun. Biol.* **4**, 1064 (2021).
15. Sakaue, S. et al. A cross-population atlas of genetic associations for 220 human phenotypes. *Nat. Genet.* **53**, 1415–1424 (2021).
16. Graham, S. E. et al. The power of genetic diversity in genome-wide association studies of lipids. *Nature* **600**, 675–679 (2021).
17. Agrawal, S. et al. Inherited basis of visceral, abdominal subcutaneous and gluteofemoral fat depots. *Nat. Commun.* **13**, 3771 (2022).
18. Warren, H. et al. Genome-wide analysis in over 1 million individuals reveals over 2,000 independent genetic signals for blood pressure. Preprint at https://doi.org/10.21203/rs.3.rs-1409164/v1 (2022).
19. Broadaway, K. A. et al. Loci for insulin processing and secretion provide insight into type 2 diabetes risk. *Am. J. Hum. Genet.* **110**, 284–299 (2023).
20. Kim, H. et al. High-throughput genetic clustering of type 2 diabetes loci reveals heterogeneous mechanistic pathways of metabolic disease. *Diabetologia* **66**, 495–507 (2023).
21. Smith, K. et al. Multi-ancestry polygenic mechanisms of type 2 diabetes elucidate disease processes and clinical heterogeneity. *Nat. Med.* (in press).
22. Ford, E. S., Li, C. & Sattar, N. Metabolic syndrome and incident diabetes: current state of the evidence. *Diabetes Care* **31**, 1898–1904 (2008).
23. Marott, S. C., Nordestgaard, B. G., Tybjærg-Hansen, A. & Benn, M. Components of the metabolic syndrome and risk of type 2 diabetes. *J. Clin. Endocrinol. Metab.* **101**, 3212–3221 (2016).
24. Jo, A. & Mainous, A. G. 3rd Informational value of percent body fat with body mass index for the risk of abnormal blood glucose: a nationally representative cross-sectional study. *BMJ Open* **8**, e019200 (2018).
25. Domcke, S. et al. A human cell atlas of fetal chromatin accessibility. *Science* **370**, eaba7612 (2020).
26. Zhang, K. et al. A single-cell atlas of chromatin accessibility in the human genome. *Cell* **184**, 5985–6001 (2021).

Article

27. Li, Y. E. et al. A comparative atlas of single-cell chromatin accessibility in the human brain. *Science* **382**, eadf7044 (2023).

28. Bertrand, P. P. & Bertrand, R. L. Serotonin release and uptake in the gastrointestinal tract. *Auton. Neurosci.* **153**, 47–57 (2010).

29. Lund, M. L. et al. Enterochromaffin 5-HT cells—a major target for GLP-1 and gut microbial metabolites. *Mol. Metab.* **11**, 70–83 (2018).

30. Finucane, H. K. et al. Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nat. Genet.* **47**, 1228–1235 (2015).

31. Akiyama, M. et al. Genome-wide association study identifies 112 new loci for body mass index in the Japanese population. *Nat. Genet.* **49**, 1458–1467 (2017).

32. Suyama, S. & Yada, T. New insight into GABAergic neurons in the hypothalamic feeding regulation. *J. Physiol. Sci.* **68**, 717–722 (2018).

33. Matikainen-Ankney, B. A. et al. Nucleus accumbens D1 receptor-expressing spiny projection neurons control food motivation and obesity. *Biol. Psychiatry* **93**, 512–523 (2023).

34. Kim, J. A., Montagnani, M., Koh, K. K. & Quon, M. J. Reciprocal relationships between insulin resistance and endothelial dysfunction: molecular and pathophysiological mechanisms. *Circulation* **113**, 1888–1904 (2006).

35. van Walree, E. S. et al. Disentangling genetic risks for metabolic syndrome. *Diabetes* **71**, 2447–2457 (2022).

36. Wojcik, G. L. et al. Genetic analyses of diverse populations improves discovery for complex traits. *Nature* **570**, 514–518 (2019).

37. Petersen, R. E. et al. Genome-wide association studies in ancestrally diverse populations: opportunities, methods, pitfalls, and recommendations. *Cell* **179**, 589–603 (2019).

38. Perry, J. R. B. et al. Stratifying type 2 diabetes cases by BMI identifies genetic risk variants in *LAMA1* and enrichment for risk variants in lean compared with obese cases. *PLoS Genet.* **8**, e1002741 (2012).

39. Aragam, K. G. et al. Discovery and systematic characterization of risk variants and genes for coronary artery disease in over a million participants. *Nat. Genet.* **54**, 1803–1815 (2022).

40. UK Prospective Diabetes Study (UKPDS) Group. Intensive blood-glucose control with sulphonylureas or insulin compared with conventional treatment and risk of complications in patients with type 2 diabetes (UKPDS 33). *Lancet* **352**, 837–853 (1998).

41. Pe'er, I., Yelensky, R., Altshuler, D. & Daly, M. J. Estimation of the multiple testing burden for genomewide association studies of nearly all common variants. *Genet. Epidemiol.* **32**, 381–385 (2008).

42. Kanai, M., Tanaka, T. & Okada, Y. Empirical estimation of genome-wide significance thresholds based on the 1000 Genomes Project data set. *J. Hum. Genet.* **61**, 861–866 (2016).

43. Carlson, J., Henn, B. M., Al-Hindi, R. R. & Ramachandran, S. Counter the weaponization of genetics research by extremists. *Nature* **610**, 444–447 (2022).

44. Lewis, A. C. F. et al. Getting genetic ancestry right for science and society. *Science* **376**, 250–252 (2022).

45. Moller, J. B. et al. Ethnic differences in insulin sensitivity, β-cell function, and hepatic extraction between Japanese and Caucasians: a minimal model analysis. *J. Clin. Endocrinol. Metab.* **99**, 4273–4280 (2014).

46. Yabe, D., Seino, Y., Fukushima, M. & Seino, S. β cell dysfunction versus insulin resistance in the pathogenesis of type 2 diabetes in East Asians. *Curr. Diab. Rep.* **15**, 602 (2015).

47. Moller, J. B. et al. Body composition is the main determinant for the difference in type 2 diabetes pathophysiology between Japanese and Caucasians. *Diabetes Care* **37**, 796–804 (2014).

48. Matsunaga, H. et al. Transethnic meta-analysis of genome-wide association studies identifies three new loci and characterizes population-specific differences for coronary artery disease. *Circ. Genom. Precis. Med.* **13**, e002670 (2020).

49. Koyama, S. et al. Population-specific and trans-ancestry genome-wide analyses identify distinct and shared genetic risk loci for coronary artery disease. *Nat. Genet.* **52**, 1169–1177 (2020).

50. Morris, A. P. et al. Trans-ethnic kidney function association study reveals putative causal genes and effects on kidney-specific disease aetiologies. *Nat. Commun.* **10**, 29 (2019).

Ken Suzuki[1,2,3,302], Konstantinos Hatzikotoulas[4,302 ✉], Lorraine Southam[4,302], Henry J. Taylor[5,6,7,302], Xianyong Yin[8,9,302], Kim M. Lorenz[10,11,12,302], Ravi Mandla[13,14,302], Alicia Huerta-Chagoya[13], Giorgio E. M. Melloni[15], Stavroula Kanoni[16], Nigel W. Rayner[4], Ozvan Bocher[4], Ana Luiza Arruda[4,17,18], Kyuto Sonehara[3,19,20,21], Shinichi Namba[3], Simon S. K. Lee[22], Michael H. Preuss[22], Lauren E. Petty[23], Philip Schroeder[13,14], Brett Vanderwerff[8], Mart Kals[24], Fiona Bragg[25,26], Kuang Lin[25], Xiuqing Guo[27], Weihua Zhang[28,29], Jie Yao[27], Young Jin Kim[30], Mariaelisa Graff[31], Fumihiko Takeuchi[32], Jana Nano[33], Amel Lamri[34,35], Masahiro Nakatochi[36], Sanghoon Moon[30], Robert A. Scott[37], James P. Cook[38], Jung-Jin Lee[39], Ian Pan[40], Daniel Taliun[8], Esteban J. Parra[41], Jin-Fang Chai[42], Lawrence F. Bielak[43], Yasuharu Tabara[44], Yang Hai[27], Gudmar Thorleifsson[45], Niels Grarup[46], Tamar Sofer[47,48,49], Matthias Wuttke[50], Chloé Sarnowski[51], Christian Gieger[33,52,53],

Darryl Nousome[54], Stella Trompet[55,56], Soo-Heon Kwak[57], Jirong Long[58], Meng Sun[59], Lin Tong[60], Wei-Min Chen[61], Suraj S. Nongmaithem[62], Raymond Noordam[56], Victor J. Y. Lim[42], Claudia H. T. Tam[63,64], Yoonjung Yoonie Joo[65,66], Chien-Hsiun Chen[67], Laura M. Raffield[68], Bram Peter Prins[69], Aude Nicolas[70], Lisa R. Yanek[71], Guanjie Chen[72], Jennifer A. Brody[73], Edmond Kabagambe[58,74], Ping An[75], Anny H. Xiang[76], Hyeok Sun Choi[77], Brian E. Cade[48,78], Jingyi Tan[27], K. Alaine Broadaway[68], Alice Williamson[37,79], Zoha Kamali[80,81], Jinrui Cui[82], Manonanthini Thangam[83], Linda S. Adair[84], Adebowale Adeyemo[72], Carlos A. Aguilar-Salinas[85], Tarunveer S. Ahluwalia[86,87], Sonia S. Anand[34,35,88], Alain Bertoni[89], Jette Bork-Jensen[46], Ivan Brandslund[90,91], Thomas A. Buchanan[92], Charles F. Burant[93], Adam S. Butterworth[6,7,94,95,96], Mickaël Canouil[97,98], Juliana C. N. Chan[63,64,99,100], Li-Ching Chang[67], Miao-Li Chee[101], Ji Chen[102,103], Shyh-Huei Chen[104], Yuan-Tsong Chen[67], Zhengming Chen[25,26], Lee-Ming Chuang[105,106], Mary Cushman[107], John Danesh[6,7,69,94,95,96], Swapan K. Das[108], H. Janaka de Silva[109], George Dedoussis[110], Latchezar Dimitrov[111], Ayo P. Doumatey[72], Shufa Du[84,112], Qing Duan[68], Kai-Uwe Eckardt[113,114], Leslie S. Emery[115], Daniel S. Evans[116], Michele K. Evans[117], Krista Fischer[24,118], James S. Floyd[73], Ian Ford[119], Oscar H. Franco[120], Timothy M. Frayling[121], Barry I. Freedman[122], Pauline Genter[123], Hertzel C. Gerstein[34,35,88], Vilmantas Giedraitis[124], Clicerio González-Villalpando[125], Maria Elena González-Villalpando[125], Penny Gordon-Larsen[84,112], Myron Gross[126], Lindsay A. Guare[127], Sophie Hackinger[69], Liisa Hakaste[128,129], Sohee Han[30], Andrew T. Hattersley[130], Christian Herder[52,131,132], Momoko Horikoshi[133], Annie-Green Howard[112,134], Willa Hsueh[135], Mengna Huang[40,136], Wei Huang[137], Yi-Jen Hung[138,139], Mi Yeong Hwang[140], Chii-Min Hwu[141,142], Sahoko Ichihara[143], Mohammad Arfan Ikram[120], Martin Ingelsson[124], Md. Tariqul Islam[144], Masato Isono[32], Hye-Mi Jang[140], Farzana Jasmine[60], Guozhi Jiang[63,64], Jost B. Jonas[145], Torben Jørgensen[146,147,148], Frederick K. Kamanu[15], Fouad R. Kandeel[149], Anuradhani Kasturiratne[150], Tomohiro Katsuya[151,152], Varinderpal Kaur[14], Takahisa Kawaguchi[44], Jacob M. Keaton[5,58,111], Abel N. Kho[153,154], Chiea-Chuen Khor[155], Muhammad G. Kibriya[60], Duk-Hwan Kim[156], Florian Kronenberg[157], Johanna Kuusisto[158], Kristi Läll[24], Leslie A. Lange[159], Kyung Min Lee[160,161], Myung-Shik Lee[162,163], Nanette R. Lee[164], Aaron Leong[165,166], Liming Li[167,168], Yun Li[68], Ruifang Li-Gao[169], Symen Lighthart[120], Cecilia M. Lindgren[170,171,172], Allan Linneberg[146,173], Ching-Ti Liu[174], Jianjun Liu[155,175], Adam E. Locke[176,177,300], Tin Louie[115], Jian'an Luan[37], Andrea O. Luk[63,64], Xi Luo[178], Jun Lv[167,168], Julie A. Lynch[160,161], Valeriya Lyssenko[179,180], Shiro Maeda[133,181,182], Vasiliki Mamakou[133,181,182], Sohail Rafik Mansuri[62,184], Koichi Matsuda[185], Thomas Meitinger[186,187,188], Olle Melander[83], Andres Metspalu[24], Huan Mo[5], Andrew D. Morris[189], Filipe A. Moura[15], Jerry L. Nadler[190], Michael A. Nalls[70,191,192], Uma Nayak[61], Ioanna Ntalla[16], Yukinori Okada[3,19,20,21,193,194], Lorena Orozco[195], Sanjay R. Patel[196], Snehal Patil[9], Pei Pei[168], Mark A. Pereira[197], Annette Peters[33,52,198], Fraser J. Pirie[199], Hannah G. Polikowsky[23], Bianca Porneala[166], Gauri Prasad[200,201], Laura A. Rasmussen-Torvik[66], Alexander P. Reiner[202], Michael Roden[52,131,132], Rebecca Rohde[31], Katheryn Roll[27], Charumathi Sabanayagam[101,203,204], Kevin Sandow[27], Alagu Sankareswaran[62,184], Naveed Sattar[205], Sebastian Schönherr[157], Mohammad Shahriar[60], Botong Shen[117], Jinxiu Shi[137], Dong Mun Shin[140], Nobuhiro Shojima[2], Jennifer A. Smith[43,206], Wing Yee So[63,100], Alena Stančáková[158], Valgerdur Steinthorsdottir[45], Adrienne M. Stilp[115], Konstantin Strauch[207,208,209], Kent D. Taylor[27], Barbara Thorand[33,52], Unnur Thorsteinsdottir[45,210], Brian Tomlinson[63,211], Tam C. Tran[5], Fuu-Jen Tsai[212], Jaakko Tuomilehto[213,214,215,216], Teresa Tusie-Luna[217,218], Miriam S. Udler[13,14,165], Adan Valladares-Salgado[219], Rob M. van Dam[42,175], Jan B. van Klinken[220,221,222], Rohit Varma[223], Niels Wacher-Rodarte[224], Eleanor Wheeler[37], Ananda R. Wickremasinghe[150], Ko Willems van Dijk[220,221,225], Daniel R. Witte[226,227], Chittaranjan S. Yajnik[228], Ken Yamamoto[229], Kenichi Yamamoto[3,193,230], Kyungheon Yoon[140], Canqing Yu[167,168], Jian-Min Yuan[231,232], Salim Yusuf[34,35,88], Matthew Zawistowski[9], Liang Zhang[101], Wei Zheng[58], VA Million Veteran Program*, AMED GRIFIN Diabetes Initiative Japan*, Biobank Japan Project*, Penn Medicine BioBank*, Regeneron Genetics Center*, Genes & Health Research Team*, eMERGE Consortium*, International Consortium of Blood Pressure (ICBP)*, Meta-Analyses of Glucose and Insulin-Related Traits Consortium (MAGIC)*, Leslie J. Raffel[233], Michiya Igase[234], Eli Ipp[123], Susan Redline[48,78,235], Yoon Shin Cho[77], Lars Lind[236], Michael A. Province[75], Myriam Fornage[237], Craig L. Hanis[238], Erik Ingelsson[239,240], Alan B. Zonderman[117], Bruce M. Psaty[73,241,242], Ya-Xing Wang[243], Charles N. Rotimi[72], Diane M. Becker[71], Fumihiko Matsuda[44], Yongmei Liu[89,244], Mitsuhiro Yokota[245], Sharon L. R. Kardia[43], Patricia A. Peyser[43], James S. Pankow[197], James C. Engert[246,247], Amélie Bonnefond[97,98,248], Philippe Froguel[97,98,248], James G. Wilson[249], Wayne H. H. Sheu[139,142,250], Jer-Yuarn Wu[67], M. Geoffrey Hayes[251,252,253], Ronald C. W. Ma[63,64,99,100], Tien-Yin Wong[101,203,204], Dennis O. Mook-Kanamori[169], Tiinamaija Tuomi[83,128,129,254], Giriraj R. Chandak[62,255], Francis S. Collins[5], Dwaipayan Bharadwaj[256], Guillaume Paré[35,257], Michèle M. Sale[61,304], Habibul Ahsan[60], Ayesha A. Motala[199], Xiao-Ou Shu[58], Kyong-Soo Park[57,258], J. Wouter Jukema[55,259], Miguel Cruz[219], Yii-Der Ida Chen[27], Stephen S. Rich[260], Roberta McKean-Cowdin[54], Harald Grallert[33,52,261], Ching-Yu Cheng[101,203,204], Mohsen Ghanbari[120], E-Shyong Tai[42,175,262], Josee Dupuis[174,263], Norihiro Kato[32], Markku Laakso[158], Anna Köttgen[50], Woon-Puay Koh[264,265], Donald W. Bowden[111,266,267], Colin N. A. Palmer[268], Jaspal S. Kooner[29,269,270,271], Charles Kooperberg[202], Simin Liu[40,136,272], Kari E. North[31], Danish Saleheen[273,274,275], Torben Hansen[46], Oluf Pedersen[46], Nicholas J. Wareham[37], Juyoung Lee[140], Bong-Jo Kim[140], Iona Y. Millwood[25,26], Robin G. Walters[25,26], Kari Stefansson[45], Emma Ahlqvist[83], Mark O. Goodarzi[82], Karen L. Mohlke[68], Claudia Langenberg[37,276,277], Christopher A. Haiman[278], Ruth J. F. Loos[22,46,279], Jose C. Florez[13,14,165], Daniel J. Rader[12,280,281,282], Marylyn D. Ritchie[12,283,284], Sebastian Zöllner[9,285], Reedik Mägi[24], Nicholas A. Marston[15], Christian T. Ruff[15], David A. van Heel[286], Sarah Finer[287], Joshua C. Denny[5,288], Toshimasa Yamauchi[2,289], Takashi Kadowaki[2,289], John C. Chambers[28,29,269,290], Maggie C. Y. Ng[111,267,291], Xueling Sim[42], Jennifer E. Below[23], Philip S. Tsao[239,292,293], Kyong-Mi Chang[10,294], Mark I. McCarthy[170,295,296,301,303], James B. Meigs[13,165,166,303], Anubha Mahajan[170,295,301,303], Cassandra N. Spracklen[297,303], Josep M. Mercader[13,14,78,303], Michael Boehnke[9,303], Jerome I. Rotter[27,303], Marijana Vujkovic[10,294,298,303], Benjamin F. Voight[10,11,12,281,303 ✉], Andrew P. Morris[1,4,24,303 ✉] & Eleftheria Zeggini[4,299,303 ✉]

[1]Centre for Genetics and Genomics Versus Arthritis, Centre for Musculoskeletal Research, Division of Musculoskeletal and Dermatological Sciences, University of Manchester, Manchester, UK. [2]Department of Diabetes and Metabolic Diseases, Graduate School of

Medicine, University of Tokyo, Tokyo, Japan. [3]Department of Statistical Genetics, Osaka University Graduate School of Medicine, Suita, Japan. [4]Institute of Translational Genomics, Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg, Germany. [5]Center for Precision Health Research, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD, USA. [6]British Heart Foundation Cardiovascular Epidemiology Unit, Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK. [7]Heart and Lung Research Institute, University of Cambridge, Cambridge, UK. [8]Department of Epidemiology, School of Public Health, Nanjing Medical University, Nanjing, China. [9]Department of Biostatistics and Center for Statistical Genetics, University of Michigan, Ann Arbor, MI, USA. [10]Corporal Michael J. Crescenz VA Medical Center, Philadelphia, PA, USA. [11]Department of Systems Pharmacology and Translational Therapeutics, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA, USA. [12]Department of Genetics, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA, USA. [13]Programs in Metabolism and Medical and Population Genetics, Broad Institute of Harvard and MIT, Cambridge, MA, USA. [14]Diabetes Unit and Center for Genomic Medicine, Massachusetts General Hospital, Boston, MA, USA. [15]TIMI Study Group, Division of Cardiovascular Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA. [16]William Harvey Research Institute, Barts and the London School of Medicine and Dentistry, Queen Mary University of London, London, UK. [17]Graduate School of Experimental Medicine, Technical University of Munich, Munich, Germany. [18]Munich School for Data Science, Helmholtz Munich, Neuherberg, Germany. [19]Department of Genome Informatics, Graduate School of Medicine, University of Tokyo, Tokyo, Japan. [20]Integrated Frontier Research for Medical Science Division, Institute for Open and Transdisciplinary Research Initiatives, Osaka University, Suita, Japan. [21]Laboratory for Systems Genetics, RIKEN Center for Integrative Medical Sciences, Yokohama, Japan. [22]Charles Bronfman Institute for Personalized Medicine, Icahn School of Medicine at Mount Sinai, New York, NY, USA. [23]Department of Medicine, Vanderbilt University Medical Center, Nashville, TN, USA. [24]Estonian Genome Centre, Institute of Genomics, University of Tartu, Tartu, Estonia. [25]Nuffield Department of Population Health, University of Oxford, Oxford, UK. [26]Medical Research Council Population Health Research Unit, University of Oxford, Oxford, UK. [27]Institute for Translational Genomics and Population Sciences, Department of Pediatrics, Lundquist Institute for Biomedical Innovation at Harbor-UCLA Medical Center, Torrance, CA, USA. [28]Department of Epidemiology and Biostatistics, Imperial College London, London, UK. [29]Department of Cardiology, Ealing Hospital, London NorthWest Healthcare NHS Trust, London, UK. [30]Division of Genome Science, Department of Precision Medicine, National Institute of Health, Cheongju-si, South Korea. [31]Department of Epidemiology, Gillings School of Global Public Health, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA. [32]Department of Gene Diagnostics and Therapeutics, Research Institute, National Center for Global Health and Medicine, Tokyo, Japan. [33]Institute of Epidemiology, Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg, Germany. [34]Department of Medicine, McMaster University, Hamilton, Ontario, Canada. [35]Population Health Research Institute, Hamilton Health Sciences and McMaster University, Hamilton, Ontario, Canada. [36]Public Health Informatics Unit, Department of Integrated Health Sciences, Nagoya University Graduate School of Medicine, Nagoya, Japan. [37]MRC Epidemiology Unit, Institute of Metabolic Science, University of Cambridge School of Clinical Medicine, Cambridge, UK. [38]Department of Health Data Science, University of Liverpool, Liverpool, UK. [39]Division of Translational Medicine and Human Genetics, University of Pennsylvania, Philadelphia, PA, USA. [40]Department of Epidemiology, Brown University School of Public Health, Providence, RI, USA. [41]Department of Anthropology, University of Toronto at Mississauga, Mississauga, Ontario, Canada. [42]Saw Swee Hock School of Public Health, National University of Singapore and National University Health System, Singapore, Singapore. [43]Department of Epidemiology, School of Public Health, University of Michigan, Ann Arbor, MI, USA. [44]Center for Genomic Medicine, Kyoto University Graduate School of Medicine, Kyoto, Japan. [45]deCODE Genetics, Amgen, Reykjavik, Iceland. [46]Novo Nordisk Foundation Center for Basic Metabolic Research, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark. [47]Department of Biostatistics, Harvard University, Boston, MA, USA. [48]Division of Sleep and Circadian Disorders, Brigham and Women's Hospital, Boston, MA, USA. [49]Department of Medicine, Harvard University, Boston, MA, USA. [50]Institute of Genetic Epidemiology, Department of Data Driven Medicine, Faculty of Medicine and Medical Center, University of Freiburg, Freiburg, Germany. [51]Department of Epidemiology, Human Genetics and Environmental Sciences, University of Texas Health Science Center at Houston School of Public Health, Houston, TX, USA. [52]German Center for Diabetes Research (DZD), Neuherberg, Germany. [53]Research Unit of Molecular Epidemiology, Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg, Germany. [54]Department of Population and Public Health Sciences, Keck School of Medicine of USC, Los Angeles, CA, USA. [55]Department of Cardiology, Leiden University Medical Center, Leiden, The Netherlands. [56]Section of Gerontology and Geriatrics, Department of Internal Medicine, Leiden University Medical Center, Leiden, The Netherlands. [57]Department of Internal Medicine, Seoul National University Hospital, Seoul, South Korea. [58]Division of Epidemiology, Department of Medicine, Institute for Medicine and Public Health, Vanderbilt Genetics Institute, Vanderbilt University Medical Center, Nashville, TN, USA. [59]Nuffield Department of Surgical Sciences, University of Oxford, Oxford, UK. [60]Institute for Population and Precision Health (IPPH), Biological Sciences Division, University of Chicago, Chicago, IL, USA. [61]Department of Public Health Sciences and Center for Public Health Genomics, University of Virginia School of Medicine, Charlottesville, VA, USA. [62]Genomic Research on Complex Diseases (GRC-Group), CSIR-Centre for Cellular and Molecular Biology (CSIR-CCMB), Hyderabad, India. [63]Department of Medicine and Therapeutics, Chinese University of Hong Kong, Hong Kong, China. [64]Chinese University of Hong Kong–Shanghai Jiao Tong University Joint Research Centre in Diabetes Genomics and Precision Medicine, Chinese University of Hong Kong, Hong Kong, China. [65]Samsung Advanced Institute for Health Sciences & Technology (SAIHST), Sungkyunkwan University, Samsung Medical Center, Seoul, South Korea. [66]Department of Preventive Medicine, Northwestern University Feinberg School of Medicine, Chicago, IL, USA. [67]Institute of Biomedical Sciences, Academia Sinica, Taipei, Taiwan. [68]Department of Genetics, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA. [69]Department of Human Genetics, Wellcome Sanger Institute, Wellcome Genome Campus, Hinxton, UK. [70]Laboratory of Neurogenetics, National Institute on Aging, National Institutes of Health, Bethesda, MD, USA. [71]Department of Medicine, Johns Hopkins University School of Medicine, Baltimore, MD, USA. [72]Center for Research on Genomics and Global Health, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD, USA. [73]Cardiovascular Health Research Unit, Department of Medicine, University of Washington, Seattle, WA, USA. [74]Division of Academics, Ochsner Health, New Orleans, LA, USA. [75]Division of Statistical Genomics, Washington University School of Medicine, St Louis, MO, USA. [76]Department of Research and Evaluation, Division of Biostatistics Research, Kaiser Permanente of Southern California, Pasadena, CA, USA. [77]Department of Biomedical Science, Hallym University, Chuncheon, South Korea. [78]Harvard Medical School, Boston, MA, USA. [79]Metabolic Research Laboratories, Wellcome Trust–Medical Research Council Institute of Metabolic Science, Department of Clinical Biochemistry, University of Cambridge, Cambridge, UK. [80]Department of Epidemiology, University of Groningen, University Medical Centre Groningen, Groningen, The Netherlands. [81]Department of Bioinformatics, Isfahan University of Medical Sciences, Isfahan, Iran. [82]Department of Medicine, Division of Endocrinology, Diabetes and Metabolism, Cedars-Sinai Medical Center, Los Angeles, CA, USA. [83]Lund University Diabetes Centre, Department of Clinical Sciences, Lund University, Skåne University Hospital, Malmö, Sweden. [84]Department of Nutrition, Gillings School of Global Public Health, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA. [85]Unidad de Investigación en Enfermedades Metabólicas y Departamento de Endocrinología y Metabolismo, Instituto Nacional de Ciencias Médicas y Nutrición Salvador Zubirán, Mexico City, Mexico. [86]Steno Diabetes Center Copenhagen, Herlev, Denmark. [87]Bioinformatics Center, Department of Biology, University of Copenhagen, Copenhagen, Denmark. [88]Department of Health Research Methods, Evidence and Impact, McMaster University, Hamilton, Ontario, Canada. [89]Department of Epidemiology and Prevention, Division of Public Health Sciences, Wake Forest School of Medicine, Winston-Salem, NC, USA. [90]Institute of Regional Health Research, University of Southern Denmark, Odense, Denmark. [91]Department of Clinical Biochemistry, Vejle Hospital, Vejle, Denmark. [92]Department of Medicine, Division of Endocrinology and Diabetes, Keck School of Medicine of USC, Los Angeles, CA, USA. [93]Department of Internal Medicine, University of Michigan, Ann Arbor, MI, USA. [94]British Heart Foundation Centre of Research Excellence, School of Clinical Medicine, Addenbrooke's Hospital, University of Cambridge, Cambridge, UK. [95]Health Data Research UK Cambridge, Wellcome Genome Campus, University of Cambridge, Hinxton, UK. [96]National Institute for Health and Care Research (NIHR) Blood and Transplant Unit (BTRU) in Donor Health and Behaviour, Heart and Lung Research Institute, University of Cambridge, Cambridge, UK. [97]Inserm U1283, CNRS UMR 8199, European Genomic Institute for Diabetes (EGID), Institut Pasteur de Lille, Lille University Hospital, Lille, France. [98]University of Lille, Lille, France. [99]Li Ka Shing Institute of Health Sciences, Chinese University of Hong Kong, Hong Kong, China. [100]Hong Kong Institute of Diabetes and Obesity, Chinese University of Hong Kong, Hong Kong, China. [101]Singapore Eye Research Institute, Singapore National Eye Centre, Singapore, Singapore. [102]Exeter Centre of Excellence in Diabetes (ExCEeD), Exeter Medical School, University of Exeter, Exeter, UK. [103]Wellcome Sanger Institute, Wellcome Genome Campus, Hinxton, UK. [104]Department of Biostatistics and Data Science, Wake Forest School of Medicine, Winston-Salem, NC, USA. [105]Division of Endocrinology and Metabolism, Department of Internal Medicine, National Taiwan University Hospital, Taipei, Taiwan. [106]Institute of Epidemiology and Preventive Medicine, National Taiwan University, Taipei, Taiwan. [107]Department of Medicine, University of Vermont, Colchester, VT, USA. [108]Section of Endocrinology and Metabolism, Department of Internal Medicine, Wake Forest School of Medicine, Winston-Salem, NC, USA. [109]Department of Medicine, Faculty of Medicine, University of Kelaniya, Ragama, Sri Lanka. [110]Department of Nutrition and Dietetics, Harokopio University of Athens, Athens, Greece. [111]Center for Genomics and Personalized Medicine Research, Wake Forest School of Medicine, Winston-Salem, NC, USA. [112]Carolina Population Center, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA. [113]Department of Nephrology and Medical Intensive Care Medicine, Charité–Universitätsmedizin Berlin, Berlin, Germany. [114]Department of Nephrology and Hypertension, Friedrich-Alexander-Universität Erlangen-Nürnberg, Erlangen, Germany. [115]Department of Biostatistics, University of Washington, Seattle, WA, USA. [116]California Pacific Medical Center Research Institute, San Francisco, CA, USA. [117]Laboratory of Epidemiology and Population Sciences, National Institute on Aging, National Institutes of Health, Baltimore, MD, USA. [118]Institute of Mathematics and Statistics, University of Tartu, Tartu, Estonia. [119]Robertson Centre for Biostatistics, University of Glasgow, Glasgow, UK. [120]Department of Epidemiology, Erasmus MC University Medical Center, Rotterdam, The Netherlands. [121]Genetics of Complex Traits, University of Exeter Medical School, University of Exeter, Exeter, UK. [122]Department of Internal Medicine, Wake Forest School of Medicine, Winston-Salem, NC, USA. [123]Department of Medicine, Division of Endocrinology and Metabolism, Lundquist Research Institute at Harbor-UCLA Medical Center, Torrance, CA, USA. [124]Department of Public Health and Caring Sciences, Uppsala University, Uppsala, Sweden. [125]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. [126]Department of Laboratory Medicine and Pathology, University of Minnesota, Minneapolis, MN, USA. [127]Genomics and Computational

Article

Biology Graduate Group, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA, USA. [128]Institute for Molecular Medicine Finland (FIMM), University of Helsinki, Helsinki, Finland. [129]Folkhalsan Research Center, Helsinki, Finland. [130]University of Exeter Medical School, University of Exeter, Exeter, UK. [131]Institute for Clinical Diabetology, German Diabetes Center, Leibniz Center for Diabetes Research at Heinrich Heine University Düsseldorf, Düsseldorf, Germany. [132]Department of Endocrinology and Diabetology, Medical Faculty and University Hospital Düsseldorf, Heinrich Heine University Düsseldorf, Düsseldorf, Germany. [133]Laboratory for Genomics of Diabetes and Metabolism, RIKEN Center for Integrative Medical Sciences, Yokohama, Japan. [134]Department of Biostatistics, Gillings School of Global Public Health, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA. [135]Department of Internal Medicine, Diabetes and Metabolism Research Center, Ohio State University Wexner Medical Center, Columbus, OH, USA. [136]Center for Global Cardiometabolic Health, Brown University, Providence, RI, USA. [137]Shanghai-MOST Key Laboratory of Health and Disease Genomics, Shanghai Institute for Biomedical and Pharmaceutical Technologies, Shanghai, China. [138]Division of Endocrine and Metabolism, Tri-Service General Hospital Songshan Branch, Taipei, Taiwan. [139]School of Medicine, National Defense Medical Center, Taipei, Taiwan. [140]Division of Genome Science, Department of Precision Medicine, National Institute of Health, Cheongju-si, Korea. [141]Section of Endocrinology and Metabolism, Department of Medicine, Taipei Veterans General Hospital, Taipei, Taiwan. [142]School of Medicine, National Yang Ming Chiao Tung University, Taipei, Taiwan. [143]Department of Environmental and Preventive Medicine, Jichi Medical University School of Medicine, Shimotsuke, Japan. [144]University of Chicago Research Bangladesh, Dhaka, Bangladesh. [145]Institute of Molecular and Clinical Ophthalmology Basel, Basel, Switzerland. [146]Center for Clinical Research and Prevention, Bispebjerg and Frederiksberg Hospital, Frederiksberg, Denmark. [147]Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark. [148]Faculty of Medicine, Aalborg University, Aalborg, Denmark. [149]Department of Clinical Diabetes, Endocrinology and Metabolism, Department of Translational Research and Cellular Therapeutics, City of Hope, Duarte, CA, USA. [150]Department of Public Health, Faculty of Medicine, University of Kelaniya, Ragama, Sri Lanka. [151]Department of Clinical Gene Therapy, Osaka University Graduate School of Medicine, Osaka, Japan. [152]Department of Geriatric and General Medicine, Graduate School of Medicine, Osaka University, Osaka, Japan. [153]Division of General Internal Medicine and Geriatrics, Department of Medicine, Northwestern University Feinberg School of Medicine, Chicago, IL, USA. [154]Center for Health Information Partnerships, Institute for Public Health and Medicine, Northwestern University Feinberg School of Medicine, Chicago, IL, USA. [155]Genome Institute of Singapore, Agency for Science, Technology and Research, Singapore, Singapore. [156]Department of Molecular Cell Biology, Sungkyunkwan University School of Medicine, Suwon, South Korea. [157]Institute of Genetic Epidemiology, Medical University of Innsbruck, Innsbruck, Austria. [158]Institute of Clinical Medicine, Internal Medicine, University of Eastern Finland and Kuopio University Hospital, Kuopio, Finland. [159]Department of Medicine, University of Colorado Denver, Anschutz Medical Campus, Aurora, CO, USA. [160]VA Salt Lake City Health Care System, Salt Lake City, UT, USA. [161]Department of Internal Medicine, University of Utah School of Medicine, Salt Lake City, UT, USA. [162]Soochunhyang Institute of Medi-bio Science and Division of Endocrinology, Department of Internal Medicine, Soochunhyang University College of Medicine, Cheonan, South Korea. [163]Department of Medicine, Samsung Medical Center, Sungkyunkwan University School of Medicine, Seoul, South Korea. [164]USC-Office of Population Studies Foundation, University of San Carlos, Cebu City, Philippines. [165]Department of Medicine, Harvard Medical School, Boston, MA, USA. [166]Division of General Internal Medicine, Massachusetts General Hospital, Boston, MA, USA. [167]Department of Epidemiology and Biostatistics, School of Public Health, Peking University, Beijing, China. [168]Peking University Center for Public Health and Epidemic Preparedness and Response, Beijing, China. [169]Department of Clinical Epidemiology, Leiden University Medical Center, Leiden, The Netherlands. [170]Wellcome Centre for Human Genetics, Nuffield Department of Medicine, University of Oxford, Oxford, UK. [171]Program in Medical and Population Genetics, Broad Institute, Cambridge, MA, USA. [172]Big Data Institute, Li Ka Shing Centre for Health Information and Discovery, University of Oxford, Oxford, UK. [173]Department of Clinical Medicine, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark. [174]Department of Biostatistics, Boston University School of Public Health, Boston, MA, USA. [175]Department of Medicine, Yong Loo Lin School of Medicine, National University of Singapore and National University Health System, Singapore, Singapore. [176]McDonnell Genome Institute, Washington University School of Medicine, St Louis, MO, USA. [177]Department of Medicine, Division of Genomics and Bioinformatics, Washington University School of Medicine, St Louis, MO, USA. [178]Department of Biostatistics and Data Science, University of Texas Health Science Center at Houston School of Public Health, Houston, TX, USA. [179]Department of Clinical Sciences, Diabetes and Endocrinology, Lund University Diabetes Centre, Malmö, Sweden. [180]Department of Clinical Science, Center for Diabetes Research, University of Bergen, Bergen, Norway. [181]Department of Advanced Genomic and Laboratory Medicine, Graduate School of Medicine, University of the Ryukyus, Nishihara, Japan. [182]Division of Clinical Laboratory and Blood Transfusion, University of the Ryukyus Hospital, Nishihara, Japan. [183]Dromokaiteio Psychiatric Hospital, National and Kapodistrian University of Athens, Athens, Greece. [184]Academy of Scientific and Innovative Research (AcSIR), Ghaziabad, India. [185]Computational Biology and Medical Sciences, Graduate School of Frontier Sciences, University of Tokyo, Tokyo, Japan. [186]Institute of Human Genetics, Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg, Germany. [187]Institute of Human Genetics, Technical University Munich, Munich, Germany. [188]German Centre for Cardiovascular Research (DZHK), Partner Site Munich Heart Alliance, Munich, Germany. [189]Usher Institute to the Population Health Sciences and Informatics, University of Edinburgh, Edinburgh, UK. [190]Department of Medicine and Pharmacology, New York Medical College, Valhalla, NY, USA. [191]Data Tecnica International, Glen Echo, MD, USA. [192]Center for Alzheimer's and Related Dementias, National Institutes of Health, Bethesda, MD, USA. [193]Laboratory of Statistical Immunology, Immunology Frontier Research Center (WPI-IFReC), Osaka University, Suita, Japan. [194]Premium Research Institute for Human Metaverse Medicine (WPI-PRIMe), Osaka University, Suita, Japan. [195]Instituto Nacional de Medicina Genómica, Mexico City, Mexico. [196]Division of Pulmonary, Allergy, and Critical Care Medicine, Department of Medicine, University of Pittsburgh, Pittsburgh, PA, USA. [197]Division of Epidemiology and Community Health, School of Public Health, University of Minnesota, Minneapolis, MN, USA. [198]Institute for Medical Information Processing, Biometry and Epidemiology, Ludwig-Maximilians-Universität München, Munich, Germany. [199]Department of Diabetes and Endocrinology, Nelson R. Mandela School of Medicine, College of Health Sciences, University of KwaZulu-Natal, Durban, South Africa. [200]Academy of Scientific and Innovative Research, CSIR-Human Resource Development Campus, Ghaziabad, India. [201]Genomics and Molecular Medicine Unit, CSIR-Institute of Genomics and Integrative Biology, New Delhi, India. [202]Fred Hutchinson Cancer Research Center, Seattle, WA, USA. [203]Ophthalmology and Visual Sciences Academic Clinical Program (Eye ACP), Duke-NUS Medical School, Singapore, Singapore. [204]Department of Ophthalmology, Yong Loo Lin School of Medicine, National University of Singapore and National University Health System, Singapore, Singapore. [205]School of Cardiovascular and Metabolic Health, University of Glasgow, Glasgow, UK. [206]Survey Research Center, Institute for Social Research, University of Michigan, Ann Arbor, MI, USA. [207]Institute of Genetic Epidemiology, Helmholtz Zentrum Munchen, German Research Center for Environmental Health, Neuherberg, Germany. [208]Institute for Medical Biostatistics, Epidemiology, and Informatics (IMBEI), University Medical Center, Johannes Gutenberg University, Mainz, Germany. [209]Chair of Genetic Epidemiology, Institute of Medical Information Processing, Biometry, and Epidemiology, Faculty of Medicine, Ludwig-Maximilians-Universität München, Munich, Germany. [210]Faculty of Medicine, University of Iceland, Reykjavik, Iceland. [211]Faculty of Medicine, Macau University of Science and Technology, Macau, China. [212]Department of Medical Genetics and Medical Research, China Medical University Hospital, Taichung, Taiwan. [213]Population Health Unit, Finnish Institute for Health and Welfare, Helsinki, Finland. [214]National School of Public Health, Madrid, Spain. [215]Department of Public Health, University of Helsinki, Helsinki, Finland. [216]Diabetes Research Group, King Abdulaziz University, Jeddah, Saudi Arabia. [217]Unidad de Biología Molecular y Medicina Genómica, Instituto Nacional de Ciencias Médicas y Nutrición Salvador Zubirán, Mexico City, Mexico. [218]Departamento de Medicina Genómica y Toxiología Ambiental, Instituto de Investigaciones Biomédicas, UNAM, Mexico City, Mexico. [219]Unidad de Investigacion Medica en Bioquimica, Hospital de Especialidades, Centro Medico Nacional Siglo XXI, Instituto Mexicano del Seguro Social, Mexico City, Mexico. [220]Einthoven Laboratory for Experimental Vascular Medicine, Leiden University Medical Center, Leiden, The Netherlands. [221]Department of Human Genetics, Leiden University Medical Center, Leiden, The Netherlands. [222]Department of Clinical Chemistry, Laboratory of Genetic Metabolic Disease, Amsterdam University Medical Center, Amsterdam, The Netherlands. [223]Southern California Eye Institute, CHA Hollywood Presbyterian Hospital, Los Angeles, CA, USA. [224]Unidad de Investigación Médica en Epidemiologia Clinica, Hospital de Especialidades, Centro Medico Nacional Siglo XXI, Instituto Mexicano del Seguro Social, Mexico City, Mexico. [225]Department of Internal Medicine, Division of Endocrinology, Leiden University Medical Center, Leiden, The Netherlands. [226]Department of Public Health, Aarhus University, Aarhus, Denmark. [227]Danish Diabetes Academy, Odense, Denmark. [228]Diabetology Research Centre, King Edward Memorial Hospital and Research Centre, Pune, India. [229]Department of Medical Biochemistry, Kurume University School of Medicine, Kurume, Japan. [230]Department of Pediatrics, Osaka University Graduate School of Medicine, Suita, Japan. [231]Division of Cancer Control and Population Sciences, UPMC Hillman Cancer Center, University of Pittsburgh, Pittsburgh, PA, USA. [232]Department of Epidemiology, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, PA, USA. [233]Department of Pediatrics, Division of Genetic and Genomic Medicine, UCI Irvine School of Medicine, Irvine, CA, USA. [234]Department of Anti-Aging Medicine, Ehime University Graduate School of Medicine, Touon, Japan. [235]Division of Pulmonary, Critical Care, and Sleep Medicine, Beth Israel Deaconess Medical Center, Boston, MA, USA. [236]Department of Medical Sciences, Uppsala University, Uppsala, Sweden. [237]Institute of Molecular Medicine, University of Texas Health Science Center at Houston School of Public Health, Houston, TX, USA. [238]Human Genetics Center, University of Texas Health Science Center at Houston, Houston, TX, USA. [239]Department of Medicine, Stanford University School of Medicine, Stanford, CA, USA. [240]Department of Medical Sciences, Molecular Epidemiology and Science for Life Laboratory, Uppsala University, Uppsala, Sweden. [241]Department of Epidemiology, University of Washington, Seattle, WA, USA. [242]Department of Health Systems and Population Health, University of Washington, Seattle, WA, USA. [243]Beijing Institute of Ophthalmology, Ophthalmology and Visual Sciences Key Laboratory, Beijing Tongren Hospital, Capital Medical University, Beijing, China. [244]Department of Medicine, Division of Cardiology, Duke University School of Medicine, Durham, NC, USA. [245]Kurume University School of Medicine, Kurume, Japan. [246]Department of Medicine, McGill University, Montreal, Quebec, Canada. [247]Department of Human Genetics, McGill University, Montreal, Quebec, Canada. [248]Department of Metabolism, Digestion and Reproduction, Imperial College London, London, UK. [249]Division of Cardiovascular Medicine, Beth Israel Deaconess Medical Center, Boston, MA, USA. [250]Division of Endocrinology and Metabolism, Department of Medicine, Taichung Veterans General Hospital, Taichung, Taiwan. [251]Division of Endocrinology, Metabolism and Molecular Medicine, Department of Medicine, Northwestern University Feinberg School of Medicine, Chicago, IL, USA. [252]Center for Genetic Medicine, Northwestern University Feinberg School of Medicine, Chicago, IL, USA. [253]Department of Anthropology, Northwestern University,

Evanston, IL, USA. [254]Department of Endocrinology, Helsinki University Hospital, Helsinki, Finland. [255]Science and Engineering Research Board (SERB), Department of Science and Technology, Ministry of Science and Technology, Government of India, New Delhi, India. [256]Systems Genomics Laboratory, School of Biotechnology, Jawaharlal Nehru University, New Delhi, India. [257]Department of Pathology and Molecular Medicine, McMaster University, Hamilton, Ontario, Canada. [258]Department of Molecular Medicine and Biopharmaceutical Sciences, Graduate School of Convergence Science and Technology, Seoul National University, Seoul, South Korea. [259]Netherlands Heart Institute, Utrecht, The Netherlands. [260]Center for Public Health Genomics, University of Virginia School of Medicine, Charlottesville, VA, USA. [261]Research Unit of Molecular Epidemiology, Helmholtz Zentrum München, German Research Center for Environmental Health, Munich, Germany. [262]Duke-NUS Medical School, Singapore, Singapore. [263]Department of Epidemiology, Biostatistics and Occupational Health, McGill University, Montreal, Quebec, Canada. [264]Singapore Institute for Clinical Sciences, Agency for Science Technology and Research (A*STAR), Singapore, Singapore. [265]Healthy Longevity Translational Research Programme, Yong Loo Lin School of Medicine, National University of Singapore, Singapore, Singapore. [266]Center for Diabetes Research, Wake Forest School of Medicine, Winston-Salem, NC, USA. [267]Department of Biochemistry, Wake Forest School of Medicine, Winston-Salem, NC, USA. [268]Pat Macpherson Centre for Pharmacogenetics and Pharmacogenomics, University of Dundee, Dundee, UK. [269]Imperial College Healthcare NHS Trust, Imperial College London, London, UK. [270]MRC-PHE Centre for Environment and Health, Imperial College London, London, UK. [271]National Heart and Lung Institute, Imperial College London, London, UK. [272]Department of Medicine, Brown University Alpert School of Medicine, Providence, RI, USA. [273]Department of Medicine, Columbia University Irving Medical Center, New York, NY, USA. [274]Department of Cardiology, Columbia University Irving Medical Center, New York, NY, USA. [275]Center for Non-Communicable Diseases, Karachi, Pakistan. [276]Computational Medicine, Berlin Institute of Health at Charité–Universitätsmedizin, Berlin, Germany. [277]Precision Healthcare University Research Institute, Queen Mary University of London, London, UK. [278]Department of Preventive Medicine, Keck School of Medicine of USC, Los Angeles, CA, USA. [279]The Mindich Child Health and Development Institute, Ichan School of Medicine at Mount Sinai, New York, NY, USA. [280]Division of Translational Medicine and Therapeutics, Department of Medicine, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA, USA. [281]Institute for Translational Medicine and Therapeutics, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA, USA. [282]Department of Pediatrics, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA, USA. [283]Center for Precision Medicine, University of Pennsylvania - Perelman School of Medicine, Philadelphia, PA, USA. [284]Institute for Biomedical Informatics, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA, USA. [285]Department of Psychiatry, University of Michigan, Ann Arbor, MI, USA. [286]Blizard Institute, Queen Mary University of London, London, UK. [287]Institute for Population Health Sciences, Barts and the London School of Medicine and Dentistry, Queen Mary University of London, London, UK. [288]All of Us Research Program, National Institutes of Health, Bethesda, MD, USA. [289]Toranomon Hospital, Tokyo, Japan. [290]Lee Kong Chian School of Medicine, Nanyang Technological University, Singapore, Singapore. [291]Vanderbilt Genetics Institute, Division of Genetic Medicine, Vanderbilt University Medical Center, Nashville, TN, USA. [292]VA Palo Alto Health Care System, Palo Alto, CA, USA. [293]Stanford Cardiovascular Institute, Stanford University School of Medicine, Stanford, CA, USA. [294]Department of Medicine, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA, USA. [295]Oxford Centre for Diabetes, Endocrinology and Metabolism, Radcliffe Department of Medicine, University of Oxford, Oxford, UK. [296]Oxford NIHR Biomedical Research Centre, Churchill Hospital, Oxford University Hospitals NHS Foundation Trust, Oxford, UK. [297]Department of Biostatistics and Epidemiology, University of Massachusetts Amherst, Amherst, MA, USA. [298]Department of Biostatistics, Epidemiology and Informatics, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA, USA. [299]TUM School of Medicine and Health, Technical University of Munich and Klinikum Rechts der Isar, Munich, Germany. [300]Present address: Regeneron Genetics Center, Tarrytown, NY, USA. [301]Present address: Genentech, South San Francisco, CA, USA. [302]These authors contributed equally: Ken Suzuki, Konstantinos Hatzikotoulas, Lorraine Southam, Henry J. Taylor, Xianyong Yin, Kim M. Lorenz, Ravi Mandla. [303]These authors jointly supervised this work: Mark I. McCarthy, James B. Meigs, Anubha Mahajan, Cassandra N. Spracklen, Josep M. Mercader, Michael Boehnke, Jerome I. Rotter, Marijana Vujkovic, Benjamin F. Voight, Andrew P. Morris, Eleftheria Zeggini. [304]Deceased: Michèle M. Sale. *Lists of authors and their affiliations appear at the end of the paper. ✉e-mail: konstantinos.hatzikotoulas@helmholtz-munich.de; bvoight28@gmail.com; andrew.morris-5@manchester.ac.uk; eleftheria.zeggini@helmholtz-munich.de

VA Million Veteran Program
Kim M. Lorenz[10,11,12,302], Kyung Min Lee[160,161], Julie A. Lynch[160,161], Philip S. Tsao[239,292,293], Kyong-Mi Chang[10,294], Marijana Vujkovic[10,294,298,303] & Benjamin F. Voight[10,11,12,281,303]

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

AMED GRIFIN Diabetes Initiative Japan
Ken Suzuki[1,2,3,302], Kyuto Sonehara[3,19,20,21], Shinichi Namba[3], Momoko Horikoshi[133], Shiro Maeda[133,181,182], Koichi Matsuda[185], Yukinori Okada[3,19,20,21,193,194], Nobuhiro Shojima[2], Kenichi Yamamoto[3,193,230], Toshimasa Yamauchi[2] & Takashi Kadowaki[2,289]

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

Biobank Japan Project
Koichi Matsuda[185]

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

Penn Medicine BioBank
Lindsay A. Guare[127], Daniel J. Rader[12,280,281,282], Marylyn D. Ritchie[12,283,284], Marijana Vujkovic[10,294,298,303] & Benjamin F. Voight[10,11,12,281,303]

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

Regeneron Genetics Center

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

Genes & Health Research Team
Stavroula Kanona[16], David A. van Heel[286] & Sarah Finer[287]

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

eMERGE Consortium
Yoonjung Yoonie Joo[65,66], Abel N. Kho[153,154] & M. Geoffrey Hayes[251,252,253]

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

International Consortium of Blood Pressure (ICBP)
Zoha Kamali[80,81]

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

Meta-Analyses of Glucose and Insulin-Related Traits Consortium (MAGIC)
K. Alaine Broadaway[68], Alice Williamson[37,79], Karen L. Mohlke[68] & Claudia Langenberg[37,276,277]

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

Article

Methods

Study-level analyses

Within each study, we assigned individuals to ancestry groups using self-report and genetic background (Supplementary Tables 1 and 2). Any individuals not assigned to an ancestry group were excluded as population outliers. Within each ancestry group-specific GWAS, we conducted quality control of genotype data and imputed up to reference panels from the Trans-Omics for Precision Medicine Program[51], Haplotype Reference Consortium[52], 1000 Genomes Project (phase 1, March 2012 release; phase 3, October 2014 release)[53,54], or population-specific whole-genome sequencing[55–61] (Supplementary Table 3). Studies imputed to reference panels mapped to GRCh38 (hg38) were lifted back to hg19 using the UCSC LiftOver tool (https://genome.ucsc.edu/cgi-bin/hgLiftOver). We excluded SNVs with poor imputation quality and/or minor allele count (MAC) < 5 (Supplementary Table 3).

Within each ancestry group-specific GWAS, we tested for association of each SNV with T2D through generalized linear (mixed) modelling, under an additive dosage of the minor allele, with adjustment for age and sex (where appropriate), and additional study-specific covariates (Supplementary Table 3). We used different strategies to account for population stratification and/or kinship: (i) exclude closely related individuals and adjust for principal components derived from a genetic relatedness matrix (GRM) as additional covariates; or (ii) incorporate a random effect for the GRM (Supplementary Table 3). Allelic effects and corresponding standard errors that were estimated from a linear mixed model were converted to the log-odds scale[62]. We corrected study-level association summary statistics for residual structure by the LD-score regression intercept[63] (Supplementary Table 3) using an LD reference that we derived from ancestry-matched haplotypes from continental groups in the 1000 Genomes Project (phase 3, October 2014 release)[54]. We matched AFA GWASs to the 'African' continental group and HIS GWASs to the 'American' continental group.

Multi-ancestry meta-analyses

We analysed autosomal bi-allelic SNVs that overlap reference panels from the 1000 Genomes Project (phase 3, October 2014 release)[54] and the Haplotype Reference Consortium[52]. We considered SNVs with MAF > 0.5% in at least one of the five continental groups in the 1000 Genomes Project (phase 3, October 2014 release)[54]. We excluded SNVs that differed in allele frequency by more than 20% when comparing reference panels in the same subsets of haplotypes.

We used meta-regression, implemented in MR-MEGA[10], to aggregate association summary statistics across GWASs. MR-MEGA models allelic effect heterogeneity that is correlated with genetic ancestry by including axes of genetic variation as covariates in the meta-regression model to capture diversity between GWASs. We used SNVs reported in all studies to construct a distance matrix of differences in mean effect allele frequency between each pair of GWASs. We implemented multi-dimensional scaling of the distance matrix to obtain three principal components that represent axes of genetic variation to separate GWASs across ancestry groups (Extended Data Fig. 1).

For each SNV, we aggregated inverse-variance weighted allelic effects across GWASs through linear regression, including three axes of genetic variation as covariates. We tested for: (i) association with T2D allowing for ancestry-correlated allelic effect heterogeneity between GWASs; (ii) ancestry-correlated allelic effect heterogeneity between GWASs (defined by the axes of genetic variation); and (iii) residual allelic effect heterogeneity between GWASs. MR-MEGA is a meta-regression approach, and therefore does not produce an allelic effect estimate because this is allowed to vary with the axes of genetic variation. Consequently, we also aggregated association summary statistics across GWASs through fixed-effects meta-analysis (inverse-variance weighting of allelic effects) using METAL[64]. To assess the extent of residual structure between GWASs, we calculated the genomic control inflation factor[65] for the multi-ancestry meta-regression and the fixed-effects meta-analysis. We considered only those SNVs reported in at least five GWASs for downstream interrogation.

Defining T2D signals and loci

We identified all SNVs attaining genome-wide significance ($P < 5 \times 10^{-8}$) for association with T2D from the multi-ancestry meta-regression. Clumps were formed around index variants, which were selected using a greedy algorithm in PLINK v.1.9 (ref. 66), after ranking SNVs by ascending P value. SNVs less than 5 Mb from an index SNV were assigned to the clump if $r^2 > 0.05$ in at least one of the five continental groups from the 1000 Genomes Project (phase 3, October 2014 release)[54]. Index SNVs separated by less than 1 Mb were assigned to the same locus. Each locus was then defined as mapping 500 kb up- and downstream of index SNVs contained within it. We considered the locus to have been previously reported if it contained variants discovered in published large-scale T2D GWASs at genome-wide significance.

Ancestry-group-specific meta-analyses

We aggregated association summary statistics across GWASs from the same ancestry group through fixed-effects meta-analysis (inverse-variance weighting of allelic effects) using METAL[64]. We estimated the mean effect allele frequency across GWASs from each ancestry group, weighted by the effective sample size of the study. We generated forest plots of association summary statistics of index SNVs across ancestry groups using the R package meta (https://cran.r-project.org/package=meta/).

Defining clusters of T2D index SNVs with distinct cardiometabolic profiles

We considered cardiometabolic-related quantitative phenotypes that are used to define T2D status and/or are associated with risk of T2D or complications. We excluded phenotypes for which GWAS summary statistics were available only after imputation to reference panels from the International HapMap Project[67] because they did not provide sufficient coverage of SNVs included in the multi-ancestry meta-analysis. We considered the largest available GWAS meta-analysis (ancestry-specific or multi-ancestry) that provided the following association summary statistics for each SNV: effect allele, other allele, allelic effect and corresponding standard error (Supplementary Table 5). We re-aligned the effect estimate to the T2D risk allele from the fixed-effects multi-ancestry meta-analysis, denoted β_{ij} for the jth index SNV and the ith phenotype. We then calculated a sample size corrected z-score, given by $Z_{ij} = \beta_{ij}/(\sqrt{N_i}s_{ij})$, where s_{ij} is the standard error of the effect estimate of the jth index SNV and the ith phenotype, and N_i is the maximum sample size reported for the ith phenotype. Where association summary statistics were not reported, the z-score was set as 'missing'.

We conducted k-means clustering of index SNVs with imputation of missing z-scores using the R package ClustImpute (https://cran.r-project.org/package=ClustImpute). For a pre-defined number of clusters, ClustImpute replaces missing z-scores at random from the marginal distribution for the phenotype in the first iteration and performs k-means clustering. In subsequent iterations, missing z-scores are updated, conditional on the current cluster assignment, so that correlations between phenotypes are considered. At each iteration, penalizing weights are imposed on imputed values and successively decreased (to zero) as the missing data imputation improves. Finally, we determined the 'optimal' number of clusters according to the majority rule across 27 indices of cluster performance[68], implemented in the R package NbClust (https://cran.r-project.org/package=NbClust).

We tested for association of the ith phenotype with index SNVs across clusters in a linear regression model, given by $E(Z_{ij}) = \sum_k \gamma_{ik} C_{jk}$, where C_{jk} is an indicator variable that takes the value 1 if the jth index SNV was assigned to the kth cluster and 0 otherwise. The strength or direction of the association of each phenotype with each cluster was then

presented in a heat map, in which the 'temperature' was defined by the direction of the regression coefficient γ_{ik} and the corresponding $-\log_{10} P$ value. Regression models were fitted using the glm function in R.

We extracted cardiometabolic phenotype z-scores from the final imputed dataset from ClustImpute. We calculated the Euclidean distance between the jth SNV and kth cluster centroid as

$$\delta_{jk} = \sqrt{\sum_i (Z_{ij} - \mu_{ik})^2},$$

where Z_{ij} and μ_{ik} are the z-score of the jth SNV and the location of the kth cluster centroid for the ith cardiometabolic phenotype. To assess cluster disparity, we also performed principal components analysis of cardiometabolic phenotype z-scores from the final imputed dataset using the R package factoextra (https://cran.r-project.org/package=factoextra).

Cluster-specific associations of index SNVs with T2D

We tested for association of T2D with index SNVs across clusters in a linear regression model, given by $E(\beta_j) = \sum_k \gamma_k C_{jk}$, where C_{jk} is an indicator variable that takes the value 1 if the jth index SNV was assigned to the kth cluster and 0 otherwise, and weighted by the inverse of the variance of the allelic effect. We tested for heterogeneity in cluster effects on T2D by comparing the deviance of this model with that of $E(\beta_j) = \gamma_0$, again weighted by the inverse of the variance of the allelic effect. To compare associations between previously reported clusters and previously unreported clusters, we replaced C_{jk} with an indicator variable that takes the value 1 if the jth index SNV was assigned to a previously reported cluster and 0 otherwise. Regression models were fitted using the glm function in R.

Enrichment of T2D associations for cell-type-specific regions of open chromatin within clusters

For each T2D association signal, we defined 'null' SNVs that mapped within 50 kb of the index SNV and were not in LD ($r^2 > 0.05$) with the index SNV in any of the five continental groups from the 1000 Genomes Project (phase 3, October 2014 release)[54]. We defined an indicator variable, Y_j, taking the value 1 if the jth SNV is an index SNV and 0 if the jth SNV is a null SNV. We mapped index SNVs and null SNVs to genic regions defined by the Ensembl Project (release 104)[69], including protein-coding exons, and 3′ UTRs and 5′ UTRs. We defined indicator variables, G_j^{EXON}, G_j^{3UTR} and G_j^{5UTR}, which each take the value 1 if the jth SNV mapped to the respective genic annotation and 0 otherwise. We also mapped index SNVs and null SNVs to ATAC-seq peaks from single-cell atlases of chromatin accessibility (CATLAS and DESCARTES) for: 222 cell types derived from 30 human adult and 15 human fetal tissues[25,26]; and 106 cell types derived from human brain[27]. We defined an indicator variable, X_{ij}, that takes the value 1 if the jth SNV mapped to an ATAC-seq peak for the ith cell type and 0 otherwise.

Within each cluster, we modelled enrichment of T2D associations for ATAC-seq peaks in the ith cell type, after accounting for genic annotations, in a Firth bias-reduced logistic regression, given by

$$f^{-1}(Y_j) = \alpha_0 + \alpha_{EXON}G_j^{EXON} + \alpha_{3UTR}G_j^{3UTR} + \alpha_{5UTR}G_j^{5UTR} + \theta_i X_{ij},$$

where f is the logit link function. In this expression, α_0 is an intercept, α_{EXON}, α_{3UTR} and α_{5UTR} are log fold enrichments of genic annotations, and θ_i is the log fold enrichment of ATAC-seq peaks in the ith cell type. We conducted a test of enrichment of the ith cell type by comparing the deviances of models in which $\theta_i = 0$ and θ_i is unconstrained. We identified cell types with significant evidence of enrichment ($P < 0.00023$, Bonferroni correction for 222 cell types in adult and fetal tissues; $P < 0.00047$, Bonferroni correction for 106 cell types in the brain). All models were fitted using the R package logistf (https://cran.r-project.org/package=logistf).

Contribution of each axis of genetic variation to ancestry-correlated heterogeneity

For each index SNV, we calculated a z-score (beta/SE) for association with each axis of variation by aligning the effect from the meta-regression model to the T2D risk allele. For each index SNV, we identified the axis of genetic variation with the strongest association (greatest magnitude z-score).

Differences in ancestry-correlated heterogeneity between mechanistic clusters

We tested for differences in z-scores (beta/SE) for association of index SNVs in each cluster with the ith axis of genetic variation by comparing two linear models by ANOVA: (i) $f^{-1}(Z_{ij}) = \tau_{0i}$; and (ii) $f^{-1}(Z_{ij}) = \sum_k \tau_{ki}C_{jk}$. In these expressions: f is the identity link function; Z_{ij} is the z-score for the jth index SNV; C_{jk} is an indicator variable that takes the value 1 if the jth index SNV was assigned to the kth cluster and 0 otherwise; and τ_{0i} and τ_{ki} are regression coefficients. Regression models were fitted using the glm function in R.

Effect of BMI on ancestry-correlated and residual heterogeneity in allelic effects between GWASs

For each index SNV, we aggregated inverse-variance weighted allelic effects across GWASs by linear regression, implemented in MR-MEGA[10], including as covariates: (i) three axes of genetic variation; (ii) mean BMI in controls; and (iii) mean BMI in T2D cases. After adjustment for BMI, we tested for: (i) ancestry-correlated allelic effect heterogeneity between GWASs; and (ii) residual allelic effect heterogeneity between GWASs. After adjustment, as outlined above, we re-assessed: (i) the contribution of each axis of genetic variation to ancestry-correlated heterogeneity; and (ii) the difference in ancestry-correlated heterogeneity between mechanistic clusters.

Cluster-specific partitioned PS analyses of vascular outcomes and age of T2D onset

We tested for association of cluster-specific components of the partitioned PS and an overall PS with T2D-related macrovascular outcomes (CAD, ischaemic stroke and peripheral artery disease), microvascular complications (ESDN and proliferative diabetic retinopathy) and age of T2D onset in participants from the All of Us Research Program (AoURP; AFA, EUR and HIS ancestry groups), Biobank Japan (BBJ; EAS ancestry group), and Genes & Health (G&H; SAS ancestry group). Cohort descriptions and details of sequencing and genotyping, quality control and phenotype derivation are provided in the Supplementary Methods.

We conducted analyses separately for each ancestry group in AoURP, BBJ and G&H. For each ancestry, we performed analyses for macrovascular outcomes using all individuals, irrespective of T2D status, and for microvascular complications in individuals with T2D only. For each analysis, we calculated the overall PS and cluster-specific partitioned PS for each individual, with each index SNV weighted by the allelic log-OR from the ancestry-specific meta-analyses. We did not include index SNVs with MAF < 1% in the PS. We also excluded index SNVs with poor imputation quality ($r^2 < 0.7$) in BBJ and G&H, and those with extreme deviation from Hardy–Weinberg equilibrium ($P < 10^{-6}$) in AoURP. We standardized the overall PS and each cluster-specific component of the partitioned PS to have mean zero and unit variance. We tested for association with each vascular outcome through generalized linear regression and with age of T2D onset through linear regression. For each outcome, we considered a model including the overall PS and then each cluster-specific component the partitioned PS adjusted for the overall PS. All association analyses were conducted using the glm function in R.

We adjusted association analyses with vascular outcomes for age, sex and the first 20 principal components. In BBJ, we also adjusted for recruitment phase and status of the registered common diseases

Article

(other than T2D) to account for ascertainment. We further adjusted analyses of macrovascular outcomes for T2D status. We also further adjusted analyses of microvascular complications for duration of T2D. In AoURP, we defined age as age at last hospital visit. In BBJ, we defined age as age at first record. In G&H, we defined age as age at diagnosis for T2D cases and age at last follow-up for controls. For CAD, we also conducted sensitivity analyses by including, as an additional covariate, a CAD PS from the largest published multi-ancestry CAD GWAS[39]. The PS was constructed from index SNVs for 241 conditionally independent CAD associations, weighted by the multi-ancestry allelic log-OR (ancestry-specific effects were not available), and standardized to have mean zero and unit variance. We adjusted association analyses with age of T2D onset for sex and the first 20 principal components. In BBJ, we also adjusted for recruitment phase and status of the registered common diseases (other than T2D) to account for ascertainment.

For each outcome, we aggregated association summary statistics from each cluster-specific component of the partitioned PS and the overall PS across ancestries through random-effects meta-analyses. All meta-analyses were conducted using the R package meta (https://cran.r-project.org/package=meta).

Cluster-specific partitioned PS analyses of clinical outcomes

We tested for association of cardiovascular and kidney-related clinical outcomes in EUR individuals with T2D in prospective GWASs from six clinical trials from the Thrombolysis in Myocardial Infarction (TIMI) Study Group (https://timi.org/). Trial descriptions and details of genotyping and quality control are provided in the Supplementary Methods.

Within each trial, we calculated the overall PS and cluster-specific components of the partitioned PS for each individual, with each index SNV weighted by the allelic log-OR from the European ancestry-specific meta-analysis. We standardized the overall PS and each cluster-specific component of the partitioned PS to have mean zero and unit variance. Data from the six trials were subsequently pooled, and we considered the following clinical outcomes in patients with T2D only: myocardial infarction, ischaemic stroke, cardiovascular death, hospitalization for heart failure, atrial fibrillation, acute limb ischaemia, peripheral revascularization, end-stage renal disease or renal death and albuminuria. We tested for association of each cluster-specific component of the partitioned PS with each clinical outcome under a Cox proportional hazards model, including age, sex, the first ten principal components and the overall PS as covariates. All association analyses were conducted using the coxph function with Efron ties handling from the R package survival (https://cran.r-project.org/package=survival).

Ethics statement
Study-level ethics statements are provided in the Supplementary Note.

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

Data availability
Genome-wide association summary statistics from the multi-ancestry meta-analysis and ancestry-specific meta-analyses reported in this study are available through the DIAGRAM Consortium website (http://www.diagram-consortium.org/downloads.html).

Code availability
Analyses were conducted using publicly available software: the UCSC LiftOver tool (https://genome.ucsc.edu/cgi-bin/hgLiftOver), MR-MEGA v.0.2 (https://genomics.ut.ee/en/tools), METAL v.2011-03-25 (https://genome.sph.umich.edu/wiki/METAL), PLINK v.1.9 (https://www.cog-genomics.org/plink/1.9/), Beagle 4.1 (https://faculty.washington.edu/browning/beagle/b4_1.html), SNPTEST v.2.5.6 (https://www.well.ox.ac.uk/~gav/snptest/), GWAMA v.2.2.2 (https://genomics.ut.ee/en/tools), EIGENSOFT v.7.2.1 (https://www.hsph.harvard.edu/alkes-price/software/), PLINK v.2.0 (https://www.cog-genomics.org/plink/2.0/), SHAPEIT4 (https://odelaneau.github.io/shapeit4/), Minimac4 (https://genome.sph.umich.edu/wiki/Minimac4), KING v.2.3 (https://www.kingrelatedness.com/) and EAGLE v.2.4 (https://alkesgroup.broadinstitute.org/Eagle/#Xeagle2). Analyses were also conducted using the following R packages: meta (https://cran.r-project.org/package=meta), ClustImpute (https://cran.r-project.org/package=ClustImpute), NbClust (https://cran.r-project.org/package=NbClust), factoextra (https://cran.r-project.org/package=factoextra) and logistf (https://cran.r-project.org/package=logistf).

51. Taliun, D. et al. Sequencing of 53,831 diverse genomes from the NHLBI TOPMed Program. *Nature* **590**, 290–299 (2021).
52. McCarthy, S. et al. A reference panel of 64,976 haplotypes for genotype imputation. *Nat. Genet.* **48**, 1279–1283 (2016).
53. The 1000 Genomes Project Consortium. An integrated map of genetic variation from 1,092 human genomes. *Nature* **491**, 56–65 (2012).
54. The 1000 Genomes Project Consortium. A global reference for human genetic variation. *Nature* **526**, 68–74 (2015).
55. 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).
56. Jónsson, H. et al. Whole genome characterization of sequence diversity of 15,220 Icelanders. *Sci. Data* **4**, 170115 (2017).
57. Moon, S. et al. The Korea Biobank Array: design and identification of coding variants associated with blood biochemical traits. *Sci. Rep.* **9**, 1382 (2019).
58. Akiyama, M. et al. Characterizing rare and low-frequency height-associated variants in the Japanese population. *Nat. Commun.* **10**, 4393 (2019).
59. Gurdasani, D. et al. The African Genome Variation Project shapes medical genetics in Africa. *Nature* **517**, 327–332 (2015).
60. Gurdasani, D. et al. Uganda genome resource enables insights into population history and genomic discovery in Africa. *Cell* **179**, 984–1002 (2019).
61. Kurki, M. I. et al. FinnGen provides genetic insights from a well-phenotyped isolated population. *Nature* **613**, 508–518 (2023).
62. 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 (2017).
63. Bulik-Sullivan, B. K. et al. LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat. Genet.* **47**, 291–295 (2015).
64. Willer, C. J., Li, Y. & Abecasis, G. R. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **26**, 2190–2191 (2010).
65. Devlin, B. & Roeder, K. Genomic control for association studies. *Biometrics* **55**, 997–1004 (1999).
66. Chang, C. C. et al. Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* **4**, 7 (2015).
67. The International HapMap Consortium. The International HapMap Project. *Nature* **426**, 789–796 (2003).
68. Charrad, M., Ghazzali, N., Boiteau, V. & Niknafs, A. NbClust: an R package for determining the relevant number of clusters in a data set. *J. Stat. Softw.* **61**, 1–36 (2014).
69. Cunningham, F. et al. Ensembl 2022. *Nucleic Acids Res.* **50**, D988–D995 (2022).

Author contributions Central analysis group: K. Suzuki, K.H., L.S., H.J.T., X.Y., K. M. Lorenz, R. Mandla, A.H.-C., N.W.R., O.B., A.L.A., A. Mahajan, C.N.S., J.M.M., M.B., J.I.R., M.V., B.F.V., A.P.M., E.Z. PS analysis: K. Suzuki, K.H., L.S., H.J.T., G.E.M.M., S.K., A.P.M. Study-level primary analysis: K. Suzuki, H.J.T., X.Y., K. M. Lorenz, R. Mandla, A.H.-C., K. Sonehara, S.N., S.S.K.L., M.H.P., L.E.P., P.S., B.V., M.K., F.B., K. Lin, X.G., W. Zhang, J.Y., Y.J.K., M. Graff, F.T., J.N., A. Lamri, M.N., S. Moon, R.A.S., J.P.C., J.-J.L., I.P., D.T., E.J.P., J.-F.C., L.F.B., Y.T., Y.H., G.T., N.G., T.S., M.W., C. Sarnowski, C.G.,

D.N., S.T., S.-H.K., J. Long, M. Sun, L.T., W.-M.C., S.S.N., R.N., V.J.Y.L., C.H.T.T., Y.Y.J., C.-H.C., L.M.R., B.P.P., A.N., L.R.Y., G.C., J.A.B., E.K., P.A., A.H.X., H.S.C., B.E.C., J. Tan, M.C.Y.N., X.S., A. Mahajan, C.N.S., A.P.M. Study-level sample collection, phenotyping, genotyping, and additional analysis: L.S.A., A.A., C.A.A.-S., T.S.A., S.S.A., A. Bertoni, J.B.-J., I.B., T.A.B., C.F.B., A.S.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, J. Danesh, 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., O.H.F., T.M.F., B.I.F., P.G., H.C.G., V.G., C.G.-V., M.E.G.-V., P.G.-L., M. Gross, L.A.G., S. Hackinger, S. Han, A.T.H., C.H., M. Horikoshi, 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., T.J., F.K.K., F.R.K., A. Kasturiratne, T. Katsuya, V.K., T. Kawaguchi, J.M.K., A.N.K., C.-C.K., M.G.K., D.-H.K., F.K., J.K., K. Läll, L.A.L., K.M. Lee, M.-S.L., N.R.L., A. Leong, L. Li, Y. Liu, R.L.-G., S. Lithgart, C.M.L., A. Linneberg, C.-T.L., J. Liu, A.E.L., T.L., J. Luan, A.O.L., X.L., J.Lv, J.A.L., V.L., S. Maeda, V.M., S.R.M., K.M., T.M., A. Metspalu, H.M., A.D.M., F.A.M., J.L.N., M.A.N., U.N., I.N., Y.O., L.O., S.R.P., S.P., P.P., M.A. Pereira, A.P., F.J.P., H.G.P., B.P., G. Prasad, L.J.R.-T., A.P.R., M.R., R.R., K.R., C. Sabanayagam, K. Sandow, A. Sankareswaran, N. Sattar, S.S., M. Shahriar, B.S., J.S., D.M.S., N. Shojima, J.A.S., W.Y.S., A. Stančáková, V.S., A.M.S., K. Strauch, K.D.T., B. Thorand, U.T., B. Tomlinson, T.C.T., F.-J.T., J. Tuomilehto, T.T.-L., M.S.U., A.V., R.M.v.D., J.B.v.K., R.V., N.W.-R., E.W., A.R.W., K.W.v.D., D.R.W., C.S.Y., Ken Yamamoto, Kenichi Yamamoto, K. Yoon, C.Y., J.-M.Y., S.Y., M.Z., L.Z., W. Zheng, J.S.P., T.-Y.W., F.S.C., M.O.G., K.L.M., J.C.F., J.E.B., M.I.M., M.V. Study-level principal investigator: L.J.R., M. Igase, E. Ipp, S.R., Y.S.C., L. Lind, M. A. Province, M.F., C.L.H., E. Ingelsson, A.B.Z., B.M.P., Y.-X.W., C.N.R., D.M.B., F.M., Y. Liu, M.Y., S.L.R.K., P.A.P., J.S.P., J.C.E., A. Bonnefond, P.F., J.G.W., W.H.H.S., J.-Y.W., M.G.H., R.C.W.M., T.-Y.W., D.O.M.-K., T.T., 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, Y.-D.I.C., S.S.R., R.M.-C., H.G., C.-Y.C., M.G., E.-S.T., J. Dupuis, N.K., M.L., A. Köttgen, W.-P.K., D.W.B., C.N.A.P., J.S.K., C.K., S. Liu, K.E.N., D.S., T.H., O.P., N.J.W., J. Lee, B.-J.K., I.Y.M., R.G.W., K. Stefansson, K.L.M., C.L., C.A.H., R.J.F.L., J.C.F., D.J.R., M.D.R., S.Z., R. Mägi, N.A.M., C.T.R., D.A.v.H., S.F., J.C.D., T.Y., T. Kadowaki, J.C.C., J.E.B., P.S.T., K.-M.C., M.I.M., J.B.M., J.M.M., M.B., J.I.R., M.V., B.F.V., A.P.M., E.Z. MAGIC analysis: K.A.B., A.W., K.L.M., C.L. ICBP analysis: Z.K. GUARDIAN Consortium analysis: J. Cui, M.O.G. GLP-1 GWAS analysis: M.T., L.H., O.M., T.T., E.A. DIAMANTE Consortium ancestry-specific data co-ordination: J.C.C., M.C.Y.N., X.S., J.E.B., A. Mahajan. Manuscript preparation: K. Suzuki, K.H., L.S., H.J.T., X.Y., K. M. Lorenz, R. Mandla, M.I.M., J.B.M., A. Mahajan, C.N.S., J.M.M., M.B., J.I.R., M.V., B.F.V., A.P.M., E.Z. T2DGGI steering group: M.I.M., J.B.M., A. Mahajan, C.N.S., J.M.M., M.B., J.I.R., M.V., B.F.V., A.P.M., E.Z.

Extended Data Fig. 1 | Axes of genetic variation separating GWASs of T2D across ancestry groups. We used SNVs that were reported in all studies to construct a distance matrix of mean effect allele frequency differences between each pair of GWASs. We implemented multi-dimensional scaling of the distance matrix to principal components that represent axes of genetic variation. The first three axes of genetic variation (PC1, PC2 and PC3) from multi-dimensional scaling of the Euclidean distance matrix between populations are sufficient to separate GWASs from six ancestry groups: African American (AFA), East Asian (EAS), European (EUR), Hispanic (HIS), South African (SAF), and South Asian (SAS). Variance explained by each axis: PC1 90.7%; PC2 6.5%; PC3 1.0%.

directed -log$_{10}$ P-value
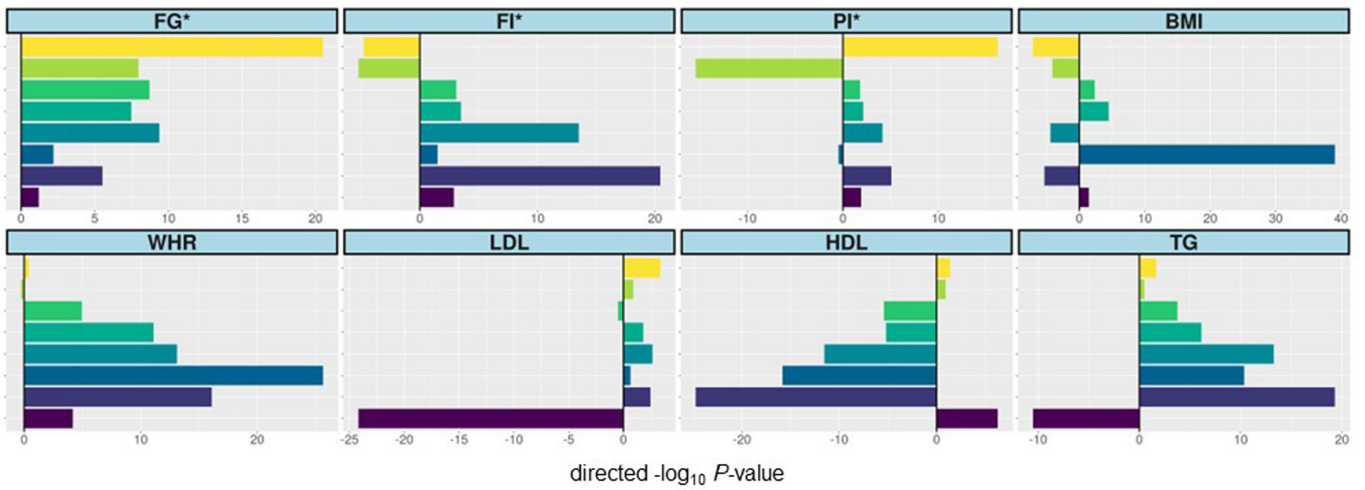
Legend:
- Beta cell +PI
- Beta cell -PI
- Residual glycaemic
- Body fat
- Metabolic syndrome
- Obesity
- Lipodystrophy
- Liver/lipid metabolism

Extended Data Fig. 2 | Cluster-specific associations of index SNVs with defining cardiometabolic phenotypes. Each bar presents the −log$_{10}$ P value for association, with effect direction aligned to the T2D risk allele. FG: fasting glucose. FI: fasting insulin. PI: proinsulin. BMI: body mass index. WHR: waist–hip ratio. LDL: low-density lipoprotein cholesterol. HDL: high-density lipoprotein cholesterol. TG: triglycerides. *Trait adjusted for BMI.
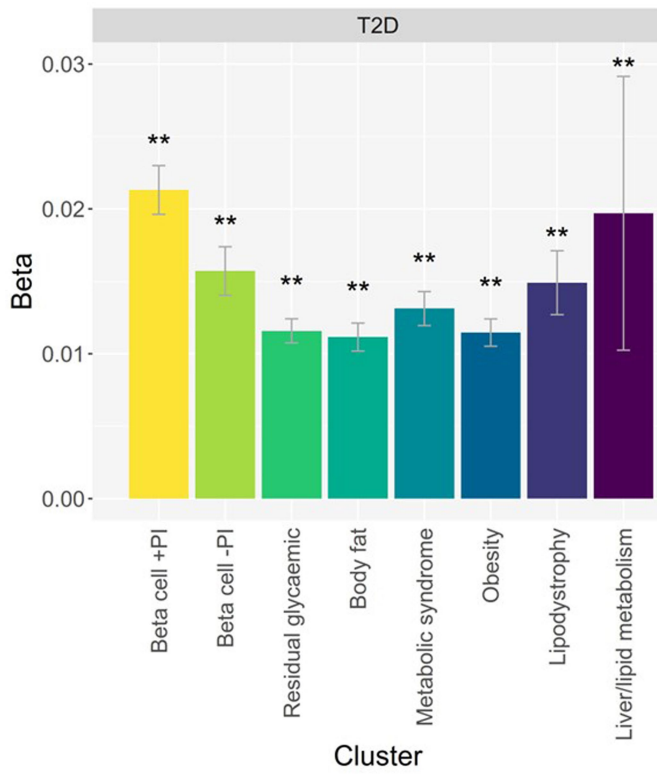
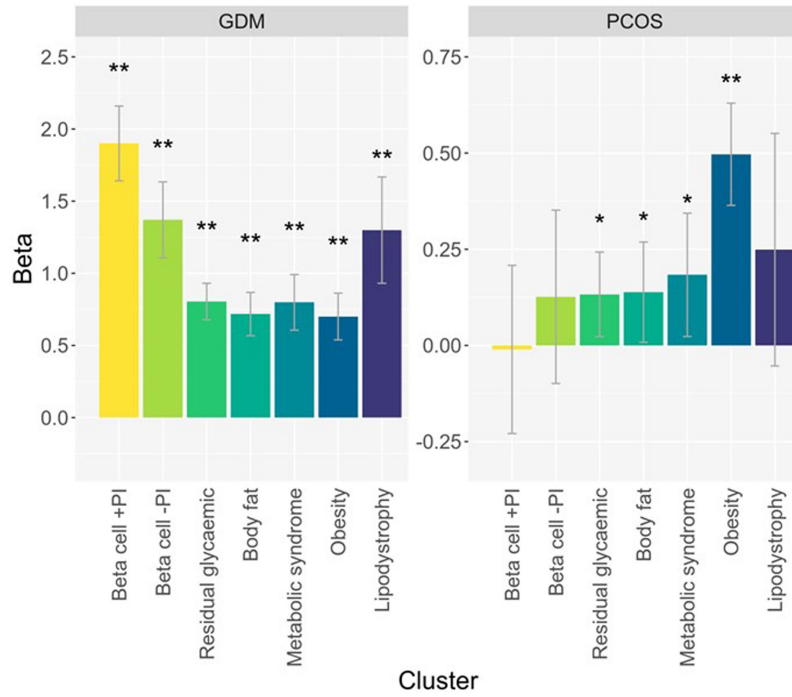**Extended Data Fig. 3 | Cluster-specific associations of index SNVs with T2D.**
The height of each bar corresponds to the log-odds ratio (beta), and the grey
bar shows the 95% confidence interval. *$P < 0.05$, nominal association.
**$P < 0.0063$, Bonferroni correction for eight clusters. Exact P values are
presented in Supplementary Table 9.
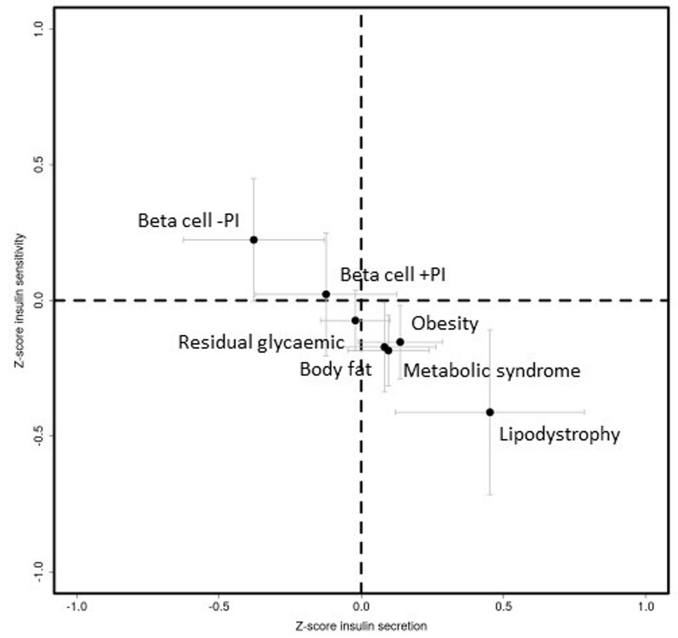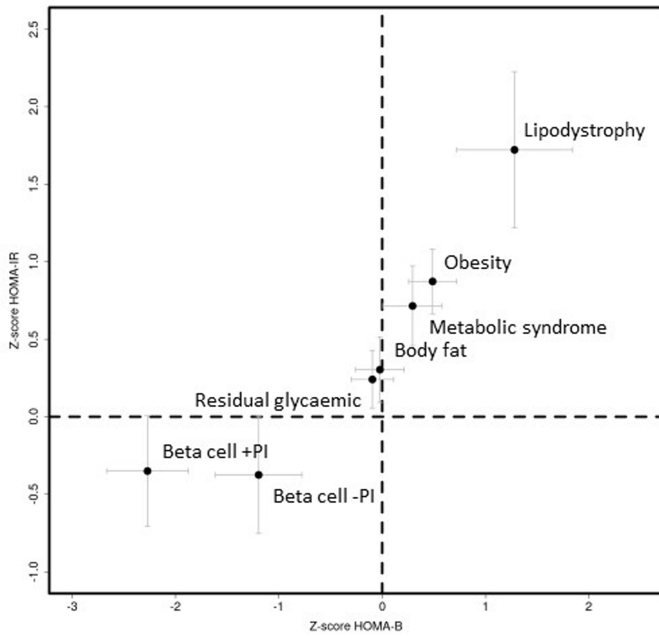
Extended Data Fig. 4 | Cluster-specific associations of T2D risk alleles at index SNVs with insulin-related endophenotypes. Measures of insulin secretion and insulin sensitivity were derived from hyperinsulinaemic-euglycaemic clamp assessments and oral glucose tolerance tests in up to 1,316 Mexican American participants without diabetes. Homeostatic model assessment measures of beta-cell function (HOMA-B) and insulin resistance (HOMA-IR) were obtained from 36,466 non-diabetic individuals of European ancestry. Each point corresponds to the cluster-specific mean z-score for each trait, and grey bars represent 95% confidence intervals. The liver and lipid metabolism cluster has been removed for ease of presentation.
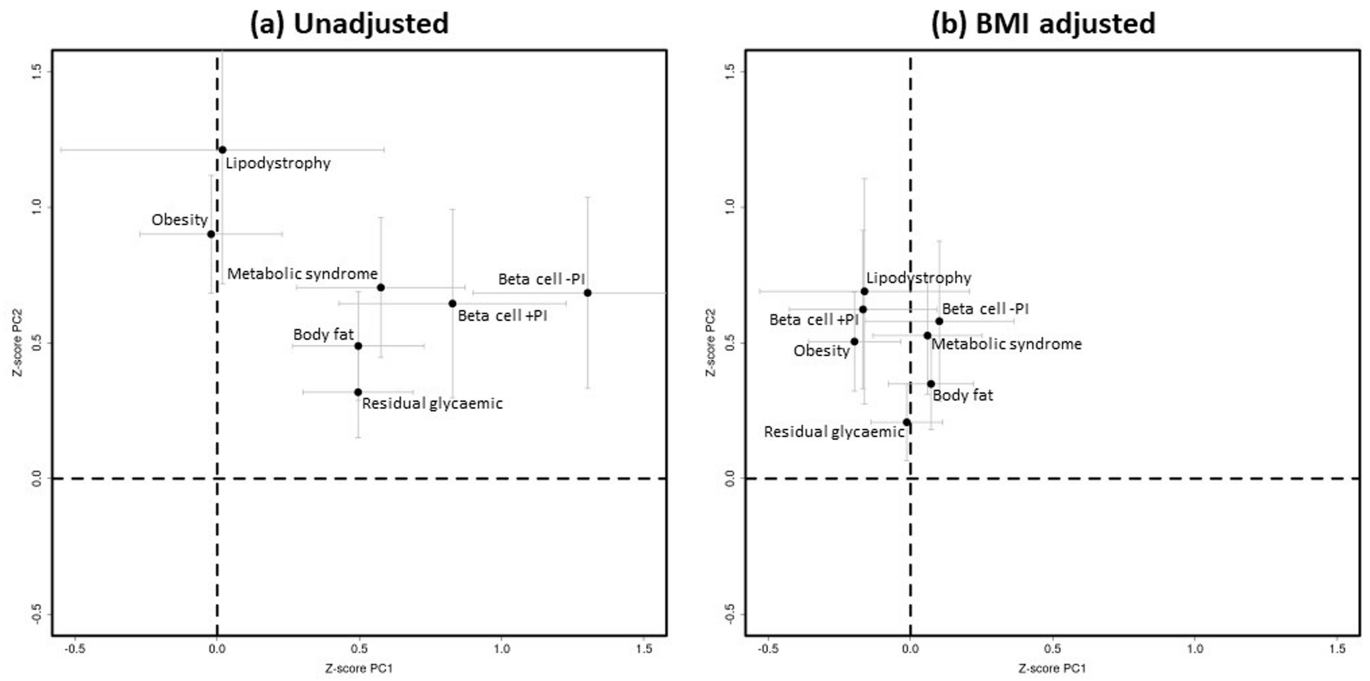
Extended Data Fig. 5 | Cluster-specific associations of T2D risk alleles at index SNVs with insulin-resistance-related disorders. Association with gestational diabetes mellitus (GDM) was assessed in 5,485 cases and 347,856 female controls of diverse ancestry. Association with polycystic ovary syndrome (PCOS) was assessed in 10,074 cases and 103,164 female controls of European ancestry. The height of each bar corresponds to the mean z-score, and the grey bar shows the 95% confidence interval. The liver and lipid metabolism cluster has been removed for ease of presentation. *P < 0.05, nominal association. **P < 0.0063, Bonferroni correction for eight clusters. Exact P values are presented in Supplementary Table 12.
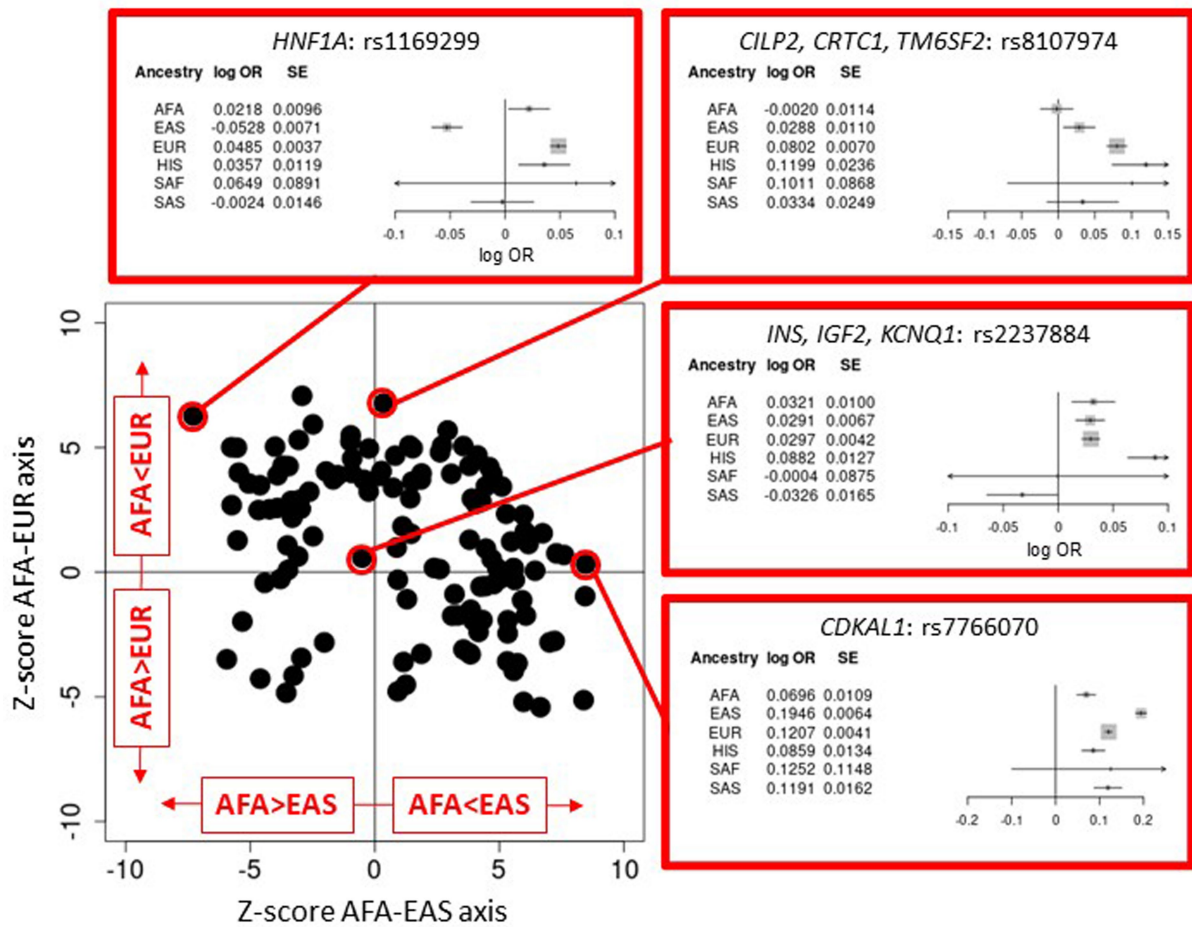
Extended Data Fig. 6 | Ancestry-correlated heterogeneity is driven by differences in allelic effect sizes between AFA, EAS and EUR ancestry groups. In the scatter plot, index SNVs with significant evidence ($P_{HET} < 3.9 \times 10^{-5}$, Bonferroni correction for 1,289 signals) for ancestry-correlated heterogeneity are plotted according to their association (z-score) with the first two axes of genetic variation. The first axis represents differences in allelic effect sizes between AFA/EUR GWASs and EAS GWASs (AFA–EAS axis), whilst the second axis represents differences in effect size between AFA/EAS GWASs and EUR GWASs (AFA–EUR axis). The forest plots present examples of ancestry-correlated heterogeneity at index SNVs. In each forest plot, the allelic log-odds ratio (OR) from each ancestry group-specific fixed-effects meta-analysis is given by the black tick mark, the 95% confidence interval is given by the horizontal line, and the weight (inverse-variance) of each ancestry group by the grey box. AFA: African American ancestry group. EAS: East Asian ancestry group. EUR: European ancestry group. HIS: Hispanic ancestry group. SAF: South African ancestry group. SAS: South Asian ancestry group.
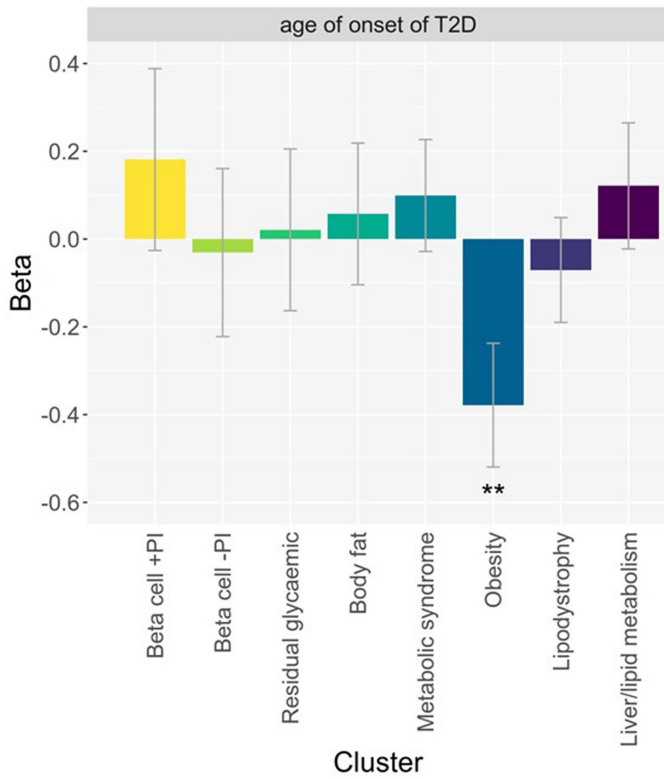
Extended Data Fig. 7 | Cluster-specific associations of index SNVs with the first two axes of genetic variation in T2D cases and controls. a, Unadjusted for BMI. **b**, Adjusted for study-level mean BMI. Each point corresponds to a cluster, plotted according to the mean z-score for association with the first two axes of genetic variation (PC1 and PC2) on the x axis and y axis, respectively. Grey bars correspond to 95% confidence intervals. The liver and lipid metabolism cluster has been removed for ease of presentation.
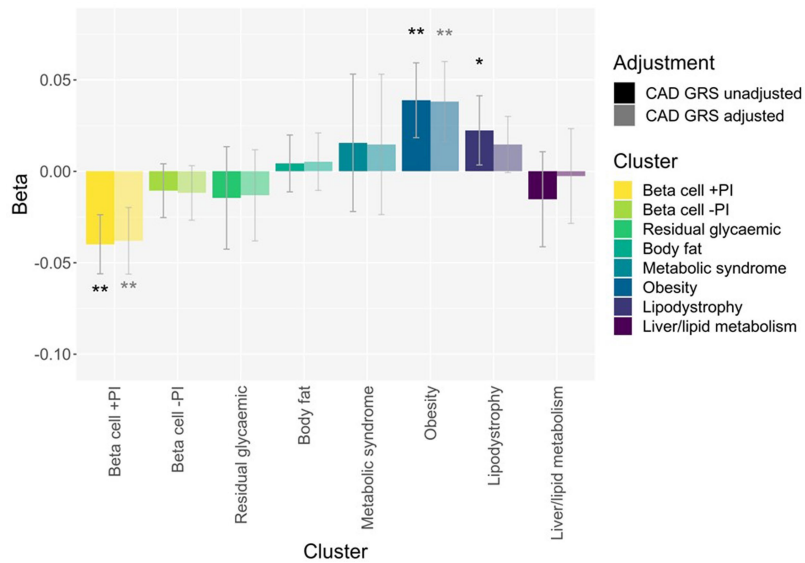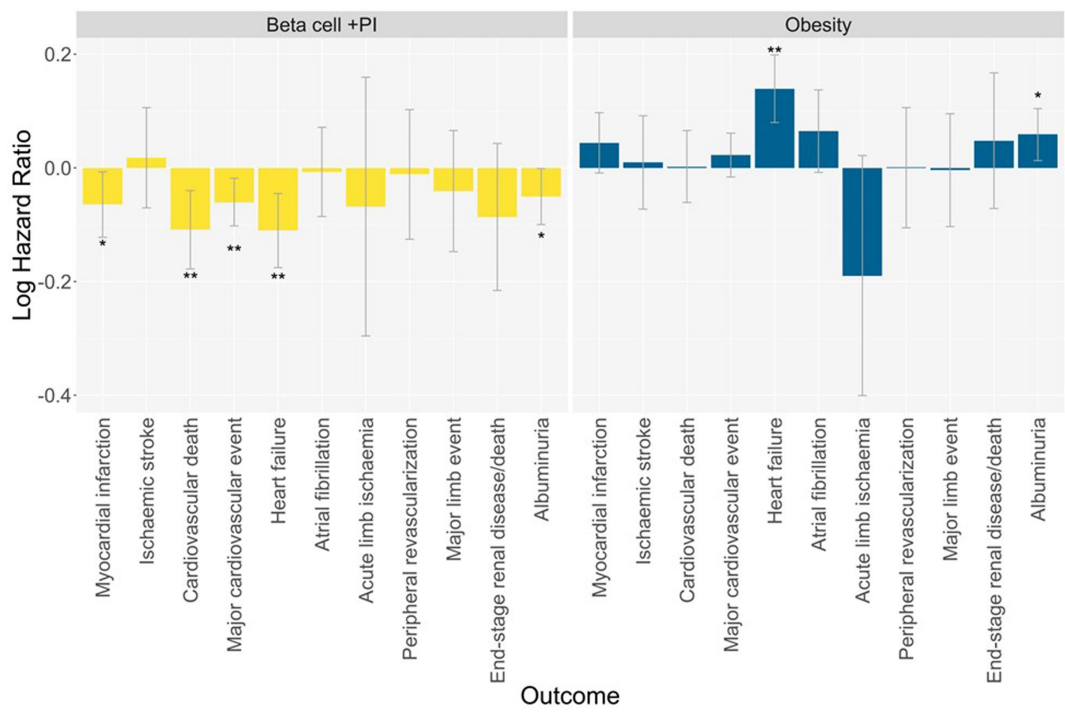
Extended Data Fig. 8 | Associations of cluster-specific components of the partitioned PS with CAD in up to 279,552 individuals across diverse ancestry groups. The panel summarizes the associations of each cluster-specific component of the partitioned PS with CAD, with and without adjustment for a previously published multi-ancestry CAD PS. The height of each bar corresponds to the log-OR (beta) per standard deviation of the PS, and the grey bar shows the 95% confidence interval. Analyses were undertaken in all individuals, with adjustment for T2D status. *$P < 0.05$, nominal association. **$P < 0.0063$, Bonferroni correction for eight clusters. Exact P values are presented in Supplementary Tables 21 and 22.

Extended Data Fig. 9 | Associations of cluster-specific components of the partitioned PS with T2D age of onset in up to 30,288 individuals across diverse ancestry groups. The panel summarizes the associations of each cluster-specific component of the partitioned PS with age of onset. The height of each bar corresponds to years (beta) per standard deviation of the PS, and the grey bar shows the 95% confidence interval. A negative effect corresponds to earlier age of onset. *$P < 0.05$, nominal association. **$P < 0.0063$, Bonferroni correction for eight clusters. Exact P values are presented in Supplementary Table 23.

Extended Data Fig. 10 | Associations of the beta cell +PI and obesity cluster-specific components of the partitioned PS with vascular outcomes in up to 29,827 EUR individuals with T2D from six clinical trials from the TIMI Study Group. Major cardiovascular event is defined as myocardial infarction, ischaemic stroke, or cardiovascular death. Major limb event is defined as acute limb ischaemia or peripheral revascularization. The height of each bar corresponds to the log-hazard ratio per standard deviation of the PS, and the grey bar shows the 95% confidence interval. *$P < 0.05$, nominal association. **$P < 0.0063$, Bonferroni correction for eight clusters. Exact P values are presented in Supplementary Table 24.

Corresponding author(s): Andrew Morris, Konstantinos Hatzikotoulas, Benjamin Voight, Eleftheria Zeggini

Last updated by author(s): Nov 17, 2023

Reporting Summary

Nature Portfolio 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 Portfolio 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 were used for data collection.
Data analysis	Analyses were conducted using publicly available software: UCSC liftOver tool (https://genome.ucsc.edu/cgi-bin/hgLiftOver), MR-MEGA v0.2 (https://genomics.ut.ee/en/tools), METAL v2011-03-25 (https://genome.sph.umich.edu/wiki/METAL), PLINKv1.9 (https://www.cog-genomics.org/plink/1.9/), Beagle 4.1 (https://faculty.washington.edu/browning/beagle/b4_1.html), SNPTEST v2.5.6 (https://www.well.ox.ac.uk/~gav/snptest/), GWAMA v2.2.2 (https://genomics.ut.ee/en/tools), EIGENSOFT v7.2.1 (https://www.hsph.harvard.edu/alkes-price/software/), PLINKv2.0 (https://www.cog-genomics.org/plink/2.0/), SHAPEIT4 (https://odelaneau.github.io/shapeit4/), Minimac4 (https://genome.sph.umich.edu/wiki/Minimac4), KING v2.3 (https://www.kingrelatedness.com/), and EAGLE v2.4 (https://alkesgroup.broadinstitute.org/Eagle/#Xeagle2). Analyses were also conducted using the following R packages: meta (https://cran.r-project.org/package=meta), ClustImpute (https://cran.r-project.org/package=ClustImpute), NbClust (https://cran.r-project.org/package=NbClust), factoextra (https://cran.r-project.org/package=factoextra), and logistf (https://cran.r-project.org/package=logistf).

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 Portfolio guidelines for submitting code & software for further information.

Data

Data availability. Genome-wide association summary statistics from the multi-ancestry meta-analysis and ancestry-specific meta-analyses reported in this study are available through the DIAGRAM Consortium website (http://www.diagram-consortium.org/downloads.html).

Human research participants

Reporting on sex and gender	The numbers of males and females for each contributing study are reported in Supplementary Table 3. Sex-stratified analyses were not conducted.
Population characteristics	Characteristics are presented for each contributing study in Supplementary Table 3.
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.

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	Our discovery GWAS meta-analysis and polygenic score test GWAS brought together the largest sample size of type 2 diabetes cases and (population) controls that was available to the Type 2 Diabetes Global Genomics Initiative. Discovery GWAS meta-analysis. With our sample size of 428,452 T2D cases and 2,107,149 controls, at a genome-wide significance threshold ($p<5\times10^{-8}$), under an additive genetic model of homogeneous effects across ancestry groups, we had ≥80% power to detect association of SNVs with MAF ≥5% and OR ≥1.035 or MAF ≥0.5% and OR ≥1.107. Polygenic score test GWAS. We aggregated 30,288 T2D cases and 249,264 controls from the All of Us Research Program, Biobank Japan, and Genes & Health, who were not included as part of the discovery meta-analysis. We also consider 29,827 individuals with T2D from six clinical trials from the Thrombolysis in Myocardial Infarction (TIMI) Study Group.
Data exclusions	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.
Replication	We did not conduct a formal replication analysis since we had already brought together all GWAS data available to the Type 2 Diabetes Global Genomics Initiative. The polygenic score test GWAS were not used to replicate association signals from the discovery meta-analysis because sample overlap can lead to increased false positive error rates in polygenic score analyses. All reported association signals from the discovery meta-analysis were checked to confirm that effects were not driven by false positives in single studies.
Randomization	Randomization was not performed. Within each study, covariates were adjusted for to account for potential confounding. Covariate adjustments are reported in Supplementary Table 3.
Blinding	Group allocation was not relevant to this study, so blinding was not necessary.

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
☒ ☐	Clinical data
☒ ☐	Dual use research of concern

Methods

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