ARTICLES
https://doi.org/10.1038/s41588-021-00852-9
nature genetics

The trans-ancestral genomic architecture of glycemic traits

Glycemic traits are used to diagnose and monitor type 2 diabetes and cardiometabolic health. To date, most genetic studies of glycemic traits have focused on individuals of European ancestry. Here we aggregated genome-wide association studies comprising up to 281,416 individuals without diabetes (30% non-European ancestry) for whom fasting glucose, 2-h glucose after an oral glucose challenge, glycated hemoglobin and fasting insulin data were available. Trans-ancestry and single-ancestry meta-analyses identified 242 loci (99 novel; $P < 5 \times 10^{-8}$), 80% of which had no significant evidence of between-ancestry heterogeneity. Analyses restricted to individuals of European ancestry with equivalent sample size would have led to 24 fewer new loci. Compared with single-ancestry analyses, equivalent-sized trans-ancestry fine-mapping reduced the number of estimated variants in 99% credible sets by a median of 37.5%. Genomic-feature, gene-expression and gene-set analyses revealed distinct biological signatures for each trait, highlighting different underlying biological pathways. Our results increase our understanding of diabetes pathophysiology by using trans-ancestry studies for improved power and resolution.

Fasting glucose (FG), 2-h glucose after an oral glucose challenge (2hGlu), and glycated hemoglobin (HbA1c) are glycemic traits that are used to diagnose diabetes[1]. In addition, HbA1c is the most commonly used biomarker to monitor glucose control in patients with diabetes. Fasting insulin (FI) reflects a combination of insulin secretion and insulin resistance, both of which are components of type 2 diabetes (T2D); it also reflects insulin clearance[2]. Collectively, all four glycemic traits are useful to better understand T2D pathophysiology[3–5] and cardiometabolic outcomes[6].

To date, genome-wide association studies (GWAS) and analyses of Metabochip and exome arrays have identified more than 120 loci associated with glycemic traits in individuals without diabetes[7–15]. However, despite considerable differences in the prevalence of T2D risk factors across ancestries[16–18], most GWAS of glycemic traits have insufficient representation of individuals of non-European ancestry. Additionally, they have limited resolution for fine-mapping of causal variants and for the identification of effector transcripts. Here we present large-scale trans-ancestry meta-analyses of GWAS for four glycemic traits in individuals without diabetes. We aimed to identify additional glycemic-trait-associated loci; investigate the portability of loci and genetic scores across ancestries; leverage differences in effect allele frequency (EAF), effect size and linkage disequilibrium (LD) across diverse populations to conduct fine-mapping and aid the identification of causal variants and/or effector transcripts; and compare the genetic architecture of glycemic traits to further identify the cell types and target tissues that are influenced the most by the traits that inform T2D pathophysiology.

Results

Study design and definitions. To identify loci associated with glycemic traits (FG, 2hGlu, FI and HbA1c), we aggregated GWAS in up to 281,416 individuals without diabetes, approximately 30% of whom were of non-European ancestry (13% East Asian, 7% Hispanic, 6% African American, 3% South Asian and 2% sub-Saharan African (Ugandan data were only available for HbA1c)). Each cohort imputed data to the 1000 Genomes Project reference panel[19] (phase 1 v.3, March 2012 or later; Methods, Supplementary Table 1, Extended Data Fig. 1, Supplementary Note). Up to around 49.3 million variants were directly genotyped or imputed, with between 38.6 million (2hGlu) and 43.5 million variants (HbA1c) available for analysis after exclusions based on minor allele count (MAC) < 3 and imputation quality (imputation r^2 or INFO score < 0.40) in each cohort. FG, 2hGlu and FI analyses were adjusted for body-mass index (BMI)[15] but for simplicity they are abbreviated as FG, 2hGlu and FI (Methods).

We first performed trait-specific fixed-effect meta-analyses within each ancestry using METAL[20] (Methods). We defined 'single-ancestry lead' variants as the strongest trait-associated variants ($P < 5 \times 10^{-8}$) within a 1 Mb region in an ancestry (Table 1). Within each ancestry and each autosome, we used approximate conditional analyses in genome-wide complex trait analysis (GCTA)[21,22] to identify 'single-ancestry index variants' ($P < 5 \times 10^{-8}$) that exert conditionally distinct effects on the trait (Table 1, Methods, Supplementary Note). This approach identified 124 FG, 15 2hGlu, 48 FI and 139 HbA1c variants that were significant in at least one ancestry (Supplementary Table 2).

Next, we conducted trait-specific trans-ancestry meta-analyses using MANTRA (Methods, Supplementary Table 1, Supplementary Note) to identify genome-wide significant 'trans-ancestry lead variants', defined as the most-significant trait-associated variant across all ancestries ($\log_{10}$[Bayes factor (BF)] > 6, equivalent to $P < 5 \times 10^{-8}$)[23] (Table 1, Methods). Here, we present trans-ancestry results as our primary results (Supplementary Table 2).

Causal variants are expected to affect related glycemic traits and may be shared across ancestries. Therefore, we combined all single-ancestry lead variants, single-ancestry index variants and/or trans-ancestry lead variants (for any trait) mapping within 500 kb of each other into a single 'trans-ancestry locus' bounded by 500 kb flanking sequences (Table 1, Extended Data Fig. 2). As defined in Table 1, a trans-ancestry locus may contain multiple causal variants that affect one or more glycemic traits, exerting their effect in one or more ancestry.

Glycemic trait locus discovery. Trans-ancestry meta-analyses identified 235 trans-ancestry loci, of which 59 contained lead variants for more than one trait. In addition, we identified seven 'single-ancestry loci' that did not contain any trans-ancestry lead variants (Table 1, Supplementary Table 2). Of the 242 combined loci, 99 (including 6 of the 7 single-ancestry loci) had not previously been associated with any of the four glycemic traits or with T2D at

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

Table 1 | Glossary of terms

Term	Definition
Effect allele	The effect allele was the allele defined by METAL based on trans-ancestry FG results and aligned such that the same allele was kept as the effect allele across all ancestries and traits, irrespective of its allele frequency or effect size for that particular ancestry and trait. In this way, the effect allele is not necessarily the trait-increasing allele.
Single-ancestry lead variant	The variant with the smallest *P* value among all variants with $P<5\times10^{-8}$ within a 1 Mb region, based on the analysis of a single trait in a single ancestry.
Single-ancestry index variants	Variants identified by GCTA of each autosome as exerting conditionally distinct effects on a given trait in a given ancestry ($P<5\times10^{-8}$). As defined, these include the single-ancestry lead variants.
Trans-ancestry lead variant	The variant identified by trans-ancestry meta-analysis of a given trait that has the strongest association for that trait ($\log_{10}$[BF] > 6, which is broadly equivalent to $P<5\times10^{-8}$) within a 1 Mb region.
Single-ancestry locus	The 1 Mb region centered on a single-ancestry lead variant that does not contain a lead variant identified in the trans-ancestry meta-analysis (that is, does not contain a trans-ancestry lead variant).
Signal	Conditionally independent association between a trait and a set of variants in LD with each other and that is noted by the corresponding index variant.
Trans-ancestry locus	A genomic interval that contains trans-ancestry trait-specific lead variants, with or without additional single-ancestry index variants, for one or more traits. This region is defined by starting at the telomere of each chromosome and selecting the first single-ancestry index variant or trans-ancestry lead variant for any trait. If other trans-ancestry lead variants or single-ancestry index variants mapped within 500 kb of the first signal, they were merged into the same locus. This process was repeated until there were no more signals within 500 kb of the previous variant. A 500 kb interval was added to the beginning of the first signal, and the end of the last signal to establish the final boundary of the trans-ancestry locus (Extended Data Fig. 2). As defined, a trans-ancestry locus may not have a single lead trans-ancestry variant, but may instead contain multiple trans-ancestry lead variants, one for each trait.

This study combined analyses of trait associations across multiple correlated glycemic traits and across multiple ancestries, which has presented challenges in our ability to apply commonly used terms with clarity. For this reason, we define terms often used in the field with variable meaning and provide definitions for new terms used in this study.

the time of analysis (Fig. 1, Supplementary Table 3, Supplementary Note). However, based on recent East Asian and trans-ancestry T2D GWAS meta-analyses[23–27], the lead variants at 27 of the 99 novel glycemic trait loci have strong evidence of association with T2D ($P<10^{-4}$; 13 loci with $P<5\times10^{-8}$), suggesting that they are also important in T2D pathophysiology (Supplementary Tables 2 and 4).

Of the six single-ancestry novel loci, three were unique to individuals of non-European ancestry (Supplementary Table 3). An association with individuals of African American ancestry for FI (lead variant rs12056334) near LOC100128993 (an uncharacterized RNA gene; Supplementary Note), an association with individuals of African American ancestry for FG (lead variant rs61909476) near *ETS1* and an association with individuals of Hispanic descent for FG (lead variant rs12315677) within *PIK3C2G* (Supplementary Table 3) were found. Despite broadly similar EAFs across ancestries, rs61909476 was significantly associated with FG only in individuals of African American descent (EAF ≈ 7%, $\beta=0.0812$ mmol l^{-1}, s.e. = 0.01 mmol l^{-1}, $P=3.9\times10^{-8}$ compared with EAF = 10–17%, $\beta=0$–0.002 mmol l^{-1}, s.e. = 0.003–0.017 mmol l^{-1}, $P=0.44$–0.95 in all other ancestries; Supplementary Table 2, Supplementary Note). The nearest protein-coding gene, *ETS1*, encodes a transcription factor that is expressed in mouse pancreatic β-cells, and its overexpression decreases glucose-stimulated insulin secretion in mouse islets[28]. Located within the *PIK3C2G* gene, rs12315677 has an 84% EAF in individuals of Hispanic descent (70–94% in other ancestries) and is significantly associated with FG in this ancestry alone ($\beta=0.0387$ mmol l^{-1}, s.e. = 0.0075 mmol l^{-1}, $P=4.0\times10^{-8}$ compared with $\beta=-0.0128$–0.010 mmol l^{-1}, s.e. = 0.003–0.018 mmol l^{-1}, $P=0.14$–0.76 in all other ancestries; Supplementary Note). In mice, deletion of *Pik3c2g* leads to a phenotype characterized by reduced glycogen storage in the liver, hyperlipidemia, adiposity and insulin resistance with increasing age or after a high-fat diet[29]. Instances of similar EAFs but differing effect sizes between populations could be due to genotype-by-environment or other epistatic effects. Alternatively, lower imputation accuracy in smaller sample sizes could deflate effect sizes, although the imputation quality for these variants was good (average $r^2=0.81$). Finally, the variants detected here may be in LD with ancestry-specific causal variants that were not investigated here that differ in frequency across ancestries. However, we could not find evidence of rarer alleles in the cognate populations from the 1000 Genomes Project (Supplementary Table 5). The final three single-ancestry loci were identified in individuals of European ancestry (Supplementary Note).

Next, by rescaling the standard errors of allelic effect sizes to artificially boost the sample size of the European meta-analysis to match that of trans-ancestry meta-analysis, we determined that 21 of the novel trans-ancestry loci would not have been discovered with an equivalent sample size that consisted exclusively of individuals of European ancestry (Supplementary Note). Their discovery was due to the higher EAF and/or larger effect size in populations of non-European ancestry. In particular, two loci (near LINC00885 and MIR4278) contain single-ancestry lead variants associated with East Asian and African American ancestry, respectively, suggesting that these specific ancestries may be driving the trans-ancestry discovery (Supplementary Tables 2,3). Combined with the three single-ancestry non-European loci described above, our results show that 24% (24 out of 99) of the novel loci were discovered due to the contribution of participants of non-European ancestry, strengthening the argument for expanding genetic studies in diverse populations.

Allelic architecture of glycemic traits. Single-ancestry and trans-ancestry results combined increased the number of established loci for FG to 102 (182 signals, 53 novel loci), FI to 66 (95 signals, 49 novel loci), 2hGlu to 21 (28 signals, 11 novel loci) and HbA1c to 127 (218 signals, 62 novel loci) (Supplementary Table 2), with considerable overlap across traits (Extended Data Fig. 3). We also detected ($P<0.05$ or $\log_{10}$[BF] > 0) most (around 90%) of the previously established glycemic signals, 70–88% of which attained genome-wide significance (Supplementary Note, Supplementary Table 6). Given that analyses for FG, FI and 2hGlu were performed adjusted for BMI, we confirmed that collider bias did not influence more than 98% of discovered signals[30] (Supplementary Note). As expected, given the greater power due to increased sample sizes,

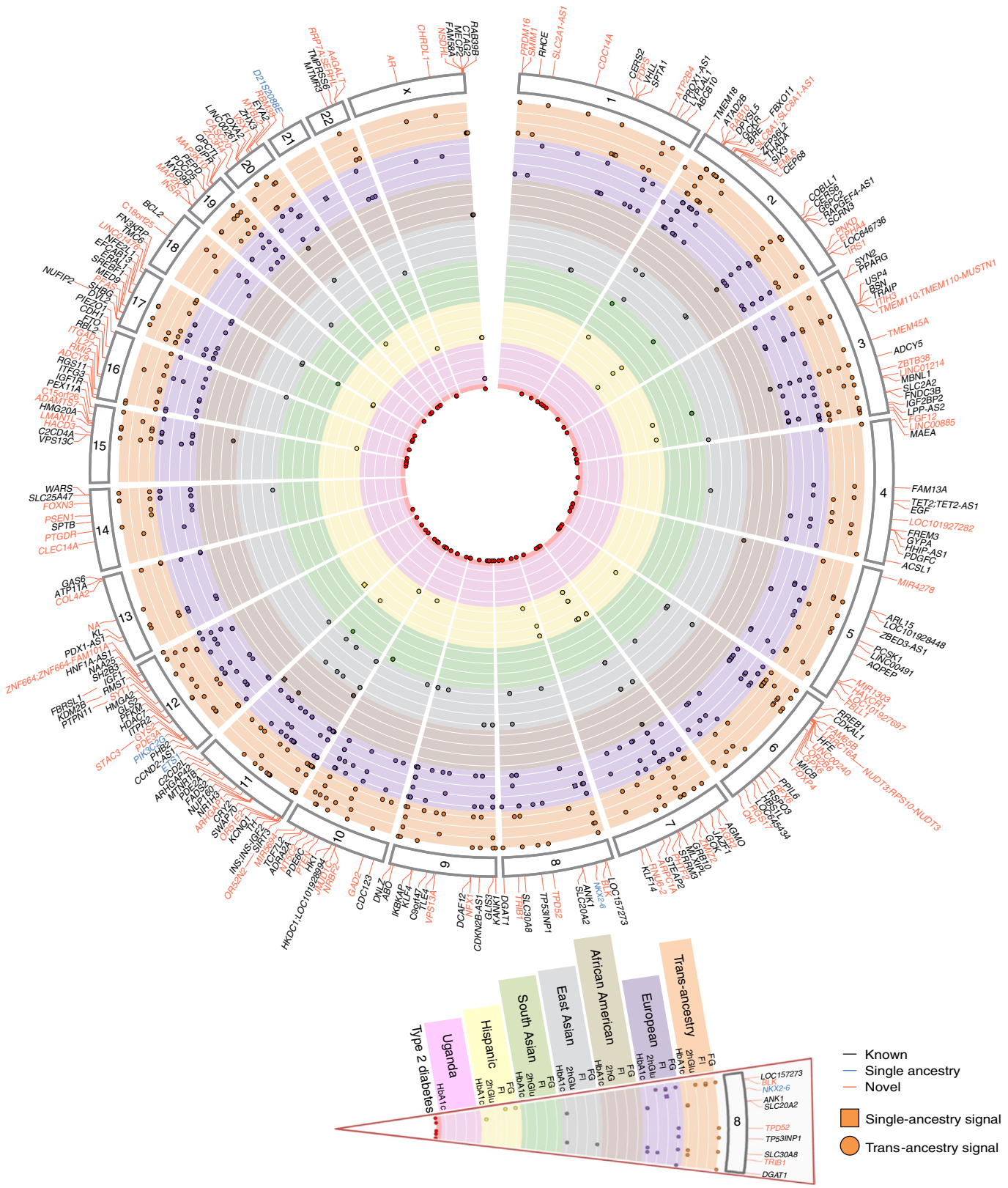


Fig. 1 | Summary of all 242 loci identified in this study. The 235 trans-ancestry loci are shown in orange (novel) or black (established) along with seven single-ancestry loci (blue) represented by the nearest gene. Each locus is mapped to the corresponding chromosome (outer segment). Each set of rows shows the results from the trans-ancestry analysis (orange) and each of the ancestries: European (purple), African American (tan), East Asian (gray), South Asian (green), Hispanic (yellow), Ugandan (pink). Loci with a corresponding signal associated with T2D are represented by red circles in the middle of the plot. *TMEM110* is also known as *STIMATE*; *FAM101A* is also known as *RFLNA*; *PDX1-AS1* is also known as *PLUT*; *LRRC16A* is also known as *CARMIL1*; *FAM65B* is also known as *RIPOR2*; C15orf26 is also known as *CFAP161*; *FAM58A* is also known as *CCNQ*; *IKBKAP* is also known as *ELP1*; *AQPEP* is also known as *LVRN*; WARS is also known as *WARS1*; *ITFG3* is also known as *FAM234A*; *BRE* is also known as *BABAM2*; *NA* is also known as *XK*.

new association signals tended to have smaller effect sizes and/or EAFs in individuals of European ancestry compared with established signals (Extended Data Fig. 4).

Characterization of lead variants across ancestries. To better understand the transferability of trans-ancestry lead variants across ancestries, we investigated the pairwise EAF correlation and the pairwise summarized heterogeneity of effect sizes between ancestries[31] (Methods, Supplementary Note). Consistent with population history and evolution, these results demonstrated considerable EAF correlation ($\rho^2 > 0.70$) between populations of European and Hispanic, European and South Asian, and Hispanic and South Asian ancestry, which was consistent across all four traits, and between individuals of African American and Ugandan descent for HbA1c (Extended Data Fig. 5). Despite high EAF correlations, some pairwise comparisons exhibited strong evidence for effect size heterogeneity between ancestries that was less consistent between traits (Extended Data Fig. 5). However, sensitivity analyses demonstrated that, across all comparisons, the evidence for heterogeneity is driven by a small number of variants, with between 81.5% (for HbA1c) and 85.7% (for FG) of trans-ancestry lead variants showing no evidence for trans-ancestry heterogeneity ($P > 0.05$) (Supplementary Note).

Trait variance explained by associated loci. The trait variance explained by genome-wide significant loci was assessed using only the single-ancestry variants or a combination of single-ancestry and trans-ancestry variants (Supplementary Table 7) with β values extracted from the relevant single-ancestry meta-analysis results (Methods). The variance explained was assessed by linear regression in a subset of the contributing cohorts (Methods, Supplementary Tables 8–11). In general, the approach that explained the most variance was one in which trans-ancestry lead variants that had $P < 0.1$ in the relevant single-ancestry meta-analysis were combined with single-ancestry variants that were not in LD with the trans-ancestry variants (LD $r^2 < 0.1$) (Fig. 2, list C in Supplementary Tables 8–11). With this approach, the mean variance in the trait distribution explained was between 0.7% (2hGlu in European ancestry) and 6% (HbA1c in African American ancestry). The European-based estimates explained more variance relative to previous estimates of 2.8% for FG and 1.7% for HbA1c[32] (Supplementary Note).

Transferability of European-ancestry-derived polygenic scores. To investigate the transferability of polygenic scores across ancestries we used the PRS-CSauto software[33] to first build polygenic scores (PGSs) for each glycemic trait based on the data from individuals of European ancestry. However, the training set for 2hGlu was too small; therefore, this trait was excluded. To build the PGSs, for each trait we first removed five of the largest European cohorts from the European ancestry meta-analysis. These five cohorts were meta-analyzed and used as our European ancestry test dataset, for each trait. The remaining European ancestry cohorts were also meta-analyzed and used as the training dataset, from which we derived a PGS for each trait (Methods). We used PRS-CSauto to revise the effect size estimates for the variants in the score (obtained from the training European datasets) based on the LD of the test population. PRS-CSauto does not have LD reference panels for South Asian or Hispanic ancestry and as such we were unable to test the transferability of the PGS to those populations. The 'gtx' package[34] (Methods) was used to obtain the R^2 for each test population (Fig. 3, Supplementary Table 12). Consistent with other complex traits[35], the European-ancestry-derived PGS had greater predictive power for test data of individuals of European ancestry than for data from other ancestry groups.

Fine-mapping. We fine-mapped, 231 trans-ancestry and six single-ancestry autosomal loci (Supplementary Table 2, Supplementary Note). Using FINEMAP with ancestry-specific LD and an average LD matrix across ancestries, we conducted fine-mapping both within (161 loci with single-ancestry lead variants) and across ancestries (231 loci) for each trait (Methods). Because 59 of the 231 trans-ancestry loci were associated with more than one trait, we conducted trans-ancestry fine-mapping for a total of 305 locus–trait associations. Of these 305 locus–trait combinations, FINEMAP estimated the presence of a single causal variant at 186 loci (61%), whereas multiple distinct causal variants were implicated at 126 loci (39%), for a total of 464 causal variants (Fig. 4a).

Credible sets for causal variants. At each locus, we next constructed credible sets (CSs) for each causal variant that account for at least 99% of the posterior probability of association (PPA). We identified 21 locus–trait associations (at 19 loci) for which the 99% CS included a single variant and we highlight four examples (Fig. 4b, Methods, Supplementary Note, Supplementary Table 13).

At *MTNR1B* and *SIX3* we identified, respectively, rs10830963 (PPA > 0.999, for both HbA1c and FG) and rs12712928 (PPA = 0.997, for FG) as the likely causal variants. Previous studies confirm for both loci that these variants affect transcriptional activity[36–38] (Supplementary Note). At a locus near *PFKM* associated with HbA1c, trans-ancestry fine-mapping identified rs12819124 (PPA > 0.999) as the likely causal variant. This variant has previously been associated with mean corpuscular hemoglobin[39], suggesting an effect on HbA1c through red blood cells (RBCs; Supplementary Note). At *HBB*, we identified rs334 (PPA > 0.999; Glu7Val) as the likely causal variant associated with HbA1c. rs334 is a causal variant of sickle-cell anemia[40], was previously associated with urinary albumin-to-creatinine ratio in individuals of Caribbean Hispanic ancestry[41], severe malaria in a study with a population of Tanzanian ancestry[42], hematocrit and mean corpuscular volume in populations of Hispanic/Latino descent[43] and RBC distribution in individuals of Ugandan ancestry[44]; all of these results point to a variant effect on HbA1c through non-glycemic pathways.

The remaining locus–trait associations with a single variant in the 99% CS (Supplementary Table 13) point to variants that could be prioritized for functional follow-up to elucidate the effect on glycemic trait physiology.

At an additional 156 locus–trait associations, trans-ancestry fine-mapping identified 99% CSs with 50 or fewer variants (Fig. 4b, Supplementary Table 13). Consistent with the potential for more than 1 causal variant in a locus, 74 locus–trait associations contained 88 variants with PPA > 0.90 that were strong candidate causal variants (Supplementary Table 14). For example, 10 are coding variants including several missense variants, such as the *HBB* Glu7Val variant mentioned above, *GCKR* Leu446Pro, *RREB1* Asp1771Asn, *G6PC2* Pro324Ser, *GLP1R* Ala316Thr and *TMPRSS6* Val736Ala, each of which have been proposed or shown to affect gene function[12,45–49]. We additionally identified *AMPD3* Val311Leu (PPA = 0.989) and *TMC6* Trp125Arg (PPA > 0.999) variants associated with HbA1c that were previously detected in an exome array analysis but had not been fine-mapped with certainty due to the absence of backbone GWAS data[50]. Our fine-mapping data now suggest that these variants are likely causal and identify their cognate genes as effector transcripts.

Finally, we evaluated the resolution obtained in the trans-ancestry versus single-ancestry fine-mapping (Methods, Supplementary Note). We compared the number of variants in 99% CS across 98 locus–trait associations that—as suggested by FINEMAP—had a single causal variant in both trans-ancestry and single-ancestry analyses. Fine-mapping within and across ancestries was conducted using the same set of variants. At 8 out of 98 locus–trait associations, single-ancestry fine-mapping identified a single variant in the CSs. In addition, at 72 of the 98 locus–trait associations, the number of variants in the 99% CSs was smaller in the trans-ancestry

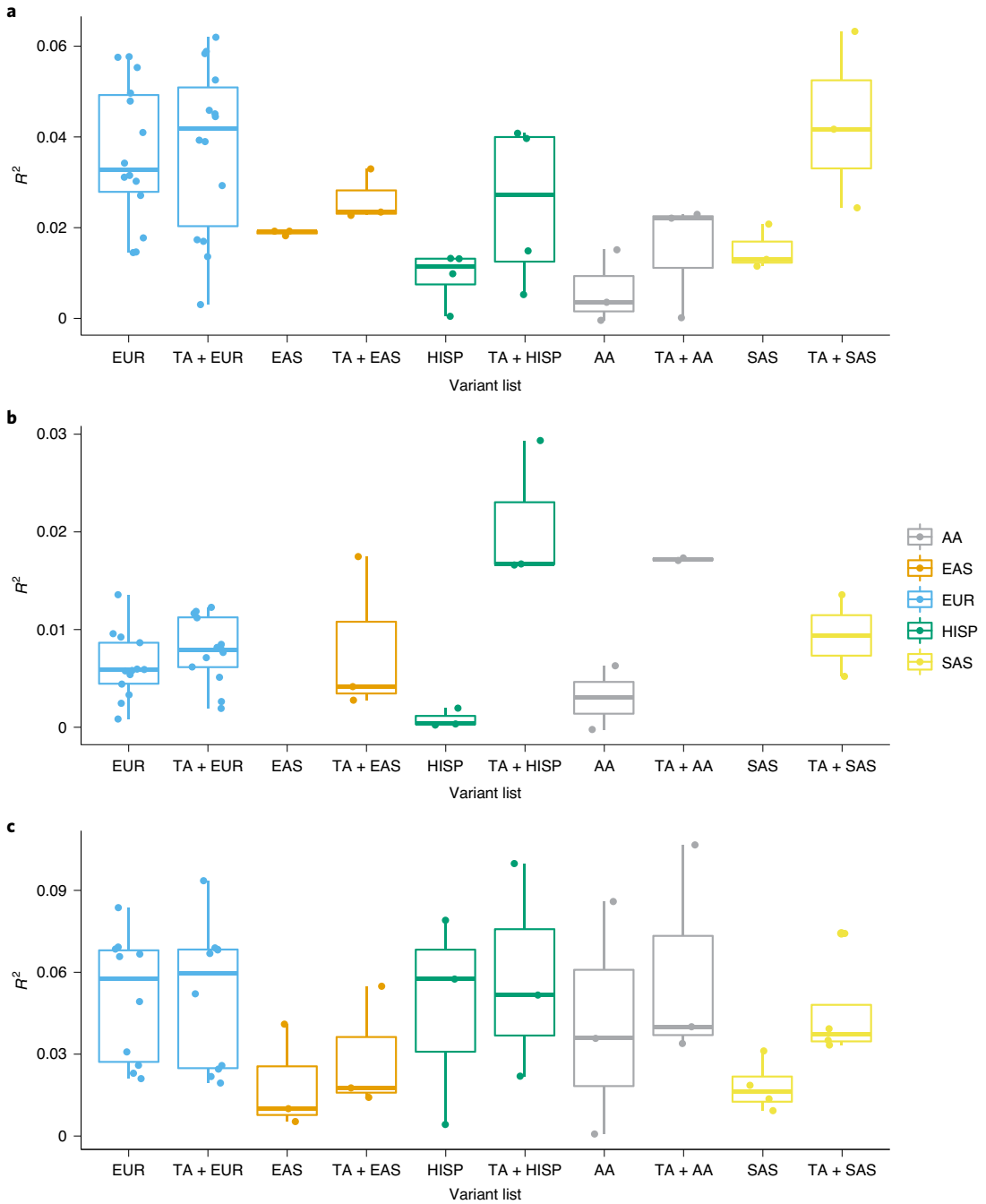


Fig. 2 | Trait variance explained by associated loci. a–c, Results from an analysis of trait variance explained by associated loci for FG (**a**), FI (**b**) and HbA1c (**c**). The box plots show the maximum, first quartile, median, third quartile and minimum of trait variance explained when using a genetic score with single-ancestry lead and index variants (European (EUR), African American (AA), East Asian (EAS), Hispanic (HISP) and Southeast Asian (SAS) ancestry) or a combination of trans-ancestry (TA) lead variants for individual traits and single-ancestry lead and index variants (TA + EUR, TA + AA, TA + EAS, TA + HISP and TA + SAS). Variance explained in each ancestry is in different colors. Data points represent the variance explained in individual cohorts used in this analysis. Adjusted R^2 was estimated in 1–11 cohorts with sample sizes ranging from 489 to 9,758 (Supplementary Tables 8–11).

fine-mapping (Fig. 4c), which likely reflects the larger sample size and differences in LD structure, EAFs and effect sizes across diverse populations. To quantify the estimated improvement in fine-mapping resolution that is attributable to the multi-ancestry GWAS, we then compared 99% CS sizes from the trans-ancestry fine-mapping to single-ancestry-specific data emulating the same total sample size by rescaling the standard errors (Methods). Of the 72 locus–trait associations with estimated improved fine-mapping in trans-ancestry analysis, resolution at 38 (53%) was improved because of the larger sample size in the trans-ancestry fine-mapping analysis (Fig. 4c), and this estimated improved resolution would likely have been obtained in a European-only fine-mapping effort with equivalent sample size. However, at 34 (47%) loci, the inclusion of samples from multiple diverse populations yielded the estimated improved resolution. On average, ancestry differences led to a reduction in the median number of variants in the 99% CSs from

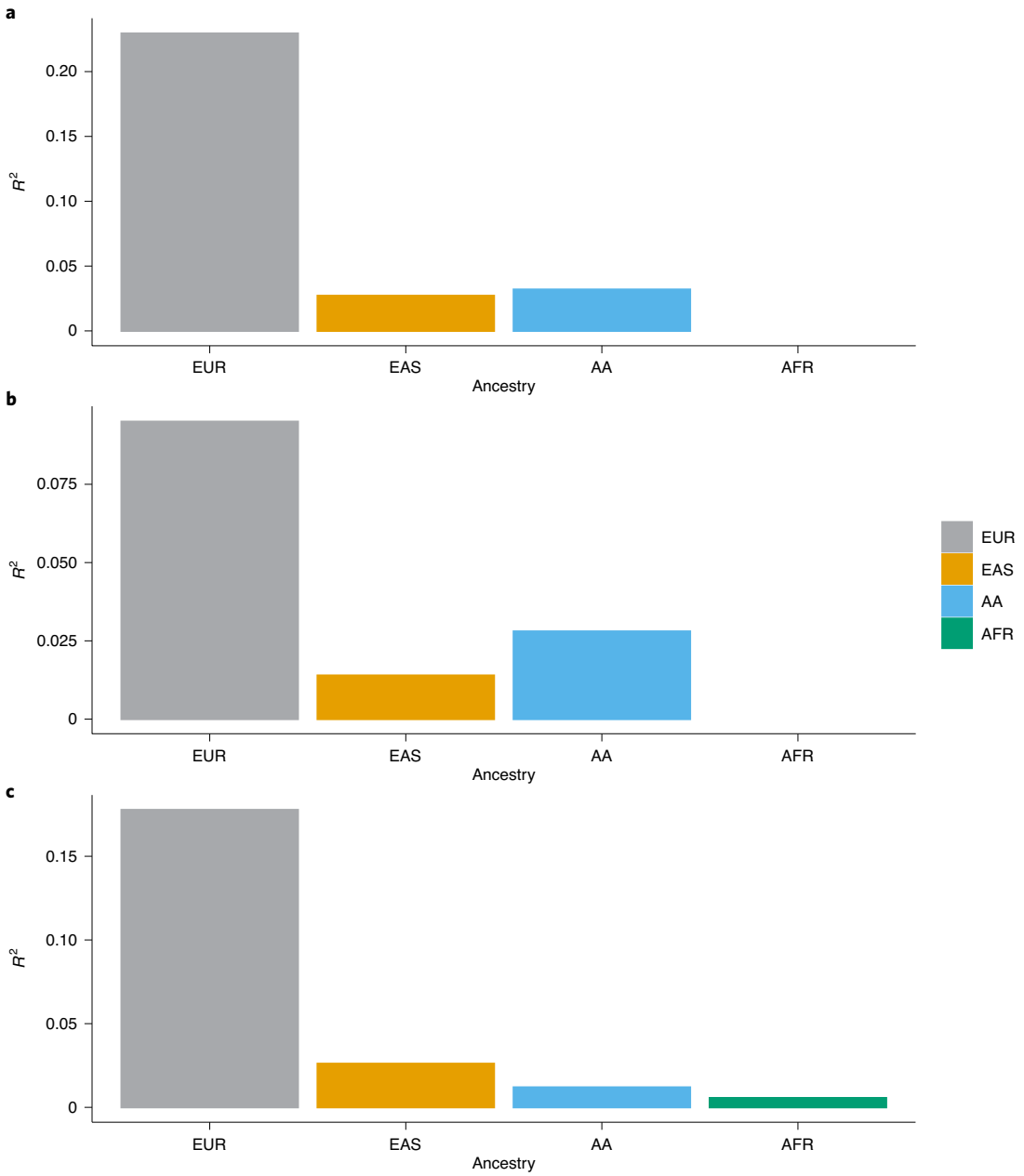


Fig. 3 | Transferability of PGSs across ancestries. Trait variance explained by polygenic scores for FG (**a**), FI (**b**) and HbA1c (**c**). For each trait, the bar plots represent the trait variance explained when using a European-ancestry-derived PGS in European, East Asian, African American and African test datasets. Variance explained (the height of each bar) in each ancestry is shown in different colors.

24 to 15 variants (37.5% median reduction; Fig. 4c), demonstrating the value of conducting fine-mapping analyses across ancestries.

HbA1c signal classification. HbA1c-associated variants can exert their effects on HbA1c levels through both glycemic and non-glycemic pathways[7,51] and their correct classification can affect T2D diagnostic accuracy[7,52]. Using previous association results for other glycemic, RBC and iron traits, as well as a fuzzy clustering approach, we classified variants into their most likely mode of action (Methods, Supplementary Note). Of the 218 HbA1c-associated variants, 27 (12%) could not be characterized due to missing data and 23 (11%) could not be classified into a 'known' class (Supplementary Note). The remaining signals were classified as principally: (1) glycemic ($n=53$; 24%); (2) affecting iron levels and/or iron metabolism ($n=12$; 6%); or (3) RBC traits ($n=103$; 47%). A genetic risk score (GRS) composed of all HbA1c-associated signals was strongly associated with T2D risk (odds ratio (OR) = 2.4, 95% confidence interval (CI) = 2.3–2.5, $P=2.7\times10^{-298}$). However, when using partitioned GRSs composed of these different classes of variants (Methods), we found that the T2D association was mainly driven by variants that influenced HbA1c through glycemic pathways (OR = 2.6, 95% CI = 2.5–2.8, $P=2.3\times10^{-250}$), with weaker evidence of an association (despite the larger number of variants in the GRS) and a more modest risk (OR = 1.4, 95% CI = 1.2–1.7, $P=4.7\times10^{-4}$) imparted by signals in the mature RBC cluster that were not glycemic (that is, for which those specific variants had $P>0.05$ for FI, 2hGlu and FG) (Extended Data Fig. 6, Supplementary Note). This is in contrast with our previous finding in which we found no significant association between a risk score of non-glycemic variants and T2D[7]. Our current results could be partly driven by cases of T2D

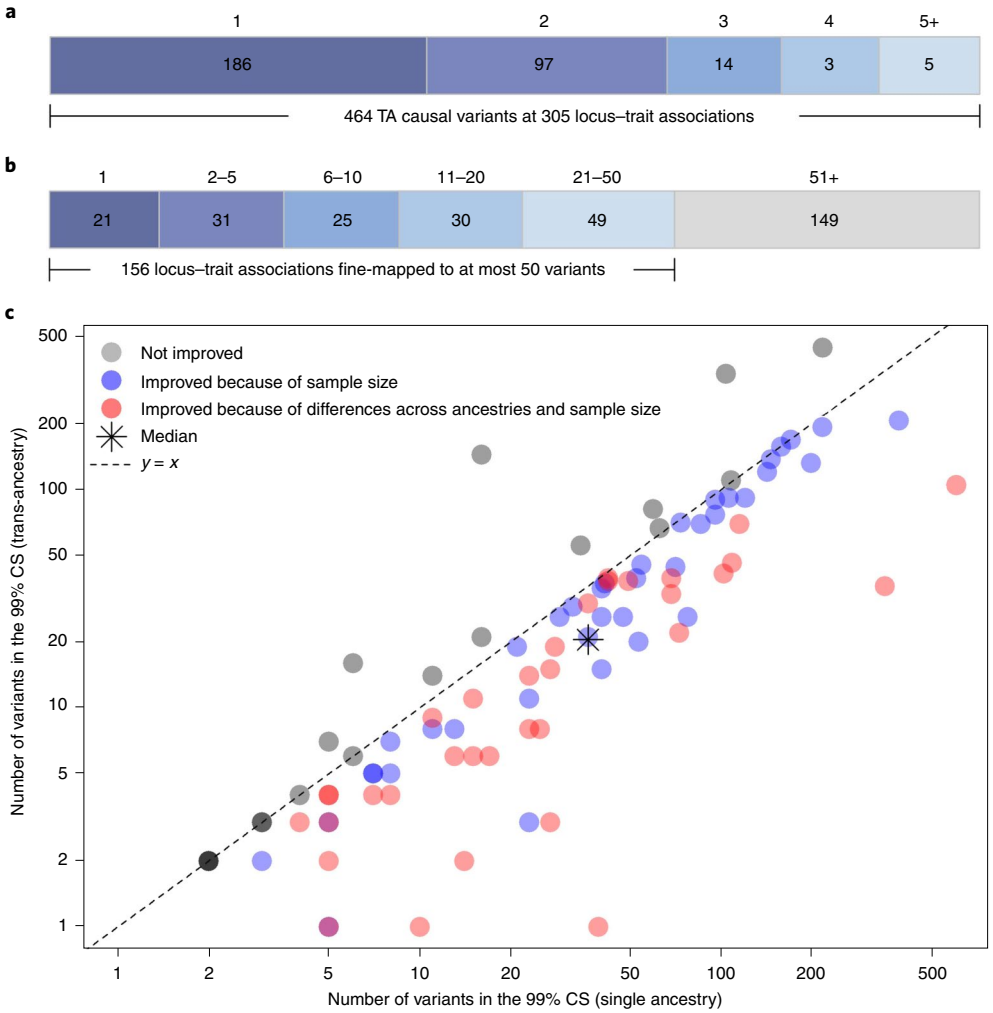


Fig. 4 | Trans-ancestry fine-mapping. a, Number of plausible causal variants at each locus–trait association derived from FINEMAP. **b**, Number of variants within each 99% CS. Twenty-one locus–trait associations at 19 loci were mapped to a single variant in the 99% CS. **c**, Fine-mapping resolution. For each of the 98 locus–trait associations with a predicted single causal variant in both trans-ancestry and single-ancestry analyses, the number of variants included in the 99% CS in the single-ancestry fine-mapping (*x* axis; logarithmic scale) is plotted against those in the trans-ancestry fine-mapping (*y* axis; logarithmic scale). Trans-ancestry and single-ancestry fine-mapping analyses were based on the same set of variants. After removing eight locus–trait associations with one variant in the 99% CSs in both trans-ancestry and single-ancestry analyses, there were 18 locus–trait associations (gray) for which trans-ancestry fine-mapping did not improve the resolution of fine-mapping results (that is, the number of variants in the 99% CS did not decrease). Of the 72 locus–trait associations with improved trans-ancestry fine-mapping resolution (blue and red) further analyses in European fine-mapping emulating the total sample size in trans-ancestry fine-mapping demonstrated that 34 locus–trait associations (red) were improved because of both total sample size and differences across ancestries, whereas 38 locus–trait associations (in blue) were improved because of only the increased sample size in the original trans-ancestry fine-mapping analysis.

being diagnosed on the basis of HbA1c levels that may be influenced by the non-glycemic signals, or by glycemic effects that are not captured by FI, 2hGlu or FG measures.

Biological signatures of glycemic-trait-associated loci. To better understand distinct and shared biological signatures underlying variant–trait associations, we conducted genomic feature enrichment, expression quantitative trait loci (eQTL) co-localization, and tissue and gene-set enrichment analyses across all four traits.

Epigenomic landscape of trait-associated variants. We explored the genomic context that underlies glycemic trait loci by computing overlap enrichment for 'static' annotations such as coding regions, conserved regions and super enhancers merged across multiple cell types[53–55] using the GREGOR tool[56]. We observed that FG, FI and HbA1c signals (Supplementary Table 7) were significantly ($P < 8.4 \times 10^{-4}$, Bonferroni threshold for 59 annotations) enriched in evolutionarily conserved regions (Fig. 5a, Extended Data Fig. 7, Supplementary Table 15).

We then considered epigenomic landscapes defined in individual cell and/or tissue types. Previously, stretch enhancers (StrE; enhancer chromatin states that are ≥3 kb in length) in pancreatic islets were shown to be highly cell-specific and strongly enriched with T2D risk signals[57]. Considering StrEs across 31 cell types[38], FG and 2hGlu signals showed the highest enrichment in islets (FG, fold enrichment = 4.70, $P = 2.7 \times 10^{-24}$; 2hGlu, fold enrichment = 5.51, $P = 3.6 \times 10^{-4}$; Fig. 5a, Supplementary Table 16), highlighting the importance of islets for these traits. FI signals were enriched in skeletal muscle (fold enrichment = 3.17, $P = 7.8 \times 10^{-6}$) and adipose StrEs (fold enrichment = 3.27, $P = 1.8 \times 10^{-7}$), consistent with the idea that these tissues are targets of insulin action (Fig. 5a). StrEs in individual cell types showed higher enrichment

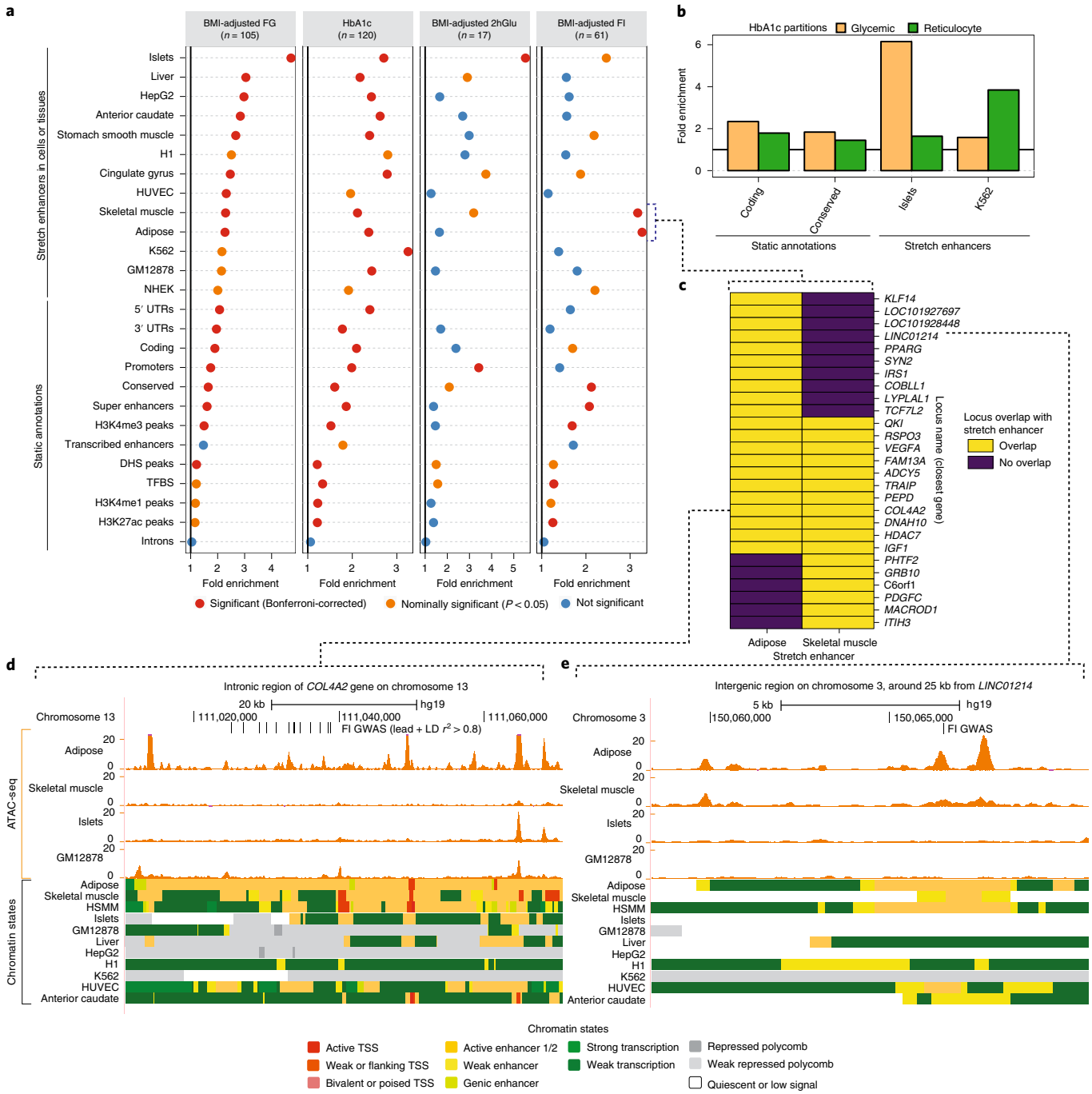


Fig. 5 | Epigenomic landscape of trait-associated variants. a, Enrichment of GWAS variants that overlap genomic regions including 'Static annotations', which are common or static across cell types, and 'Stretch enhancers' (StrE), which are identified in each tissue and/or cell type. The numbers of signals for each trait are indicated in parentheses. Enrichment was calculated using GREGOR[56]. Black line shows the null (enrichment = 1). One-sided test for significance (red) is determined after Bonferroni correction to account for 59 total annotations tested for each trait; nominal significance ($P < 0.05$) is indicated in yellow. HepG2 (hepatoma cells), H1 (embryonic stem cells), HUVEC (human umbilical vein endothelial cells), K562 (myelogenous leukemia cells), GM12878 (lymphoblastoid cells) and NEHK (normal human epidermal keratinocytes) are human immortalized cell lines. UTR, untranslated region, DHS, DNase I hypersensitivity sites; TFBS, transcription factor binding sites. H3K4me3, H3K4me1 and H3K27ac are epigenetic modifications of histone 3 lysine residues. **b**, Enrichment for HbA1c GWAS signals partitioned into the 'hard' glycemic and RBC cluster (signals from 'hard' mature RBC and reticulocyte clusters together) to overlap annotations that include StrEs in islets and the blood-derived leukemia cell line K562, respectively (additional partitioned results are shown in Supplementary Table 17). **c**, Individual FI GWAS signals that drive enrichment in adipose and skeletal muscle StrEs. C6orf1 is also known as *SMIM29*. **d**,**e**, Genome browser shots of FI GWAS signals and an intronic region of the *COL4A2* gene (**d**) and an intergenic region around 25 kb from the LINC01214 gene (**e**) showing GWAS SNPs (lead and LD $r^2 > 0.8$ proxies), assay for transposase accessible chromatin followed by sequencing (ATAC-seq) signal tracks and chromatin state annotations in different tissues and cell types. TSS, transcription start site.

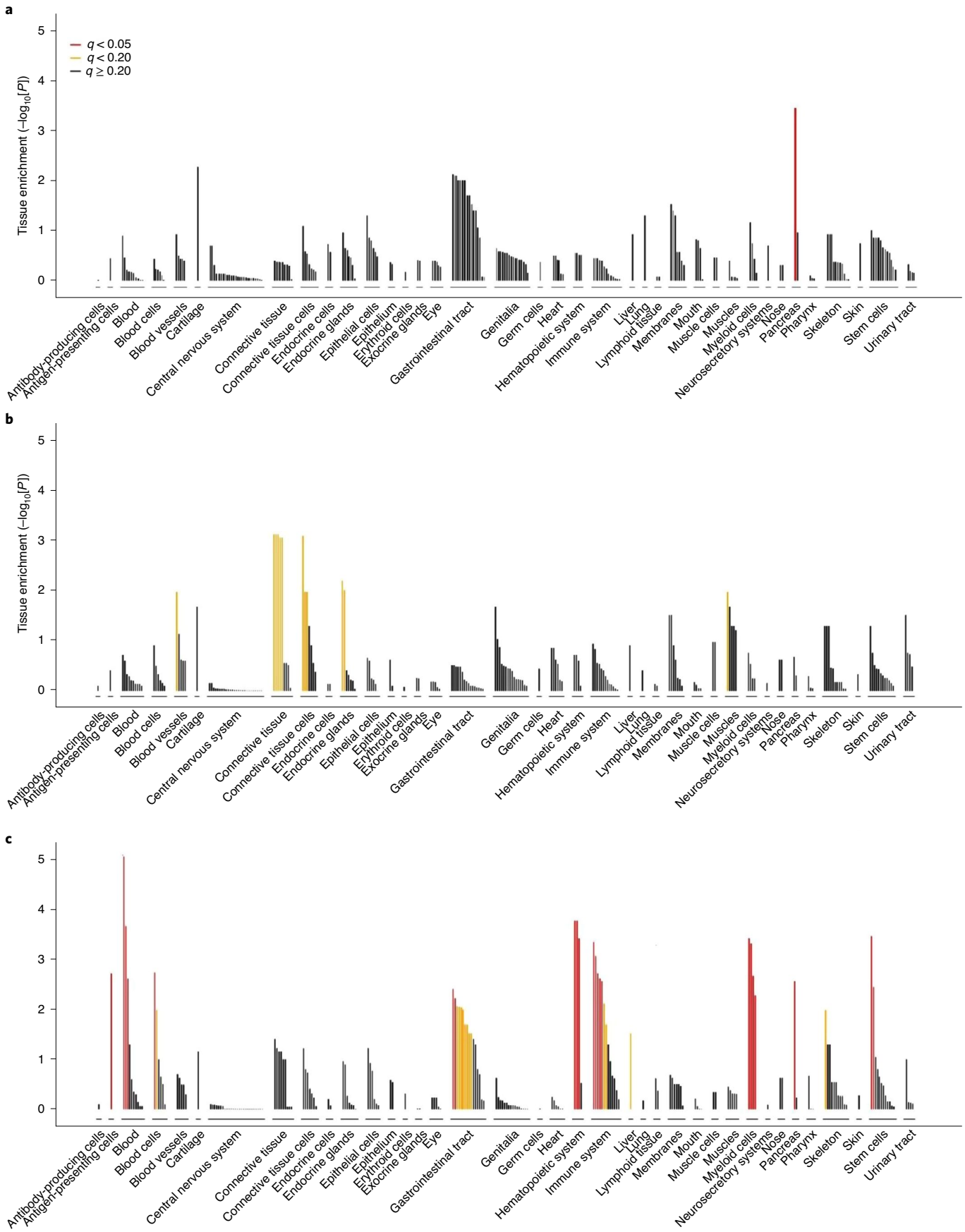


Fig. 6 | Tissues and cell types that are significantly enriched in genes in loci associated with glycemic traits. Results of tissue and cell-type enrichment analysis for FG-associated loci (**a**), FI-associated loci (**b**) and HbA1c-associated loci (**c**). FDR thresholds are shown in red ($q < 0.05$), orange ($q < 0.2$) or black ($q \geq 0.2$).

than super enhancers merged across cell types, highlighting the importance of cell-specific analyses (Fig. 5a). HbA1c signals were enriched in StrEs of multiple cell types and tissues, but have the strongest enrichment in K562 leukemia-derived cells (fold enrichment = 3.24, $P = 1.2 \times 10^{-7}$; Fig. 5a). Among the 'hard' glycemic and RBC (mature + reticulocyte) HbA1c signals, glycemic signals were enriched in islet StrEs (fold enrichment = 3.96, $P = 3.7 \times 10^{-16}$) whereas RBC signals were enriched in K562 StrEs (fold enrichment = 7.5, $P = 2.08 \times 10^{-14}$; Fig. 5b, Supplementary Table 17). These analyses suggest that these glycemic-trait-associated variants influence the function of tissue-specific enhancers.

Independent analyses with fGWAS[58] and GARFIELD[59] yielded consistent results (Extended Data Figs. 8 and 9, Supplementary Tables 16 and 18). Notably, FI signals at a lenient threshold of $P < 10^{-5}$ were enriched in liver StrEs using GARFIELD (OR = 1.92, $P = 1.7 \times 10^{-4}$) (Extended Data Fig. 9a). This suggests that liver regulatory annotations are relevant for FI GWAS signals, but that we lack the power to detect significant enrichment using the genome-wide significant loci and the current set of reference annotations.

We next explored the 27 loci that drive the FI enrichment in adipose and skeletal muscle, 11 of which overlapped with StrEs in both tissues (Fig. 5c). At the *COL4A2* locus, variants within an intronic region overlap with StrEs in adipose tissue, skeletal muscle and a human skeletal muscle myoblast (HSMM) cell line that are not shared across other cell or tissue types. Among these, rs9555695 (in the 99% CS) also overlaps with accessible chromatin regions in adipose (Fig. 5d). At a narrow signal with no proxy variants (LD $r^2 > 0.7$ in individuals of European ancestry), the lead trans-ancestry variant rs62271373 (PPA = 0.94), which is located in an intergenic region around 25 kb from the LINC01214 gene, overlaps with StrEs that are specific to adipose and HSMM and an active enhancer chromatin state in skeletal muscle (Fig. 5e). Collectively, the tissue-specific epigenomic signatures at GWAS signals provide an opportunity to nominate tissues in which these variants are likely to be active. This map may help future efforts to deconvolute GWAS signals into tissue-specific disease pathology.

Co-localization of GWAS and eQTLs. Among the 99 novel glycemic trait loci, we identified co-localized eQTLs at 34 loci in blood, pancreatic islets, subcutaneous or visceral adipose, skeletal muscle or liver, providing suggestive evidence of causal genes (Supplementary Table 19). The co-localized eQTLs include several genes that have previously been reported at glycemic trait loci[60–62]: *ADCY5*, *CAMK1D, IRS1*, *JAZF1* and *KLF14*. For some additional loci, the co-localized genes have previous evidence for a role in glycemic regulation. For example, the lead trans-ancestry variant and likely causal variant—rs1799815 (PPA = 0.993)—that is associated with FI is the strongest variant associated with expression of *INSR*, which encodes the insulin receptor, in subcutaneous adipose from METSIM ($P = 2 \times 10^{-9}$) and GTEx ($P = 5 \times 10^{-6}$) datasets. The A allele at rs1799815 is associated with higher FI and lower expression of *INSR*, which is consistent with the relationship between insulin resistance and reduced INSR function[63]. In a second example, rs841572, which is the trans-ancestry lead variant associated with FG, has the highest PPA (PPA = 0.535) among the 20 variants in the 99% CS and is in strong LD ($r^2 = 0.87$) with the lead eQTL variant (rs841576, also in the 99% CS) associated with *SLC2A1* expression in blood (eQTLGen, $P = 1 \times 10^{-8}$). *SLC2A1* (which is also known as *GLUT1*) encodes the major glucose transporter in brain, placenta and erythrocytes, and is responsible for glucose entry into the brain[64]. rs841572-A is associated with lower FG and lower *SLC2A1* expression. Although rare missense variants in *SLC2A1* are an established cause of seizures and epilepsy[65], our data suggest that *SLC2A1* variants also affect plasma glucose levels within a population. These co-localized signals provide possible regulatory mechanisms for variant effects on genes that influence glycemic traits.

The co-localized eQTLs also provide insights into the mechanisms of action of glycemic trait loci. For example, rs9884482 (in the 99% CS) is associated with FI and *TET2* expression in subcutaneous adipose ($P = 2 \times 10^{-20}$); rs9884482 is in high LD ($r^2 = 0.96$ in individuals of European ancestry) with the lead *TET2* eQTL variant (rs974801). *TET2* encodes a DNA demethylase that can affect transcriptional repression[66]. Tet2 expression in adipose is reduced after diet-induced insulin resistance in mice[67], and knockdown of Tet2 blocked adipogenesis[67,68]. Furthermore, in human adipose tissue, rs9884482-C was associated with lower *TET2* expression and higher FI. In a second example, rs617948 is associated with HbA1c (in the 99% CS) and is the lead variant associated with *C2CD2L* expression in blood (eQTLGen, $P = 3 \times 10^{-96}$). *C2CD2L* (which is also known as *TMEM24*) encodes a protein that regulates pulsatile insulin secretion and facilitates release of insulin pool reserves[69,70]. rs617948-G was associated with higher HbA1c and lower *C2CD2L*, providing evidence for a role for this insulin secretion protein in glucose homeostasis. Our HbA1c 'soft' clustering assigned this signal to both the 'unknown' (0.51 probability) and 'reticulocyte' (0.42 probability) clusters. rs617948 is strongly associated with HbA1c ($P < 6.8 \times 10^{-8}$), but not with FG, FI or 2hGlu ($P > 0.05$; Supplementary Table 20, Supplementary Note). This suggests that there is an effect of this variant on reticulocyte biology and on insulin secretion, potentially influencing HbA1c levels through different tissues and providing a plausible explanation for the classification as 'unknown'.

Tissue expression. Consistent with effector transcript expression analysis using GTEx data[50], we found considerable differences in tissue expression across the glycemic trait signals. FG signals were enriched for genes expressed in the pancreas (false-discovery rate (FDR) < 0.05), whereas there was an insufficient number of significant associations in 2hGlu to identify enrichment for any tissue or cell type at a threshold of FDR < 0.2. FI signals were enriched in connective tissue and cells (which includes adipose tissue), endocrine glands, blood cells and muscles (FDR < 0.2) and HbA1c signals were significantly enriched in genes expressed in the pancreas, hemic and immune system (FDR < 0.05) (Fig. 6, Supplementary Table 21). Consistent with previous analysis[50], FI enrichment in connective tissue was driven by adipose tissue (subcutaneous and visceral), whereas the newly described enrichment in endocrine glands was driven by the adrenal glands and cortex (Supplementary Table 21). In addition to enrichment in genes expressed in glycemic-related tissues, HbA1c signals were enriched in genes expressed in the blood, consistent with the role of RBCs in this trait and our previous results[50].

The association between FI signals and genes expressed in adrenal glands is notable, suggesting a possible direct role for these genes in insulin resistance. These genes could influence cortisol levels, which may contribute to insulin resistance and FI levels through impaired insulin receptor signaling in peripheral tissues, as well as influencing the distribution of body fat, stimulating lipolysis and affecting other indirect mechanisms[71,72].

Gene-set analyses. Next, we performed gene-set analysis using DEPICT (Data-driven Expression-Prioritized Integration for Complex Traits) (Methods). In agreement with previous results[50], we found distinct gene sets that were enriched (FDR < 0.05) in each glycemic trait except for 2hGlu, which had insufficient associations to have power in this analysis. FG-associated variants highlighted gene sets that are involved in metabolism and gene sets that are involved in general cellular functions, such as 'cytoplasmic vesicle membrane' and 'circadian clock' (Fig. 7a). By contrast, in addition to metabolism-related gene sets, FI-associated variants highlighted pathways that are related to growth, cancer and reproduction (Fig. 7b). This is consistent with the role of insulin as a mitogenic

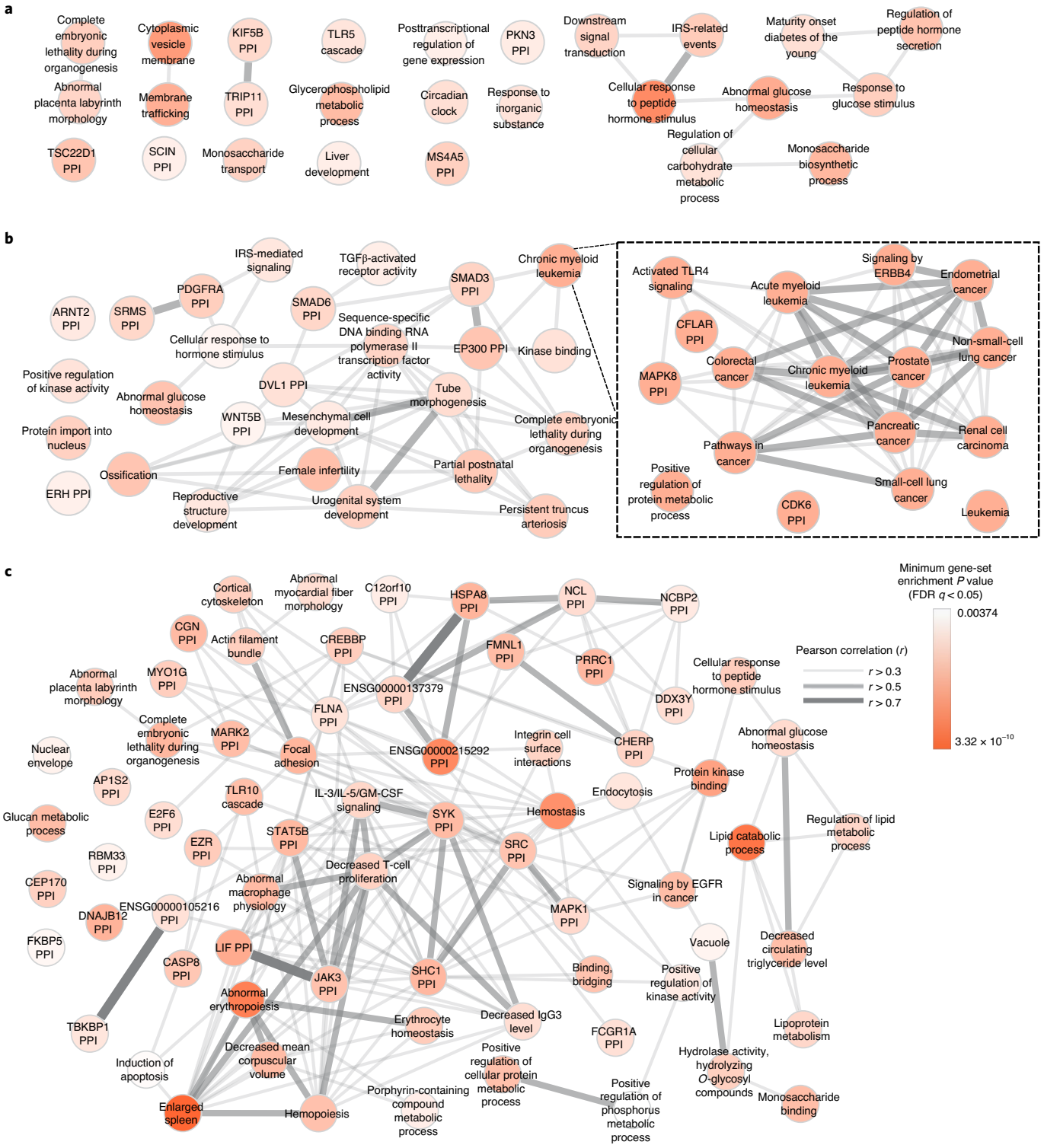


Fig. 7 | Gene-set enrichment analyses. a–c, Results from affinity-propagation clustering of significantly enriched gene sets (FDR $q < 0.05$) identified by DEPICT for FG (**a**), FI (**b**) and HbA1c (**c**). Each node is a meta-gene set that is represented by an example gene set within the meta-gene set. For example, in **b** 'chronic myeloid leukemia' is an example gene set that represents a much broader meta-gene set relating to cancer. Inset: magnification of the broader meta-gene set related to cancer, of which chronic myeloid leukemia is representative. Similarities between the meta-gene sets are represented by Pearson correlation coefficients ($r > 0.3$). The nodes are colored according to the minimum gene-set enrichment P value (FDR $q < 0.05$) of the gene sets in that meta-gene set. IRS, insulin receptor substrate; PPI, protein–protein interaction network; GM-CSF, granulocyte–macrophage colony-stimulating factor.

hormone, and with epidemiological links between insulin and certain types of cancer[73] and reproductive disorders such as polycystic ovary syndrome[74]. HbA1c-associated variants highlighted many gene sets (Fig. 7c), including those linked to metabolism and hematopoiesis, again recapitulating our postulated effects of variants on glucose and RBC biology. Additional pathways from

HbA1c-associated variants also highlighted previous 'CREBP protein–protein interactions' and lipid biology related to T2D[75] and HbA1c[76], respectively, and potential new biological pathways through which variants may influence HbA1c.

Discussion

Here we describe a large glycemic-trait meta-analysis of GWAS in which 30% of the population was composed of participants of East Asian, Hispanic, African American, South Asian and sub-Saharan African ancestry. This effort identified 242 loci (235 trans-ancestry and seven single-ancestry), which jointly explained between 0.7% (2hGlu in individuals of European ancestry) and 6% (HbA1c in individuals of African American ancestry) of the variance in glycemic traits in any given ancestry. Although 114 out of 242 loci are associated with T2D ($P < 10^{-4}$; 83 loci with $P < 5 \times 10^{-8}$; Supplementary Table 4), the absence of strong evidence of association for the remaining loci ($P \geq 10^{-4}$) suggests that for alleles with a frequency above 5% we can exclude T2D-associated OR ≥ 1.07 with 80% power ($\alpha = 5 \times 10^{-8}$; and OR ≥ 1.05 for $\alpha = 10^{-4}$) given a current study of 228,499 cases of T2D and 1,178,783 control individuals[27]. We identified 486 signals that were associated with glycemic traits, of which eight have minor allele frequency (MAF) < 1% and 45 have 1% ≤ MAF < 5% in all ancestries, highlighting that 89% of signals identified are common in at least one ancestry studied.

A key aim of our study was to evaluate the added advantage of including population diversity in genetic discovery and fine-mapping efforts. In addition to the larger sample size included in the trans-ancestry meta-analysis, we were able to estimate the contribution of data from individuals of non-European ancestry in locus discovery and fine-mapping resolution. We found that 24 of the 99 newly discovered loci owe their discovery to the inclusion of data from participants of East Asian, Hispanic, African American, South Asian and sub-Saharan African ancestry, due to differences in EAF and effect sizes across ancestries.

Comparison of 295 trans-ancestry lead variants (315 locus–trait associations) across ancestries demonstrated that between 81.5% (HbA1c) and 85.7% (FG) of the trans-ancestry lead variants showed no evidence of trans-ancestry heterogeneity in allelic effects ($P > 0.05$).

Given sample size and power limitations, genome-wide significant trait-associated variants in a single-ancestry analysis explain only a modest proportion of trait variance in that ancestry (Fig. 2). We demonstrate that trans-ancestry lead variants explain more trait variance than the ancestry-specific variants (Fig. 2). This shows that even though some trans-ancestry lead variants are not genome-wide significant in all ancestries, they contribute to the genetic architecture of the trait in most ancestries.

We evaluated the transferability of glycemic-trait PGSs derived from data from individuals of European ancestry to other ancestries. In agreement with other traits[35,77,78], we confirm that PGS derived from data from participants of European ancestry perform much worse when the test dataset is from a different ancestry. Each trait-specific PGS improves trait variance explained by between 3.5-fold (HbA1c) and 6-fold (FG) in the European dataset (Fig. 3, Supplementary Table 12) compared with a score built from only trans-ancestry lead variants and European index variants (Fig. 2, Supplementary Tables 9–12).

Despite development of approaches to derive polygenic risk scores[79], we note the difficulty in using summary level data to build a PGS in one ancestry and then apply it to test datasets of a different ancestry. Although PRS-CSauto[33] is able to use summary-level data, revision of the effect size estimates to account for LD required reference panels that matched the ancestry of the test dataset. However, the current software lacks appropriate reference panels for many ancestries, precluding its broad application. Future developments of trans-ancestry PGSs are required for improved cross-ancestry performance.

We show that fine-mapping resolution is improved in trans-ancestry, compared with single-ancestry fine-mapping efforts. In around 50% of our loci, we showed that the improvement was due to differences in EAF, effect size or LD structure between ancestries, and not only due to the overall increased sample size that was available for trans-ancestry fine-mapping. By performing trans-ancestry fine-mapping, and co-localizing GWAS signals with eQTL signals and coding variants, we identified new candidate causal genes. Taken together, these results motivate continued expansion of genetic and genomic efforts in diverse populations to improve our understanding of these traits in groups that are disproportionally affected by T2D.

Given data on four different glycemic traits and their use in the diagnosis and monitoring of T2D and metabolic health, we also sought to characterize biological features underlying these traits. We show that despite considerable sharing of loci across the four traits, each trait is also characterized by unique features based on StrE, gene expression and gene-set signatures. Combining genetic data from these traits with T2D data will further elucidate pathways that drive normal physiology and pathophysiology, and help to further develop useful predictive scores for disease classification and management[4,5].

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41588-021-00852-9.

Received: 7 July 2020; Accepted: 22 March 2021;
Published online: 31 May 2021

References

1. *Use of Glycated Haemoglobin (HbA1c) in the Diagnosis of Diabetes Mellitus: Abbreviated Report of a WHO Consultation* Report No. WHO/NMH/CHP/CPM/11.1 (World Health Organization, 2011).
2. Goodarzi, M. O. et al. Fasting insulin reflects heterogeneous physiological processes: role of insulin clearance. *Am. J. Physiol. Endocrinol. Metab.* **301**, E402–E408 (2011).
3. Dimas, A. S. et al. Impact of type 2 diabetes susceptibility variants on quantitative glycemic traits reveals mechanistic heterogeneity. *Diabetes* **63**, 2158–2171 (2014).
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. The Emerging Risk Factors Collaboration Diabetes mellitus, fasting blood glucose concentration, and risk of vascular disease: a collaborative meta-analysis of 102 prospective studies. *Lancet* **375**, 2215–2222 (2010).
7. Wheeler, E. et al. Impact of common genetic determinants of hemoglobin A1c on type 2 diabetes risk and diagnosis in ancestrally diverse populations: a transethnic genome-wide meta-analysis. *PLoS Med.* **14**, e1002383 (2017).
8. Dupuis, J. et al. New genetic loci implicated in fasting glucose homeostasis and their impact on type 2 diabetes risk. *Nat. Genet.* **42**, 105–116 (2010).
9. Manning, A. K. et al. A genome-wide approach accounting for body mass index identifies genetic variants influencing fasting glycemic traits and insulin resistance. *Nat. Genet.* **44**, 659–669 (2012).
10. Walford, G. A. et al. Genome-wide association study of the modified Stumvoll insulin sensitivity index identifies *BCL2* and *FAM19A2* as novel insulin sensitivity loci. *Diabetes* **65**, 3200–3211 (2016).
11. Horikoshi, M. et al. Discovery and fine-mapping of glycaemic and obesity-related trait loci using high-density imputation. *PLoS Genet.* **11**, e1005230 (2015).
12. Mahajan, A. et al. Identification and functional characterization of *G6PC2* coding variants influencing glycemic traits define an effector transcript at the *G6PC2-ABCB11* locus. *PLoS Genet.* **11**, e1004876 (2015).

13. Hwang, J. Y. et al. Genome-wide association meta-analysis identifies novel variants associated with fasting plasma glucose in East Asians. *Diabetes* **64**, 291–298 (2015).
14. Chen, P. et al. Multiple nonglycemic genomic loci are newly associated with blood level of glycated hemoglobin in East Asians. *Diabetes* **63**, 2551–2562 (2014).
15. Scott, R. A. et al. Large-scale association analyses identify new loci influencing glycemic traits and provide insight into the underlying biological pathways. *Nat. Genet.* **44**, 991–1005 (2012).
16. Spanakis, E. K. & Golden, S. H. Race/ethnic difference in diabetes and diabetic complications. *Curr. Diabetes Rep.* **13**, 814–823 (2013).
17. Tillin, T. et al. Insulin resistance and truncal obesity as important determinants of the greater incidence of diabetes in Indian Asians and African Caribbeans compared with Europeans: the Southall And Brent REvisited (SABRE) cohort. *Diabetes Care* **36**, 383–393 (2013).
18. Whincup, P. H. et al. Early emergence of ethnic differences in type 2 diabetes precursors in the UK: the Child Heart and Health Study in England (CHASE Study). *PLoS Med.* **7**, e1000263 (2010).
19. The 1000 Genomes Project Consortium A global reference for human genetic variation. *Nature* **526**, 68–74 (2015).
20. Willer, C. J., Li, Y. & Abecasis, G. R. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **26**, 2190–2191 (2010).
21. Yang, J., Lee, S. H., Goddard, M. E. & Visscher, P. M. GCTA: a tool for genome-wide complex trait analysis. *Am. J. Hum. Genet.* **88**, 76–82 (2011).
22. Yang, J. et al. Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. *Nat. Genet.* **44**, 369–375 (2012).
23. Wellcome Trust Case Control Consortium Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* **447**, 661–678 (2007).
24. Mahajan, A. et al. Trans-ancestry genetic study of type 2 diabetes highlights the power of diverse populations for discovery and translation. Preprint at *medRxiv* https://doi.org/10.1101/2020.09.22.20198937 (2020).
25. Mahajan, A. et al. Fine-mapping type 2 diabetes loci to single-variant resolution using high-density imputation and islet-specific epigenome maps. *Nat. Genet.* **50**, 1505–1513 (2018).
26. Spracklen, C. N. et al. Identification of type 2 diabetes loci in 433,540 East Asian individuals. *Nature* **582**, 240–245 (2020).
27. 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).
28. Luo, Y. et al. Transcription factor Ets1 regulates expression of thioredoxin-interacting protein and inhibits insulin secretion in pancreatic beta-cells. *PLoS ONE* **9**, e99049 (2014).
29. Braccini, L. et al. PI3K-C2γ is a Rab5 effector selectively controlling endosomal Akt2 activation downstream of insulin signalling. *Nat. Commun.* **6**, 7400 (2015).
30. Aschard, H., Vilhjálmsson, B. J., Joshi, A. D., Price, A. L. & Kraft, P. Adjusting for heritable covariates can bias effect estimates in genome-wide association studies. *Am. J. Hum. Genet.* **96**, 329–339 (2015).
31. Lee, J. J. et al. Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals. *Nat. Genet.* **50**, 1112–1121 (2018).
32. Nolte, I. M. et al. Missing heritability: is the gap closing? An analysis of 32 complex traits in the Lifelines Cohort Study. *Eur. J. Hum. Genet.* **25**, 877–885 (2017).
33. Ge, T., Chen, C. Y., Ni, Y., Feng, Y. A. & Smoller, J. W. Polygenic prediction via Bayesian regression and continuous shrinkage priors. *Nat. Commun.* **10**, 1776 (2019).
34. Dastani, Z. et al. Novel loci for adiponectin levels and their influence on type 2 diabetes and metabolic traits: a multi-ethnic meta-analysis of 45,891 individuals. *PLoS Genet.* **8**, e1002607 (2012).
35. Martin, A. R. et al. Clinical use of current polygenic risk scores may exacerbate health disparities. *Nat. Genet.* **51**, 584–591 (2019).
36. Gaulton, K. J. et al. Genetic fine mapping and genomic annotation defines causal mechanisms at type 2 diabetes susceptibility loci. *Nat. Genet.* **47**, 1415–1425 (2015).
37. Spracklen, C. N. et al. Identification and functional analysis of glycemic trait loci in the China Health and Nutrition Survey. *PLoS Genet.* **14**, e1007275 (2018).
38. Varshney, A. et al. Genetic regulatory signatures underlying islet gene expression and type 2 diabetes. *Proc. Natl Acad. Sci. USA* **114**, 2301–2306 (2017).
39. Kichaev, G. et al. Leveraging polygenic functional enrichment to improve GWAS power. *Am. J. Hum. Genet.* **104**, 65–75 (2019).
40. Shriner, D. & Rotimi, C. N. Whole-genome-sequence-based haplotypes reveal single origin of the sickle allele during the Holocene wet phase. *Am. J. Hum. Genet.* **102**, 547–556 (2018).
41. Kramer, H. J. et al. African ancestry-specific alleles and kidney disease risk in Hispanics/Latinos. *J. Am. Soc. Nephrol.* **28**, 915–922 (2017).
42. Ravenhall, M. et al. Novel genetic polymorphisms associated with severe malaria and under selective pressure in North-eastern Tanzania. *PLoS Genet.* **14**, e1007172 (2018).
43. Hodonsky, C. J. et al. Genome-wide association study of red blood cell traits in Hispanics/Latinos: The Hispanic Community Health Study/Study of Latinos. *PLoS Genet.* **13**, e1006760 (2017).
44. Gurdasani, D. et al. Uganda genome resource enables insights into population history and genomic discovery in Africa. *Cell* **179**, 984–1002 (2019).
45. Rees, M. G. et al. Cellular characterisation of the *GCKR* P446L variant associated with type 2 diabetes risk. *Diabetologia* **55**, 114–122 (2012).
46. Bonomo, J. A. et al. The ras responsive transcription factor *RREB1* is a novel candidate gene for type 2 diabetes associated end-stage kidney disease. *Hum. Mol. Genet.* **23**, 6441–6447 (2014).
47. Wessel, J. et al. Low-frequency and rare exome chip variants associate with fasting glucose and type 2 diabetes susceptibility. *Nat. Commun.* **6**, 5897 (2015).
48. Scott, R. A. et al. A genomic approach to therapeutic target validation identifies a glucose-lowering GLP1R variant protective for coronary heart disease. *Sci. Transl. Med.* **8**, 341ra76 (2016).
49. Nai, A. et al. TMPRSS6 rs855791 modulates hepcidin transcription in vitro and serum hepcidin levels in normal individuals. *Blood* **118**, 4459–4462 (2011).
50. Ng, N. H. J. et al. Tissue-specific alteration of metabolic pathways influences glycemic regulation. Preprint at *bioRxiv* https://doi.org/10.1101/790618 (2019).
51. Soranzo, N. et al. Common variants at 10 genomic loci influence hemoglobin A_{1C} levels via glycemic and nonglycemic pathways. *Diabetes* **59**, 3229–3239 (2010).
52. Sarnowski, C. et al. Impact of rare and common genetic variants on diabetes diagnosis by hemoglobin A1c in multi-ancestry cohorts: the trans-omics for precision medicine program. *Am. J. Hum. Genet.* **105**, 706–718 (2019).
53. Kundaje, A. et al. Integrative analysis of 111 reference human epigenomes. *Nature* **518**, 317–330 (2015).
54. Nagel, M. et al. Meta-analysis of genome-wide association studies for neuroticism in 449,484 individuals identifies novel genetic loci and pathways. *Nat. Genet.* **50**, 920–927 (2018).
55. Savage, J. E. et al. Genome-wide association meta-analysis in 269,867 individuals identifies new genetic and functional links to intelligence. *Nat. Genet.* **50**, 912–919 (2018).
56. Schmidt, E. M. et al. GREGOR: evaluating global enrichment of trait-associated variants in epigenomic features using a systematic, data-driven approach. *Bioinformatics* **31**, 2601–2606 (2015).
57. Parker, S. C. et al. Chromatin stretch enhancer states drive cell-specific gene regulation and harbor human disease risk variants. *Proc. Natl Acad. Sci. USA* **110**, 17921–17926 (2013).
58. Pickrell, J. K. Joint analysis of functional genomic data and genome-wide association studies of 18 human traits. *Am. J. Hum. Genet.* **94**, 559–573 (2014).
59. Iotchkova, V. et al. GARFIELD classifies disease-relevant genomic features through integration of functional annotations with association signals. *Nat. Genet.* **51**, 343–353 (2019).
60. van de Bunt, M. et al. Transcript expression data from human islets links regulatory signals from genome-wide association studies for type 2 diabetes and glycemic traits to their downstream effectors. *PLoS Genet.* **11**, e1005694 (2015).
61. Civelek, M. et al. Genetic regulation of adipose gene expression and cardio-metabolic traits. *Am. J. Hum. Genet.* **100**, 428–443 (2017).
62. Scott, L. J. et al. The genetic regulatory signature of type 2 diabetes in human skeletal muscle. *Nat. Commun.* **7**, 11764 (2016).
63. Ben Harouch, S., Klar, A. & Falik Zaccai, T. C. in *GeneReviews* (eds Adam, M. P. et al.) (Univ. of Washington, 1993).
64. Agus, D. B. et al. Vitamin C crosses the blood-brain barrier in the oxidized form through the glucose transporters. *J. Clin. Invest.* **100**, 2842–2848 (1997).
65. Wolking, S. et al. Focal epilepsy in glucose transporter type 1 (Glut1) defects: case reports and a review of literature. *J. Neurol.* **261**, 1881–1886 (2014).
66. Guallar, D. et al. RNA-dependent chromatin targeting of TET2 for endogenous retrovirus control in pluripotent stem cells. *Nat. Genet.* **50**, 443–451 (2018).
67. Bian, F. et al. TET2 facilitates PPARγ agonist-mediated gene regulation and insulin sensitization in adipocytes. *Metabolism* **89**, 39–47 (2018).
68. Yoo, Y. et al. TET-mediated hydroxymethylcytosine at the *Pparγ* locus is required for initiation of adipogenic differentiation. *Int. J. Obes.* **41**, 652–659 (2017).

69. Lees, J. A. et al. Lipid transport by TMEM24 at ER-plasma membrane contacts regulates pulsatile insulin secretion. *Science* **355**, eaah6171 (2017).
70. Pottekat, A. et al. Insulin biosynthetic interaction network component, TMEM24, facilitates insulin reserve pool release. *Cell Rep.* **4**, 921–930 (2013).
71. Androulakis, I. I. et al. Patients with apparently nonfunctioning adrenal incidentalomas may be at increased cardiovascular risk due to excessive cortisol secretion. *J. Clin. Endocrinol. Metab.* **99**, 2754–2762 (2014).
72. Altieri, B. et al. Adrenocortical tumors and insulin resistance: what is the first step? *Int. J. Cancer* **138**, 2785–2794 (2016).
73. Johansson, M. et al. The influence of obesity-related factors in the etiology of renal cell carcinoma-A Mendelian randomization study. *PLoS Med.* **16**, e1002724 (2019).
74. Diamanti-Kandarakis, E. & Dunaif, A. Insulin resistance and the polycystic ovary syndrome revisited: an update on mechanisms and implications. *Endocr. Rev.* **33**, 981–1030 (2012).
75. The DIAbetes Genetics Replication And Meta-analysis (DIAGRAM) Consortium Large-scale association analysis provides insights into the genetic architecture and pathophysiology of type 2 diabetes. *Nat. Genet.* **44**, 981–990 (2012).
76. Leong, A. et al. Mendelian randomization analysis of hemoglobin A_{1C} as a risk factor for coronary artery disease. *Diabetes Care* **42**, 1202–1208 (2019).
77. Duncan, L. et al. Analysis of polygenic risk score usage and performance in diverse human populations. *Nat. Commun.* **10**, 3328 (2019).
78. Mostafavi, H. et al. Variable prediction accuracy of polygenic scores within an ancestry group. *eLife* **9**, e48376 (2020).
79. Choi, S. W., Mak, T. S. & O'Reilly, P. F. Tutorial: a guide to performing polygenic risk score analyses. *Nat. Protoc.* **15**, 2759–2772 (2020).

Ji Chen[1,2,320], Cassandra N. Spracklen[3,4,320], Gaëlle Marenne[2,5,320], Arushi Varshney[6,320], Laura J. Corbin[7,8,320], Jian'an Luan[9], Sara M. Willems[9], Ying Wu[3], Xiaoshuai Zhang[9,10], Momoko Horikoshi[11,12,13], Thibaud S. Boutin[14], Reedik Mägi[15], Johannes Waage[16], Ruifang Li-Gao[17], Kei Hang Katie Chan[18,19,20], Jie Yao[21], Mila D. Anasanti[22], Audrey Y. Chu[23], Annique Claringbould[24], Jani Heikkinen[22], Jaeyoung Hong[25], Jouke-Jan Hottenga[26,27], Shaofeng Huo[28], Marika A. Kaakinen[22,29], Tin Louie[30], Winfried März[31,32,33], Hortensia Moreno-Macias[34], Anne Ndungu[12], Sarah C. Nelson[30], Ilja M. Nolte[35], Kari E. North[36], Chelsea K. Raulerson[3], Debashree Ray[37], Rebecca Rohde[36], Denis Rybin[25], Claudia Schurmann[38,39], Xueling Sim[40,41,42], Lorraine Southam[2,43], Isobel D. Stewart[9], Carol A. Wang[44], Yujie Wang[36], Peitao Wu[25], Weihua Zhang[45,46], Tarunveer S. Ahluwalia[16,47,48], Emil V. R. Appel[49], Lawrence F. Bielak[50], Jennifer A. Brody[51], Noël P. Burtt[52], Claudia P. Cabrera[53,54], Brian E. Cade[55,56], Jin Fang Chai[40], Xiaoran Chai[57,58], Li-Ching Chang[59], Chien-Hsiun Chen[59], Brian H. Chen[60], Kumaraswamy Naidu Chitrala[61], Yen-Feng Chiu[62], Hugoline G. de Haan[17], Graciela E. Delgado[33], Ayse Demirkan[29,63], Qing Duan[3,64], Jorgen Engmann[65], Segun A. Fatumo[66,67,68], Javier Gayán[69], Franco Giulianini[23], Jung Ho Gong[18], Stefan Gustafsson[70], Yang Hai[71], Fernando P. Hartwig[7,72], Jing He[73], Yoriko Heianza[74], Tao Huang[75], Alicia Huerta-Chagoya[76,77], Mi Yeong Hwang[78], Richard A. Jensen[51], Takahisa Kawaguchi[79], Katherine A. Kentistou[80,81], Young Jin Kim[78], Marcus E. Kleber[33], Ishminder K. Kooner[46], Shuiqing Lai[18], Leslie A. Lange[82], Carl D. Langefeld[83], Marie Lauzon[21], Man Li[84], Symen Ligthart[63], Jun Liu[63,85], Marie Loh[45,86], Jirong Long[87], Valeriya Lyssenko[88,89], Massimo Mangino[90,91], Carola Marzi[92,93], May E. Montasser[94], Abhishek Nag[12], Masahiro Nakatochi[95], Damia Noce[96], Raymond Noordam[97], Giorgio Pistis[98], Michael Preuss[38,99], Laura Raffield[3], Laura J. Rasmussen-Torvik[100], Stephen S. Rich[101,102], Neil R. Robertson[11,12], Rico Rueedi[103,104], Kathleen Ryan[94], Serena Sanna[24,98], Richa Saxena[105,106,107], Katharina E. Schraut[80,81], Bengt Sennblad[108], Kazuya Setoh[79], Albert V. Smith[109,110], Thomas Sparsø[49], Rona J. Strawbridge[111,112], Fumihiko Takeuchi[113], Jingyi Tan[21], Stella Trompet[97,114], Erik van den Akker[115,116,117], Peter J. van der Most[35], Niek Verweij[118,119], Mandy Vogel[120], Heming Wang[55,56], Chaolong Wang[121,122], Nan Wang[123,124], Helen R. Warren[53,54], Wanqing Wen[87], Tom Wilsgaard[125], Andrew Wong[126], Andrew R. Wood[1], Tian Xie[35], Mohammad Hadi Zafarmand[127,128], Jing-Hua Zhao[129], Wei Zhao[50], Najaf Amin[63,85], Zorayr Arzumanyan[21], Arne Astrup[130], Stephan J. L. Bakker[131], Damiano Baldassarre[132,133], Marian Beekman[115], Richard N. Bergman[134], Alain Bertoni[135], Matthias Blüher[136], Lori L. Bonnycastle[137], Stefan R. Bornstein[138], Donald W. Bowden[139], Qiuyin Cai[73], Archie Campbell[140,141], Harry Campbell[80], Yi Cheng Chang[59,142,143], Eco J. C. de Geus[26,27], Abbas Dehghan[63], Shufa Du[144], Gudny Eiriksdottir[110], Aliki Eleni Farmaki[145,146],

Mattias Frånberg[112], Christian Fuchsberger[96], Yutang Gao[147], Anette P. Gjesing[49], Anuj Goel[12,148], Sohee Han[78], Catharina A. Hartman[149], Christian Herder[150,151,152], Andrew A. Hicks[96], Chang-Hsun Hsieh[153,154], Willa A. Hsueh[155], Sahoko Ichihara[156], Michiya Igase[157], M. Arfan Ikram[63], W. Craig Johnson[30], Marit E. Jørgensen[47,158], Peter K. Joshi[80], Rita R. Kalyani[159], Fouad R. Kandeel[160], Tomohiro Katsuya[161,162], Chiea Chuen Khor[122], Wieland Kiess[120], Ivana Kolcic[163], Teemu Kuulasmaa[164], Johanna Kuusisto[165], Kristi Läll[15], Kelvin Lam[21], Deborah A. Lawlor[7,8], Nanette R. Lee[166,167], Rozenn N. Lemaitre[51], Honglan Li[168], Lifelines Cohort Study*, Shih-Yi Lin[169,170], Jaana Lindström[171], Allan Linneberg[172,173], Jianjun Liu[122,174], Carlos Lorenzo[175], Tatsuaki Matsubara[176], Fumihiko Matsuda[79], Geltrude Mingrone[177], Simon Mooijaart[97], Sanghoon Moon[78], Toru Nabika[178], Girish N. Nadkarni[38], Jerry L. Nadler[179], Mari Nelis[15], Matt J. Neville[11,180], Jill M. Norris[181], Yasumasa Ohyagi[182], Annette Peters[93,183,184], Patricia A. Peyser[50], Ozren Polasek[163,185], Qibin Qi[186], Dennis Raven[149], Dermot F. Reilly[187], Alex Reiner[188], Fernando Rivideneira[189], Kathryn Roll[21], Igor Rudan[190], Charumathi Sabanayagam[57,191], Kevin Sandow[21], Naveed Sattar[192], Annette Schürmann[93,193], Jinxiu Shi[194], Heather M. Stringham[41,42], Kent D. Taylor[21], Tanya M. Teslovich[195], Betina Thuesen[172], Paul R. H. J. Timmers[80,196], Elena Tremoli[133], Michael Y. Tsai[197], Andre Uitterlinden[189], Rob M. van Dam[40,174,198], Diana van Heemst[97], Astrid van Hylckama Vlieg[17], Jana V. van Vliet-Ostaptchouk[35], Jagadish Vangipurapu[199], Henrik Vestergaard[49,200], Tao Wang[186], Ko Willems van Dijk[201,202,203], Tatijana Zemunik[204], Gonçalo R. Abecasis[42], Linda S. Adair[144,205], Carlos Alberto Aguilar-Salinas[206,207,208], Marta E. Alarcón-Riquelme[209,210], Ping An[211], Larissa Aviles-Santa[212], Diane M. Becker[213], Lawrence J. Beilin[214], Sven Bergmann[103,104,215], Hans Bisgaard[16], Corri Black[216], Michael Boehnke[41,42], Eric Boerwinkle[217,218], Bernhard O. Böhm[219,220], Klaus Bønnelykke[16], D. I. Boomsma[26,27], Erwin P. Bottinger[38,221,222], Thomas A. Buchanan[124,223,224], Mickaël Canouil[225,226], Mark J. Caulfield[53,54], John C. Chambers[45,46,86,227,228], Daniel I. Chasman[23,229], Yii-Der Ida Chen[21], Ching-Yu Cheng[57,191], Francis S. Collins[137], Adolfo Correa[230], Francesco Cucca[98], H. Janaka de Silva[231], George Dedoussis[232], Sölve Elmståhl[233], Michele K. Evans[234], Ele Ferrannini[235], Luigi Ferrucci[236], Jose C. Florez[107,237,238], Paul W. Franks[89,239], Timothy M. Frayling[1], Philippe Froguel[225,226,240], Bruna Gigante[241], Mark O. Goodarzi[242], Penny Gordon-Larsen[144,205], Harald Grallert[92,93], Niels Grarup[49], Sameline Grimsgaard[125], Leif Groop[243,244], Vilmundur Gudnason[110,245], Xiuqing Guo[21], Anders Hamsten[112], Torben Hansen[49], Caroline Hayward[196], Susan R. Heckbert[246], Bernardo L. Horta[72], Wei Huang[194], Erik Ingelsson[247], Pankow S. James[248], Marjo-Ritta Jarvelin[249,250,251,252], Jost B. Jonas[253,254,255], J. Wouter Jukema[114,256], Pontiano Kaleebu[257], Robert Kaplan[186,188], Sharon L. R. Kardia[50], Norihiro Kato[113], Sirkka M. Keinanen-Kiukaanniemi[258,259], Bong-Jo Kim[78], Mika Kivimaki[260], Heikki A. Koistinen[261,262,263], Jaspal S. Kooner[46,227,228,264], Antje Körner[120], Peter Kovacs[136,265], Diana Kuh[126], Meena Kumari[266], Zoltan Kutalik[104,267], Markku Laakso[165], Timo A. Lakka[268,269,270], Lenore J. Launer[61], Karin Leander[271], Huaixing Li[28], Xu Lin[28], Lars Lind[272], Cecilia Lindgren[12,273,274], Simin Liu[18], Ruth J. F. Loos[38,99], Patrik K. E. Magnusson[275], Anubha Mahajan[12,319], Andres Metspalu[15], Dennis O. Mook-Kanamori[17,276], Trevor A. Mori[214], Patricia B. Munroe[53,54], Inger Njølstad[125], Jeffrey R. O'Connell[94], Albertine J. Oldehinkel[149], Ken K. Ong[9], Sandosh Padmanabhan[277], Colin N. A. Palmer[278], Nicholette D. Palmer[139], Oluf Pedersen[49], Craig E. Pennell[44], David J. Porteous[140,279], Peter P. Pramstaller[96], Michael A. Province[211], Bruce M. Psaty[51,246,280], Lu Qi[281], Leslie J. Raffel[282], Rainer Rauramaa[270], Susan Redline[55,56], Paul M. Ridker[23,283], Frits R. Rosendaal[17], Timo E. Saaristo[284,285], Manjinder Sandhu[286], Jouko Saramies[287], Neil Schneiderman[288], Peter Schwarz[93,138,289],

Laura J. Scott[41,42], Elizabeth Selvin[37], Peter Sever[264], Xiao-ou Shu[87], P. Eline Slagboom[115], Kerrin S. Small[90], Blair H. Smith[290], Harold Snieder[35], Tamar Sofer[238,291], Thorkild I. A. Sørensen[7,8,49,292], Tim D. Spector[90], Alice Stanton[293], Claire J. Steves[90,294], Michael Stumvoll[136], Liang Sun[28], Yasuharu Tabara[79], E. Shyong Tai[40,174,295], Nicholas J. Timpson[7,8], Anke Tönjes[136], Jaakko Tuomilehto[296,297,298], Teresa Tusie[77,299], Matti Uusitupa[300], Pim van der Harst[24,118], Cornelia van Duijn[63,85], Veronique Vitart[196], Peter Vollenweider[301], Tanja G. M. Vrijkotte[127], Lynne E. Wagenknecht[302], Mark Walker[303], Ya X. Wang[254], Nick J. Wareham[9], Richard M. Watanabe[123,124,224], Hugh Watkins[12,148], Wen B. Wei[304], Ananda R. Wickremasinghe[305], Gonneke Willemsen[26,27], James F. Wilson[80,196], Tien-Yin Wong[57,191], Jer-Yuarn Wu[59], Anny H. Xiang[306], Lisa R. Yanek[213], Loïc Yengo[307], Mitsuhiro Yokota[308], Eleftheria Zeggini[2,43,309], Wei Zheng[87], Alan B. Zonderman[61], Jerome I. Rotter[21], Anna L. Gloyn[11,12,180,310], Mark I. McCarthy[11,12,180,311,319], Josée Dupuis[25], James B. Meigs[107,238,312], Robert A. Scott[9], Inga Prokopenko[22,29], Aaron Leong[229,313,314], Ching-Ti Liu[25], Stephen C. J. Parker[6,315,321], Karen L. Mohlke[3,321], Claudia Langenberg[9,321], Eleanor Wheeler[2,9,321], Andrew P. Morris[12,316,317,318,321], Inês Barroso[1,2,9,321] ✉ and The Meta-Analysis of Glucose and Insulin-related Traits Consortium (MAGIC)*

[1]Exeter Centre of Excellence for Diabetes Research (EXCEED), Genetics of Complex Traits, University of Exeter Medical School, University of Exeter, Exeter, UK. [2]Department of Human Genetics, Wellcome Sanger Institute, Cambridge, UK. [3]Department of Genetics, University of North Carolina, Chapel Hill, NC, USA. [4]Department of Biostatistics and Epidemiology, University of Massachusetts, Amherst, MA, USA. [5]Inserm, Univ Brest, EFS, UMR 1078, GGB, Brest, France. [6]Department of Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, MI, USA. [7]MRC Integrative Epidemiology Unit, University of Bristol, Bristol, UK. [8]Department of Population Health Sciences, Bristol Medical School, University of Bristol, Bristol, UK. [9]MRC Epidemiology Unit, Institute of Metabolic Science, University of Cambridge, Cambridge, UK. [10]Department of Biostatistics, School of Public Health, Shandong University, Jinan, China. [11]Oxford Centre for Diabetes, Endocrinology and Metabolism, Radcliffe Department of Medicine, University of Oxford, Oxford, UK. [12]Wellcome Centre for Human Genetics, University of Oxford, Oxford, UK. [13]Laboratory for Genomics of Diabetes and Metabolism, RIKEN Centre for Integrative Medical Sciences, Yokohama, Japan. [14]Medical Research Council Human Genetics Unit, Institute for Genetics and Molecular Medicine, Edinburgh, UK. [15]Estonian Genome Center, Institute of Genomics, University of Tartu, Tartu, Estonia. [16]COPSAC, Copenhagen Prospective Studies on Asthma in Childhood, Herlev and Gentofte Hospital, University of Copenhagen, Copenhagen, Denmark. [17]Department of Clinical Epidemiology, Leiden University Medical Center, Leiden, the Netherlands. [18]Department of Epidemiology, Brown University School of Public Health, Brown University, Providence, RI, USA. [19]Department of Biomedical Sciences, City University of Hong Kong, Hong Kong SAR, China. [20]Department of Electrical Engineering, City University of Hong Kong, Hong Kong SAR, China. [21]The Institute for Translational Genomics and Population Sciences, Department of Pediatrics, The Lundquist Institute for Biomedical Innovation at Harbor-UCLA Medical Center, Torrance, CA, USA. [22]Department of Metabolism, Digestion and Reproduction, Imperial College London, London, UK. [23]Division of Preventive Medicine, Brigham and Women's Hospital, Boston, MA, USA. [24]Department of Genetics, University of Groningen, University Medical Center Groningen, Groningen, the Netherlands. [25]Department of Biostatistics, Boston University School of Public Health, Boston, MA, USA. [26]Department of Biological Psychology, Faculty of Behaviour and Movement Sciences, Vrije Universiteit Amsterdam, Amsterdam, the Netherlands. [27]Amsterdam Public Health Research Institute, Amsterdam University Medical Center, Amsterdam, the Netherlands. [28]CAS Key Laboratory of Nutrition, Metabolism and Food Safety, Shanghai Institute of Nutrition and Health, University of Chinese Academy of Sciences, Chinese Academy of Sciences, Shanghai, China. [29]Section of Statistical Multi-omics, Department of Clinical and Experimental Research, University of Surrey, Guildford, UK. [30]Department of Biostatistics, University of Washington, Seattle, WA, USA. [31]SYNLAB Academy, SYNLAB Holding Deutschland GmbH, Mannheim, Germany. [32]Clinical Institute of Medical and Chemical Laboratory Diagnostics, Medical University Graz, Graz, Austria. [33]Vth Department of Medicine (Nephrology, Hypertensiology, Rheumatology, Endocrinology, Diabetology), Medical Faculty Mannheim, Heidelberg University, Mannheim, Baden-Württemberg, Germany. [34]Department of Economics, Metropolitan Autonomous University, Mexico City, Mexico. [35]Department of Epidemiology, University of Groningen, University Medical Center Groningen, Groningen, the Netherlands. [36]CVD Genetic Epidemiology Computational Laboratory, Gillings School of Global Public Health, University of North Carolina, Chapel Hill, NC, USA. [37]Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, USA. [38]The Charles Bronfman Institute for Personalized Medicine, Icahn School of Medicine at Mount Sinai, New York, NY, USA. [39]HPI Digital Health Center, Digital Health and Personalized Medicine, Hasso Plattner Institute, Potsdam, Germany. [40]Saw Swee Hock School of Public Health, National Univeristy of Singapore and National University Health System, Singapore, Singapore. [41]Center for Statistical Genetics, University of Michigan, Ann Arbor, MI, USA. [42]Department of Biostatistics, School of Public Health, University of Michigan, Ann Arbor, MI, USA. [43]Institute of Translational Genomics, Helmholtz Zentrum München–German Research Center for Environmental Health, Neuherberg, Germany. [44]School of Medicine and Public Health, College of Health, Medicine and Wellbeing, The University of Newcastle, Newcastle, New South Wales, Australia. [45]Department of Epidemiology and Biostatistics, Imperial College London, London, UK. [46]Department of Cardiology, Ealing Hospital, London North West Healthcare NHS Trust, London, UK. [47]Steno Diabetes Center Copenhagen, Gentofte, Denmark. [48]The Bioinformatics Centre, Department of Biology, University of Copenhagen, Copenhagen, Denmark. [49]Novo Nordisk Foundation Center for Basic Metabolic Research, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark. [50]Department of Epidemiology, School of Public Health, University of Michigan, Ann Arbor, MI, USA. [51]Department of Medicine, Cardiovascular Health Research Unit, University of Washington, Seattle, WA, USA. [52]Metabolism Program, Program in Medical and Population Genetics, Broad Institute, Cambridge, MA, USA. [53]Department of Clinical Pharmacology, William Harvey Research Institute, Barts and The London School of Medicine and Dentistry, Queen Mary University of London, London, UK. [54]NIHR Barts Cardiovascular Biomedical Research Centre, Queen Mary University of London, London, UK. [55]Department of Medicine, Sleep and Circadian Disorders, Brigham and Women's Hospital, Boston, MA, USA. [56]Department of Medicine, Sleep Medicine, Harvard Medical School, Boston, MA, USA. [57]Ocular

Epidemiology, Singapore Eye Research Institute, Singapore National Eye Centre, Singapore, Singapore. [58]Department of Ophthalmology, National University of Singapore and National University Health System, Singapore, Singapore. [59]Institute of Biomedical Sciences, Academia Sinica, Taipei, Taiwan. [60]Department of Epidemiology, The Herbert Wertheim School of Public Health and Human Longevity Science, University of California San Diego, La Jolla, CA, USA. [61]Laboratory of Epidemiology and Population Sciences, National Institute on Aging, National Institutes of Health, Baltimore, MD, USA. [62]Institute of Population Health Sciences, National Health Research Institutes, Miaoli, Taiwan. [63]Department of Epidemiology, Erasmus Medical Center, Rotterdam, the Netherlands. [64]Department of Statistics, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA. [65]Institute of Cardiovascular Science, University College London, London, UK. [66]Uganda Medical Informatics Centre (UMIC), MRC/UVRI and London School of Hygiene & Tropical Medicine (Uganda Research Unit), Entebbe, Uganda. [67]London School of Hygiene & Tropical Medicine, London, UK. [68]H3Africa Bioinformatics Network (H3ABioNet) Node, Centre for Genomics Research and Innovation, NABDA/FMST, Abuja, Nigeria. [69]Bioinfosol, Sevilla, Spain. [70]Molecular Epidemiology and Science for Life Laboratory, Department of Medical Sciences, Uppsala University, Uppsala, Sweden. [71]Department of Statistics, The University of Auckland, Science Center, Auckland, New Zealand. [72]Postgraduate Program in Epidemiology, Federal University of Pelotas, Pelotas, Brazil. [73]Department of Medicine, Epidemiology, Vanderbilt University Medical Center, Nashville, TN, USA. [74]Department of Epidemiology, Tulane University Obesity Research Center, Tulane University, New Orleans, LA, USA. [75]Department of Epidemiology and Biostatistics, School of Public Health, Peking University, Beijing, China. [76]Molecular Biology and Genomic Medicine Unit, National Council for Science and Technology, Mexico City, Mexico. [77]Molecular Biology and Genomic Medicine Unit, National Institute of Medical Sciences and Nutrition, Mexico City, Mexico. [78]Division of Genome Science, Department of Precision Medicine, National Institute of Health, Cheongju, South Korea. [79]Center for Genomic Medicine, Kyoto University Graduate School of Medicine, Kyoto, Japan. [80]Centre for Global Health Research, Usher Institute, University of Edinburgh, Edinburgh, UK. [81]Centre for Cardiovascular Sciences, Queen's Medical Research Institute, University of Edinburgh, Edinburgh, UK. [82]Department of Medicine, Divison of Biomedical Informatics and Personalized Medicine, University of Colorado Anschutz Medical Campus, Denver, CO, USA. [83]Department of Biostatistics and Data Science, Wake Forest School of Medicine, Winston-Salem, NC, USA. [84]Department of Medicine, Division of Nephrology and Hypertension, University of Utah, Salt Lake City, UT, USA. [85]Nuffield Department of Population Health, University of Oxford, Oxford, UK. [86]Lee Kong Chian School of Medicine, Nanyang Technological University, Singapore, Singapore. [87]Division of Epidemiology, Department of Medicine, Vanderbilt Epidemiology Center, Vanderbilt University Medical Center, Nashville, TN, USA. [88]Department of Clinical Science, Center for Diabetes Research, University of Bergen, Bergen, Norway. [89]Department of Clinical Sciences, Lund University Diabetes Centre, Lund University, Malmo, Sweden. [90]Department of Twin Research and Genetic Epidemiology, School of Life Course Sciences, King's College London, London, UK. [91]NIHR Biomedical Research Centre, Guy's and St Thomas' NHS Foundation Trust, London, UK. [92]Institute of Epidemiology, Research Unit of Molecular Epidemiology, Helmholtz Zentrum München Research Center for Environmental Health, Neuherberg, Germany. [93]German Center for Diabetes Research (DZD), Neuherberg, Germany. [94]Department of Medicine, Division of Endocrinology, Diabetes and Nutrition, University of Maryland School of Medicine, Baltimore, MD, USA. [95]Public Health Informatics Unit, Department of Integrated Sciences, Nagoya University Graduate School of Medicine, Nagoya, Japan. [96]Institute for Biomedicine, Eurac Research, Bolzano, Italy. [97]Department of Internal Medicine, Section of Gerontology and Geriatrics, Leiden University Medical Center, Leiden, the Netherlands. [98]Istituto di Ricerca Genetica e Biomedica (IRGB), Consiglio Nazionale delle Ricerche (CNR), Monserrato, Italy. [99]The Mindich Child Health and Development Institute for Personalized Medicine, Icahn School of Medicine at Mount Sinai, New York, NY, USA. [100]Department of Preventive Medicine, Northwestern University Feinberg School of Medicine, Chicago, IL, USA. [101]Center for Public Health Genomics, University of Virginia, Charlottesville, VA, USA. [102]Department of Public Health Sciences, University of Virginia, Charlottesville, VA, USA. [103]Department of Computational Biology, University of Lausanne, Lausanne, Switzerland. [104]Swiss Institute of Bioinformatics, Lausanne, Switzerland. [105]Center for Genomic Medicine, Massachusetts General Hospital, Harvard Medical School, Boston, MA, USA. [106]Department of Anesthesia, Critical Care and Pain Medicine, Massachusetts General Hospital, Boston, MA, USA. [107]Program in Medical and Population Genetics, Broad Institute, Cambridge, MA, USA. [108]Department of Cell and Molecular Biology, National Bioinformatics Infrastructure Sweden, Science for Life Laboratory, Uppsala University, Uppsala, Sweden. [109]Department of Biostatistics, University of Michigan, Ann Arbor, MI, USA. [110]Icelandic Heart Association, Kopavogur, Iceland. [111]Institute of Health and Wellbeing, University of Glasgow, Glasgow, UK. [112]Department of Medicine Solna, Cardiovascular Medicine, Karolinska Institutet, Stockholm, Sweden. [113]National Center for Global Health and Medicine, Tokyo, Japan. [114]Department of Cardiology, Leiden University Medical Center, Leiden, the Netherlands. [115]Department of Biomedical Data Sciences, Molecular Epidemiology, Leiden University Medical Center, Leiden, the Netherlands. [116]Department of Pattern Recognition and Bioinformatics, Delft University of Technology, Delft, the Netherlands. [117]Department of Biomedical Data Sciences, Leiden Computational Biology Center, Leiden University Medical Center, Leiden, the Netherlands. [118]Department of Cardiology, University of Groningen, University Medical Center Groningen, Groningen, the Netherlands. [119]Genomics PLC, Oxford, UK. [120]Center of Pediatric Research, University Children's Hospital Leipzig, University of Leipzig Medical Center, Leipzig, Germany. [121]Department of Epidemiology and Biostatistics, School of Public Health, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China. [122]Genome Institute of Singapore, Agency for Science, Technology and Research, Singapore, Singapore. [123]Department of Preventive Medicine, Keck School of Medicine of University of Southern California, Los Angeles, CA, USA. [124]University of Southern California Diabetes and Obesity Research Institute, Keck School of Medicine of University of Southern California, Los Angeles, CA, USA. [125]Department of Community Medicine, Faculty of Health Sciences, UIT the Arctic University of Norway, Tromsø, Norway. [126]MRC Unit for Lifelong Health and Ageing at University College London, London, UK. [127]Department of Public Health, Amsterdam Public Health Research Institute, Amsterdam University Medical Center, Amsterdam, the Netherlands. [128]Department of Clinical Epidemiology, Biostatistics, and Bioinformatics, Amsterdam Public Health Research Institute, Amsterdam University Medical Center, Amsterdam, the Netherlands. [129]Department of Public Health and Primary Care, School of Clinical Medicine, University of Cambridge, Cambridge, UK. [130]Department of Nutrition, Exercise, and Sports, Faculty of Science, University of Copenhagen, Copenhagen, Denmark. [131]Department of Internal Medicine, University of Groningen, University Medical Center Groningen, Groningen, the Netherlands. [132]Department of Medical Biotechnology and Translational Medicine, University of Milan, Milan, Italy. [133]Centro Cardiologico Monzino, IRCCS, Milan, Italy. [134]Diabetes and Obesity Research Institute, Cedars-Sinai Medical Center, Los Angeles, CA, USA. [135]Department of Epidemiology and Prevention, Division of Public Health Sciences, Wake Forest School of Medicine, Winston-Salem, NC, USA. [136]Medical Department III–Endocrinology, Nephrology, Rheumatology, University of Leipzig Medical Center, Leipzig, Germany. [137]Medical Genomics and Metabolic Genetics Branch, National Human Genome Research Institute, National Institues of Health, Bethesda, MD, USA. [138]Department for Prevention and Care of Diabetes, Faculty of Medicine Carl Gustav Carus, Technische Universität Dresden, Dresden, Germany. [139]Department of Biochemistry, Wake Forest School of Medicine, Winston-Salem, NC, USA. [140]Centre for Genomic and Experimental Medicine, Institute of Genetics and Molecular Medicine, University of Edinburgh, Western General Hospital, Edinburgh, UK. [141]Usher Institute, University of Edinburgh, Edinburgh, UK. [142]Department of Internal Medicine, National Taiwan University Hospital, Taipei, Taiwan. [143]Graduate Institute of Medical Genomics and Proteomics, National Taiwan University, Taipei, Taiwan. [144]Department of Nutrition, Gillings School of Global Public Health, University of North Carolina, Chapel Hill, NC, USA. [145]Department of Population Science and Experimental Medicine, Institute of Cardiovascular Science, University College London, London, UK. [146]Department of Nutrition and Dietetics, School of Health Science and Education, Harokopio University of Athens, Athens, Greece. [147]Department of Epidemiology, Shanghai Cancer Institute, Shanghai, China. [148]Division of Cardiovascular Medicine, Radcliffe Department of Medicine, University of Oxford, Oxford, UK. [149]Department of Psychiatry, Interdisciplinary Center Psychopathy and Emotion Regulation, University of Groningen, University Medical Center Groningen, Groningen, the Netherlands. [150]Institute for Clinical Diabetology, German Diabetes Center, Leibniz Center for Diabetes Research at Heinrich Heine University Düsseldorf, Düsseldorf,

Germany. [151]Division of Endocrinology and Diabetology, Medical Faculty, Heinrich Heine University Düsseldorf, Düsseldorf, Germany. [152]German Center for Diabetes Research (DZD), Düsseldorf, Germany. [153]Internal Medicine, Endocrine and Metabolism, Tri-Service General Hospital, Taipei, Taiwan. [154]School of Medicine, National Defense Medical Center, Taipei, Taiwan. [155]Internal Medicine, Endocrinology, Diabetes and Metabolism, Diabetes and Metabolism Research Center, The Ohio State University Wexner Medical Center, Columbus, OH, USA. [156]Department of Environmental and Preventive Medicine, Jichi Medical University School of Medicine, Shimotsuke, Japan. [157]Department of Anti-aging Medicine, Ehime University Graduate School of Medicine, Toon, Japan. [158]National Institute of Public Health, University of Southern Denmark, Odense, Denmark. [159]Department of Medicine, Endocrinology, Diabetes and Metabolism, Johns Hopkins University School of Medicine, Baltimore, MD, USA. [160]Clinical Diabetes, Endocrinology and Metabolism, Translational Research and Cellular Therapeutics, Beckman Research Institute of the City of Hope, Duarte, CA, USA. [161]Department of Clinical Gene Therapy, Osaka University Graduate School of Medicine, Suita, Japan. [162]Department of Geriatric and General Medicine, Osaka University Graduate School of Medicine, Suita, Japan. [163]Department of Public Health, University of Split School of Medicine, Split, Croatia. [164]Institute of Biomedicine, Bioinformatics Center, Univeristy of Eastern Finland, Kuopio, Finland. [165]Department of Medicine, University of Eastern Finland and Kuopio University Hospital, Kuopio, Finland. [166]USC-Office of Population Studies Foundation, University of San Carlos, Cebu City, the Philippines. [167]Department of Anthropology, Sociology and History, University of San Carlos, Cebu City, the Philippines. [168]State Key Laboratory of Oncogene and Related Genes and Department of Epidemiology, Shanghai Cancer Institute, Renji Hospital, Shanghai Jiaotong University School of Medicine, Shanghai, China. [169]Center for Geriatrics and Gerontology, Taichung Veterans General Hospital, Taichung, Taiwan. [170]National Defense Medical Center, National Yang-Ming University, Taipei, Taiwan. [171]Diabetes Prevention Unit, National Institute for Health and Welfare, Helsinki, Finland. [172]Center for Clinical Research and Prevention, Bispebjerg and Frederiksberg Hospital, Copenhagen, Denmark. [173]Department of Clinical Medicine, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark. [174]Yong Loo Lin School of Medicine, National University of Singapore and National University Health System, Singapore, Singapore. [175]Department of Medicine, University of Texas Health Sciences Center, San Antonio, TX, USA. [176]Department of Internal Medicine, Aichi Gakuin University School of Dentistry, Nagoya, Japan. [177]Department of Diabetes, Diabetes, and Nutritional Sciences, James Black Centre, King's College London, London, UK. [178]Department of Functional Pathology, Shimane University School of Medicine, Izumo, Japan. [179]Department of Medicine and Pharmacology, New York Medical College School of Medicine, Valhalla, NY, USA. [180]Oxford NIHR Biomedical Research Centre, Oxford University Hospitals NHS Foundation Trust, Oxford, UK. [181]Colorado School of Public Health, University of Colorado Anschutz Medical Campus, Aurora, CO, USA. [182]Department of Geriatric Medicine and Neurology, Ehime University Graduate School of Medicine, Toon, Japan. [183]Institute of Epidemiology, Helmholtz Zentrum München Research Center for Environmental Health, Neuherberg, Germany. [184]Institute for Medical Information Processing, Biometry and Epidemiology, Ludwig-Maximilians University Munich, Munich, Germany. [185]Gen-Info, Zagreb, Croatia. [186]Department of Epidemiology and Population Health, Albert Einstein College of Medicine, New York, NY, USA. [187]Genetics and Pharmacogenomics, Merck Sharp & Dohme, Kenilworth, NJ, USA. [188]Department of Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, WA, USA. [189]Department of Internal Medicine, Erasmus Medical Center, Rotterdam, the Netherlands. [190]Centre for Global Health, The Usher Institute, University of Edinburgh, Edinburgh, UK. [191]Ophthalmology & Visual Sciences Academic Clinical Program (Eye ACP), Duke-NUS Medical School, Singapore, Singapore. [192]BHF Glasgow Cardiovascular Research Centre, Institute of Cardiovascular and Medical Sciences, University of Glasgow, Glasgow, UK. [193]Department of Experimental Diabetology, German Institute of Human Nutrition Potsdam-Rehbruecke, Nuthetal, Germany. [194]Department of Genetics, Shanghai-MOST Key Laboratory of Health and Disease Genomics, Chinese National Human Genome Center at Shanghai (CHGC) and Shanghai Academy of Science & Technology (SAST), Shanghai, China. [195]Sarepta Therapeutics, Cambridge, MA, USA. [196]Medical Research Council Human Genetics Unit, Institute for Genetics and Cancer, University of Edinburgh, Edinburgh, UK. [197]Department of Laboratory Medicine and Pathology, University of Minnesota, Minneapolis, MN, USA. [198]Department of Nutrition, Harvard T. H. Chan School of Public Health, Boston, MA, USA. [199]Institute of Clinical Medicine, Internal Medicine, University of Eastern Finland, Kuopio, Finland. [200]Department of Medicine, Bornholms Hospital, Rønne, Denmark. [201]Department of Internal Medicine, Division of Endocrinology, Leiden University Medical Center, Leiden, the Netherlands. [202]Laboratory for Experimental Vascular Medicine, Leiden University Medical Center, Leiden, the Netherlands. [203]Department of Human Genetics, Leiden University Medical Center, Leiden, the Netherlands. [204]Department of Human Biology, University of Split School of Medicine, Split, Croatia. [205]Carolina Population Center, University of North Carolina, Chapel Hill, NC, USA. [206]Department of Endocrinology and Metabolism, Instituto Nacional de Ciencias Medicas y Nutricion, Mexico City, Mexico. [207]Unidad de Investigación de Enfermedades Metabólicas, Instituto Nacional de Ciencias Médicas y Nutrición and Tec Salud, Mexico City, Mexico. [208]Instituto Tecnológico y de Estudios Superiores de Monterrey Tec Salud, Monterrey, Mexico. [209]Department of Medical Genomics, Pfizer/University of Granada/Andalusian Government Center for Genomics and Oncological Research (GENYO), Granada, Spain. [210]Institute for Environmental Medicine, Chronic Inflammatory Diseases, Karolinska Institutet, Solna, Sweden. [211]Department of Genetics, Division of Statistical Genomics, Washington University School of Medicine, St Louis, MO, USA. [212]Clinical and Health Services Research, National Institute on Minority Health and Health Disparities, Bethesda, MD, USA. [213]Department of Medicine, General Internal Medicine, Johns Hopkins University School of Medicine, Baltimore, MD, USA. [214]Medical School, Royal Perth Hospital Unit, University of Western Australia, Perth, Western Australia, Australia. [215]Department of Integrative Biomedical Sciences, University of Cape Town, Cape Town, South Africa. [216]Aberdeen Centre for Health Data Science, School of Medicine, Medical Sciences and Nutrition, University of Aberdeen, Aberdeen, UK. [217]Human Genetics Center, School of Public Health, The University of Texas Health Science Center at Houston, Houston, TX, USA. [218]Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX, USA. [219]Division of Endocrinology and Diabetes, Graduate School of Molecular Endocrinology and Diabetes, University of Ulm, Ulm, Germany. [220]LKC School of Medicine, Nanyang Technological University, Singapore and Imperial College London, UK, Singapore, Singapore. [221]Hasso Plattner Institute for Digital Health at Mount Sinai, Icahn School of Medicine at Mount Sinai, New York, NY, USA. [222]Digital Health Center, Hasso Plattner Institut, University Potsdam, Potsdam, Germany. [223]Department of Medicine, Keck School of Medicine of University of Southern California, Los Angeles, CA, USA. [224]Department of Physiology and Neuroscience, Keck School of Medicine of University of Southern California, Los Angeles, CA, USA. [225]INSERM UMR 1283/CNRS UMR 8199, European Institute for Diabetes (EGID), Université de Lille, Lille, France. [226]INSERM UMR 1283/CNRS UMR 8199, European Institute for Diabetes (EGID), Institut Pasteur de Lille, Lille, France. [227]Imperial College Healthcare NHS Trust, Imperial College London, London, UK. [228]MRC-PHE Centre for Environment and Health, Imperial College London, London, UK. [229]Harvard Medical School, Boston, MA, USA. [230]Department of Medicine, Jackson Heart Study, University of Mississippi Medical Center, Jackson, MS, USA. [231]Department of Medicine, Faculty of Medicine, University of Kelaniya, Ragama, Sri Lanka. [232]Department of Nutrition and Dietetics, School of Health Science and Education, Harokopio University of Athens, Kallithea, Greece. [233]Department of Clinical Sciences, Lund University, Malmö, Sweden. [234]Laboratory of Epidemiology and Population Sciences, National Institute on Aging Intramural Research Program, National Institutes of Health, Baltimore, MD, USA. [235]CNR Institute of Clinical Physiology, Pisa, Italy. [236]Intramural Research Program, National Institute of Aging, Baltimore, MD, USA. [237]Diabetes Unit and Center for Genomic Medicine, Massachusetts General Hospital, Boston, MA, USA. [238]Department of Medicine, Harvard Medical School, Boston, MA, USA. [239]Department of Public Health and Clinical Medicine, Umeå University, Umeå, Sweden. [240]Department of Genomics of Common Disease, Imperial College London, London, UK. [241]Department of Medicine, Cardiovascular Medicine, Karolinska Institutet, Stockholm, Sweden. [242]Department of Medicine, Division of Endocrinology, Diabetes and Metabolism, Cedars-Sinai Medical Center, Los Angeles, CA, USA. [243]Diabetes Centre, Lund University, Lund, Sweden. [244]Finnish Institute of Molecular Medicine, Helsinki University, Helsinki, Finland. [245]Faculty of Medicine, School of Health Sciences, University of Iceland, Reykjavik, Iceland. [246]Department of Epidemiology, Cardiovascular Health Research Unit, University of Washington, Seattle, WA, USA. [247]Department of Medicine, Division of Cardiovascular Medicine, Stanford University School of

Medicine, Stanford University, Stanford, CA, USA. [248]Division of Epidemiology and Community Health, University of Minnesota, Minneapolis, MN, USA. [249]Department of Epidemiology and Biostatistics, MRC-PHE Centre for Environment and Health, School of Public Health, Imperial College London, London, UK. [250]Center for Life Course Health Research, Faculty of Medicine, University of Oulu, Oulu, Finland. [251]Unit of Primary Health Care, Oulu Univerisity Hospital, OYS, Oulu, Finland. [252]Department of Life Sciences, College of Health and Life Sciences, Brunel University London, London, UK. [253]Department of Ophthalmology, Medical Faculty Mannheim, Heidelberg University, Mannheim, Germany. [254]Beijing Institute of Ophthalmology, Beijing Ophthalmology and Visual Science Key Lab, Beijing Tongren Eye Center, Beijing Tongren Hospital, Capital Medical University, Beijing, China. [255]Institute of Molecular and Clinical Ophthalmology Basel IOB, Basel, Switzerland. [256]Netherlands Heart Institute, Utrecht, the Netherlands. [257]MRC/UVRI and LSHTM (Uganda Research Unit), Entebbe, Uganda. [258]Faculty of Medicine, Institute of Health Sciences, University of Oulu, Oulu, Finland. [259]Unit of General Practice, Oulu University Hospital, Oulu, Finland. [260]Department of Epidemiology and Public Health, University College London, London, UK. [261]Department of Public Health Solutions, Finnish Institute for Health and Welfare, Helsinki, Finland. [262]Department of Medicine, University of Helsinki and Helsinki University Central Hospital, Helsinki, Finland. [263]Minerva Foundation Institute for Medical Research, Helsinki, Finland. [264]National Heart and Lung Institute, Imperial College London, London, UK. [265]IFB Adiposity Diseases, University of Leipzig Medical Center, Leipzig, Germany. [266]Institute for Social and Economic Research, University of Essex, Colchester, UK. [267]University Institute of Primary Care and Public Health, Division of Biostatistics, University of Lausanne, Lausanne, Switzerland. [268]Institute of Biomedicine, School of Medicine, University of Eastern Finland, Kuopio, Finland. [269]Department of Clinical Physiology and Nuclear Medicine, Kuopio University Hospital, Kuopio, Finland. [270]Foundation for Research in Health Exercise and Nutrition, Kuopio Research Institute of Exercise Medicine, Kuopio, Finland. [271]Institute of Environmental Medicine, Cardiovascular and Nutritional Epidemiology, Karolinska Institutet, Stockholm, Sweden. [272]Department of Medical Sciences, University of Uppsala, Uppsala, Sweden. [273]Big Data Institute, Nuffield Department of Medicine, University of Oxford, Oxford, UK. [274]Nuffield Department of Women's and Reproductive Health, University of Oxford, Oxford, UK. [275]Department of Medical Epidemiology and Biostatistics and the Swedish Twin Registry, Karolinska Institutet, Stockholm, Sweden. [276]Department of Public Health and Primary Care, Leiden University Medical Center, Leiden, the Netherlands. [277]Institute of Cardiovascular and Medical Sciences, University of Glasgow, Glasgow, UK. [278]Division of Population Health and Genomics, School of Medicine, University of Dundee, Ninewells Hospital and Medical School, Dundee, UK. [279]Centre for Cognitive Ageing and Cognitive Epidemiology, University of Edinburgh, Edinburgh, UK. [280]Department of Health Services, Cardiovascular Health Research Unit, University of Washington, Seattle, WA, USA. [281]Department of Epidemiology, Tulane University School of Public Health and Tropical Medicine, New Orleans, LA, USA. [282]Department of Pediatrics, Genetic and Genomic Medicine, University of California, Irvine, Irvine, CA, USA. [283]Havard Medical School, Boston, MA, USA. [284]Tampere, Finnish Diabetes Association, Tampere, Finland. [285]Pirkanmaa Hospital District, Tampere, Finland. [286]Department of Medicine, University of Cambridge, Cambridge, UK. [287]South Karelia Central Hospital, Lappeenranta, Finland. [288]Department of Psychology, University of Miami, Miami, FL, USA. [289]Paul Langerhans Institute Dresden of the Helmholtz Center Munich, University Hospital and Faculty of Medicine, Dresden, Germany. [290]Division of Population Health and Genomics, Ninewells Hospital and Medical School, University of Dundee, Dundee, UK. [291]Division of Sleep and Circadian Disorders, Brigham and Women's Hospital, Boston, MA, USA. [292]Department of Public Health, Section of Epidemiology, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark. [293]Department of Molecular and Cellular Therapeutics, Royal College of Surgeons in Ireland, Dublin, Ireland. [294]Department of Ageing and Health, Guy's and St Thomas' NHS Foundation Trust, London, UK. [295]Cardiovascular and Metabolic Disease Signature Research Program, Duke-NUS Medical School, Singapore, Singapore. [296]Department of Public Health Solutions, National Institute for Health and Welfare, Helsinki, Finland. [297]Department of Public Health, University of Helsinki, Helsinki, Finland. [298]Saudi Diabetes Research Group, King Abdulaziz University, Jeddah, Saudi Arabia. [299]Department of Genomic Medicine and Environmental Toxicology, Instituto de Investigaciones Biomedicas, Universidad Nacional Autonoma de Mexico, Mexico City, Mexico. [300]Department of Public Health and Clinical Nutrition, University of Eastern Finland, Kuopio, Finland. [301]Department of Medicine, Internal Medicine, Lausanne University Hospital (CHUV), Lausanne, Switzerland. [302]Department of Public Health Sciences, Wake Forest School of Medicine, Winston-Salem, NC, USA. [303]Faculty of Medical Sciences, Newcastle University, Newcastle upon Tyne, UK. [304]Beijing Tongren Eye Center, Beijing Key Laboratory of Intraocular Tumor Diagnosis and Treatment, Beijing Ophthalmology & Visual Sciences Key Lab, Beijing Tongren Hospital, Capital Medical University, Beijing, China. [305]Department of Public Health, Faculty of Medicine, University of Kelaniya, Ragama, Sri Lanka. [306]Department of Research and Evaluation, Kaiser Permanente of Southern California, Pasadena, CA, USA. [307]Institute for Molecular Bioscience, The University of Queensland, St Lucia, Queensland, Australia. [308]Kurume University School of Medicine, Kurume, Japan. [309]TUM School of Medicine, Technical University of Munich and Klinikum Rechts der Isar, Munich, Germany. [310]Department of Pediatrics, Division of Endocrinology, Stanford School of Medicine, Stanford, CA, USA. [311]Wellcome Centre for Human Genetics, Nuffield Department of Medicine, University of Oxford, Oxford, UK. [312]Department of Medicine, Division of General Internal Medicine, Massachusetts General Hospital, Boston, MA, USA. [313]Department of Medicine, General Internal Medicine, Massachusetts General Hospital, Boston, MA, USA. [314]Department of Medicine, Diabetes Unit and Endocrine Unit, Massachusetts General Hospital, Boston, MA, USA. [315]Department of Human Genetics, University of Michigan, Ann Arbor, MI, USA. [316]Centre for Genetics and Genomics Versus Arthritis, Division of Musculoskeletal and Dermatological Sciences, The University of Manchester, Manchester, UK. [317]Centre for Musculoskeletal Research, Division of Musculoskeletal and Dermatological Sciences, The University of Manchester, Manchester, UK. [318]Department of Biostatistics, University of Liverpool, Liverpool, UK. [319]Present address: Genentech, South San Francisco, CA, USA. [320]These authors contributed equally: Ji Chen, Cassandra N. Spracklen, Gaëlle Marenne, Arushi Varshney, Laura J. Corbin. [321]These authors jointly supervised this work: Stephen C. J. Parker, Karen L. Mohlke, Claudia Langenberg, Eleanor Wheeler, Andrew P. Morris, Inês Barroso. *Lists of authors and their affiliations appear at the end of the paper. ✉e-mail: ines.barroso@exeter.ac.uk

Lifelines Cohort Study

Ilja M. Nolte[35], Tian Xie[35], Jana V. van Vliet-Ostaptchouk[35] and Harold Snieder[35]

The Meta-Analysis of Glucose and Insulin-related Traits Consortium (MAGIC)

Ji Chen[1,2,320], Cassandra N. Spracklen[3,4,320], Gaëlle Marenne[2,5,320], Arushi Varshney[6,320], Laura J. Corbin[7,8,320], Jian'an Luan[9], Sara M. Willems[9], Ying Wu[3], Xiaoshuai Zhang[9,10], Momoko Horikoshi[11,12,13], Thibaud S. Boutin[14], Reedik Mägi[15], Johannes Waage[16], Ruifang Li-Gao[17], Kei Hang Katie Chan[18,19,20], Jie Yao[21], Mila D. Anasanti[22], Audrey Y. Chu[23], Annique Claringbould[24],

Jani Heikkinen[22], Jaeyoung Hong[25], Jouke-Jan Hottenga[26,27], Shaofeng Huo[28], Marika A. Kaakinen[22,29], Tin Louie[30], Winfried März[31,32,33], Hortensia Moreno-Macias[34], Anne Ndungu[12], Sarah C. Nelson[30], Ilja M. Nolte[35], Kari E. North[36], Chelsea K. Raulerson[3], Debashree Ray[37], Rebecca Rohde[36], Denis Rybin[25], Claudia Schurmann[38,39], Xueling Sim[40,41,42], Lorraine Southam[2,43], Isobel D. Stewart[9], Carol A. Wang[44], Yujie Wang[36], Peitao Wu[25], Weihua Zhang[45,46], Tarunveer S. Ahluwalia[16,47,48], Emil V. R. Appel[49], Lawrence F. Bielak[50], Jennifer A. Brody[51], Noël P. Burtt[52], Claudia P. Cabrera[53,54], Brian E. Cade[55,56], Jin Fang Chai[40], Xiaoran Chai[57,58], Li-Ching Chang[59], Chien-Hsiun Chen[59], Brian H. Chen[60], Kumaraswamy Naidu Chitrala[61], Yen-Feng Chiu[62], Hugoline G. de Haan[17], Graciela E. Delgado[33], Ayse Demirkan[29,63], Qing Duan[3,64], Jorgen Engmann[65], Segun A. Fatumo[66,67,68], Javier Gayán[69], Franco Giulianini[23], Jung Ho Gong[18], Stefan Gustafsson[70], Yang Hai[71], Fernando P. Hartwig[7,72], Jing He[73], Yoriko Heianza[74], Tao Huang[75], Alicia Huerta-Chagoya[76,77], Mi Yeong Hwang[78], Richard A. Jensen[51], Takahisa Kawaguchi[79], Katherine A. Kentistou[80,81], Young Jin Kim[78], Marcus E. Kleber[33], Ishminder K. Kooner[46], Shuiqing Lai[18], Leslie A. Lange[82], Carl D. Langefeld[83], Marie Lauzon[21], Man Li[84], Symen Ligthart[63], Jun Liu[63,85], Marie Loh[45,86], Jirong Long[87], Valeriya Lyssenko[88,89], Massimo Mangino[90,91], Carola Marzi[92,93], May E. Montasser[94], Abhishek Nag[12], Masahiro Nakatochi[95], Damia Noce[96], Raymond Noordam[97], Giorgio Pistis[98], Michael Preuss[38,99], Laura Raffield[3], Laura J. Rasmussen-Torvik[100], Stephen S. Rich[101,102], Neil R. Robertson[11,12], Rico Rueedi[103,104], Kathleen Ryan[94], Serena Sanna[24,98], Richa Saxena[105,106,107], Katharina E. Schraut[80,81], Bengt Sennblad[108], Kazuya Setoh[79], Albert V. Smith[109,110], Thomas Sparsø[49], Rona J. Strawbridge[111,112], Fumihiko Takeuchi[113], Jingyi Tan[21], Stella Trompet[97,114], Erik van den Akker[115,116,117], Peter J. van der Most[35], Niek Verweij[118,119], Mandy Vogel[120], Heming Wang[55,56], Chaolong Wang[121,122], Nan Wang[123,124], Helen R. Warren[53,54], Wanqing Wen[87], Tom Wilsgaard[125], Andrew Wong[126], Andrew R. Wood[1], Tian Xie[35], Mohammad Hadi Zafarmand[127,128], Jing-Hua Zhao[129], Wei Zhao[50], Najaf Amin[63,85], Zorayr Arzumanyan[21], Arne Astrup[130], Stephan J. L. Bakker[131], Damiano Baldassarre[132,133], Marian Beekman[115], Richard N. Bergman[134], Alain Bertoni[135], Matthias Blüher[136], Lori L. Bonnycastle[137], Stefan R. Bornstein[138], Donald W. Bowden[139], Qiuyin Cai[73], Archie Campbell[140,141], Harry Campbell[80], Yi Cheng Chang[59,142,143], Eco J. C. de Geus[26,27], Abbas Dehghan[63], Shufa Du[144], Gudny Eiriksdottir[110], Aliki Eleni Farmaki[145,146], Mattias Frånberg[112], Christian Fuchsberger[96], Yutang Gao[147], Anette P. Gjesing[49], Anuj Goel[12,148], Sohee Han[78], Catharina A. Hartman[149], Christian Herder[150,151,152], Andrew A. Hicks[96], Chang-Hsun Hsieh[153,154], Willa A. Hsueh[155], Sahoko Ichihara[156], Michiya Igase[157], M. Arfan Ikram[63], W. Craig Johnson[30], Marit E. Jørgensen[47,158], Peter K. Joshi[80], Rita R. Kalyani[159], Fouad R. Kandeel[160], Tomohiro Katsuya[161,162], Chiea Chuen Khor[122], Wieland Kiess[120], Ivana Kolcic[163], Teemu Kuulasmaa[164], Johanna Kuusisto[165], Kristi Läll[15], Kelvin Lam[21], Deborah A. Lawlor[7,8], Nanette R. Lee[166,167], Rozenn N. Lemaitre[51], Honglan Li[168], Shih-Yi Lin[169,170], Jaana Lindström[171], Allan Linneberg[172,173], Jianjun Liu[122,174], Carlos Lorenzo[175], Tatsuaki Matsubara[176], Fumihiko Matsuda[79], Geltrude Mingrone[177], Simon Mooijaart[97], Sanghoon Moon[78], Toru Nabika[178], Girish N. Nadkarni[38], Jerry L. Nadler[179], Mari Nelis[15], Matt J. Neville[11,180], Jill M. Norris[181], Yasumasa Ohyagi[182], Annette Peters[93,183,184], Patricia A. Peyser[50], Ozren Polasek[163,185], Qibin Qi[186], Dennis Raven[149], Dermot F. Reilly[187], Alex Reiner[188], Fernando Rivideneira[189], Kathryn Roll[21], Igor Rudan[190], Charumathi Sabanayagam[57,191], Kevin Sandow[21], Naveed Sattar[192], Annette Schürmann[93,193], Jinxiu Shi[194], Heather M. Stringham[41,42], Kent D. Taylor[21], Tanya M. Teslovich[195], Betina Thuesen[172], Paul R. H. J. Timmers[80,196], Elena Tremoli[133], Michael Y. Tsai[197], Andre Uitterlinden[189], Rob M. van Dam[40,174,198], Diana van Heemst[97],

Astrid van Hylckama Vlieg[17], Jana V. van Vliet-Ostaptchouk[35], Jagadish Vangipurapu[199], Henrik Vestergaard[49,200], Tao Wang[186], Ko van Willems van Dijk[201,202,203], Tatijana Zemunik[204], Gonçalo R. Abecasis[42], Linda S. Adair[144,205], Carlos Alberto Aguilar-Salinas[206,207,208], Marta E. Alarcón-Riquelme[209,210], Ping An[211], Larissa Aviles-Santa[212], Diane M. Becker[213], Lawrence J. Beilin[214], Sven Bergmann[103,104,215], Hans Bisgaard[16], Corri Black[216], Michael Boehnke[41,42], Eric Boerwinkle[217,218], Bernhard O. Böhm[219,220], Klaus Bønnelykke[16], D. I. Boomsma[26,27], Erwin P. Bottinger[38,221,222], Thomas A. Buchanan[124,223,224], Mickaël Canouil[225,226], Mark J. Caulfield[53,54], John C. Chambers[45,46,86,227,228], Daniel I. Chasman[23,229], Yii-Der Ida Chen[21], Ching-Yu Cheng[57,191], Francis S. Collins[137], Adolfo Correa[230], Francesco Cucca[98], H. Janaka de Silva[231], George Dedoussis[232], Sölve Elmståhl[233], Michele K. Evans[234], Ele Ferrannini[235], Luigi Ferrucci[236], Jose C. Florez[107,237,238], Paul W. Franks[89,239], Timothy M. Frayling[1], Philippe Froguel[225,226,240], Bruna Gigante[241], Mark O. Goodarzi[242], Penny Gordon-Larsen[144,205], Harald Grallert[92,93], Niels Grarup[49], Sameline Grimsgaard[125], Leif Groop[243,244], Vilmundur Gudnason[110,245], Xiuqing Guo[21], Anders Hamsten[112], Torben Hansen[49], Caroline Hayward[196], Susan R. Heckbert[246], Bernardo L. Horta[72], Wei Huang[194], Erik Ingelsson[247], Pankow S. James[248], Marjo-Ritta Jarvelin[249,250,251,252], Jost B. Jonas[253,254,255], J. Wouter Jukema[114,256], Pontiano Kaleebu[257], Robert Kaplan[186,188], Sharon L. R. Kardia[50], Norihiro Kato[113], Sirkka M. Keinanen-Kiukaanniemi[258,259], Bong-Jo Kim[78], Mika Kivimaki[260], Heikki A. Koistinen[261,262,263], Jaspal S. Kooner[46,227,228,264], Antje Körner[120], Peter Kovacs[136,265], Diana Kuh[126], Meena Kumari[266], Zoltan Kutalik[104,267], Markku Laakso[165], Timo A. Lakka[268,269,270], Lenore J. Launer[61], Karin Leander[271], Huaixing Li[28], Xu Lin[28], Lars Lind[272], Cecilia Lindgren[12,273,274], Simin Liu[18], Ruth J. F. Loos[38,99], Patrik K. E. Magnusson[275], Anubha Mahajan[12,319], Andres Metspalu[15], Dennis O. Mook-Kanamori[17,276], Trevor A. Mori[214], Patricia B. Munroe[53,54], Inger Njølstad[125], Jeffrey R. O'Connell[94], Albertine J. Oldehinkel[149], Ken K. Ong[9], Sandosh Padmanabhan[277], Colin N. A. Palmer[278], Nicholette D. Palmer[139], Oluf Pedersen[49], Craig E. Pennell[44], David J. Porteous[140,279], Peter P. Pramstaller[96], Michael A. Province[211], Bruce M. Psaty[51,246,280], Lu Qi[281], Leslie J. Raffel[282], Rainer Rauramaa[270], Susan Redline[55,56], Paul M. Ridker[23,283], Frits R. Rosendaal[17], Timo E. Saaristo[284,285], Manjinder Sandhu[286], Jouko Saramies[287], Neil Schneiderman[288], Peter Schwarz[93,138,289], Laura J. Scott[41,42], Elizabeth Selvin[37], Peter Sever[264], Xiao-ou Shu[87], P. Eline Slagboom[115], Kerrin S. Small[90], Blair H. Smith[290], Harold Snieder[35], Tamar Sofer[238,291], Thorkild I. A. Sørensen[7,8,49,292], Tim D. Spector[90], Alice Stanton[293], Claire J. Steves[90,294], Michael Stumvoll[136], Liang Sun[28], Yasuharu Tabara[79], E. Shyong Tai[40,174,295], Nicholas J. Timpson[7,8], Anke Tönjes[136], Jaakko Tuomilehto[296,297,298], Teresa Tusie[77,299], Matti Uusitupa[300], Pim van der Harst[24,118], Cornelia van Duijn[63,85], Veronique Vitart[196], Peter Vollenweider[301], Tanja G. M. Vrijkotte[127], Lynne E. Wagenknecht[302], Mark Walker[303], Ya X. Wang[254], Nick J. Wareham[9], Richard M. Watanabe[123,124,224], Hugh Watkins[12,148], Wen B. Wei[304], Ananda R. Wickremasinghe[305], Gonneke Willemsen[26,27], James F. Wilson[80,196], Tien-Yin Wong[57,191], Jer-Yuarn Wu[59], Anny H. Xiang[306], Lisa R. Yanek[213], Loïc Yengo[307], Mitsuhiro Yokota[308], Eleftheria Zeggini[2,43,309], Wei Zheng[87], Alan B. Zonderman[61], Jerome I. Rotter[21], Anna L. Gloyn[11,12,180,310], Mark I. McCarthy[11,12,180,311,319], Josée Dupuis[25], James B. Meigs[107,238,312], Robert A. Scott[9], Inga Prokopenko[22,29], Aaron Leong[229,313,314], Ching-Ti Liu[25], Stephen C. J. Parker[6,315,321], Karen L. Mohlke[3,321], Claudia Langenberg[9,321], Eleanor Wheeler[2,9,321], Andrew P. Morris[12,316,317,318,321] and Inês Barroso[1,2,9,321]

Methods

Study design and participants. This study included trait data from four glycemic traits: FG, FI, 2hGlu and glycated HbA1c. The total number of contributing cohorts ranged from 41 (2hGlu) to 131 (FG), and the maximum sample size for each trait ranged from 85,916 (2hGlu) to 281,416 (FG) (Supplementary Table 1). Self-identified ancestry was initially defined at the cohort level, but within each cohort ancestry was confirmed with genetic data with ancestry outliers removed (Supplementary Table 1). Overall, participants of European ancestry dominated the sample size for all traits, representing between 68.0% (HbA1c) and 73.8% (2hGlu) of the overall sample size. Individuals of African American ancestry represented between 1.7% (2hGlu) and 5.9% (FG) of participants; individuals of Hispanic ancestry represented between 6.8% (FG) and 14.6% (2hGlu) of participants; individuals of East Asian ancestry represented between 9.9% (2hGlu) and 15.4% (HbA1c) of participants; and individuals of South Asian ancestry represented between 0% (no contribution to 2hGlu) and 4.4% (HbA1c) of participants. Data from participants of Ugandan ancestry were only available for the HbA1c analysis and represented 2% of participants.

Phenotypes. Analyses included data for FG and 2hGlu measured in mmol l^{-1}, FI measured in pmol l^{-1} and HbA1c as a percentage (where possible, studies reported HbA1c as a National Glycohemoglobin Standardization Program percentage). Similar to previous MAGIC efforts[7], individuals were excluded if they had type 1 diabetes or T2D (defined according to a diagnosis by a physician); reported use of diabetes-relevant medication(s); or had a FG ≥ 7 mmol l^{-1}, 2hGlu ≥ 11.1 mmol l^{-1} or HbA1c ≥ 6.5%, as described in Supplementary Table 1. 2hGlu measurements were obtained 120 min after a glucose challenge using an oral glucose-tolerance test. Measurements of FG and FI obtained from whole blood were corrected to plasma levels using the correction factor 1.13 as previously described[80].

Genotyping, quality control and imputation. Each participating cohort performed study-level quality control (QC), imputation and association analyses following a shared analysis plan. Cohorts were genotyped using commercially available genome-wide arrays or the Illumina CardioMetabochip (Metabochip) array[81] (Supplementary Table 1). Before imputation, each cohort performed stringent sample and variant QC to ensure only high-quality variants were kept in the genotype scaffold for imputation. Sample QC checks included removing samples with a low call rate less than 95%, extreme heterozygosity, sex mismatch with X chromosome variants, duplicates, first- or second-degree relatives (unless by design) or ancestry outliers. After sample QC, cohorts applied variant QC thresholds for call rate (less than 95%), Hardy–Weinberg equilibrium $P < 1 \times 10^{-6}$ and MAF. Full details of QC thresholds and exclusions for the participating cohorts are available in Supplementary Table 1.

Imputation was performed up to the 1000 Genomes Project phase 1 (v.3) cosmopolitan reference panel[82], with a small number of cohorts imputing up to the 1000 Genomes Project phase 3 panel[19] or population-specific reference panels (Supplementary Table 1).

Study-level association analyses. Each of the glycemic traits (FG, natural log-transformed FI and 2hGlu) were regressed on BMI (except for HbA1c), study-specific covariates and principal components (unless implementing a linear mixed model). Analyses for FG, FI and 2hGlu were adjusted for BMI as we had previously shown that this did not materially affected the results for FG and 2hGlu but improved our ability to detect FI-associated loci[15]. For simplicity, we refer to the traits as FG, FI and 2hGlu. For a discussion on collider bias, see Supplementary Note section 2c. Both the raw and rank-based inverse-normal transformed residuals from the regression were tested for association with genetic variants using SNPTEST[23] or Mach2Qtl[83,84]. Poorly imputed variants, defined as imputation $r^2 < 0.4$ or INFO score < 0.4, were excluded from downstream analyses (Supplementary Table 1). After study-level QC, approximately 12,229,036 variants (GWAS cohorts) and 1,999,204 variants (Metabochip cohorts) were available for analysis (Supplementary Table 1).

Centralized QC. Each contributing cohort shared their summary statistic results with the central analysis group, who performed additional QC using EasyQC[85]. Allele-frequency estimates were compared to estimates from 1000 Genomces Project phase 1 reference panel[82], and variants were excluded from downstream analyses if there was a MAF difference greater than 0.2 for populations of African American, European, Hispanic and East Asian ancestry compared with populations of African, European, Mexican and Asian ancestry from 1000 Genomes Project phase 1, respectively, or a MAF difference of more than 0.4 for individuals of South Asian ancestry compared with the population of European ancestry. At this stage, additional variants were excluded from each cohort file if they met one of the following criteria: were tri-allelic; had a MAC < 3; demonstrated a standard error of the effect size ≥ 10; or were missing an effect estimate, standard error or imputation quality. All data that passed QC (approximately 12,186,053 variants from GWAS cohorts and 1,998,657 variants from Metabochip cohorts) were available for downstream meta-analyses.

Single-ancestry meta-analyses. Single-ancestry meta-analyses were performed within each ancestry group using the fixed-effects inverse-variance meta-analysis implemented in METAL[20]. We applied a double-genomic control correction[15,86] to both the study-specific GWAS results and the single-ancestry meta-analysis results. Study-specific Metabochip results were corrected by genomic control using 4,973 SNPs included on the Metabochip array for replication of associations with QT interval, a phenotype that is not correlated with the glycemic traits being analyzed[15].

Identification of single-ancestry index variants. To identify distinct association index variants across each chromosome within each ancestry (Table 1), we performed approximate conditional analyses implemented in GCTA[21] using the --cojo-slct option (autosomes) and distance-based clumping (X chromosome). LD correlations for GCTA were estimated from a representative cohort from each ancestry: Women's Genome Health Study (European); China Health and Nutrition Survey (East Asian); Singapore Indian Eye Study (South Asian); BioMe (African American); Study of Latinos (Hispanic) and Uganda (for itself). The results from the GCTA were comparable when using alternative cohorts as the LD reference. For any index variant with a QC flag that caused reason for concern, we performed manual inspection of forest plots to decide whether the signal was likely to be real (Supplementary Note). Among 335 single-ancestry index variants across all traits, this manual inspection was done for 40 signals of which 32 passed and 8 failed after inspection. Thus, a total of 327 single-ancestry index variants passed and 8 failed.

Trans-ancestry meta-analyses. To leverage power across all ancestries, we also conducted trait-specific trans-ancestry meta-analysis by combining the single-ancestry meta-analysis results using MANTRA[87] (Supplementary Note). We defined $\log_{10}[BF] > 6$ as genome-wide significant, approximately comparable to $P < 5 \times 10^{-8}$.

Manual curation of trans-ancestry lead variants. To ensure that trans-ancestry lead variants were robust, we performed manual inspection of forest plots by at least two authors, for any variants with flags that indicated possible QC issues (Supplementary Note). Of 463 trans-ancestry lead variants across all traits, 184 passed without inspection, 131 passed after inspection and 148 failed after inspection.

Comparison of trans-ancestry lead variants across ancestries. For each pair of ancestries, we calculated Pearson's correlations in EAFs for each trans-ancestry lead variant. The pairwise summarized heterogeneity of effect sizes between ancestries was then tested using a joint *F*-test of heterogeneity[31]. The test statistic is the sum of Cochran *Q*-statistics for heterogeneity across all trans-ancestry signals. Under the null hypothesis, the statistics follows a χ^2 distribution with *n* degrees of freedom, where *n* is the number of the trans-ancestry lead variants.

LD-pruned variant lists. Several downstream analyses (for example, genomic feature enrichment, genetic scores and estimation of variance explained by associated variants) require independent LD-pruned variants ($r^2 < 0.1$) to avoid double-counting variants that might otherwise be in LD with each other and that do not provide additional 'independent' evidence. Therefore, for these analyses we generated different lists of either trans-ancestry or single-ancestry LD-pruned ($r^2 < 0.1$) variants, retaining—in each case—the variant with the strongest evidence of association (Supplementary Table 7). Subsequently, we combined trans-ancestry and single-ancestry variant lists and conducted further LD pruning. For some analyses, we took the trans-ancestry-pruned variant list and added single-ancestry signals if the LD $r^2 < 0.1$, whereas for others we started with the single-ancestry pruned lists and supplemented with trans-ancestry lead variants if the LD $r^2 < 0.1$. One exception was the list used for eQTL co-localizations, which included all single-ancestry European signals (without LD pruning) and supplemented with any additional trans-ancestry lead variants (starting from the variants with the most significant *P* values) with LD $r^2 < 0.1$ for data from individuals of European ancestry with any of the variants already in list, and that reached at least $P < 1 \times 10^{-5}$ in the meta-analysis of individuals of European ancestry.

Trait variance explained by associated loci. To determine how much of the phenotypic variance of each trait could be explained by the corresponding trait-associated loci, variants were combined in a series of weighted genetic scores. The analysis was performed in a subset of the cohorts included in the discovery GWAS (with representation from each ancestry) and in a smaller number of independent cohorts (European ancestry only). Up to three different genetic scores were derived per trait (and for each ancestry) to evaluate the potential for the trans-ancestry meta-GWAS-identified loci to provide additional information above and beyond that contributed by the ancestry-specific meta-analysis results. These genetic scores comprised: list A, single-ancestry signals; list B, single-ancestry signals plus trans-ancestry signals; and list C, trans-ancestry signals plus single-ancestry signals (Supplementary Table 7). In the case of the cohorts of individuals of European ancestry that contributed to the GWAS, we used a previously published method[32] to adjust the effect sizes (β values) from the GWAS for the contribution of that cohort, providing sets of cohort-specific effect sizes that were then used to generate the genetic scores. The association between each

genetic score and its corresponding trait was tested by linear regression and the adjusted R^2 from the model was extracted as an estimate of the variance explained.

Transferability of PGSs across ancestries. We used the PRS-CSauto[33] software to first build PGSs derived from data from individuals of European ancestry for each glycemic trait (FG, FI, 2hGlu and HbA1c) on the basis of the summary statistics. However, PRS-CSauto does not perform well when the training dataset is relatively small and the genetic architecture is sparse[33]. As a consequence, 2hGlu was excluded from this analysis. For each trait, to obtain training and test datasets for populations of European ancestry, we first removed all cohorts only genotyped on the Metabochip that were not included in this analysis. From the remaining cohorts we then removed five of the largest cohorts of European ancestry that contributed to the respective meta-analysis of data of populations of European ancestry. For each trait, these five cohorts were meta-analyzed and used as the test dataset of individuals of European ancestry. Subsequently, the remaining cohorts comprising individuals of European ancestry were also meta-analyzed and used as the training dataset of individuals of European ancestry. For each of the other ancestries, cohorts only genotyped on the Metabochip were also removed, and the remaining cohorts were meta-analyzed, and used as the test datasets of populations of non-European ancestry. Variants that had MAF < 0.05 or that were missing in over half of the individuals in the training dataset were removed[33,88]. The PGS for each trait was built using PRS-CSauto with default settings[33] with the effect size estimates based on the training dataset of individuals of European ancestry being revised based on an LD reference panel that matched the test dataset. The proportion of the trait variance explained by the PGS derived from data from individuals of European ancestry (R^2) was estimated using the R package 'gtx'[34] on the basis of the revised effect sizes and summary statistics from the test dataset for each ancestry.

Fine-mapping. Of the 242 loci identified in this study, 237 were autosomal loci that we took forward for fine-mapping (Supplementary Table 2). We used the Bayesian fine-mapping method FINEMAP[89] (v.1.1) to refine association signals and attempt to identify likely causal variants at each locus. FINEMAP estimates the maximum number of causal variants at each locus, calculates the posterior probability of each variant being causal and proposes the most likely configuration of causal variants. The posterior probabilities of the configurations in each locus were used to construct 99% CSs.

We performed both single-ancestry and trans-ancestry fine-mapping. In both analyses, only data from cohorts genotyped on GWAS arrays were used, and analyses were limited to trans-ancestry lead variants and other single-ancestry lead variants that were present in at least 90% of the samples for each trait. For the single-ancestry fine-mapping, FINEMAP estimates the number of causal variants in a region up to a maximum number, which we set to be two plus the number of distinct signals identified from the GCTA signal selection. FINEMAP uses single-ancestry and trait-specific z-scores from the fixed-effect meta-analysis in METAL[20] and an ancestry-specific LD reference, which we created from a subset of cohorts (combined sample size of more than 30% of the sample size for that ancestry), weighting each cohort by sample size. In the trans-ancestry fine-mapping analysis, FINEMAP was similarly used to estimate the number of causal variants starting with two, and trait-specific z-scores and LD maps were generated from the sample-size-weighted average of those used in the single-ancestry fine-mapping. The maximum number of causal variants was iteratively increased by one until it was larger than the number of causal variants supported by data (Bayes factor), which was the estimated maximum number of causal variants used in the final run of the fine-mapping analysis.

To compare fine-mapping results obtained from the single-ancestry and trans-ancestry efforts, analyses were limited to fine-mapping regions with evidence for a single likely causal variant in both, enabling a straightforward comparison of CSs (Supplementary Note). To ensure any difference in the fine-mapping results was not driven by different sets of variants being present in the different analyses, we repeated the single-ancestry fine-mapping limited to the same set of variants used in the trans-ancestry fine-mapping. The fine-mapping resolution was assessed on the basis of comparisons of the 99% CSs in terms of the number of variants included in the set and length of the region. To assess whether the improvement in the trans-ancestry fine-mapping was due to differences in LD, increased sample size or both, we repeated the trans-ancestry fine-mapping mimicking the sample size present in the single-ancestry fine-mapping by dividing the standard errors by the square root of the sample size ratio and compared the results with those from the single-ancestry fine-mapping.

Functional annotation of trait-associated variants. *HbA1c signal classification.* There were 218 HbA1c-associated signals from either the single-ancestry (that is all GCTA signals from any ancestry) or trans-ancestry meta-analyses. To classify these signals in terms of their likely mode of action (that is, glycemic, erythrocytic or other[7]), we examined association summary statistics for the lead variants at the 218 signals in other large datasets of individuals of European ancestry for 19 additional traits: three glycemic traits from this study (FG, 2hGlu and FI); seven mature RBC traits[90,91] (RBC count, mean corpuscular volume, hematocrit, mean corpuscular hemoglobin, mean corpuscular hemoglobin concentration, hemoglobin concentration and RBC distribution width); five reticulocyte traits (reticulocyte count, reticulocyte fraction of RBCs, immature fraction of reticulocytes, high light-scatter reticulocyte count and high light-scatter percentage of RBCs)[90,91], and four iron traits (serum iron, transferrin, transferrin saturation and ferritin)[92]. Of the 218 HbA1c signals, data were available for the lead ($n = 183$) or proxy (European LD $r^2 > 0.8$, $n = 8$) variants for 191 signals.

The additional traits were clustered using hierarchical clustering to ensure biologically related traits would cluster together (Supplementary Note). We then used a non-negative matrix factorization[93] process to cluster the HbA1c signals. Each cluster was labeled as glycemic, reticulocyte, mature RBC or iron-related based on the strength of association of the signals in the cluster to the glycemic, reticulocyte, mature RBC and iron traits (Supplementary Note). To verify that our cluster naming was correct, we used HbA1c association results conditioned on either FG or iron traits or T2D association results (Supplementary Note).

HbA1c GRSs and T2D risk. We constructed GRSs for each cluster of HbA1c-associated signals (based on hard clustering) and tested the association of each cluster with T2D risk using samples from the UK Biobank. Pairs of HbA1c signals in LD (European $r^2 > 0.10$) were LD-pruned by removing the signal with the less-significant P value of association with HbA1c. The GRS for each cluster was calculated on the basis of the logarithm of the ORs from the latest T2D study summary statistics[94] and UK Biobank genotypes imputed in the Haplotype Reference Consortium[19]. From 487,409 UK Biobank samples (age between 46 and 82 years; 55% female), we excluded participants for the following reasons: 373 with mismatched sex; 9 not used in the kinship calculation; 78,365 individuals of non-European ancestry; and 138,504 with missing T2D status, age or sex information. We further removed 26,896 related participants (kinship > 0.088, preferentially removing individuals with the largest number of relatives and control individuals for whom a case of T2D was related to that control individual). Individuals with T2D were defined as: (1) a history of diabetes without metformin or insulin treatment; (2) self-reported diagnosis of T2D; or (3) diagnosis of T2D in a national registry ($n = 17{,}022$; age between 47 and 79 years; 36% female). Control individuals were participants without a history of T2D ($n = 226{,}240$; age between 46 and 82 years; 56% female). We tested for associations between each GRS and T2D using logistic regression including covariates for age, sex and the first five principal components. The significance of the associations was evaluated by a bootstrap approach to incorporate the variance of each HbA1c-associated signal in the T2D summary data. To do this, we generated the GRS of each cluster 200 times by resampling the logarithm of the OR of each signal with T2D. For each non-glycemic class that had a GRS that was significantly associated with T2D, we performed sensitivity analyses to evaluate whether the association was driven by variants that also belonged to a glycemic cluster when using a soft clustering approach (the signals were classified as also glycemic in the soft clustering or had an association $P \leq 0.05$ with any of the three glycemic traits).

Chromatin states. To identify genetic variants within association signals that overlapped predicted chromatin states, we used a previously published, 13-chromatin-state model that included 31 diverse tissues, including pancreatic islets, skeletal muscle, adipose and liver[38]. In brief, this model was generated from cell and tissue chromatin immunoprecipitation–sequencing data for H3K27ac, H3K27me3, H3K36me3, H3K4me1 and H3K4me3, and input control from a diverse set of publicly available data[53,57,95,96] using the ChromHMM program[97]. As reported previously[38], StrEs were defined as contiguous enhancer chromatin state (active enhancer 1 and 2, genic enhancer and weak enhancer) segments that were longer than 3 kb (ref. [57]).

Enrichment of genetic variants in genomic features. We used GREGOR (v.1.2.1) to calculate the enrichment of GWAS variants that overlapped static annotations and StrEs[56]. To calculate the enrichment of glycemic-trait-associated variants in these annotations, we used the filtered list of trait-associated variants as described above (Supplementary Table 7) as input. To calculate the enrichment of sub-classified HbA1c variants, we included the list of loci characterized as glycemic, another list of loci characterized as reticulocyte or mature RBC—which collectively represented the RBC fraction—along with lists of iron-related or unclassified loci (Supplementary Table 17). We used the following parameters in GREGOR enrichment analyses: European r^2 threshold (for inclusion of variants in LD with the lead variant) = 0.8, LD window size = 1 Mb, and minimum neighbour number = 500.

We used fGWAS (v.0.3.6)[58] to calculate the enrichment of glycemic-trait-associated variants in static annotations and StrEs using summary-level GWAS results. We used the default fGWAS parameters for enrichment analyses for individual annotations for each trait. For each annotation, the model provided the natural logarithm of the maximum likelihood estimate of the enrichment parameter. Annotations were considered to be significantly enriched if the $\log_2$[parameter estimate] value and respective 95% confidence intervals were above zero or significantly depleted if the $\log_2$[parameter estimate] value and respective 95% confidence intervals were below zero.

We tested the enrichment of trait-associated variants in static annotations and StrEs with GARFIELD (v.2)[59]. We formatted annotation overlap files as required

by the tool; prepared input data at two GWAS thresholds—a threshold of 1×10^{-5} and a more stringent threshold of 1×10^{-8}—by pruning and clumping with default parameters (garfield-prep-chr script). We calculated enrichment in each individual annotation using garfield-test.R with --c option set to 0. We also calculated the effective number of annotations using the garfield-Meff-Padj.R script. We used the effective number of annotations for each trait to obtain Bonferroni-corrected significance thresholds for enrichment of each trait.

eQTL analyses. To aid in the identification of candidate casual genes associated with the European-only and trans-ancestry association signals, we examined whether any of the lead variants associated with glycemic traits (Supplementary Table 7) were also associated with the expression level (FDR < 5%) of nearby transcripts located within 1 Mb using existing eQTL datasets of blood, subcutaneous adipose, visceral adipose, skeletal muscle and pancreatic islet samples[60,61,98–101]. The LD was estimated from the collected cohort pairwise LD information, where available, and otherwise from the samples of individuals of European ancestry from 1000 Genomes Project phase 3. GWAS and eQTL signals likely co-localize when the GWAS variant and the variant most strongly associated with the expression level of the corresponding transcript (eSNP) exhibit high pairwise LD ($r^2 > 0.8$; 1000 Genomes Project phase 3, European ancestry). For these signals, we conducted reciprocal conditional analyses to test associations between the GWAS variant and transcript level when the eSNP was also included in the model, and vice versa. We report GWAS and eQTL signals as co-localized if the association for the eSNP was not significant (FDR ≥ 5%) when conditioned on the GWAS variant; we also report signals from the eQTLGen whole-blood meta-analysis data that meet only the LD threshold because conditional analysis was not possible.

Tissue and gene-set analysis. We performed enrichment analysis using DEPICT v.3, which was specifically developed for the imputed meta-analysis data of the 1000 Genomes Project[102] to identify cell types and tissues in which genes of trait-associated variants were strongly expressed, and to detect enrichment of gene sets or pathways. DEPICT data included human gene-expression data for 19,987 genes in 10,968 reconstituted gene sets, and 209 tissues and/or cell types. Because gene-expression data in DEPICT is based on samples of individuals of European ancestry and LD, we selected trait-associated variants with $P < 10^{-5}$ in the meta-analysis of data of individuals of European ancestry and tested for enrichment of signals in each reconstituted gene set, and each tissue or cell type. Enrichment results with FDR < 0.05 were considered to be significant. We ran DEPICT on the basis of the association results for all traits among: (1) cohorts with genome-wide data; or (2) all cohorts (genome-wide and Metabochip cohorts). Because results were broadly consistent between the two approaches, we present results from the analysis that contained all cohorts as it had greater statistical power.

Statistics and reproducibility. Sample size. No statistical method was used to predetermine sample size. We aimed to bring together the largest possible sample size with GWAS data from individuals of diverse ancestries (European, Hispanic, African American, East Asian, South Asian and sub-Saharan African) without diabetes and with data for one or more of the following traits: FG, FI, 2hGlu and HbA1c. The sample sizes were 281,416 (FG), 213,650 (FI), 215,977 (HbA1c) and 85,916 (2hGlu) (Supplementary Table 1). Our sample size was sufficiently powered to detect common variant associations for each of the glycemic traits and was able to detect associations at 242 loci.

Randomization and blinding. This is a study of continuous traits and there were therefore no experiments to randomize and no 'outcome' to which investigators needed to be blinded to.

Data exclusions. Before conducting this study, we identified reasons for which data should be excluded from the analysis at either the cohort or summary level; these exclusions are as follows. Sample QC checks included removing samples with low call rate less than 95%, extreme heterozygosity, sex mismatch with X chromosome variants, duplicates, first- or second-degree relatives (unless by design) or ancestry outliers. Following sample QC, cohorts applied variant QC thresholds for call rate (less than 95%), Hardy–Weinberg equilibrium $P < 1 \times 10^{-6}$ and MAF. Full details of QC thresholds and exclusions by participating cohorts are available in Supplementary Table 1. Each contributing cohort shared their summary statistic results with the central analysis group, who performed additional QC using EasyQC. Allele-frequency estimates were compared with estimates from the 1000 Genomes Project phase 1 reference panel, and variants were excluded from downstream analyses if there was a MAF difference of more than 0.2 for populations of African American, European, Hispanic and East Asian ancestry compared with populations of African, European, Mexican and Asian ancestry from 1000 Genomes Project phase 1, respectively, or a MAF difference of more than 0.4 for individuals of South Asian ancestry compared with populations of European ancestry. At this stage, additional variants were excluded from each cohort file if they met one of the following criteria: were tri-allelic; had a MAC < 3; demonstrated a standard error of the effect size ≥ 10; imputation $r^2 < 0.4$ or INFO score < 0.4; or were missing an effect estimate, standard error or imputation quality.

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

Data availability

Ancestry-specific and overall meta-analysis summary level results are available through the MAGIC website (https://www.magicinvestigators.org/). Summary statistics are also available through the GWAS catalog (https://www.ebi.ac.uk/gwas/) with the following accession codes: GCST90002225, GCST90002226, GCST90002227, GCST90002228, GCST90002229, GCST90002230, GCST90002231, GCST90002232, GCST90002233, GCST90002234, GCST90002235, GCST90002236, GCST90002237, GCST90002238, GCST90002239, GCST90002240, GCST90002241, GCST90002242, GCST90002243, GCST90002244, GCST90002245, GCST90002246, GCST90002247 and GCST90002248.

Code availability

Source code implementing the methods described in the paper are publicly available at https://doi.org/10.5281/zenodo.4607311.

References

80. D'Orazio, P. et al. Approved IFCC recommendation on reporting results for blood glucose (abbreviated). *Clin. Chem.* **51**, 1573–1576 (2005).
81. Voight, B. F. et al. The metabochip, a custom genotyping array for genetic studies of metabolic, cardiovascular, and anthropometric traits. *PLoS Genet.* **8**, e1002793 (2012).
82. The 1000 Genomes Project Consortium An integrated map of genetic variation from 1,092 human genomes. *Nature* **491**, 56–65 (2012).
83. Li, Y., Willer, C. J., Ding, J., Scheet, P. & Abecasis, G. R. MaCH: using sequence and genotype data to estimate haplotypes and unobserved genotypes. *Genet. Epidemiol.* **34**, 816–834 (2010).
84. Pei, Y. F., Zhang, L., Li, J. & Deng, H. W. Analyses and comparison of imputation-based association methods. *PLoS ONE* **5**, e10827 (2010).
85. Winkler, T. W. et al. Quality control and conduct of genome-wide association meta-analyses. *Nat. Protoc.* **9**, 1192–1212 (2014).
86. Devlin, B. & Roeder, K. Genomic control for association studies. *Biometrics* **55**, 997–1004 (1999).
87. Morris, A. P. Transethnic meta-analysis of genomewide association studies. *Genet. Epidemiol.* **35**, 809–822 (2011).
88. Bulik-Sullivan, B. K. et al. LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat. Genet.* **47**, 291–295 (2015).
89. Benner, C. et al. FINEMAP: efficient variable selection using summary data from genome-wide association studies. *Bioinformatics* **32**, 1493–1501 (2016).
90. Astle, W. J. et al. The allelic landscape of human blood cell trait variation and links to common complex disease. *Cell* **167**, 1415–1429 (2016).
91. Canela-Xandri, O., Rawlik, K. & Tenesa, A. An atlas of genetic associations in UK Biobank. *Nat. Genet.* **50**, 1593–1599 (2018).
92. Benyamin, B. et al. Novel loci affecting iron homeostasis and their effects in individuals at risk for hemochromatosis. *Nat. Commun.* **5**, 4926 (2014).
93. Binesh, N. & Rezghi, M. Fuzzy clustering in community detection based on nonnegative matrix factoriztion with two novel evaluation criteria. *Appl. Soft Comput.* **69**, 689–703 (2018).
94. Scott, R. A. et al. An expanded genome-wide association study of type 2 diabetes in Europeans. *Diabetes* **66**, 2888–2902 (2017).
95. Ernst, J. et al. Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature* **473**, 43–49 (2011).
96. Mikkelsen, T. S. et al. Comparative epigenomic analysis of murine and human adipogenesis. *Cell* **143**, 156–169 (2010).
97. Ernst, J. & Kellis, M. ChromHMM: automating chromatin-state discovery and characterization. *Nat. Methods* **9**, 215–216 (2012).
98. GTEx Consortium Genetic effects on gene expression across human tissues. *Nature* **550**, 204–213 (2017).
99. Zhernakova, D. V. et al. Identification of context-dependent expression quantitative trait loci in whole blood. *Nat. Genet.* **49**, 139–145 (2017).
100. Westra, H. J. et al. Systematic identification of *trans* eQTLs as putative drivers of known disease associations. *Nat. Genet.* **45**, 1238–1243 (2013).
101. Joehanes, R. et al. Integrated genome-wide analysis of expression quantitative trait loci aids interpretation of genomic association studies. *Genome Biol.* **18**, 16 (2017).
102. Pers, T. H. et al. Biological interpretation of genome-wide association studies using predicted gene functions. *Nat. Commun.* **6**, 5890 (2015).

Acknowledgements

We thank all investigators, staff members and study participants for their contribution to all participating studies. The funders had no role in study design, data collection, analysis, decision to publish or preparation of the manuscript. The authors received

no specific funding for this work. A full list of funding as well as individual and study acknowledgments appears in the Supplementary Note.

Author contributions

Project coordination: I.B. Writing group: J.C., C.N.S., G. Marenne, A.V., L.J.C., S.C.J.P., K.L.M., C. Langenberg, E.W., A.P.M. and I.B. Central analysis group: J.C., C.N.S., G. Marenne, A.V., L.J.C, J'an Luan, S.W., Y. Wu, X.Z., M.H., T.S.B., R.M., J.W., A.P., R.L.-G., K.H.K.C., J.Y., M.D.A., A.Y.C., A. Claringbould, J. Heikkinen, J. Hong, J.-J.H., S. Huo, M.A.K., T.L., W.M., H.M.-M., A. Ndungu, S.C.N., K.N., C.K.R., D. Ray, R. Rohde, D. Rybin, C. Schurmann, X.S., L.S., I.D.S., C.A.W., Y. Wang, P.W., W. Zhang, J.I.R., A.L.G., M.I.M., J.D., J.B.M., R.A.S., I.P., A. Leong, C.-T.L., S.C.J.P., K.L.M., C. Langenberg, E.W., A.P.M. and I.B. Cohort analysts: T.S.A., E.V.R.A., L.F.B., J.A.B., N.P.B., C.P.C., B.E.C., J.C., X.C., L.-C.C., C.-H.C., B.H.C., K.C., Y.-F.C., H.G.d.H., G.E.D., A. Demirkan, Q.D., J.E., S.A.F., J.G., F.G., J.G., S. Gustafsson, Y. Hai, F.P.H., J.-J.H., Y. Heianza, T. Huang, A.H.-C., M.H., R.A.J., T. Kawaguchi, K.A.K., Y.K., M.E.K., I.K.K., S. Lai, L.A.L., C.D.L., M. Lauzon, M. Li, S. Ligthart, J. Liu, M. Loh, J. Long, V.L., M.M., C.M., M.E.M., A. Nag, M. Nakatochi, D.N., R.N., G.P., M.P., L.R., L.J.R.-T., S.S.R., N.R.R., R. Rueedi, K. Ryan, S.S., R.S., K.E.S., B.S., J. He, K. Setoh, A.V.S., L.S., T. Sparsø, R.J.S., F.T., J. Tan, S.T., E.v.d.A., P.J.v.d.M., N.V., M.V., H. Wang, C.W., N.W., H.R.W., W.W., T. Wilsgaard, A.W., A. R. Wood, T.X., M.Z., J.-H.Z. and W. Zhao. Cohort genotyping and phenotyping: N.A., Z.A., A.A., S.J.L.B., D.B., M. Beekman, R.N.B., A.B., M. Blüher, L.L.B., S.R.B., D.W.B., Q.C., A. Campbell, H.C., Y.-F.C., E.J.C.d.G., A. Dehghan, S.D., G.E., A.F., M.F., C.F., Y.G., A.P.G., A.G., S. Han, C.A.H., C.-H.H., A.A.H., C. Herder, Y.C.C., W.A.H., S.I., M.I., M.A.I., W.C.J., M.E.J., P.K.J., R.R.K., F.R.K., T. Katsuya, C.K., W.K., I.K., T. Kuulasmaa, J.K., K. Läll, K. Lam, D.A.L., N.R.L., R.N.L., Honglan Li, S.-Y.L., J. Lindström, A. Linneberg, J. Liu, C. Lorenzo, T.M., F.M., G. Mingrone, S.M., S.M., T.N., G.N.N., J.L.N., M. Nelis, M.J.N., J.M.N., Y.O., A.P., P.A.P., O. Polasek, Q.Q., D. Raven, D.F.R., A.R., F.R., K. Roll, I.R., C. Sabanayagam, K. Sandow, N. Sattar, A. Schürmann, J. Shi, H.M.S., K.D.T., T.M.T., B.T., P.R.H.J.T., E.T., M.Y.T., A.U., R.M.v.D., D.v.H., A.v.H.V., J.V.v.V.-O., J.V., H.V., T. Wang, T.-Y.W., K.W.v.D. and T.Z. Cohort oversight and/or principal investigator: G.R.A., L.S.A., C.A.A.-S., M.E.A.-R., P.A., L.A.-S., D.M.B., L.J.B., S.B., H.B., C.B., M. Boehnke, E.B., B.O.B., K.B., D.I.B., E.P.B., T.A.B., M.C., M.J.C., J.C.C., D.I.C., Y.-D.I.C., C.-Y.C., F.S.C., A. Correa, F.C., H.G.d.H., G.D., S.E., M.K.E., E.F., L.F., J.C.F., P.W.F., T.M.F., P.F., B.G., M.O.G., P.G.-L., H.G., N.G., S. Grimsgaard, L.G., V.G., X.G., A.H., T. Hansen, C. Hayward, S.R.H., B.L.H., W.H., E.I., P.S.J., M.-R.J., J.B.J., J.W.J., P. Kaleebu, R.K., S.L.R.K., N.K., S.M.K.-K., B.-J.K., M. Kivimaki, H.A.K., J.S.K., A.K., P. Kovacs, D.K., M. Kumari, Z.K., M. Laakso, T.A.L., L.J.L., K. Leander, Huaixing Li, X.L., L.L., C. Lindgren, S. Liu, R.J.F.L., P.K.E.M., A. Mahajan, A. Metspalu, D.O.M.-K., T.A.M., P.B.M., I.N., J.R.O., A.J.O., K.K.O., S.P., C.N.A.P., N.D.P., O. Pedersen, C.E.P., D.J.P., P.P.P., M.A.P., B.M.P., L.Q., L.J.R., R. Rauramaa, S.R., P.M.R., F.R.R., T.E.S., M. Sandhu, J. Saramies, N. Schneiderman, P. Schwarz, L.J.S., E.S., P. Sever, X-o.S., P.E.S., K.S.S., B.H.S., H.S., T. Sofer, T.I.A.S., T.D.S., A. Stanton, C.J.S., M. Stumvoll, Y.T., E.T., N.J.T., A.T., J. Tuomilehto, T.T., M.U., P.v.d.H., C.v.D., P.V., T.G.M.V., L.E.W., M.W., Y.X.W., N.J.W., R.M.W., H. Watkins, W.B.W., A. R. Wickremasinghe, G.W., J.F.W., T.-Y.W., J.-Y.W., A.H.X., L.R.Y., L.Y., M.Y., E.Z., W. Zheng, A.B.Z., J.I.R., A.L.G., M.I.M., J.D., J.B.M., R.A.S., I.P., A.L., C.-T.L., S.C.J.P., K.L.M., C. Langenberg, E.W., A.P.M. and I.B. All authors read, edited and approved the final version of the manuscript.

Competing interests

A.A. is the recipient of honoraria as a speaker for a wide range of Danish and international concerns and receives royalties from textbooks, and from popular diet and cookery books. A.A. is also co-inventor of a number of patents, including methods of inducing weight loss, treating obesity and preventing weight gain (licensee Gelesis) and biomarkers for predicting the degree of weight loss (licensee Nestec), owned by the University of Copenhagen, in accordance with Danish law. I.B. and spouse own stock in GlaxoSmithKline and Incyte Corporation. B.H.C. is now an employee of Life Epigenetics; all work was completed before employment by Life Epigenetics. A.Y.C. is now an employee of Merck & Co.; all work was completed before employment by Merck & Co. J.C.F. has received consulting honoraria from Janssen. J.G. is now an employee of F. Hoffmann-La Roche, and owns stock in Roche and GlaxoSmithKline. A.L.G. has received honoraria from Merck and Novo Nordisk. As of June 2019, A.L.G. discloses that her spouse is an employee of Genentech and hold stock options in Roche. E.I. is now an employee of GlaxoSmithKline; all work was completed before his employment by GlaxoSmithKline. W.M. has received grants and/or personal fees from the following companies/corporations: Siemens Healthineers, Aegerion Pharmaceuticals, AMGEN, AstraZeneca, Sanofi, Alexion Pharmaceuticals, BASF, Abbott Diagnostics Numares, Berlin-Chemie, Akzea Therapeutics, Bayer Vital, Bestbion dx, Boehringer Ingelheim Pharma, Immundiagnostik, Merck Chemicals, MSD Sharp and Dohme, Novartis Pharma, Olink Proteomics and Synlab Holding Deutschland. M.I.M. has served on advisory panels for Pfizer, NovoNordisk and Zoe Global, and has received honoraria from Merck, Pfizer, NovoNordisk and Eli Lilly. He holds stock options in Zoe Global and has received research funding from Abbvie, Astra Zeneca, Boehringer Ingelheim, Eli Lilly, Janssen, Merck, NovoNordisk, Pfizer, Roche, Sanofi Aventis, Servier and Takeda. He is now an employee of Genentech and a holder of Roche stock. J.B.M. has consulted for Quest Diagnostics, who is a manufacturer of an HbA1c assay. M.E.M. has received grant funding from Regeneron Pharmaceuticals. M.E.M. is also an inventor on a patent that was published by the US Patent and Trademark Office on 6 December 2018 under Publication Number US 2018-0346888, and international patent application that was published on 13 December 2018 under Publication Number WO-2018/226560; all work was completed before these competing interests arose, and are unrelated to this work. D.O.M.-K. is a part-time clinical research consultant for Metabolon. J.L.N. is a member of the Scientific Advisory Board for Veralox Therapeutics. C.N.A.P. has received research support from GlaxoSmithKline and AstraZeneca unrelated to this project. B.M.P. serves on the Steering Committee of the Yale Open Data Access Project funded by Johnson & Johnson. N. Sattar has consulted for AstraZeneca, Boehringer Ingelheim, Eli Lilly, Novo Nordisk, Napp and Sanofi, and received grant support from Boehringer Ingelheim. R.A.S. is an employee and shareholder of GlaxoSmithKline. T.D.S. is the founder of Zoe Global. J. Tuomilehto receives research support from Bayer, is a consultant for Eli Lily and holds stock in Orion Pharma and Aktivolabs.

Additional information

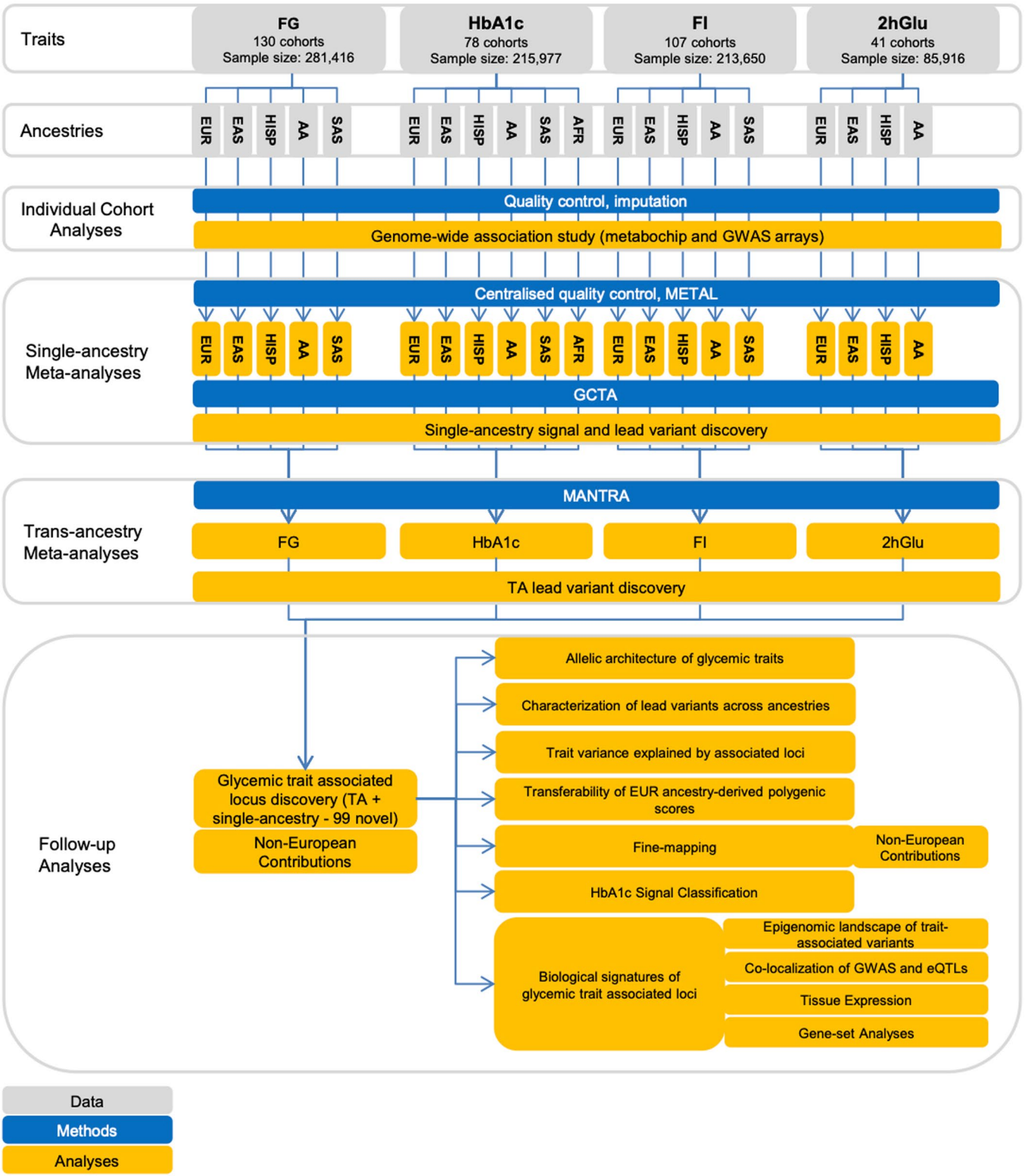
Extended data is available for this paper at https://doi.org/10.1038/s41588-021-00852-9.

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41588-021-00852-9.

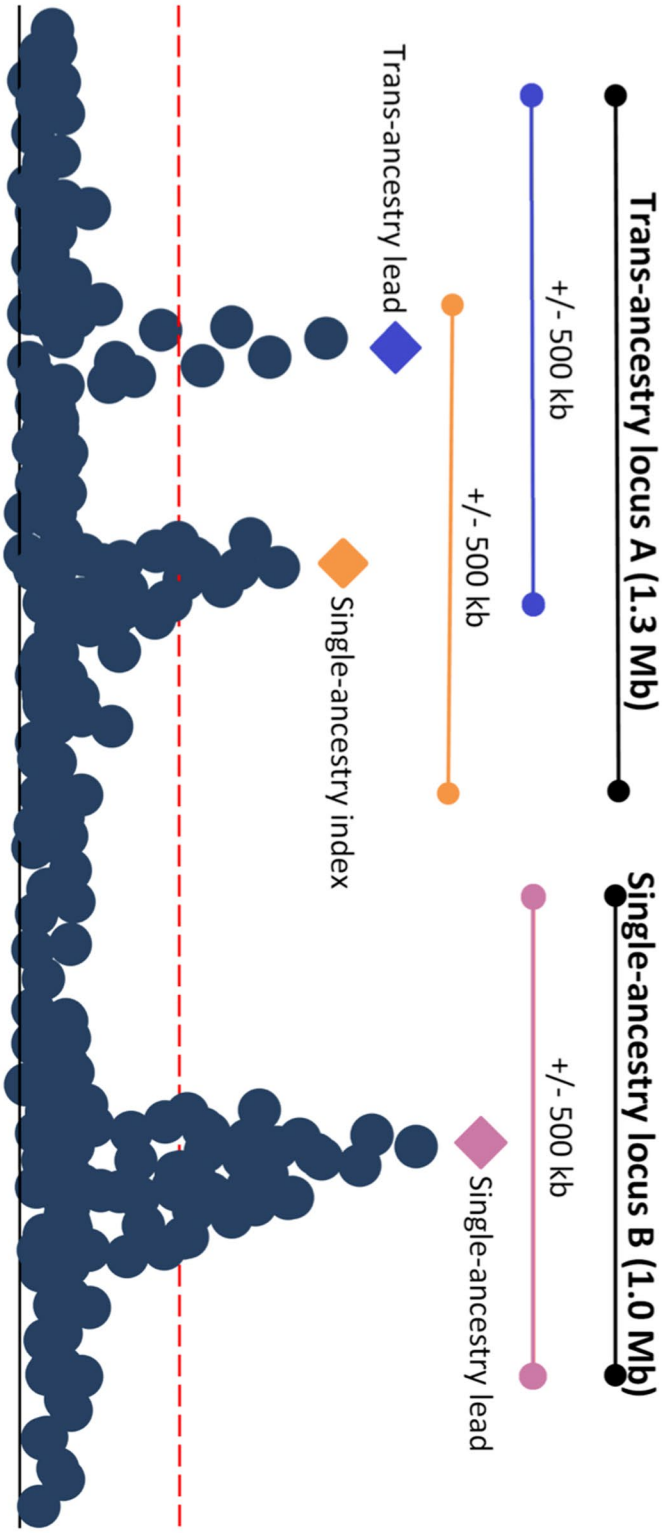
Correspondence and requests for materials should be addressed to I.B.

Peer review information *Nature Genetics* thanks Anurag Verma and Constantin Polychronakos for their contribution to the peer review of this work. Peer reviewer reports are available.

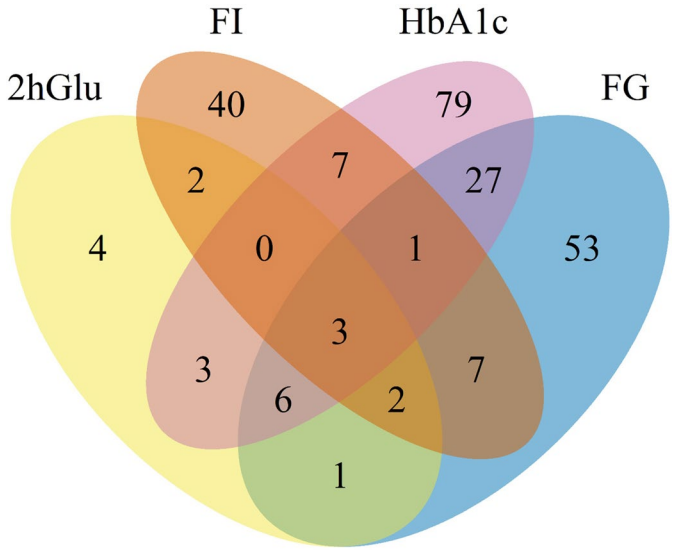
Reprints and permissions information is available at www.nature.com/reprints.



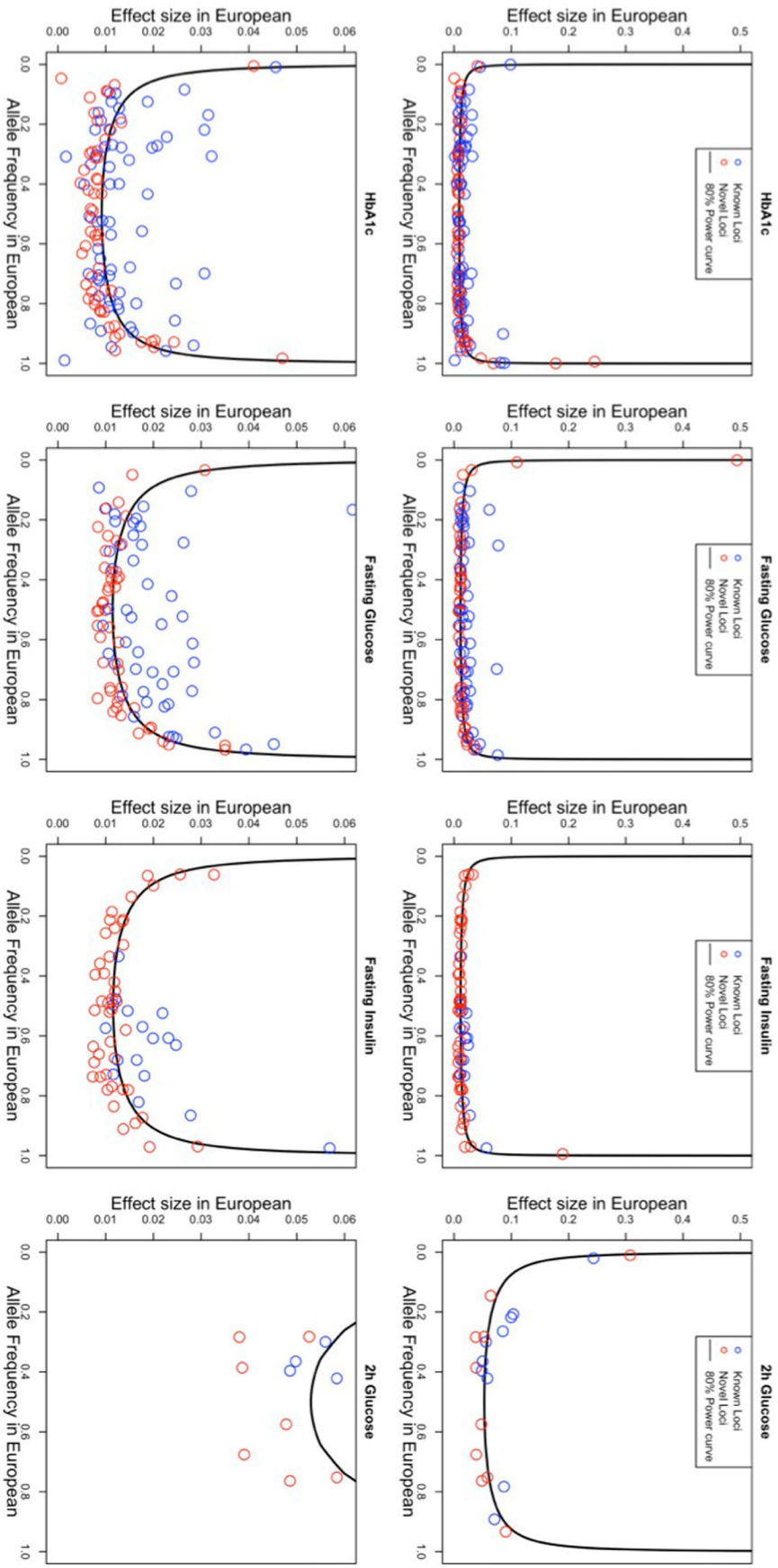
Extended Data Fig. 1 | Flow diagram of this study. The figure shows the data, key methods and main analyses included in this effort.



Extended Data Fig. 2 | Locus diagram. Trans-ancestry locus A contains a trans-ancestry lead variant for one glycemic trait represented by the blue diamond, and another single-ancestry index variant for another glycemic trait represented by the orange triangle. Single-ancestry locus B contains a single-ancestry lead variant represented by the purple square. The orange, blue and purple bars represent a +/− 500Kb window around the orange, blue, and purple variants, respectively. The black bars indicate the full locus window where trans-ancestry locus A contains trans-ancestry lead and single-ancestry index variants for two traits and single-ancestry locus B has a single-ancestry lead variant for a single trait.



Extended Data Fig. 3 | Venn diagram. Overlap of TA loci between traits.



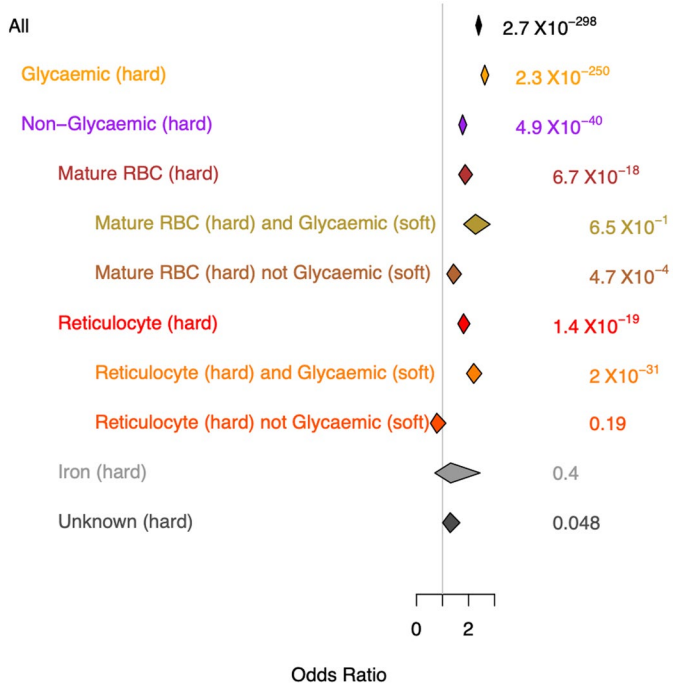
Extended Data Fig. 4 | See next page for caption.

Extended Data Fig. 4 | Allele frequency versus effect size. Allele frequency versus effect size for all signals detected through the trans-ancestry meta-analyses, for each of the four traits. Frequency and effect size are from the European meta-analyses. The power curves were computed based on the European sample size for each trait, and the mean (m) and standard deviation (sd) computed on the FENLAND study: FG, m = 4.83 mmol/l, sd=0.68; FI, m = 3.69 mmol/l, sd=0.60; 2hGlu, m = 5.30 mmol/l, sd=1.74; HbA1c, m = 5.55%, sd=0.48.

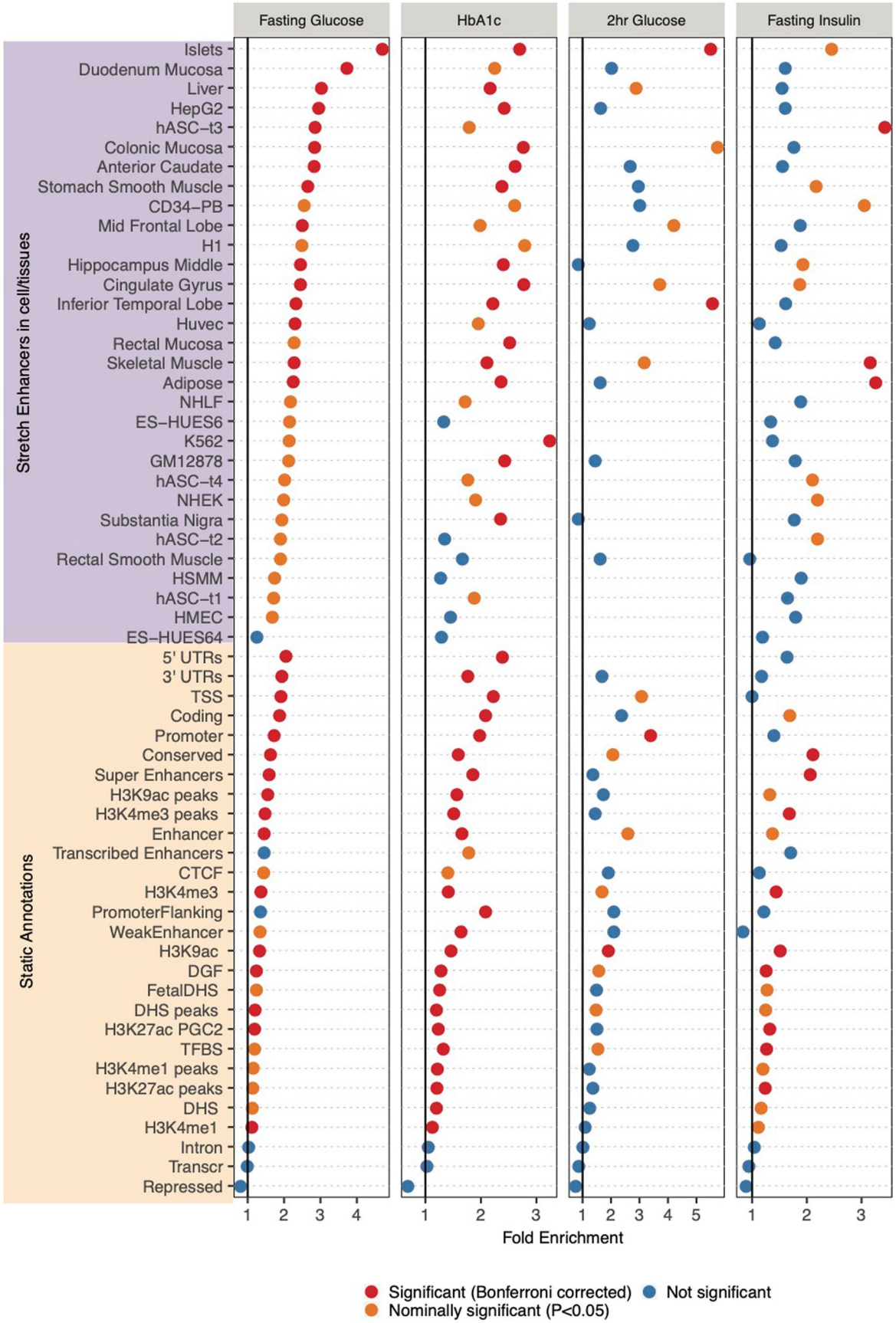
	EUR	EAS	HISP	AA	SAS	AFR
EUR	FG	$2.72X10^{-11}$	0.016	$1.20X10^{-5}$	0.16	
	HbA1c	$1.55X10^{-15}$	$1.98X10^{-7}$	$<2.2X10^{-16}$	0.017	$8.86X10^{-7}$
	FI	$1.13X10^{-6}$	$1.7X10^{-4}$	0.352	0.3	
	2hGlu	0.348	0.841	0.098		
EAS	0.36	FG	$5.9X10^{-4}$	0.0262	$7.5X10^{-4}$	
	0.35	HbA1c	0.0099	$9.24X10^{-8}$	0.057	$1.01X10^{-5}$
	0.34	FI	0.00103	0.224	0.014	
	0.19	2hGlu	0.527	0.083		
HISP	0.79	0.58	FG	0.057	0.032	
	0.88	0.55	HbA1c	$1.98X10^{-6}$	0.531	$<2.2X10^{-16}$
	0.83	0.57	FI	0.044	0.623	
	0.86	0.4	2hGlu	0.056		
AA	0.36	0.31	0.6	FG	0.084	
	0.41	0.37	0.67	HbA1c	$8.85X10^{-6}$	$<2.2X10^{-16}$
	0.53	0.09	0.59	FI	0.419	
	0.21	-0.03	0.37	2hGlu		
SAS	0.82	0.65	0.79	0.48	FG	
	0.8	0.63	0.82	0.5	HbA1c	0.046
	0.82	0.41	0.74	0.54	FI	
					2hGlu	
AFR						FG
	0.21	0.24	0.5	0.98	0.3	HbA1c
						FI
						2hGlu

R^2	0	0.2	0.4	0.6	0.8	1

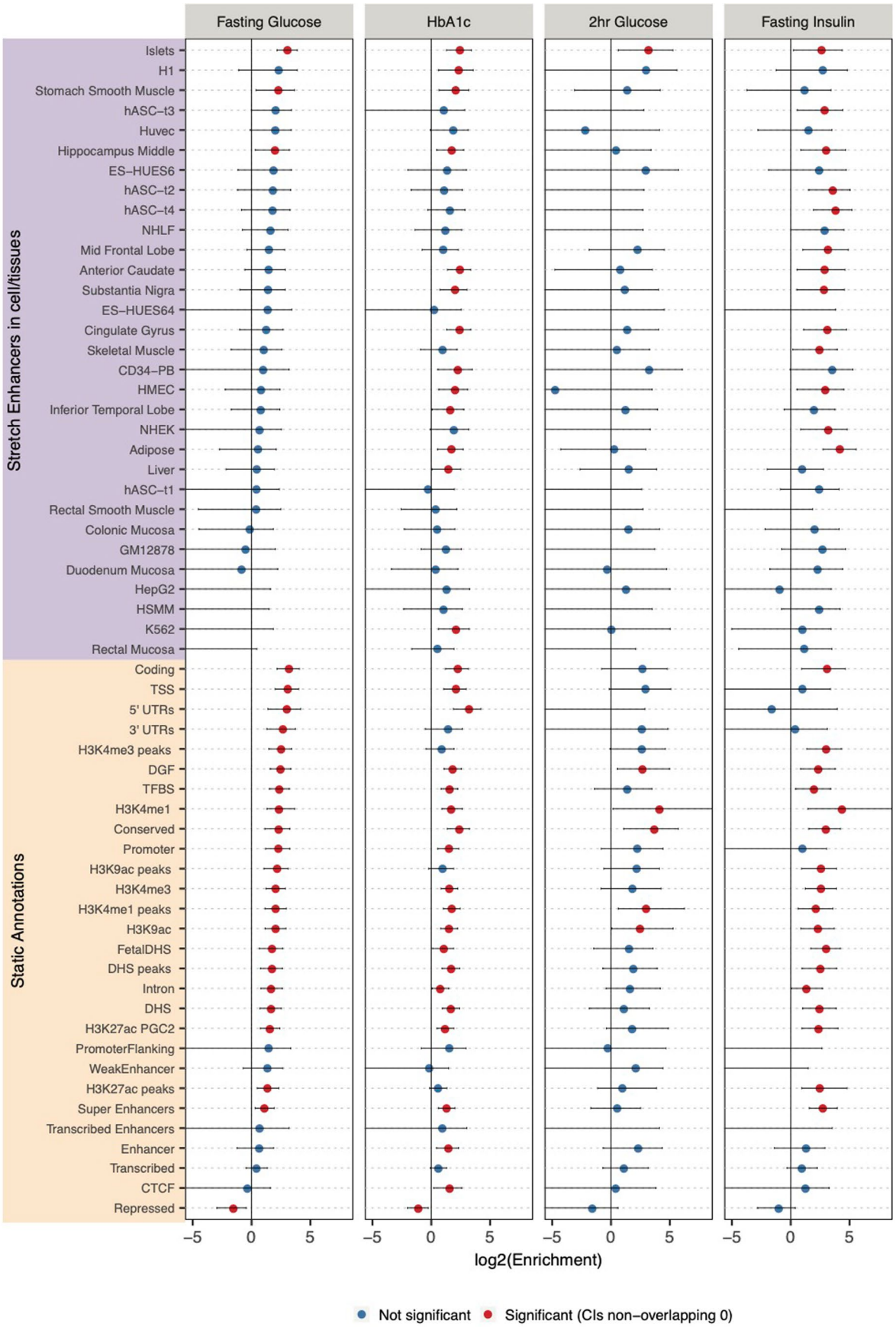
Extended Data Fig. 5 | EAF correlation and heterogeneity test. Pearson correlation of EAF on the lower tri-angle and p-value of one-side heterogeneity test without multiple testing corrections on the upper tri-angle of the trans-ancestry lead variants associated with each trait between ancestries. Correlations > 0.7 are in bold.



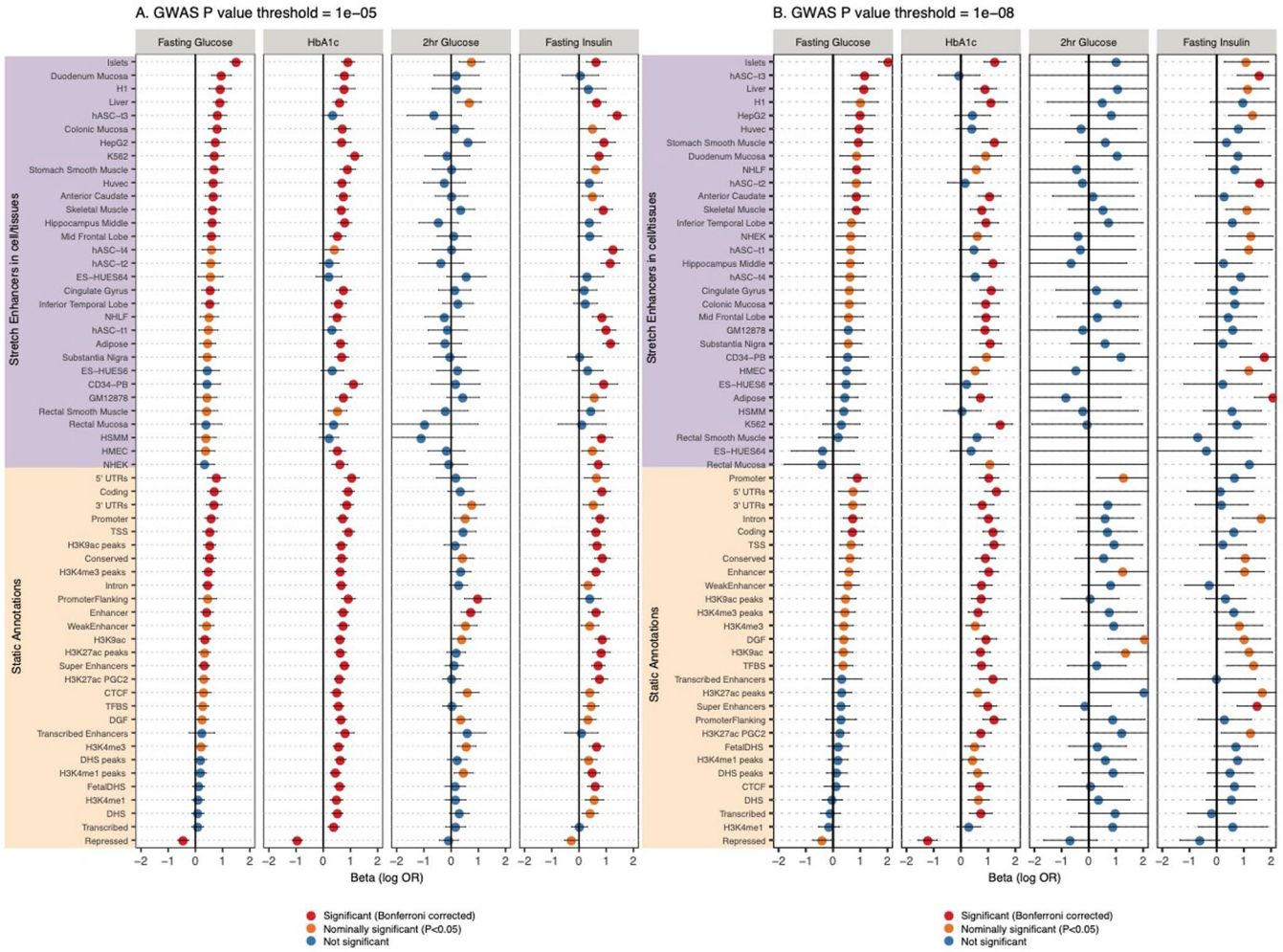
Extended Data Fig. 6 | Forest plot of T2D GRS from HbA1c variants. The p-value on the right side is from the two-side test without multiple testing corrections. Vertical points of each diamond represent the point estimate of the odds ratio. The horizontal points of each diamond represent the 95% confidence interval of the odds ratio. Figure shows the association results between HbA1c-associated variants built into a GRS for T2D by taking each HbA1c-associated variant and using a weight that corresponds to its T2D effect size (logOR) based on analysis by the DIAGRAM consortium. The overall GRS is subsequently partitioned according to the HbA1c signal classification. The overall and partitioned GRS were tested for association with T2D based on data from UK biobank.



Extended Data Fig. 7 | Enrichment of glycemic trait associated GWAS variants to overlap genomic annotations using GREGOR. Figure shows enrichment for 59 total static and stretch enhancer annotations considered. One-side test significance (red) is determined after Bonferroni correction to account for 59 total annotations tested for each trait; nominal significance ($P < 0.05$) is indicated in yellow.



Extended Data Fig. 8 | Enrichment of glycemic trait associated GWAS variants to overlap genomic annotations using fGWAS. Figure shows log2(Fold Enrichment) of GWAS variants to overlap 59 static and stretch enhancer annotations calculated. Significant enrichment (red) is considered if the 95% confidence intervals (shown by the error bars) do not overlap 0.



Extended Data Fig. 9 | Enrichment of glycemic trait associated GWAS variants to overlap genomic annotations using GARFIELD. Figure shows the β or effect size (log odds ratio) for GWAS variants to overlap 59 static and stretch enhancer annotations. GWAS variants were included at two significance thresholds, 1e-05 (A) and 1e-08 (B). One-side test significance (red) is determined after Bonferroni correction to account for effective annotations tested for each trait reported by GARFIELD (see Supplementary Note); nominal significance ($P < 0.05$) is indicated in yellow. The 95% confidence intervals are shown by the error bars.

nature research

Corresponding author(s): Inês Barroso

Last updated by author(s): 02/17/2021

Reporting Summary

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

Statistics

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

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

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

Software and code

Policy information about availability of computer code

Data collection	No software was used for data collection
Data analysis	Details are described in the Method section and Supplementary Table 1. Software and code include: Pre-imputation: EasyQC v9.2, https://www.uni-regensburg.de/medizin/epidemiologie-praeventivmedizin/genetische-epidemiologie/software/index.html Pre-phasing: SHAPEIT v1.ESHG/SHAPEIT v2.790, https://mathgen.stats.ox.ac.uk/genetics_software/shapeit/shapeit.html Imputation: IMPUTE v2.3.2, https://mathgen.stats.ox.ac.uk/impute/impute_v2.html Imputation: MaCH v1.0, http://csg.sph.umich.edu/abecasis/MaCH/download/ Imputation: MACH2 v0.3.0.0, https://github.com/riverar/mach2 Imputation: MINIMAC RELEASE STAMP 2013-07-17, https://genome.sph.umich.edu/wiki/Minimac Imputation: MINIMAC3 v1.0.6, https://genome.sph.umich.edu/wiki/Minimac3 Imputation: BEAGLE v4.0.r1399, https://faculty.washington.edu/browning/beagle/b4_0.html Imputation: PBWT Sanger server, https://www.sanger.ac.uk/tool/sanger-imputation-service/ Imputation: michigan imputation server, https://imputationserver.sph.umich.edu/index.html Association: SNPTEST v2.3.0/v2.4.0, https://mathgen.stats.ox.ac.uk/genetics_software/snptest/snptest.html Association: GWAF v2.1 (R package), https://cran.r-project.org/src/contrib/Archive/GWAF/ Association: MMAP v 2017_08_18.intel, https://mmap.github.io/ Association: Mach2Qtl v1.1.3, https://hpc.nih.gov/apps/mach2qtl.html Association: PLINK v1.07/v1.9, http://zzz.bwh.harvard.edu/plink/ Association: EPACTS v3.2.6, http://genome.sph.umich.edu/wiki/EPACTS Association: SAS v9.4, http://support.sas.com/software/94/ Association: ProbABEL v.0.4.3, https://github.com/GenABEL-Project/ProbABEL Association: LMEKIN (R package coxme v2.2-4), https://cran.r-project.org/web/packages/coxme/vignettes/lmekin.pdf

Association: Quicktest v0.94, http://toby.freeshell.org/software/quicktest.shtml
Association: GEMMA v0.95a, https://github.com/genetics-statistics/GEMMA
Association: EMMAX vbeta-07Mar2010, http://csg.sph.umich.edu//kang/emmax/download/index.html
Association: SOLAR v7.2.5, https://userinfo.surfsara.nl/systems/lisa/software/solar
Association: STATA v9.2, https://www.stata.com/stata9/
Meta-analysis: METAL v2011-03-25, https://github.com/statgen/METAL
Meta-analysis: MANTRA v1, https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3460225/
Conditional analysis: GCTA v1.26.0, https://cnsgenomics.com/software/gcta/
Fine-mapping: FINEMAPv1.1, http://www.christianbenner.com/
Enrichment analysis: GREGOR v1.4.0, http://csg.sph.umich.edu/GREGOR/
Enrichment analysis: fGWAS v0.3.6, https://github.com/joepickrell/fgwas
Enrichment analysis: GARFIELD v2, https://www.ebi.ac.uk/birney-srv/GARFIELD/
Enrichment analysis: DEPICT v1_rel194, https://data.broadinstitute.org/mpg/depict/documentation.html
Poly genetic risk score analysis: PRS-CS vApr.24.2020, https://github.com/getian107/PRScs
Poly genetic risk score analysis: gtx v0.0.8 (R package), https://www.rdocumentation.org/packages/gtx/versions/0.0.8

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

Data

Policy information about availability of data

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

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

Ancestry specific and overall meta-analyses summary level results will be available through the MAGIC website (https://www.magicinvestigators.org/) upon publication. We will share summary statistic through the GWAS catalog. Accession codes of GWAS catalog (https://www.ebi.ac.uk/gwas/): GCST90002225, GCST90002226, GCST90002227, GCST90002228, GCST90002229, GCST90002230, GCST90002231, GCST90002232, GCST90002233, GCST90002234, GCST90002235, GCST90002236, GCST90002237, GCST90002238, GCST90002239, GCST90002240, GCST90002241, GCST90002242, GCST90002243, GCST90002244, GCST90002245, GCST90002246, GCST90002247 and GCST90002248.

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	We aimed to bring together the largest possible sample size with GWAS data imputed to 1000 Genomes Project reference panel, of individuals from diverse ancestries (European, Hispanic, African American, East Asian, South Asian and sub-Saharan African) without diabetes and with data for one or more of the following traits: fasting glucose, fasting insulin, 2hr post-challenge glucose and glycated haemoglobin. The sample sizes are 281,416 (FG), 213,650 (FI), 215,977 (HbA1c) and 85,916 (2hGlu). Imputation was performed up to the 1000 Genomes Project phase 1 (v3) cosmopolitan reference panel, with a small number of cohorts imputing up to the 1000 Genomes phase 3 panel or population-specific reference panels (see Supplementary Table 1).Our sample size was sufficiently powered to detect common variant associations with each of the glycaemic traits and was able to detect associations at 242 loci.
Data exclusions	Prior to conducting this study, we identified reasons for which data should be excluded from the analysis at either the cohort or summary level; these exclusions are as follows. Sample quality control checks included removing samples with low call rate < 95%, extreme heterozygosity, sex mismatch with X chromosome variants, duplicates, first- or second-degree relatives (unless by design), or ancestry outliers. Following sample QC, cohorts applied variant QC thresholds for call rate (< 95%), Hardy-Weinberg Equilibrium (HWE) P < 1x10-6, and minor allele frequency (MAF). Full details of QC thresholds and exclusions by participating cohort are available in Supplementary Table 1. Each contributing cohort shared their summary statistic results with the central analysis group who performed additional QC using EasyQC. Allele frequency estimates were compared to estimates from 1000Gp1 reference panel, and variants were excluded from downstream analyses if there was a minor allele frequency difference > 0.2 for AA, EUR, HISP, and EAS populations against AFR, EUR, MXL, and ASN populations from 1000 Genomes Phase 1, respectively, or a minor allele frequency difference > 0.4 for SAS against EUR populations. At this stage, additional variants were excluded from each cohort file if they met one of the following criteria: were tri-allelic; had a minor allele count (MAC) < 3; demonstrated a standard error of the effect size ≥ 10; imputation r2 < 0.4 or INFO score < 0.4; or were missing an effect estimate, standard error, or imputation quality.
Replication	Because we used all data available for discovery no replication was attempted.
Randomization	This study meta-analyzed existing data and did not require randomization.

Blinding	This study meta-analyzed existing data and did not require blinding.

Reporting for specific materials, systems and methods

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

Materials & experimental systems

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

Methods

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

Human research participants

Policy information about studies involving human research participants

Population characteristics	This study included trait data from four glycaemic traits: fasting glucose (FG), fasting insulin (FI), 2hr post-challenge glucose (2hGlu), and glycated haemoglobin (HbA1c). The total numbers of contributing cohorts are 131 (FG), 107 (FI), 78 (HbA1c) and 41 (2hGlu), and the sample sizes are 281,416 (FG), 213,650 (FI), 215,977 (HbA1c) and 85,916 (2hGlu). Relevant characteristics include age, sex and BMI. Sample characteristics of each cohort is described in Supplementary Table 1.
Recruitment	Participants were originally recruited from 150 individual case-control and cohort studies totalling over 280,000 participants. Details of each participating study are in Supplementary Table 1. Individuals were excluded if they had type 1 or type 2 diabetes (defined by physician diagnosis); reported use of diabetes-relevant medication(s); or had a FG ≥7 mmol/L, 2hGlu ≥11.1mmol/L, or HbA1c ≥ 6.5%, as detailed in Supplementary Table 1. 2hGlu measures were obtained 120 minutes after a glucose challenge in an oral glucose tolerance test (OGTT). Measures for FG and FI taken from whole blood were corrected to plasma level using the correction factor 1.13. Each individual study is subject to potential bias due to its original study design. However, no individual study should impact our findings.
Ethics oversight	All studies were approved by relevant institutional review boards or regional/national ethics committees. All individuals provided informed consent. Specifically: All ABCD participants gave written informed consent for data collection of the phenotypes. Regarding the DNA collection and analysis, an opt-out procedure was used. The ABCD study protocol was approved by the Central Committee on Research Involving Human Subjects in The Netherlands, the medical ethics review committees of the participating hospitals, and the Registration Committee of the Municipality of Amsterdam. The AGES-Reykjavik Study was approved by the Icelandic National Bioethics Committee (VSN 00-063) and by the Institutional Review Board of the US National Institute on Aging, NIH. All participants signed an informed consent. Ethical approval for the ALSPAC study was obtained from the ALSPAC Ethics and Law Committee and the Local Research Ethics Committees. Consent for biological samples has been collected in accordance with the Human Tissue Act (2004). Informed consent for the use of data collected via questionnaires and clinics was obtained from participants following the recommendations of the ALSPAC Ethics and Law Committee at the time. All study protocols for the AMISH study were approved by the institutional review board at the University of Maryland Baltimore. Informed consent was obtained from each study participant. IRB approvals for ARIC were obtained at all study sites (including DCC UNC Chapel Hill). All study participants provided written informed consent. The ASCOT study protocols were reviewed and ratified by central and regional ethics review boards in the UK and by national ethics and statutory bodies in Ireland and the Nordic countries (Sweden, Denmark, Iceland, Norway, and Finland). Patients were recruited between February 1998 and May 2000. All patients provided written informed consent. The BC1936 was approved by the Ethical Committee of Copenhagen County (KA96008) and the Danish Data Protection Agency. All participants provided written informed consent. The Beijing Eye Study was approved by the Medical Ethics Committee of the Beijing Tongren Hospital and all participants gave informed written consent. The BetaGene Study was approved by the Institutional Review Boards of the University of Southern Californai and Kaiser Permanente Southern California. All participants provided writeen informed consent. The BioMe cohort was approved by the Institutional Review Board at the Icahn School of Medicine at Mount Sinai. All BioMe participants provided written, informed consent for genomic data sharing. The CAGE-GWAS1 was approved by the Institutional Review Boards at the National Center for Global Health and Medicine. All participants provided written informed consent. The CAGE-KING was approved by the ethics committees of Aichi Gakuin University, Jichi Medical University, Nagoya University, and Kyushu University; and all participants provided written informed consent. The CHNS was approved by the Institutional Review Boards at the University of North Carolina at Chapel Hill, the Chinese National Human Genome Center at Shanghai, and the Institute of Nutrition and Food Safety at the China Centers for Disease Control. All participants provided written informed consent. CHS was approved by the Institutional Review Boards at the Wake Forest University, University of California, Davis, Johns Hopkins and University of Pittsburgh. All participants provided written informed consent. The Cleveland Family Study was approved by the Institutional Review Board of Mass General Brigham (formerly Partners HealthCare). Written informed consent was obtained from all participants. Written informed consent for CLHNS was obtained from all participants, and study protocols were approved by the University of North Carolina Institute Review Board for the Protection of Human Subjects. The institutional Ethics Committee of the University of Lausanne, which afterwards became the Ethics Commission of Canton Vaud (www.cer-vd.ch) approved the

baseline CoLaus study (reference 16/03, decisions of 13th January and 10th February 2003). The approval was renewed for the first (reference 33/09, decision of 23rd February 2009), the second (reference 26/14, decision of 11th March 2014) and the third (reference PB_2018-00040, decision of 20th March 2018) follow-ups. The study was performed in agreement with the Helsinki declaration and its former amendments, and in accordance with the applicable Swiss legislation. All participants gave their signed informed consent before entering the study. The COPSAC2000 study was approved by the Local Ethics Committee (KF 01-289/96) and the Danish Data Protection Agency (2008-41-1754). All participants and parents provided written informed consent. The CROATIA_Korcula cohort was approved by the Institutional Review Board at the University of Split, Croatia. All participants provided written, informed consent. The CROATIA_Split cohort was approved by the Institutional Review Board at the University of Split, Croatia. All participants provided written, informed consent. The CROATIA_Vis cohort was approved by the Institutional Review Board at the Universities of Zagreb, Croatia and Edinburgh, Scotland. All participants provided written, informed consent. The DPS was a randomized, controlled, multicenter study carried out in Finland between the years 1993 and 2001 (ClinicalTrials.gov NCT00518167). The study protocol was approved by the Ethics Committee of the National Public Health Institute of Helsinki, Finland. The study design and procedures of the study were carried out in accordance with the principles of the Declaration of Helsinki. All study participants provided written informed consent. The DRECA studies were approved by the Ethical and Research Commission of the primary health assistance district of Seville 1992 and 2006. All participants, those in the first study and in the second follow up provided a signed informed consent. The DR's EXTRA was a randomized controlled trial between years 2005 and 2011 (ISRCTN45977199). The study protocol was approved by the Research Ethics committee of the Hospital Districk of Nothern Savo, Finland. The participants gave signed informed consent. All analyses in EGCUT were approved by the Ethics Review Committee of the University of Tartu. All participants provided written informed consent. The Ely study was approved by the Cambridge Local Research Ethics Committee (99/246). All participants in the EPIC-InterAct study gave written informed consent and ethical approval was given by the ethics committees of the International Agency for Research on Cancer and the local institutions. The EPIC-Norfolk study was approved by the Norfolk Research Ethics Committee (ref. 05/Q0101/191) and all participants gave their written consent before entering the study. The EpiHealth study was approved by the Ethics Committee of Uppsala University. Each participant gave their written informed consent. The ERF study was approved by the Institutional Review Board at the Erasmus University Medical Center, Rotterdam, the Netherlands. All participants provided written informed consent. The Family Heart Study (FamHS) was approved by the Institutional Review Board at the Washington University in St. Louis. Written informed consent including consent to participate in genetic studies was obtained from each participant. Ethical approval for the Fenland study was given by the Cambridge Local Ethics committee (ref. 04/Q0108/19) and all participants gave their written consent prior to entering the study. The Framingham Heart Study was approved by the Institutional Review Board of the Boston University Medical Center. All study participants provided written informed consent. The French adult and young studies followed ethical principles defined in the Helsinki declaration, and they were approved by local ethical committees from Comité Consultatif de Protection des Personnes se prêtant à des Recherches Biomédicales (CPPRB) of Lille - Lille Hospital (Lille, France), Hotel-Dieu hospital (France) and Bicêtre hospital (France). All participants older than 18 years signed an informed consent form. Oral assent from children or adolescents was obtained and parents (or legal guardian) signed an informed consent form." FUSION was approved by the coordinating Ethics Committee of the Hospital District of Helsinki and Uusimaa. All participants gave written informed consent. Ethical approval for the GS:SFHS study was obtained from the Tayside Committee on Medical Research Ethics (on behalf of the National Health Service. It has Research Tissue Bank approval from East of Scotland Research Ethics Service (ref ES-20-0021). The GeneSTAR study was approved by the Johns Hopkins Medicine Institutional Review Board. All participants gave written informed consent. Written informed consent for GENOA was obtained from all subjects and approval was granted by participating institutional review boards (University of Michigan, University of Mississippi Medical Center, and Mayo Clinic). The Tayside Medical Ethics Committee has approved the GoDARTS study and informed consent was obtained for all participants. The participants have consented to research on their samples and data. The data included in the MAGIC 1KG Trans-ancestry meta-analysis stems from the ADIGEN project, a subset of the original GOYA study. The ADIGEN project was approved by the Committee on Health Research Ethics for Copenhagen and Frederiksberg Districts, and the Danish Data Protection Agency. All participants gave written informed consent. The HANDSL Study has been approved by the National Institutes of Health Institutional Review Board study number 09AGN248. All participants provided written informed consent. This HCHS/SOL study was approved by the institutional review boards (IRBs) at each field center, where all participants gave written informed consent, and by the Non-Biomedical IRB at the University of North Carolina at Chapel Hill, to the HCHS/SOL Data Coordinating Center. All IRBs approving the study are: Non-Biomedical IRB at the University of North Carolina at Chapel Hill. Chapel Hill, NC; Einstein IRB at the Albert Einstein College of Medicine of Yeshiva University. Bronx, NY; IRB at Office for the Protection of Research Subjects (OPRS), University of Illinois at Chicago. Chicago, IL; Human Subject Research Office, University of Miami. Miami, FL; Institutional Review Board of San Diego State University. San Diego, CA. The Health2006 was approved by the Ethical Committee of Copenhagen County (KA20060011) and the Danish Data Protection Agency. All participants provided written informed consent. The HELIC collections include blood for DNA extraction, laboratory-based haematological and biochemical measurements, and interview-based questionnaire data. The study was approved by the Harokopio University Bioethics Committee, and informed consent was obtained from human subjects. The HTN-IR study was approved by Human Subjects Protection Institutional Review Boards at UCLA, the University of Southern California, Lundquist/LA BioMed/Harbor-UCLA and Cedars-Sinai Medical Center. The IMPROVE study was approved by all local IRBs and by the 7 independent ethics committees: 1) The Consultative Committee for the Protection of Persons in Biomedical Research, Pitié Salpetrière site, Paris, France; 2) The Medical Ethics Review Committee - Academic Hospital Groningen – Groningen, the Netherlands; 3) The Research Ethics Committee of Northern Savonia Hospital District. - Kuopio University Hospital – Kuopio, Finland; 4) The Research Ethics Committee of the University of Kuopio and Kuopio University Hospital – Kuopio, Finland; 5) The Ethical-Scientific Commission of the Niguarda Ca' Granda Hospital - Milan, Italy; 6) The Ethics Committee of the Umbrian Health Authorities - Perugia, Italy; and 7) The Research Ethics Committee / North Karolinska Hospital Administration H6 171 76 Stockholm, Sweden. All participants gave written informed consent. The Inter99 was approved by the Ethical Committee of Copenhagen County (KA98155) and the Danish Data Protection Agency. All participants provided written informed consent. The institutional review boards at the University of Colorado/Denver, UTHSC-San Antonio, Kaiser Permanente-Northern CA, UCLA, and the Wake Forest School of Medicine, approved the IRAS and IRASFS study protocol and all participants provided written informed consent. The JHS study was approved by Jackson State University, Tougaloo College, and the University of Mississippi Medical Center IRBs, and all participants provided written informed consent. All participants of KARE provided written informed consent. The study using KARE samples was approved by an institutional review board at the Korean National Institute of Health, Republic of Korea. All participants of the KORA F4 study provided informed consent, which was approved by the Ethics Committee of the Medical Association of Bavaria (Ethics Committee Number 06068). The Leiden Longevity Study protocol was approved by the ethical committee of the Leiden

University Medical Center (P01.113) and conducted according to the principles of the declaration of Helsinki. All participants provided written informed consent. The Leipzig adult study was approved by Leipzig University Ethics committee (Reg.No. 031-2006 and 017-12-23012012). Written informed consent was obtained from all participants. Informed written consent was provided by all parents and children from the age of 12 years. All participants in the Lifelines cohort study signed an informed consent. The Lifelines cohort study is conducted according to the principles of the declaration of Helsinki and following the research code of University Medical Center Groningen and approved by its Medical Ethical Committee. The Living Biobank study was approved by the National University of Singapore IRB. All participants provided written informed consent. The LOLIPOP study is approved by the local Research Ethics Committee, and all participants provided written consent for genetic studies. The LURIC study was approved by the "Landesärztekammer Rheinland-Pfalz" (#837.255.97(1394)). Informed written consent was obtained from all participants. The MACAD study was approved by Human Subjects Protection Institutional Review Boards at University of California Los Angeles, Lundquist/LA BioMed/Harbor-UCLA and Cedars-Sinai Medical Center. The MEGA study was approved by the Ethics Committee of the Leiden University Medical Center, and written informed consent was obtained from all participants. The MESA Study approved by IRBs at University of Washington, Wake Forest School of Medicine, Northwestern University, University of Minnesota, Columbia University, Johns Hopkins University, and the Univ of California at Los Angeles. All participants provided written informed consent. The METSIM study was approved by the Ethics Committee of the University of Kuopio and Kuopio University Hospital. All study participants gave written informed consent. The MICROS study was approved by the Ethics Committee of the Autonomous Province of Bolzano. All study participants gave informed written consent. The ethics committee of Kyoto University Graduate School of Medicine approved the Nagahama study, and we obtained written informed consent from all participants. The NEO study was approved by the Medical Ethical Committee of the Leiden University Medical Center. All participants gave htier written informed consent. For the NFBC1966 and NFBC1986 studies, we used data only from those participants for whom a written informed consent was obtained. The study has been approved by the ethical committees of University of Oulu and the Northern Ostrobothnia Hospital District. The NHAPC study protocol was approved by the Institutional Review Board of the Institute for Nutritional Sciences, Chinese Academy of Sciences and abided by the Declaration of Helsinki principles. Written informed consent was obtained from all participants. The NIDDM-Athero study was approved by Human Subjects Protection Institutional Review Boards at the University of Southern California, Lundquist/LA BioMed/Harbor-UCLA and Cedars-Sinai Medical Center. The NSHD study received Multi-Centre Research Ethics Committee approval (Central Manchester REC: 07/H1008/168) and informed consent was given by participants. Informed consent was obtained from all NTR participants. The study protocol was approved by the Central Ethics Committee on Research Involving Human Subjects of the VU University Medical Centre, Amsterdam. The Orkney Complex Disease Study (ORCADES) was approved by the Local Research Ethics Committee of NHS Orkney and the North of Scotland Research Ethics Committee. All participants gave written informed consent. The Ethical Review Board of the Faculty of Medicine of the Federal University of Pelotas approved the PELOTAS study, and written informed consent was obtained from all participants. The PIVUS study was approved by the Ethics Committee of Uppsala University. Each participant gave their written informed consent. PREVEND was approved by the medical ethics committee of the University Medical Center Groningen and conducted in accordance with the Helsinki Declaration guidelines. All subjects gave written informed consent. PROCARDIS study was approved by the National Research Ethics Service (NRES) London South East (MREC 99/1/02). The PROSPER study was approved by the Medical Ethics Committees of the three collaborating centers and complied with the Declaration of Helsinki. All participants gave written informed consent. The Ragama Health Study was approved by the Institutional Review Boards at the National Center for Global Health, Tokyo, Japan and the Faculty of Medicine, University of Kelaniya, Sri Lanka (P38/09/2006). All participants provided written informed consent. The RISC study was approved by the Medical Ethics Committee of each recruiting centre, and all subjects gave written informed consent. The Rotterdam study was approved by the Institutional Review Board at the Erasmus University Medical Center, Rotterdam, the Netherlands. All participants provided written informed consent. THe SardiNIA study received ethical approval from the Comitato Etico di Azienda Sanitaria Locale 8, Lanusei (2009/0016600) and from the NIH Office of Human Subject Research. the SCARF and SHEEP studies were approved by the Regional Ethical Review Board at Karolinska Institutet, Stockholm, Sweden. SIGMA study was approved by the Institutional Review Board of the Instituto Nacional de Ciencias Medicas y Nutricion Salvador Zubiran in Mexico City. All participants provided written informed consent. The SEED study followed the principles of the Declaration of Helsinki with ethics approval obtained from the Singapore Eye Research Institute (SERI) Institutional Review Board (IRB). All participants provided written informed consent. SP2 was approved by the Institutional Review Boards of the National University of Singapore and the Singapore General Hospital. All participants provided written informed consent. The SORBS study was approved by Leipzig University Ethics committee. Written informed consent was obtained from all participants. TAICHI study was performed in accordance with the tenets of the Declaration of Helsinki and approved by the Institutional Review Boards of each participating centers in the U.S. and Taiwan. The U.S. sites include Stanford University School of Medicine in Stanford, California; Hudson-Alpha Biotechnology Institute in Huntsville, Alabama; Lundquist/LABioMed/Harbor-UCLA; and Cedars-Sinai Medical Center (CSMC) in Los Angeles, California. The Taiwan sites include Taichung Veteran's General Hospitals (Taichung VGH), Taipei Veterans General Hospital (Taipei VGH), National Health Research Institutes (NHRI), Tri-Service General Hospital (TSGH), and National Taiwan University Hospital (NTUH). The Cardiometabolic Risk in Chinese (CRC) Study was reviewed and approved by the ethics committee of the Central Hospital of Xuzhou, Affiliated Hospital of Medical School of Southeast University, Nanjing, China. Written consent was obtained from each participant. The Human Research Ethics Committees at the University of Western Australia, King Edward Memorial Hospital and Princess Margaret Hospital in Perth, Australia, granted ethics approval for each follow-up in the Raine study. Parents, guardians and adolescent participants provided written informed consent either before enrolment or at data collection at each stage of follow-up. All procedures for the TRAILS cohort were approved by the Dutch Central Committee on Research Involving Human Subjects. Written informed consent, including specific consent to undertake genetic analyses, was obtained from participants and their parents or custodians. The TRIPOD Study was approved by the Institutional Review Board of the University of Southern California. All participants gave written informed consent. The Tromsø Study was approved by the Regional Committee for Medical Research Ethics. All participants gave written informed consent. The TwinGene project was approved by the regional ethics committee. All participants gave written informed consent. The TwinsUK project was approved by the ethics committee at St Thomas' Hospital London. All participants gave written informed consent. TWSC was approved by the Institutional Review Board at the Institute of Biomedicial Sciences, Academia Sinica, Taiwan. All participants provided written informed consent. UKHLS: The University of Essex Ethics Committee has approved all data collection on Understanding Society main study and innovation panel waves, including asking consent for all data linkages except to health records. Requesting consent for health record linkage was approved at Wave 1 by the National Research Ethics Service (NRES) Oxfordshire REC A (08/H0604/124), at BHPS Wave 18 by the NRES Royal Free Hospital & Medical School (08/H0720/60) and at Wave 4 by NRES Southampton REC A (11/SC/0274). Approval for the collection of biosocial data by trained nurses in Waves 2 and 3 of the main survey was

obtained from the National Research Ethics Service (Understanding Society - UK Household Longitudinal Study: A Biosocial Component, Oxfordshire A REC, Reference: 10/H0604/2). The ULSAM study was approved by the Ethics Committee of Uppsala University. Each participant gave their written informed consent. The Shanghai Breast Cancer and Shanghai Men's Health studies were approved by the IRB of the Vanderbilt University Medical Center and Shanghai Cancer Institute. All participants provided written informed consent to the study. The VIKING study was approved by the South East Scotland Research Ethics Committee. All participants gave written informed consent. The WHI project was reviewed and approved by the Fred Hutchinson Cancer Research Center (Fred Hutch) IRB in accordance with the U.S. Department of Health and Human Services regulations at 45 CFR 46 (approval number: IR# 3467-EXT). Participants provided written informed consent to participate. Additional consent to review medical records was obtained through signed written consent. Fred Hutch has an approved FWA on file with the Office for Human Research Protections (OHRP) under assurance number 0001920. In the Whitehall II study, informed consent and research ethics are renewed at each clinical examination; the most recent approval was from the University College London Hospital Committee on the Ethics of Human Research, reference 85/0938. Analysis in the WGHS was approved by the Institutional Review Board (IRB) of Brigham and Women's Hospital.

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