Genome-wide trans-ancestry meta-analysis provides insight into the genetic architecture of type 2 diabetes susceptibility

DIAbetes Genetics Replication And Meta-analysis (DIAGRAM) Consortium\textsuperscript{1,2}, Asian Genetic Epidemiology Network Type 2 Diabetes (AGEN-T2D) Consortium\textsuperscript{1,2}, South Asian Type 2 Diabetes (SAT2D) Consortium\textsuperscript{1,2}, Mexican American Type 2 Diabetes (MAT2D) Consortium\textsuperscript{1,2} & Type 2 Diabetes Genetic Exploration by Next-generation sequencing in multi-Ethnic Samples (T2D-GENES) Consortium\textsuperscript{1,2}

To further understanding of the genetic basis of type 2 diabetes (T2D) susceptibility, we aggregated published meta-analyses of genome-wide association studies (GWAS), including 26,488 cases and 83,964 controls of European, east Asian, south Asian and Mexican and Mexican American ancestry. We observed a significant excess in the directional consistency of T2D risk alleles across ancestry groups, even at SNPs demonstrating only weak evidence of association. By following up the strongest signals of association from the trans-ethnic meta-analysis in an additional 21,491 cases and 55,647 controls of European ancestry, we identified seven new T2D susceptibility loci. Furthermore, we observed considerable improvements in the fine-mapping resolution of common variant association signals at several T2D susceptibility loci. These observations highlight the benefits of trans-ethnic GWAS for the discovery and characterization of complex trait loci and emphasize an exciting opportunity to extend insight into the genetic architecture and pathogenesis of human diseases across populations of diverse ancestry.

The majority of GWAS of T2D susceptibility have been undertaken in populations of European ancestry\textsuperscript{1–5}, predominantly because of existing infrastructure, sample availability and relatively poor coverage by many of the earliest genome-wide genotyping arrays of common genetic variation in other major ethnic groups\textsuperscript{6}. However, populations of European ancestry constitute only a subset of human genetic variation and are thus insufficient to fully characterize T2D risk variants in other ethnic groups. Furthermore, the latest genome-wide genotyping arrays are less biased toward Europeans, and more recent T2D GWAS have been performed with great success in populations from other ancestry groups, including east Asians\textsuperscript{7–12}, south Asians\textsuperscript{13,14}, Mexicans and Mexican Americans\textsuperscript{15} and African Americans\textsuperscript{16}. These studies have provided initial evidence of overlap in T2D susceptibility loci between ancestry groups, as well as for coincident risk alleles at lead SNPs across diverse populations\textsuperscript{17,18}. These observations are consistent with a model in which the underlying causal variants at many of these loci are shared across ancestry groups and thus arose before migration of the human population out of Africa. Under such a model, we would expect to improve the power to detect new susceptibility loci for the disease and enhance the fine-mapping resolution of causal variants by combining GWAS across ancestry groups through trans-ethnic meta-analysis because of increased sample size and differences in the structure of linkage disequilibrium (LD) between such diverse populations\textsuperscript{6,19–21}.

In this study, we aggregated published meta-analyses of GWAS in a total of 26,488 cases and 83,964 controls from populations of European, east Asian, south Asian and Mexican and Mexican American ancestry\textsuperscript{5,11,13,15}. T2D GWAS from populations of African ancestry, which would be expected to provide the greatest potential for fine mapping of common causal variants because of less extensive LD than other ethnic groups\textsuperscript{6}, were not accessible for inclusion in our analyses. With these data, we aimed to (i) assess the evidence for excess concordance in the direction of effect of T2D risk alleles across ancestry groups; (ii) identify new T2D susceptibility loci through trans-ethnic meta-analysis and subsequent validation in an additional 21,491 cases and 55,647 controls of European ancestry; and (iii) evaluate the improvements in the fine-mapping resolution of common variant association signals in established T2D susceptibility loci through trans-ethnic meta-analysis despite the lack of GWAS from populations of African ancestry.

RESULTS

Study overview

We considered published meta-analyses of GWAS of T2D susceptibility from four major ethnic groups (Supplementary Tables 1 and 2) undertaken by (i) the DIAbetes Genetics Replication and Meta-analysis (DIAGRAM) Consortium\textsuperscript{5} (European ancestry; 12,171 cases and 56,862 controls); (ii) the Asian Genetic Epidemiology Network T2D (AGEN-T2D) Consortium\textsuperscript{11} (east Asian ancestry; 6,952 cases and 11,865 controls); (iii) the South Asian T2D (SAT2D) Consortium\textsuperscript{13} (south Asian ancestry; 5,561 cases and 14,458 controls); and (iv) the Mexican American T2D (MAT2D) Consortium\textsuperscript{15} (Mexican

\textsuperscript{1}A full list of authors and affiliations appears at the end of the paper.

\textsuperscript{2}A full list of members and affiliations appears in the Supplementary Note.

Received 14 September 2013; accepted 17 January 2014; published online 9 February 2014; doi:10.1038/ng.2897
and Mexican American ancestry; 1,804 cases and 779 controls). We obtained association summary statistics from the four available ethnic-specific meta-analyses, each of which was imputed at up to 2.5 million autosomal SNPs from Phase II/III HapMap22,23, to provide a uniform catalog of common genetic variation, defined by a minor allele frequency (MAF) of at least 5%, across ancestry groups (Online Methods). We then combined these association summary statistics across ancestry groups in a trans-ethnic fixed-effects meta-analysis (Online Methods).

### Concordance of T2D risk alleles across ancestry groups

We began by evaluating heterogeneity in allelic effects (i.e., discordance in the direction and/or magnitude of odds ratios) between ancestry groups at 69 established autosomal T2D susceptibility loci. We assessed the evidence for heterogeneity at previously reported lead SNPs on the basis of Cochran's Q statistics from the trans-ethnic meta-analysis (Online Methods and Supplementary Table 3). We observed nominal evidence of heterogeneity (Bonferroni correction, P ≤ 0.05/69 = 0.00072) at the previously reported lead SNP at just three loci. At TCF7L2 (rs7903146, P = 0.00055), the odds ratio is the largest in populations of European ancestry, although the risk allele has a consistent direction of effect across ethnicities. At PEPD (rs3786897, P = 0.00055) and KLF14 (rs13233731, P = 0.00064), however, the association signals are apparently specific to the populations of east Asian and European ancestry, respectively, despite the fact that the reported lead SNPs are common in all ethnic groups. We also observed that at 52 previously reported lead SNPs passing quality control in each of the four ethnic-specific meta-analyses, 34 showed the same direction of effect across all ancestry groups (65.4% compared with 12.5% expected by chance, binomial test P < 2.2 × 10−16). The strong evidence of homogeneity in allelic effects across ancestry groups at the majority of the previously reported lead SNPs argues against the ‘synthetic association’ hypothesis24. It is improbable that GWAS signals at most of the established T2D susceptibility loci reflect unobserved lower-frequency causal alleles with larger effects because (i) rare variants are unlikely to have arisen under a fixed-effects model, we identified 33 independent SNPs (separated by at least 500 kb) with nominal evidence of association (P ≤ 0.001) with T2D from the European ancestry meta-analysis. By aligning the effect of the T2D risk allele from the European meta-analysis into the other ancestry groups, we observed evidence of a significant excess in directional concordance between ethnicities: 57.0% with east Asian populations (binomial test P = 0.0077); 55.4% with south Asian populations (binomial test P = 0.032); and 56.6% with Mexican and Mexican American populations (binomial test P = 0.010). Using the same approach, we also observed an excess of consistency in the direction of effect between ethnicities at independent SNPs demonstrating weaker evidence of T2D association (0.001 < P ≤ 0.01) from the European meta-analysis (Table 1). In contrast, when we considered independent SNPs with no evidence of association (P > 0.5) with T2D, there was no enrichment in coincident risk alleles across ethnic groups. We repeated this analysis by identifying T2D risk alleles at SNPs with nominal evidence of association in each of the east Asian, south Asian and Mexican and Mexican American meta-analyses and assessing concordance in the direction of effect in each of the other ancestry groups (Supplementary Table 4). The evidence for an excess in concordance between T2D risk alleles across ethnicities was not as strong, particularly for the Mexican and Mexican American meta-analysis. However, this finding presumably reflects reduced power due to smaller sample sizes, and there was still significant over-representation of alleles with the same direction of effect across ancestry groups at SNPs with nominal evidence of association with the disease.

### Seven new T2D susceptibility loci at genome-wide significance

The observations from our concordance analyses are consistent with a long tail of common T2D susceptibility variants with effects that are decreasing in magnitude but that are homogeneous across ancestry groups. Under such a model, we would expect these variants to be amenable to discovery by trans-ethnic fixed-effects meta-analyses. In this study, by aggregating the published ethnic-specific meta-analyses under a fixed-effects model, we identified 33 independent SNPs (separated by at least 500 kb) with suggestive evidence of association (P < 5 × 10−8) at loci not previously reported for T2D susceptibility in any ancestry group (Supplementary Table 5 and Supplementary Fig. 1). By convention, we labeled loci according to the gene nearest to the lead SNP unless a compelling biological candidate mapped nearby. It is essential to validate partially imputed association signals with direct genotyping. Consequently, we carried forward these 33 loci for in silico follow up in a meta-analysis of an additional 21,491 cases and 55,647 controls of European ancestry5 genotyped with the Metabochip (Online Methods and Supplementary Tables 1 and 2). This custom array was designed to facilitate cost-effective replication of nominal associations for T2D and other metabolic and cardiovascular traits26. However, it provides relatively limited coverage of common genetic variation genome wide, with the result that the lead

### Table 1 Concordance in the direction of effect of T2D risk alleles

| European ancestry meta-analysis | Trans-ethnic concordance | European into Mexican and Mexican American |
|--------------------------------|--------------------------|--------------------------------------------|
|                               | Concordant SNPs/total SNPs | % Binomial test P | Concordant SNPs/total SNPs | % Binomial test P | Concordant SNPs/total SNPs | % Binomial test P |
| P ≤ 0.001                     | 180/316                   | 57.0           | 0.0077                      | 175/316       | 55.4           | 0.032                      | 179/316       | 56.6           | 0.010                      |
| 0.001 < P ≤ 0.01             | 877/1,624                 | 54.0           | 0.00068                     | 861/1,624     | 53.0           | 0.0080                     | 886/1,624     | 54.6           | 0.00013                     |
| 0.01 < P ≤ 0.5               | 2,556/5,053               | 50.6           | 0.21                        | 2,604/5,053   | 51.5           | 0.015                      | 2,588/5,053   | 51.2           | 0.043                      |
| 0.5 < P ≤ 1                  | 2,535/5,039               | 50.3           | 0.34                        | 2,532/5,039   | 50.2           | 0.37                        | 2,519/5,039   | 50.0           | 0.51                        |

The T2D risk alleles shown were identified in a meta-analysis of a GWAS of European ancestry (12,171 cases and 56,862 controls) along with those identified in meta-analyses of GWAS of east Asian (6,952 cases and 11,865 controls), south Asian (5,561 cases and 14,458 controls) and Mexican and Mexican American (1,804 cases and 779 controls) ancestry after exclusion of the 69 established autosomal susceptibility loci defined as mapping within 500 kb of the previously reported lead SNP.
SNPs, or close proxies (CEU $r^2 > 0.6$ from Phase II HapMap), were present at just 24 of the loci. We also identified poorer proxies at two additional loci, rs9505118 (SSR1-RREB1, CEU $r^2 = 0.26, P = 1.9 \times 10^{-6}$) and rs4275659 (MPHOSPH9, CEU $r^2 = 0.48, P = 5.5 \times 10^{-6}$), which nonetheless demonstrated only marginally weaker association signals than the lead SNPs (SSR1-RREB1, rs9502570, $P = 5.7 \times 10^{-6}$; MPHOSPH9, rs1727313, $P = 1.2 \times 10^{-6}$). Given that variants met our threshold for follow up from the trans-ethnic meta-analysis, we also considered them for validation.

By combining association summary statistics from the trans-ethnic discovery and European ancestry validation meta-analyses, SNPs at seven loci achieved genome-wide significance (combined meta-analysis $P < 5 \times 10^{-8}$) (Table 2 and Fig. 1). We observed no evidence of heterogeneity in allelic effects between the discovery and validation stages of the combined meta-analysis (Supplementary Table 5). As we expected, the new loci are characterized by lead SNPs that are relatively common in all ethnicities and have modest effects on T2D susceptibility that are homogeneous across ancestry groups (Supplementary Table 6). We did not harmonize adjustments for covariates within or between consortia because of variation in individual study design and recorded non-genetic risk factors. However, we observed no evidence of heterogeneity in allelic effects in the European ancestry validation meta-analysis after stratification of the studies according to covariate adjustment (Online Methods and Supplementary Table 7). These data thus provide no evidence of bias in the allelic effect estimates at lead SNPs at the new loci and suggest our results to be robust to variability in correction for potential confounders across studies.

The new loci include SNPs mapping near POU5F1-TCF19 in the major histocompatibility complex (MHC), a region of the genome that is essential to the immune response. The MHC harbors HLA class II genes, which together account for approximately half of the genetic risk to type 1 diabetes (T1D). We observed no evidence of association of T2D with tags for traditional T1D HLA risk alleles in the trans-ethnic meta-analysis: HLA-DR4 (rs608985, $P = 0.32$) and HLA-DR3 (rs2187668, $P = 0.34$). Furthermore, when we considered lead SNPs at 49 T1D susceptibility loci (Supplementary Table 8), we observed nominal evidence of association ($P < 0.05$) with T2D and the same risk allele for both diseases at just two loci (GLIS3 and 6q22.32) but not at that mapping to the MHC (rs8926645, $P = 0.33$). There is very strong evidence that T1D risk variants, particularly in the MHC, are also associated with latent autoimmune diabetes of adulthood (LADA) and, which is a late-age-onset, more indolent form of the disease that often results in a clinical misdiagnosis of T2D. Although studies contributing to the trans-ethnic meta-analysis differed in the degree to which they were able to exclude cases with LADA, the lack of association of T1D risk variants suggests that the rates of diagnostic misclassification of autoimmune diabetes were too modest to drive the T2D GWAS signal at the POU5F1-TCF19 locus.

The new loci also include SNPs mapping to ARL15 and SSR1-RREB1, which have been implicated previously at genome-wide significance in the regulation of fasting insulin (FI) and fasting glucose (FG), respectively. The lead SNPs for T2D (rs702634) and FI (rs4865796) mapping to ARL15 are closely correlated in populations of European and east Asian ancestry (CEU $r^2 = 1.00$ and CHB+JPT $r^2 = 0.87$ from Phase II HapMap). However, the lead T2D SNP (rs9505118) is independent of that for FG (rs17762454) at the SSR1-RREB1 locus (CEU and CHB+JPT $r^2 < 0.05$). The ARL15 locus has also been associated with circulating levels of adiponectin, which is an adipocyte-secreted protein that has antidiabetic effects, but the lead SNP (rs4311394) is independent of that for T2D susceptibility from the trans-ethnic meta-analysis.

To obtain a more comprehensive view of the overlap of the newly associated T2D susceptibility loci with metabolic phenotypes, we interrogated published European ancestry meta-analyses from the Meta-Analysis of Glucose and Insulin-related traits Consortium (MAGIC) Investigators, the Genetic Investigation of ANthropometric Traits (GIANT) Consortium and the Global Lipids Genetics Consortium to evaluate the effect of T2D risk alleles on glycemic traits, including homeostatic model of assessment indices of beta-cell function (HOMA-B) and insulin resistance (HOMA-IR); anthropometric measures; and plasma lipid concentrations (Online Methods and Supplementary Tables 9–11). T2D risk alleles at SSR1-RREB1 and LPP have features that indicate a primary role in susceptibility through beta-cell dysfunction: increased FG ($P = 1.0 \times 10^{-5}$ and $P = 8.6 \times 10^{-7}$, respectively) and reduced HOMA-B ($P = 0.11$ and $P = 0.011$, respectively). Conversely, the T2D risk allele mapping to ARL15 is associated with increased FI, most strongly after adjustment for body mass index (BMI) ($P = 5.0 \times 10^{-12}$), and increased HOMA-IR ($P = 0.021$) and is thus more characteristic of action through insulin resistance. This risk allele is also associated with reduced levels of high-density lipoprotein cholesterol ($P = 0.022$) and increased levels of triglycerides ($P = 0.010$), as we expected, but also with reduced BMI ($P = 5.6 \times 10^{-5}$).

To identify the most promising functional candidate transcripts among those mapping to the new susceptibility loci, we interrogated public databases and unpublished resources for expression quantitative trait loci (eQTL) from a variety of tissues (Online Methods). The lead T2D SNPs at three loci showed nominal association ($P < 10^{-5}$)
with expression and were in strong LD (CEU and CHB+JPT $r^2 > 0.8$) with the reported cis-eQTL variant: SSR1 (B cells, $P = 2.2 \times 10^{-6}$) at the SSR1-RREB1 locus; ABCB9 (liver, $P = 7.4 \times 10^{-12}$) and SETD8 (lung, $P < 2.0 \times 10^{-10}$) at the MPHOSPH9 locus; and HCG27 (monocytes, $P = 1.3 \times 10^{-6}$) at the POU5F1-TCF19 locus (Supplementary Table 12).

We also evaluated new loci for potential functional mechanisms underlying T2D susceptibility (Online Methods). We identified variants for functional annotation in pilot data from the 1000 Genomes Project\[^5\] that are in strong LD (CEU and CHB+JPT $r^2 > 0.8$) with the lead SNPs in the seven new susceptibility loci. We identified a missense variant at the POU5F1-TCF19 locus in TCF19 (rs113581344, p.Val211Met; CEU $r^2 = 0.96$ and CHB+JPT $r^2 = 0.80$ with lead SNP rs3130501), although it is predicted to be tolerated by SIFT\[^35\] (Supplementary Table 13). Lead SNPs in the new susceptibility loci were also in strong LD with variants in the UTRs of SSR1 (at the SSR1-RREB1 locus) and ABCB9, OGFOD2 and PITPNM2 (at the MPHOSPH9 locus). Variants in strong LD with the lead SNPs at two of the new susceptibility loci overlap regions of predicted regulatory function generated by the ENCODE Project\[^36\] (Supplementary Fig. 2). The lead SNP at the LPP locus maps to an enhancer region that is active in HepG2 cells. We also identified a variant at the FAF1 locus (rs58836765; CEU $r^2 = 0.89$ and CHB+JPT $r^2 = 0.80$ with lead SNP rs17106184) that overlaps a region of open chromatin activity in pancreatic islets and other cell types. This open chromatin site is in a region that is correlated with expression of ELAVL4, which has been demonstrated to regulate insulin translation in pancreatic beta cells\[^37\], highlighting this transcript as a credible candidate at the FAF1 locus. Regulatory annotations in HepG2 cells and pancreatic islets are both broadly enriched at T2D-associated variants\[^38\] and are thus supportive of these functional mechanisms for causal variant activity at both loci.

**Improved fine-mapping resolution at T2D susceptibility loci**

Given our observation that the causal variants underlying GWAS signals are shared across ancestry groups at many T2D susceptibility loci.
compared to either a fixed- or random-effects model. Simulation studies have demonstrated improved detection and localization of causal variants through trans-ethnic meta-analysis with MANTRA. This Bayesian approach has the advantage of allowing for heterogeneity in allelic odds ratios between ancestry groups arising as a result of differential patterns of LD with a shared underlying causal variant across diverse populations, which cannot be accommodated in fixed-effects meta-analysis (Online Methods). Simulation studies have demonstrated improved detection and localization of causal variants through trans-ethnic meta-analysis with MANTRA compared to either a fixed- or random-effects model.

Within each locus, we constructed credible sets of SNPs that are most likely to be causal on the basis of their statistical evidence of association from the MANTRA meta-analysis. Credible sets can be interpreted in a way similar to confidence intervals in a frequentist statistical framework. For example, assuming that a locus harbors a single causal variant that is reported in the meta-analysis, the probability that it will be contained in the 99% credible set is 0.99. Smaller credible sets, in terms of the number of SNPs they contain or the genomic interval they cover, thus correspond to fine mapping at a higher resolution. It is essential that SNP coverage be as uniform as possible across studies in the construction of credible sets, otherwise differences in association signals between variants may reflect variability in sample sizes in the meta-analysis and not true differences in the magnitude of effects on T2D susceptibility. Consequently, we did not consider loci with weaker signals of association, as they were typically characterized by large 99% credible sets in the European ancestry meta-analysis and thus might provide an overestimate of the improvement in fine-mapping resolution by combining GWAS across ancestry groups. Of the loci considered, only at MTNR1B did we not see any improvement in fine-mapping resolution in terms of the number of SNPs and the genomic interval covered by the 99% credible set after trans-ethnic meta-analysis.

Table 3 Properties of the 99% credible sets of SNPs at ten established T2D susceptibility loci

| Locus   | Chr | SNPs | Interval (bp) | Build 36 location (bp) | SNPs | Interval (bp) | Build 36 location (bp) | SNPs | Interval (bp) |
|---------|-----|------|--------------|------------------------|------|--------------|------------------------|------|--------------|
| JAZF1   | 7   | 9    | 75,685       | 28,147,081–28,222,765  | 4    | 15,667       | 28,147,081–28,162,747  | 5    | 60,018       |
| SLC30A8 | 8   | 4    | 35,488       | 118,253,964–118,289,451| 2    | 243          | 118,253,964–118,254,206| 2    | 35,245       |
| CDKAL1  | 6   | 5    | 24,244       | 20,787,688–20,811,243  | 2    | 1,549        | 20,794,552–20,796,100  | 3    | 22,695       |
| HHEX/IDE| 10  | 8    | 19,195       | 94,452,862–94,472,056  | 2    | 937          | 94,455,539–94,456,475  | 6    | 18,258       |
| TCF7L2  | 10  | 3    | 13,684       | 114,744,078–114,757,761| 2    | 2,309        | 114,746,031–114,748,339| 1    | 11,375       |
| IGF2BP2 | 3   | 17   | 32,656       | 186,980,329–187,012,984| 12   | 24,504       | 186,988,481–187,012,984| 5    | 8,152        |
| FTO     | 16  | 27   | 45,981       | 52,357,008–52,402,988  | 10   | 39,335       | 52,361,075–52,400,409  | 17   | 6,646        |
| CDKN2A/B| 9   | 3    | 2,019        | 22,122,076–22,122,094  | 1    | 1            | 22,122,076–22,122,076  | 2    | 2,018        |
| PPARG   | 3   | 23   | 265,269      | 12,106,687–12,371,955  | 21   | 265,269      | 12,106,687–12,371,955  | 2    | 0            |
| MTNR1B  | 11  | 15   | 55,032       | 92,307,378–92,362,409  | 15   | 55,032       | 92,307,378–92,362,409  | 0    | 0            |

The properties shown are based on association summary statistics from the meta-analysis of the European ancestry GWAS only (12,171 cases and 56,862 controls) and the trans-ethnic meta-analysis of European, east Asian, south Asian and Mexican and Mexican American ancestry GWAS (26,488 cases and 83,964 controls).

To assess the improvements in fine-mapping resolution by combining GWAS from diverse populations, we compared the properties of the MANTRA credible set on the basis of association summary statistics from the European ancestry–only meta-analysis and the trans-ethnic meta-analysis of European, east Asian, south Asian and Mexican and Mexican American ancestry groups. We focused on 10 autosomal loci (of the 69 previously established) that attained association with T2D susceptibility at genome-wide significance in the European ancestry meta-analysis (Table 3). We did not consider loci with weaker signals of association, as they were typically characterized by large 99% credible sets in the European ancestry meta-analysis and thus might provide an overestimate of the improvement in fine-mapping resolution by combining GWAS across ancestry groups. Of the loci considered, only at MTNR1B did we not see any improvement in fine-mapping resolution in terms of the number of SNPs and the genomic interval covered by the 99% credible set after trans-ethnic meta-analysis.

We observed the greatest enhancement in fine-mapping resolution after trans-ethnic meta-analysis at the JAZF1 locus, where the genomic interval covered by the 99% credible set was reduced from 76 kb to just 16 kb (Fig. 2 and Supplementary Fig. 3). Of the nine variants in the European 99% credible set, five were excluded after trans-ethnic meta-analysis because of low LD with the lead SNP at this locus in populations of east Asian ancestry (CHB+JPT r2 < 0.05 with rs864745). Among the variants retained in the 99% credible set after trans-ethnic meta-analysis, interrogation of predicted regulatory function from the ENCODE Project showed that rs1635852 maps to a region of open chromatin with enhancer activity that is bound by several transcription factors. This SNP has been shown previously to have allelic differences in pancreatic islet enhancer activity and is also correlated with expression of CREB5, highlighting this transcript as a credible candidate at the JAZF1 locus.

We also observed a substantial reduction in the genomic interval covered by the credible set at the SLC30A8 locus (Fig. 2 and Supplementary Fig. 3) from 35 kb (four SNPs) on the basis of the European ancestry GWAS only to less than 1 kb (two SNPs) after trans-ethnic meta-analysis. However, the lead SNP is strongly correlated with all variants in the credible set before trans-ethnic meta-analysis in both the European and east Asian ancestry groups (CEU and CHB+JPT r2 ≥ 0.8 with rs13266634), suggesting that the improved fine-mapping resolution at this locus is more likely the result of increased sample size than differences in LD structure between the populations. Encouragingly, the lead SNP after trans-ethnic meta-analysis is more clearly separated from the other SNPs.
in the credible set and is a nonsynonymous variant, p.Arg325Trp, that has an established functional role in T2D susceptibility.

We next tested variants present in the 99% credible sets at the ten loci on the basis of the European ancestry GWAS only and the trans-ethnic meta-analysis for enrichment of functional annotation compared to randomly shifted element locations (Online Methods). Variants in the trans-ethnic 99% credible sets were significantly enriched (empirical \(p < 0.05\)) for overlap with DNsaseI hypersensitive sites (DHS \(P = 0.038\)) and transcription factor binding sites (TFBS \(P = 0.0060\)). However, we observed no such enrichment in either annotation category for the European ancestry 99% credible sets (DHS \(P = 0.18\); TFBS \(P = 0.087\)). These data suggest that variants retained after trans-ethnic meta-analysis show greater potential for functional impact on T2D susceptibility through these regulatory mechanisms.

The fine-mapping intervals defined by credible sets after trans-ethnic meta-analysis are limited by the density and allele frequency spectrum of the GWAS genotyping arrays and HapMap reference panels used for imputation. Although these reference panels provide comprehensive coverage of common SNPs (MAF > 5%) across ancestry groups, imputation up to phased haplotypes from the 1000 Genomes Project for example, would allow assessment of the impact of lower-frequency variation on T2D susceptibility in diverse populations. However, we have demonstrated that for a fixed reference panel, trans-ethnic meta-analysis can improve the localization of common causal SNPs within established T2D susceptibility loci, and we have identified highly annotated variants within fine-mapping intervals defined by the 99% credible sets. We have also assessed the sensitivity of the trans-ethnic fine-mapping analysis to genotype quality at directly typed or imputed SNPs (Supplementary Table 14). We repeated MANTRA fine mapping with subsets of SNPs that passed quality control in at least 80% \((n = 88,361)\) or 90% \((n = 99,406)\) of individuals from the trans-ethnic meta-analysis. As the threshold for the reported sample size increased, the number of SNPs included in the fine-mapping analysis was reduced, but the genomic intervals covered by the 99% credible sets remained unchanged, suggesting the resolution to be relatively robust to genotype quality at common variants.

DISCUSSION

We have identified seven new loci for T2D susceptibility at genome-wide significance by combining GWAS from multiple ancestry groups. Our study has provided evidence of many more common variant loci not yet reaching genome-wide significance that contribute to the heritability of T2D susceptibility, which is in agreement with polygenic analyses in European ancestry GWAS. The effects of these common variants are modest but are homogeneous across ancestry groups and would thus be amenable to discovery through trans-ethnic meta-analysis in larger samples. We have also demonstrated improvements in the resolution of fine mapping of common variant association signals through trans-ethnic meta-analysis, even in the absence of GWAS of African ancestry, which would be expected to better refine localization because of reduced LD in these populations. Future releases of reference panels from the 1000 Genomes Project are anticipated to comprise 2,500 samples, including haplotypes of south Asian ancestry and a wider representation of populations of African descent. This panel will provide a comprehensive catalog of genetic variation with MAFs as low as 0.5%, as well as many rarer variants, across major ancestry groups, thus facilitating imputation and coverage of loci for future trans-ethnic fine-mapping efforts.

Our analyses clearly highlight the benefits of combining GWAS from multiple ancestry groups for discovery and characterization of common variant loci contributing to complex traits and emphasize an exciting opportunity to further our understanding of the biological mechanisms underlying human diseases across populations from diverse ethnicities.

METHODS

Methods and any associated references are available in the online version of the paper.
ACKNOWLEDGMENTS

Funding for the research undertaken in this study has been received from the following: the Canadian Institutes of Health Research; the European Commission (ENGAGE FP7 HEALTH-F4-2007-201413); the Medical Research Council UK (G061261); the Mexico Convocatoria (SSA/IMMS/ISSSTE-CONACyT 2012-2, clave 150352, IMSS R-2011-785-018 and CONACyT Salud-2007-C01-71068); the US National Institutes of Health (DK062370, HG003076, DK085584, DK085545, DK073541 and DK085501); and the Wellcome Trust (WT098017, WT095032, WT090367, WT098381, WT081682 and WT085475). We acknowledge the many colleagues who contributed to collection and phenotypic characterization of the clinical samples and the genotyping and analysis of the GWAS data, full details of which are provided in the contributing consortia papers11,13,15. We also thank those individuals who agreed to participate in this study.

AUTHOR CONTRIBUTIONS

Writing group: A. Mahajan, M.J.G., W. Zhang, J.E.B., K.J.G., M.H., A.D.J., I.P., E.Z., Y.Y.T., M.B., E.J.P., J.C.C., E.S.T., M.I.M. and A.P.M.

Analysis group: A. Mahajan, M.J.G., W. Zhang, J.E.B., K.J.G., T. Ferreira, M.H., A.D.J., M.C.Y.N., I.P., D.S., X.W., E.Z., Y.Y.T., M.B., E.J.P., M.I.M. and A.P.M.

DIAGRAM Consortium samples, genotyping, analysis and management: A. Mahajan, M.I., P.S., D.Z., E.Z., G.R.A., P.A., M.A., D.B., B.B., I.B., J.B., R.N.B., B.O.B., E.B., L.B., N.B., H. Campbell, J.C., S.C., G.C., H. Chen, P.S.C., F.S.C., M.C.C., D.J.C., A.T.C., R.M.V.D., D. Janes, U.D.F., G.D., P.S.D., A.D., C.D., A.S.F.D., P.J.D., M.D.C., G.V., D. Dupuis, S.E., V.E., R.E., J.G.E., T.E., E.E., T. Ferreira, J.C.F., E. Fontanillas, N.G.E, T. Forsen, C.F., R.M.F., T.M.E., P. Froguel, K.G., C. Gieger, B.G., H.G., B.G.B., I.C.G., C. Guiducci, A. Hamsten, A.T.H., C. Hayward, C. Herder, A. Hofman, O.H.I., K. Hovingh, A.B.H., F.B.H., J.H.E., S. Humphries, S.E. Hunt, D.J.H., K. Hveem, T.I., E.I., B.I., A.U., J. James, K.-H.J., A. Jonsson, H.M.K., S. Kanoni, W.H.K., S. Kathiresan, M.-K.-K., H.K., K.-T., K.-K., L. N. Klop, A. Kong, E.-K.H., P. Kraft, J. Kravic, A. Kumar, J. Kuusisto, M. Laasko, V. Lagou, T.A.L., C. Langenberg, C. Langford, R.I.L., K.L., M. Li, L.L., C.M.L, E.L., C.-T.L., S. Lobhans, R.J.F.L., J. Luan, V. Lyssenko, R.M., S. Männistö, J.B.M., O.M., A. Mestpalu, J.M.G., E.M., S. Moebus, K.L.M., D.A.M., T.W.M., M.-M., N.-M., P.N., P.M.N., I.N., M.M.N., K.R.O., C.A.N.P., J.P., M.P., S. Pechivichin, N.L.P., I.P., J.R.B.P., A.P., C.G.P., S. Potter, J.F.P., L.Q., L.R., W.R., R.R., S. Raychaudhuri, N.W.R., E.R., S. Ripatti, N.R., M.R., E.J.R., I.R., D.R., T.E.S., V. Salomaa, J. Saltevo, J. Saramies, L.J.S., R.A.S., A.V.S., B.S., S. Shah, A.R.S., G. Sigurðsson, E.S., A. Silveira, S. Snipavalaratnam, A. Stanciakova, K. Stefansson, G. Steinbach, V. Steinhoffsdottir, K. Stirups, R.J.S., H.M.S., Q.S., A.-C.S., T.M.T., B.T., G.T., U.T., E. Tikkanen, J. Trakalo, E. Tremoli, M.D.T., T.T., F. Tuomilehto, A.G.U., S.V., F.V., B.E.V., N.J.W., R.W., T.W., J.F.W., S.W., W.W., A.R. Wood, L.Y., D.Z., D.A., M.B., M.I.M. and A.P.M.

AGEN-T2D Consortium samples, genotyping, analysis and management: M.J.G., X.W., L.S.A., T.A., Y.B., Q.C., J.C.N.C., L.-C.C., T.-J.C., Y.-C.C., C.-H.C., Y.-T.C., N.H.C., Y.M.C., L.-M.C.Y., G.-B.C.H., K. Hara, A.K.H., C. Hu, F.B.H., H.W.J., T.K., N. Kato, H.-L.K., S. Kim, Y.I.K., S.H.K., J.-M.I., N.-R.L., Y.L., J.I.J., J. Long, W.L., R.C.W.M., S. Maeda, K.L.M., J.-N., E.N., P.-K.N., K.O., T.H.O., K.S., X.O.S., X.S., W.Y.S., B.T., W.T.T., F.J.W., C.T.W., Y.W., J.-W.W., Y.W., K.Y., T.Y., M. Yokota, R.Z., W. Zheng, Y.C.-J., Y.L., M. Seielstad, Y.Y.T., E.S.T. and M.I.M.

SAT2D Consortium samples, genotyping, analysis and management: W. Zhang, I.P., D.S., G.R.A., T.A., A.H.B., A.B., L.F.B., M. Caulfield, C.-K.C., M. Chidambaram, J. Danesh, D.D., P.D., A.S.D., P.F., T.M.F., P. Froguel, P. Frossard, E.G., N.H., A.K.H., N.I., A.D.J., M.I., T.J., B.M., N. Kato, P. Katulanda, A.M.K.-C., C.-K., S. Kowlessur, M.M.K.-X.L., S. Liang, S. Liu, W.-Y.L., J.L.J., D.M.R., V.M., A.C.N., J.M.P., V.R., A.R., S.D.R., M. Samuel, D.K.S., J. Scott, J. Sehmi, N.S., A.S.S., X.S., K.S.S., C.S., R.T., F.T., A.R., W. Wirthmesser, T.W.Y., M. Yang, R.Y., F.Z., P.Z.Z., J. Koonen, M. Seielstad, Y.Y.T., J.C.C., E.S.T. and M.I.M.

MAT2D Consortium samples, genotyping, analysis and management: J.E.B., G.I.B., J.E., S. Krikitha, J. Kumate, A.-V.S., N.I.C., M. Cruz, C.L.H. and E.J.P.

Project management: D.A., D.W.B., Y.S.C., N.I.C., M. Cruz, C.L.H., J. Koonen, J.-Y.L., M. Seielstad, Y.Y.T., M.B., E.J.P., J.C.C., E.S.T., M.I.M. and A.P.M.

COMPETING FINANCIAL INTERESTS

The authors declare competing financial interests: details are available in the online version of the paper.

Reprints and permissions information is available online at http://www.nature.com/reprints/index.html.
The authors of this paper are:

Anubha Mahajan1–220, Min Jin Go2–220, Weihua Zhang3–220, Jennifer E Below4–220, Kyle J Gaulton1–220, Teresa Ferreira1, Momoko Horikoshi1–5, Andrew D Johnson6, Maggie C Y Ng7–8, Inga Prokopenko1–5,9, Danish Saleheen10,11, Xu Wang12, Eleftheria Zeggini13, Goncalo R Abecasis14, Linda S Adair15, Peter Almgren16, Mustafa Atalay17, Tim Aung18,19, Damiano Baldassarre20,21, Beverley Balkau22,23, Yuqian Bao24, Tom Forsen81,93, Caroline Fox6,94, Ross M Fraser46, Timothy M Frayling95, Philippe Froguel9,48, Jorge Escobedo84, Tonu Esko85–88, Elodie Eury48, Jose C Florez87–91, Pierre Fontanillas44, Nita G Forouhi92, Graeme I Bell34,35, Rafn Benediktsson36,37, Richard N Bergman38, Philippe Frossard11, Yutang Gao96, Caroline Hayward47, George B Grant44,90,125, Mark Caulfield49, Juliana C N Chan50, Li-Ching Chang51, Tien-Jyun Chang52, Yi-Cheng Chang52, Paola Cattaneo53, Christian Herder107, Albert Hofman75, Cornelia van Duijn75,76, Tazeen Jafar117,119, Alan James31,120,121, Mustafa Atalay17, Tim Aung18,19, Damiano Baldassarre20,21, Beverley Balkau22,23, Yuqian Bao24, Anthony H Barnett25,26, Ines Barroso1,5, Andrew D Johnson6, Maggie C Y Ng7–8, Inga Prokopenko1–5,9, Danish Saleheen10,11, Xu Wang12, Eleftheria Zeggini13, Goncalo R Abecasis14, Linda S Adair15, Peter Almgren16, Mustafa Atalay17, Tim Aung18,19, Damiano Baldassarre20,21, Beverley Balkau22,23, Yuqian Bao24, Anthony H Barnett25,26, Ines Barroso1,5, Andrew D Johnson6, Maggie C Y Ng7–8, Inga Prokopenko1–5,9, Danish Saleheen10,11, Xu Wang12, Eleftheria Zeggini13, Goncalo R Abecasis14, Linda S Adair15, Peter Almgren16, Mustafa Atalay17, Tim Aung18,19, Damiano Baldassarre20,21, Beverley Balkau22,23, Yuqian Bao24, Anthony H Barnett25,26, Ines Barroso1,5, Andrew D Johnson6, Maggie C Y Ng7–8, Inga Prokopenko1–5,9, Danish Saleheen10,11, Xu Wang12, Eleftheria Zeggini13, Goncalo R Abecasis14, Linda S Adair15, Peter Almgren16, Mustafa Atalay17, Tim Aung18,19, Damiano Baldassarre20,21, Beverley Balkau22,23, Yuqian Bao24, Anthony H Barnett25,26, Ines Barroso1,5, Andrew D Johnson6, Maggie C Y Ng7–8, Inga Prokopenko1–5,9, Danish Saleheen10,11, Xu Wang12, Eleftheria Zeggini13, Goncalo R Abecasis14, Linda S Adair15, Peter Almgren16, Mustafa Atalay17, Tim Aung18,19, Damiano Baldassarre20,21, Beverley Balkau22,23, Yuqian Bao24, Anthony H Barnett25,26, Ines Barroso1,5, Andrew D Johnson6, Maggie C Y Ng7–8, Inga Prokopenko1–5,9, Danish Saleheen10,11, Xu Wang12, Eleftheria Zeggini13, Goncalo R Abecasis14, Linda S Adair15, Peter Almgren16, Mustafa Atalay17, Tim Aung18,19, Damiano Baldassarre20,21, Beverley Balkau22,23, Yuqian Bao24, Anthony H Barnett25,26, Ines Barroso1,5, Andrew D Johnson6, Maggie C Y Ng7–8, Inga Prokopenko1–5,9, Danish Saleheen10,11, Xu Wang12, Eleftheria Zeggini13, Goncalo R Abecasis14, Linda S Adair15, Peter Almgren16, Mustafa Atalay17, Tim Aung18,19, Damiano Baldassarre20,21, Beverley Balkau22,23, Yuqian Bao24, Anthony H Barnett25,26, Ines Barroso1,5, Andrew D Johnson6, Maggie C Y Ng7–8, Inga Prokopenko1–5,9, Danish Saleheen10,11, Xu Wang12, Eleftheria Zeggini13, Goncalo R Abecasis14, Linda S Adair15, Peter Almgren16, Mustafa Atalay17, Tim Aung18,19, Damiano Baldassarre20,21, Beverley Balkau22,23, Yuqian Bao24, Anthony H Barnett25,26, Ines Barroso1,5, Andrew D Johnson6, Maggie C Y Ng7–8, Inga Prokopenko1–5,9, Danish Saleheen10,11, Xu Wang12, Eleftheria Zeggini13, Goncalo R Abecasis14, Linda S Adair15, Peter Almgren16, Mustafa Atalay17, Tim Aung18,19, Damiano Baldassarre20,21, Beverley Balkau22,23, Yuqian Bao24, Anthony H Barnett25,26, Ines Barroso1,5, Andrew D Johnson6, Maggie C Y Ng7–8, Inga Prokopenko1–5,9, Danish Saleheen10,11, Xu Wang12, Eleftheria Zeggini13, Goncalo R Abecasis14, Linda S Adair15, Peter Almgren16, Mustafa Atalay17, Tim Aung18,19, Damiano Baldassarre20,21, Beverley Balkau22,23, Yuqian Bao24.
ONLINE METHODS

Ancestry-specific GWAS meta-analyses. Ancestry-specific meta-analyses have been previously performed by the DIAGRAM Consortium (12,171 cases and 56,862 controls; European ancestry)\(^2\); the AGEN-T2D Consortium (6,952 cases and 11,865 controls; east Asian ancestry)\(^3\); the SAT2D Consortium (5,561 cases and 14,458 controls; south Asian ancestry)\(^1\); and the MAT2D Consortium (1,804 cases and 779 controls; Mexican and Mexican American ancestry)\(^5\). Further details of the samples and methods employed within each ancestry group are presented in the corresponding consortium papers\(^5,11,13,15\). Briefly, individuals were assayed with a range of genotyping products, with sample and SNP quality control (QC) undertaken within each study (Supplementary Tables 1 and 2). Each GWAS scaffold was imputed up to 2.5 million autosomal SNPs using reference panels from Phase II/III HapMap\(^2,22,23\) (Supplementary Table 2). Each SNP with MAF > 1%, or MAF > 5% in the Mexican and Mexican American ancestry GWAS because of smaller sample size and passing QC was tested for association with T2D under an additive model after adjustment for study-specific covariates (Supplementary Table 2). Covariate adjustments were not harmonized within or between consortia because of variation in study design and recorded non-genetic risk factors. The results of each GWAS were corrected for population structure with genomic control\(^20\) (unless \(\lambda_{GC} < 1\)). Association summary statistics from GWAS within each ancestry group were then combined by fixed-effects meta-analysis. The results of each ethnic-specific meta-analysis were then corrected by a second round of genomic control: European ancestry (\(\lambda_{GC} = 1.10\)); east Asian ancestry (\(\lambda_{GC} = 1.05\)); south Asian ancestry (\(\lambda_{GC} = 1.02\)); Mexican and Mexican American ancestry (\(\lambda_{GC} = 1.01\)).

Trans-ethnic discovery-stage GWAS meta-analysis. Association summary statistics from each ancestry-specific meta-analysis were combined in a fixed-effects inverse-variance–weighted meta-analysis (in a total of 26,488 cases and 83,964 controls). The association results of the trans-ethnic meta-analysis were corrected by genomic control\(^50\) (\(\lambda_{GC} = 1.05\)).

Heterogeneity analyses. For each previously reported lead SNP at an established T2D susceptibility locus, we assessed heterogeneity in allelic effects between the ethnic-specific meta-analyses by means of Cochran’s Q statistic\(^31\) (Supplementary Table 3). Among the 52 SNPs passing QC in all four ethnic-specific meta-analyses, we identified those that showed the same direction of effect across all ancestry groups and evaluated the significance of the excess in concordance (12.5% expected) with a one-sided binomial test.

Concordance analyses. We identified SNPs passing QC and with MAF > 1% in all four ethnic-specific meta-analyses. We excluded variants in the 69 established autosomal T2D susceptibility loci, defined as 500 kb upstream and 500 kb downstream of the previously reported lead SNPs. We also excluded AT/TC/GC SNPs to eliminate bias due to strand misalignment between ethnic-specific meta-analyses. Among the remaining SNPs, we selected an independent subset with nominal evidence of association (\(P \leq 0.001\)) with T2D from the European ancestry meta-analysis and separated by at least 500 kb. For each independent SNP, we identified the T2D risk allele from the European ancestry meta-analysis and determined the direction of effect in the east Asian, south Asian and Mexican and Mexican American ancestry meta-analyses. We excluded 5 SNPs that had the same direction of effect for the European ancestry risk allele and the significance of the excess in concordance (50% expected) with a one-sided binomial test. We repeated this analysis for SNPs with weaker evidence of association with T2D from the European ancestry meta-analysis: 0.001 < \(P \leq 0.01\); 0.01 < \(P \leq 0.5\); and 0.5 < \(P \leq 1\) (Table 1). We then repeated these analyses using the east Asian, south Asian and Mexican and Mexican American ancestry meta-analyses, in turn, to identify subsets of independent T2D risk alleles and assessed concordance into the other ethnic groups (Supplementary Table 4).

European ancestry validation-stage meta-analysis. The previously published validation meta-analysis consisted of 21,491 cases and 55,647 controls of European ancestry from the DIAGRAM Consortium\(^5\), all genotyped with the Metabochip\(^26\) (Supplementary Table 1). We excluded the Pakistan Risk Of Myocardial Infarction Study (PROMIS) from the validation meta-analysis to avoid overlap with a subset of the same individuals contributing to the SAT2D Consortium meta-analysis\(^23\). Full details of the samples and methods employed in the validation meta-analysis are presented in the DIAGRAM Consortium paper\(^5\). Briefly, sample and SNP QC were undertaken within each study (Supplementary Table 2). Each high-quality SNP (MAF > 1%) was tested for association with T2D under an additive model after adjustment for study-specific covariates (Supplementary Table 2). Association summary statistics for each study were corrected using the genomic control inflation factor obtained from a subset of 3,598 ‘QT interval’ replication SNPs\(^2,26\) (unless \(\lambda_{QT} < 1\)). These statistics were then combined in a fixed-effects inverse-variance–weighted meta-analysis and were corrected by a second round of genomic control (\(\lambda_{QT} = 1.19\)).

Combined meta-analysis. We selected lead SNPs at 33 new loci with suggestive evidence of association (\(P < 10^{-5}\)) from the trans-ethnic discovery GWAS meta-analysis for in silico follow up in the European ancestry validation meta-analysis. Of these SNPs, 16 were genotyped directly on Metabochip, and 10 more had a proxy (CEU and CHB+JPT HapMap \(r^2 \geq 0.2\)). For these 26 SNPs, association summary statistics from the discovery and validation meta-analyses were combined in a fixed-effects inverse-variance–weighted meta-analysis (Supplementary Table 5). The combined meta-analysis consisted of 47,979 T2D cases and 139,611 controls. Heterogeneity in allelic effects between the two stages of the combined meta-analysis was assessed by means of Cochran’s Q statistic\(^31\).

Sensitivity to covariate adjustment. We identified 19 studies (11,327 cases and 31,342 controls) from the European ancestry validation meta-analysis that adjusted for only age, sex (unless male or female specific) and population structure where necessary (Supplementary Table 2): AMC-PAS; BHS; DLGOM; EAS; EGCUT; EMIL-ULM; EPIC; FUSION Stage 2; D2D2007; Dr’s Extra; HUNT; METSIM (male specific); HNR; IMPROVE; KORAGen Stage 2; PIVUS; THISEAS; ULSAM (male specific); and WARREN2. Association summary statistics from each of these studies were then combined in a fixed-effects inverse-variance–weighted meta-analysis, the results of which were subsequently corrected for genomic control (\(\lambda_{QT} = 1.12\)). The remaining six studies (10,164 cases and 24,305 controls) did not adjust for age and/or sex or include additional covariates to account for BMI or cardiovascular-related disease status (Supplementary Table 2): deCODE Stage 2; DUNDEE; GmE6S; PMB; SCARFSHEEP; and STR. Association summary statistics from each of these studies were then combined in a fixed-effects inverse-variance–weighted meta-analysis but did not require subsequent correction for genomic control (\(\lambda_{QT} = 1.00\)). We then tested for heterogeneity in allelic effects between these two sets of studies by means of Cochran’s Q statistic\(^31\) (Supplementary Table 7).

Association of lead T1D SNPs with T2D. We obtained association summary statistics with T2D from the trans-ethnic meta-analysis for previously reported lead SNPs in established T1D susceptibility loci\(^27\) (Supplementary Table 8). For each SNP, we aligned the allelic effect on T2D according to the risk allele for T1D (where reported). We also obtained association summary statistics for T1D HLA risk alleles: HLA-DR4 (rs660895) and HLA-DR3 (rs2187668).

Association of lead T2D SNPs with metabolic traits. We obtained association summary statistics (\(P\) values, directed \(Z\) scores and/or allelic effects and corresponding standard errors) for lead SNPs at new T2D susceptibility loci in published European ancestry GWAS meta-analyses of metabolic phenotypes: glycemic traits\(^3,30\), anthropometric measures\(^32,33\) and plasma lipid concentrations\(^34\). We considered glycemic traits in non-diabetic individuals from the MAGIC Investigators (Supplementary Table 9). For FG and F1 concentrations (with and without adjustment for BMI), the meta-analysis consisted of up to 133,010 and 108,557 individuals, respectively. For HOMA-B and HOMA-IR, the meta-analysis consisted of up to 37,037 individuals. We considered anthropometric measures from the GIANT Consortium (Supplementary Table 10). For BMI and waist:hip ratio adjusted for BMI, the meta-analysis consisted of 123,865 and 77,167 individuals, respectively. We then considered plasma lipid concentrations from the Global Lipids Genetics Consortium (Supplementary Table 11). For total cholesterol, high-density lipoprotein cholesterol, low-density cholesterol, high-density lipoprotein cholesterol, and low-density lipoprotein cholesterol, we used the inverse-variance weighted meta-analysis with HOMA-B and HOMA-IR, respectively. For BMI and waist:hip ratio adjusted for BMI, we used the inverse-variance weighted meta-analysis with HOMA-B and HOMA-IR, respectively.
lipoprotein cholesterol and triglycerides, the meta-analysis consisted of up to 100,184 individuals.

Expression analyses. We interrogated public databases and unpublished resources for cis-eQTL expression with lead SNPs in the new susceptibility loci in multiple tissues. Details of these resources are summarized in the Supplementary Note. The collated results from these resources met study-specific criteria for statistical significance for association with expression. For each transcript associated with the lead T2D SNP (Supplementary Table 12), we identified the cis-eQTL SNP with the strongest association with expression in the same tissue and subsequently estimated the LD between them using pilot data from the 1000 Genomes Project25 (CEU and CHB+JPT) to assess the coincidence of the signals.

Functional annotation. We identified variants in pilot data from the 1000 Genomes Project25 that are in strong LD (CEU and CHB+JPT $r^2 > 0.8$) with the lead SNPs in the new susceptibility loci for functional annotation. Identified nonsynonymous variants were interrogated for likely downstream functional consequences using SIFT35 (Supplementary Table 13). Variants were also assessed for overlap with regions of predicted regulatory function generated by the ENCODE Project36 including: ChromHMM regulatory state definitions from 9 cell lines (GM12878, HepG2, HUVEC, HMEC, HSMM, K562, NHLE, NHEK and hESC); transcription factor binding ChIP sites from 95 cell types; open chromatin (DNaseI hypersensitivity) sites from 125 cell types; transcripts correlated with open chromatin site activity; and sequence motifs from JASPAR, TRANSFAC and de novo prediction (Supplementary Fig. 2).

Fine-mapping analyses. We used MANTRA39 to fine-map T2D susceptibility loci on the basis of association summary statistics from the meta-analysis of European ancestry GWAS only5 and the trans-ethnic meta-analysis of European, east Asian, south Asian and Mexican and Mexican American ancestry GWAS5,11,13,15. MANTRA allows for trans-ethnic heterogeneity in allelic effects arising as a result of differences in the structure of LD with the causal variant in diverse populations by assigning ancestry groups to ‘clusters’ according to a Bayesian partition model of relatedness between them, as defined by pairwise genome-wide mean allele frequency differences (Supplementary Fig. 4). Evidence in favor of association of each SNP with T2D is measured by a Bayes’ factor (BF). We assume a single causal variant for T2D at each locus (defined by the region 500 kb upstream and 500 kb downstream of the lead SNP from the trans-ethnic meta-analysis). We then calculated the posterior probability that the $j$th SNP is causal among those reported in the meta-analysis:

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

In this expression, $BF_j$ denotes the BF in favor of association of the $j$th SNP, and the summation in the denominator is over all variants passing QC across the locus41. A 99% credible set of variants was then constructed by ranking all SNPs according to their BF and combining ranked SNPs until their cumulative posterior probability exceeds 0.99.

SNPs in the 99% credible sets were assessed for enrichment in ChromHMM regulatory state (enhancer, promoter and insulator), DNaseI hypersensitive and transcription factor binding sites using data from the ENCODE Project36. We performed 1,000 permutations by shifting the location of the annotation sites a random distance within 100 kb and recalculated the overlap to obtain empirical $P$ values for enrichment in each annotation category.

50. Devlin, B. & Roeder, K. Genomic control for association studies. *Biometrics* **55**, 997–1004 (1999).
51. Ioannidis, J.P. *et al.* Heterogeneity in meta-analyses of genome-wide association investigations. *PLoS ONE* **2**, e841 (2007).