

# Identification of type 2 diabetes loci in 433,540 East Asian individuals

<https://doi.org/10.1038/s41586-020-2263-3>

Received: 24 June 2019

Accepted: 2 March 2020

Published online: 6 May 2020

 Check for updates

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

Meta-analyses of genome-wide association studies (GWAS) have identified more than 240 loci that are associated with type 2 diabetes (T2D)<sup>1,2</sup>; however, most of these loci have been identified in analyses of individuals with European ancestry. Here, to examine T2D risk in East Asian individuals, we carried out a meta-analysis of GWAS data from 77,418 individuals with T2D and 356,122 healthy control individuals. In the main analysis, we identified 301 distinct association signals at 183 loci, and across T2D association models with and without consideration of body mass index and sex, we identified 61 loci that are newly implicated in predisposition to T2D. Common variants associated with T2D in both East Asian and European populations exhibited strongly correlated effect sizes. Previously undescribed associations include signals in or near *GDAP1*, *PTF1A*, *SIX3*, *ALDH2*, a microRNA cluster, and genes that affect the differentiation of muscle and adipose cells<sup>3</sup>. At another locus, expression quantitative trait loci at two overlapping T2D signals affect two genes—*NKX6-3* and *ANK1*—in different tissues<sup>4–6</sup>. Association studies in diverse populations identify additional loci and elucidate disease-associated genes, biology, and pathways.

Type 2 diabetes is a common metabolic disease that is primarily caused by insufficient insulin production and/or secretion by the pancreatic β cells, and insulin resistance in peripheral tissues<sup>7</sup>. Most genetic loci associated with T2D have been identified in populations of European (EUR) ancestry, including a recent meta-analysis of GWAS of nearly 900,000 individuals of European ancestry that identified more than 240 loci that influence risk of T2D<sup>1</sup>. Differences in allele frequency between ancestries affect the power to detect associations within a population, particularly among variants that are rare or monomorphic in one population but more frequent in another<sup>2,8</sup>. Although smaller than studies of European populations, a recent T2D meta-analysis of data from almost 200,000 Japanese individuals identified 28 additional loci<sup>2</sup>. The relative contributions of different pathways to the pathophysiology of T2D may also differ between ancestry groups. For example, the prevalence of T2D is greater in East Asian (EAS) populations than in European populations among people of similar body mass index (BMI) or waist circumference<sup>9</sup>. To identify new genetic associations and provide insight into the pathogenesis of T2D, we performed, to our knowledge, the largest meta-analysis of T2D data from East Asian individuals so far.

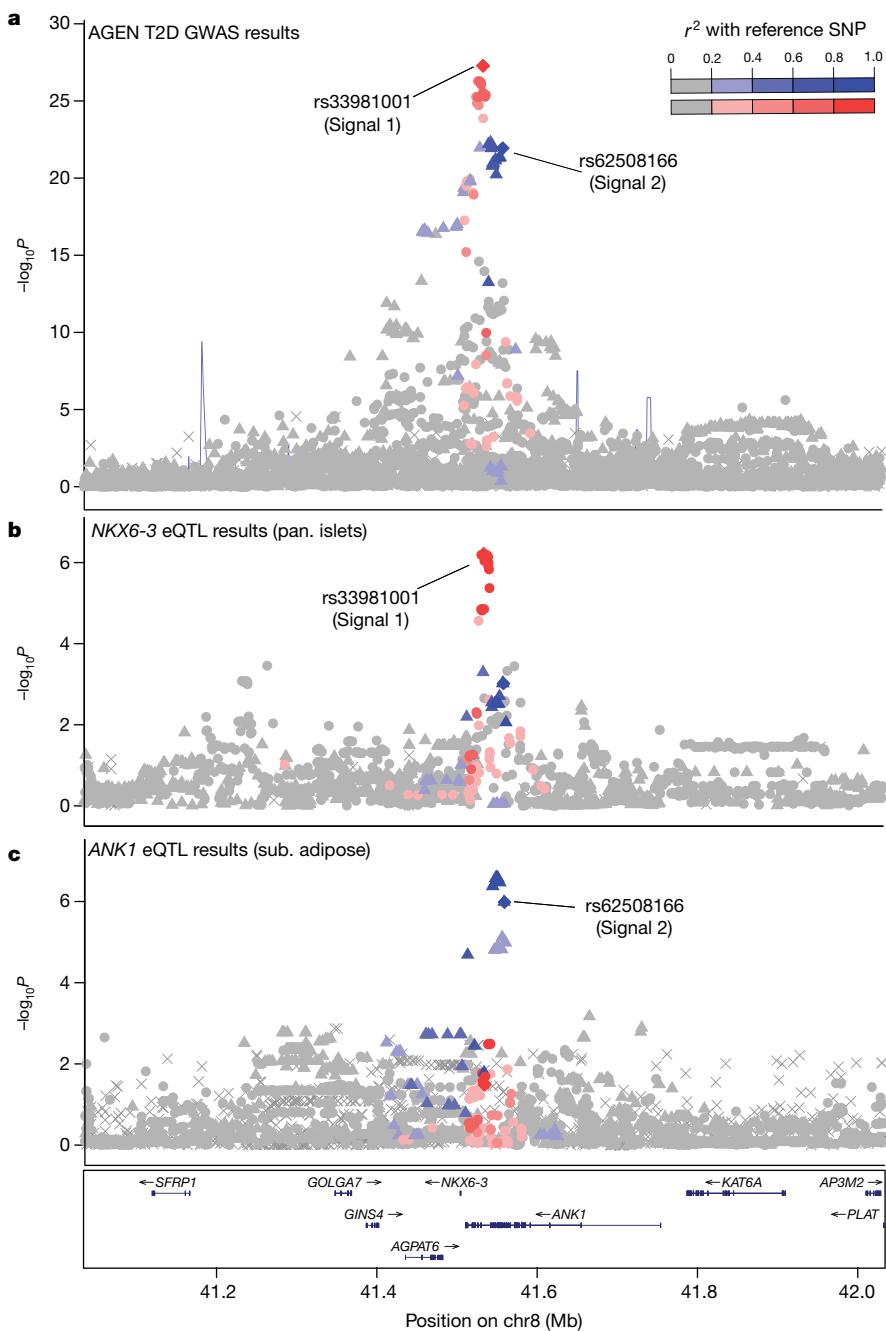
## Genetic discovery from association analyses

We conducted a fixed-effect inverse-variance weighted GWAS meta-analysis that combined 23 studies imputed to the 1000 Genomes Phase 3 reference panel from the Asian Genetic Epidemiology Network (AGEN) consortium (Supplementary Tables 1–3). We performed sex-combined T2D association without BMI adjustment for 77,418 individuals with T2D and 356,122 controls (effective sample size,  $N_{\text{eff}} = 211,793$ ). For the subset of 54,481 individuals with T2D and 224,231 controls ( $N_{\text{eff}} = 135,780$ ) for whom BMI was available, we performed additional analyses with and without BMI adjustment in sex-combined and sex-stratified models (Extended Data Fig. 1). We defined ‘lead’ variants as the strongest T2D-associated variants with  $P < 5 \times 10^{-8}$  and defined the region  $\pm 500$  kb from the lead variant as a locus. A locus was

considered ‘novel’ if the lead variant was located at least 500 kb from previously reported T2D-associated variants in any ancestry.

Using summary association statistics for about 11.8 million variants, without adjustment for BMI (Extended Data Fig. 1; Supplementary Tables 1–3), we identified lead variants associated with T2D at 183 loci, of which 51 were novel (Extended Data Table 1; Extended Data Fig. 2; Supplementary Table 4). Lead variants at all novel loci were common (minor allele frequency (MAF)  $\geq 5\%$ ; Extended Data Fig. 3), except for low-frequency variants near *GDAP1* (MAF = 2.4%), which regulates mitochondrial proteins and metabolic flux in skeletal muscle<sup>10</sup>, and *PTF1A* (MAF = 4.7%), which encodes a transcription factor that is required for pancreatic acinar cell development<sup>11</sup>. Lead variants met a stricter *P*-value threshold for significance based on Bonferroni correction for 11.8 million tests ( $P < 4.2 \times 10^{-9}$ ) at 146 of the 183 loci, including 29 of the 51 novel loci.

Using genome-wide complex trait analysis (GCTA)<sup>12</sup>, we identified 301 distinct association signals that met a locus-wide significance threshold of  $P < 1 \times 10^{-5}$  (Supplementary Table 5), 228 of which were genome-wide significant ( $P < 5 \times 10^{-8}$ ). Overall, we observed 2–4 signals at 46 loci and 5 or more signals at 12 loci. Among the ten loci with the most significant meta-analysis *P*-values of association, 7 contained 5 or more distinct signals (17 signals at *INS/IGF2/KCNQ1*; 7 signals at *CDKN2A/CDKN2B* and *GRM8/PAX4/LEP*; 5 signals at *CDKAL1/HHEX/IDE/CDC123/CAMK1D*, and *TCF7L2*; Extended Data Fig. 4; Supplementary Table 5). The seven signals at the *GRM8/PAX4/LEP* locus spanned 1.4 Mb, and, to our knowledge, no evidence of T2D association at this locus has previously been reported in populations of non-East Asian ancestry<sup>1,13</sup> (Extended Data Fig. 4c). Joint analyses confirmed independent associations (linkage disequilibrium (LD)  $r^2 = 0.0025$ ) at two previously reported *PAX4* missense variants<sup>14</sup>: rs2233580 (Arg192His: risk allele frequency (RAF) = 8.6%, odds ratio (OR) = 1.31, 95% confidence interval (CI) 1.28–1.34, GCTA *P*-value ( $P_{\text{GCTA}}$ ) =  $3.4 \times 10^{-93}$ ) and rs3824004 (Arg192Ser: RAF = 3.4%, OR = 1.24, 95% CI 1.19–1.28,  $P_{\text{GCTA}} = 1.1 \times 10^{-30}$ ). The association signals at this locus also included variants near *LEP*, which encodes leptin, a hormone that regulates appetite<sup>15</sup>; increased leptin levels are associated with obesity and T2D<sup>16</sup>.

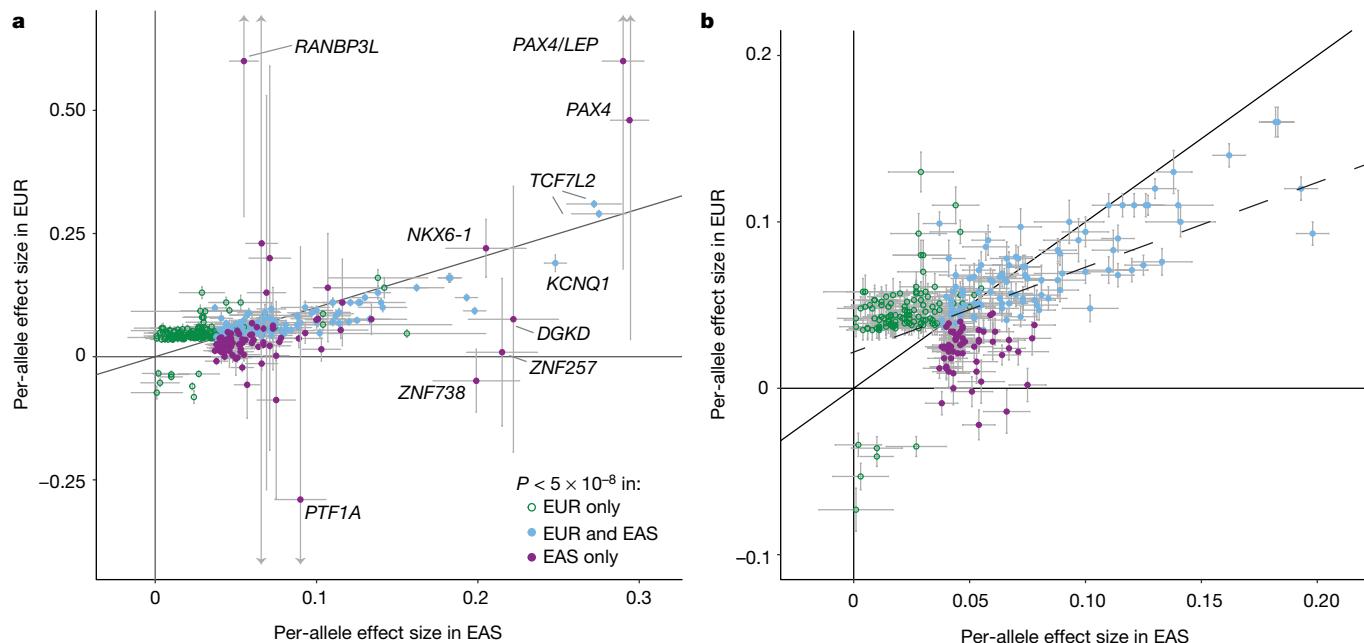


**Fig. 1 | Two distinct T2D-association signals at the *ANK1*-*NKX6-3* locus are associated with expression levels of two transcripts in two tissues.** **a**, Regional association plot for East Asian sex-combined BMI-unadjusted two-sided fixed-effect inverse-variance meta-analysis at *ANK1*-*NKX6-3* locus. Approximate conditional analysis using GCTA identified three distinct T2D-association signals ( $P < 1 \times 10^{-5}$ ) at this locus (signal 1, rs33981001,  $N_{\text{eff}} = 211,793$ ; signal 2, rs62508166,  $N_{\text{eff}} = 211,793$ ; signal 3, rs144239281,  $N_{\text{eff}} = 208,431$ , in order of strength of association), of which two signals are associated with expression levels of two transcripts in two tissues. Using 1000 Genomes (1000G) Phase 3 East Asian LD, variants are coloured red and blue for the first and second distinct signals, respectively (lead variants represented as diamonds). **b**, Variant rs12549902, in high LD (EAS LD  $r^2 = 0.80$ , EUR  $r^2 = 0.83$ ) with T2D signal 1, shows the strongest association with expression levels of *NKX6-3* in pancreatic (pan.) islets in 118 individuals<sup>16</sup>. **c**, Variant rs516946, in high LD (EAS LD  $r^2 = 0.96$ , EUR  $r^2 = 0.80$ ) with T2D signal 2, shows the strongest association with expression levels of *ANK1* in subcutaneous (sub.) adipose tissue in 770 individuals<sup>19</sup>. As rs62508166 is not available in the subcutaneous adipose tissue data set, a variant in perfect LD (rs28591316) was used and is represented by the blue diamond.

At the previously reported *ANK1*/*NKX6-3* locus<sup>1,17</sup>, we observed three distinct T2D association signals, two of which overlapped and consisted of variants spanning only about 25 kb (Fig. 1a). Given conflicting interpretations of candidate genes<sup>1,5,18</sup>, we compared the T2D-association signals identified in East Asian individuals to eQTLs reported at this locus in pancreatic islets<sup>1,18–20</sup>, subcutaneous adipose tissue<sup>6</sup>, and skeletal muscle<sup>5</sup>. At the strongest signal, the lead T2D-associated variant rs33981001 was in high LD with the lead *cis*-eQTL variant for *NKX6-3* in pancreatic islets<sup>18</sup> (rs12549902; EAS LD  $r^2 = 0.79$ , EUR  $r^2 = 0.83$ ), and the T2D risk allele was associated with decreased expression of *NKX6-3* (per-allele effect size ( $\beta$ ) =  $-0.36$ ,  $P = 6.1 \times 10^{-7}$ ; Fig. 1b). *NKX6-3* encodes a pancreatic islet transcription factor that is required for the development of  $\alpha$  and  $\beta$  cells in the pancreas<sup>21</sup> and has been shown to influence insulin secretion<sup>16</sup>. At the second T2D-association signal, rs62508166 is in high LD with the lead *cis*-eQTL variant for *ANK1* in subcutaneous adipose tissue<sup>19</sup> and skeletal muscle<sup>15</sup> (rs516946; EAS LD  $r^2 = 0.96$ , EUR  $r^2 = 0.80$ ), and the T2D risk allele is associated with increased expression

of *ANK1* (subcutaneous adipose:  $\beta = 0.20$ ,  $P = 1.8 \times 10^{-7}$ ; skeletal muscle:  $\beta = 1.01$ ,  $P = 2.8 \times 10^{-22}$ ; Fig. 1c). *ANK1* belongs to the ankyrin family of integral membrane proteins and has been shown to affect glucose uptake in skeletal muscle; changes in its expression level may lead to insulin resistance<sup>22</sup>. Together, these GWAS and eQTL signals suggest that variants within this approximately 25-kb region act to increase or decrease the expression of two genes in different tissues to increase risk of T2D.

In T2D association analyses adjusted for BMI, we identified an additional six loci, four of which have not previously been reported for T2D to our knowledge, including loci near *MYOM3/SRSF10*, *TSN*, *GRB10*, and *NID2* (Supplementary Table 4). At the *NID2* locus, the T2D risk allele is very rare or monomorphic in non-East Asian individuals and has previously demonstrated significant associations with lower BMI and higher triglycerides in East Asian individuals, consistent with a lipodystrophy phenotype<sup>23,24</sup>. The lead *GRB10* variant is in low LD (EUR  $r^2 = 0.08$ , EAS  $r^2 = 0.57$ ) with a variant associated with glucose-stimulated insulin secretion in European individuals<sup>25</sup>.



**Fig. 2 | Effect size comparison of lead variants identified in this East Asian T2D GWAS BMI-unadjusted meta-analysis and previous European T2D GWAS meta-analysis<sup>2</sup>.** For 332 unique lead variants identified from the two BMI unadjusted meta-analyses, per-allele effect sizes ( $\beta$ ) from the European meta-analysis<sup>2</sup> (y axis) were plotted against per-allele effect sizes from our East Asian meta-analysis (x axis). Effect sizes from both meta-analyses were from two-sided fixed-effect inverse-variance meta-analysis (maximal  $N_{\text{eff}} = 211,793$  for East Asian and 231,436 for European meta-analyses). Each point denotes the per-allele effect size; grey lines, s.e. **a**, All 332 lead variants. **b**, The 278 lead

variants with minor allele frequency  $\geq 5\%$  in both ancestries. Purple, significant ( $P < 5 \times 10^{-8}$ ) in the East Asian analysis only; green, significant in European analysis only; blue, significant in both analyses (see Methods and Supplementary Table 7). The dashed diagonal line represents the trend line across all plotted variants. Compared to Supplementary Table 7, 70 variants are not plotted; 31 variants were present only in the analysis of East Asian individuals (median effect size 0.065; interquartile range 0.049–0.110) and 39 variants were present only in the analysis of European individuals (median effect size 0.083; interquartile range 0.063–0.170).

Across the models with and without adjustment for BMI, correlation for the effect sizes genome-wide was higher in East Asian individuals ( $r = 0.98$ ) than in European individuals ( $r = 0.89$ ). For the 189 T2D-associated loci in East Asian individuals, the correlation increased to 0.99 (Extended Data Fig. 5). Loci with larger effects in BMI-adjusted models include *FGFR2* and *NID2*, which have been identified only in East Asian populations and associated with lipodystrophy traits or body fat distribution. These results may reflect the role of body fat distribution in insulin resistance and T2D among East Asian individuals.

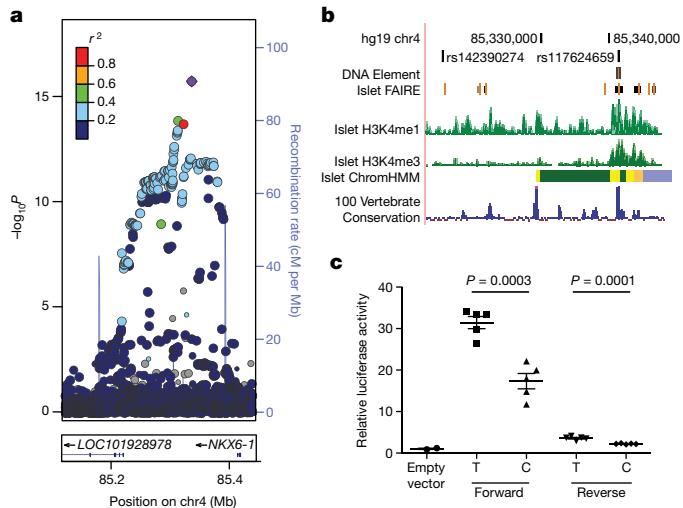
In sex-stratified analyses of males (28,027 cases and 89,312 controls) and females (27,370 cases and 135,055 controls), we identified six additional novel sex-specific loci: (i) three male-specific loci near *FOXK1*, *PDE3A*, and *IFT81* and one female-specific locus near *LMTK2* in models without adjustment for BMI, and (ii) one male-specific locus near *LINCO0851* and one female-specific locus near *CPS1* in models with adjustment for BMI (Supplementary Table 6). The lead *CPS1* variant rs1047891 (Thr1412Asn) has been reported to have a stronger effect in females than in males for cardiovascular disease and several blood metabolites<sup>26</sup>. Altogether, we identified 61 novel loci across BMI-unadjusted, BMI-adjusted, and sex-stratified models, of which 33 met a stricter  $P$  value threshold ( $P < 4.2 \times 10^{-9}$ ).

Among all T2D-associated loci, a region that spans about 2 Mb near *ALDH2* exhibited the strongest differences between sexes (rs12231737, sex-heterogeneity  $P$  value ( $P_{\text{het}}$ ) =  $2.6 \times 10^{-19}$ ), with compelling evidence of association in males ( $P_{\text{males}} = 5.8 \times 10^{-27}$ ) and no evidence for association in females ( $P_{\text{females}} = 0.19$ ) (Extended Data Fig. 6; Supplementary Table 6). This sex difference was also observed after adjusting for BMI ( $P_{\text{males, adjBMI}} = 5.2 \times 10^{-21}$ ,  $P_{\text{females, adjBMI}} = 0.053$ ). Furthermore, joint conditional analyses revealed two conditionally distinct signals (rs12231737,  $P_{\text{GCTA}} = 1.7 \times 10^{-21}$ ; rs557597782,  $P_{\text{GCTA}} = 4.9 \times 10^{-7}$ ) in males only. *ALDH2* encodes a key enzyme in alcohol metabolism that converts acetaldehyde into acetic acid. This stretch of T2D associations in males reflects a long LD block that

arose due to a recent selective sweep in East Asian individuals and results in flushing, nausea, and headache following alcohol consumption<sup>27</sup>. The most significantly associated missense variant in moderate LD with rs12231737 ( $r^2 = 0.68$ ) was rs671 (*ALDH2* Glu504Lys: RAF = 77.7%, OR = 1.17, 95% CI 1.14–1.20,  $P_{\text{males}} = 1.5 \times 10^{-24}$ ), which leads to reduced ALDH2 activity and reduced alcohol metabolism, and has been associated with cardiometabolic traits in East Asian populations. The T2D risk allele is associated with better tolerance for alcohol; increased BMI, blood pressure, and high-density lipoprotein cholesterol; and decreased low-density lipoprotein cholesterol and cardiovascular risk<sup>28–30</sup>. The strong sexual dimorphism observed at this locus may be due to differences in alcohol consumption patterns between males and females<sup>28,30</sup>, the effects of alcohol on BMI, and/or differences in the effect of alcohol on insulin sensitivity<sup>31</sup>.

## Comparing effects in EAS and EUR populations

With an effective sample size comparable to the largest study of T2D in European individuals (East Asian  $N_{\text{eff}} = 211,793$ ; European  $N_{\text{eff}} = 231,436$ )<sup>1</sup> and imputation to a dense 1000 Genomes reference panel, our results provide the most comprehensive and precise catalogue of East Asian T2D effects so far, to our knowledge, for comparisons across ancestries (Fig. 2; Supplementary Table 7). For 183 EAS T2D loci and 231 EUR T2D loci (unadjusted for BMI)<sup>1</sup>, we compared the per-allele effect sizes for the 332 variants available in both data sets (polymorphic and passed quality control), including lead variants from both ancestries at shared signals. Overall, the per-allele effect sizes between the two ancestries were moderately correlated ( $r = 0.55$ ; Fig. 2a). When the comparison was restricted to the 278 variants that are common (MAF  $\geq 5\%$ ) in both ancestries, the effect size correlation increased to  $r = 0.59$  (Fig. 2b; Extended Data Fig. 7). This effect size correlation further increased to  $r = 0.87$  for 106 variants that were significantly associated with T2D ( $P < 5 \times 10^{-8}$ ) in both ancestries. Using Cochran's heterogeneity test, 28 of 332 variants (8.4%) exhibited



**Fig. 3 | rs117624659 at the *NKX6-1* locus exhibits allelic differences in transcriptional activity.** **a**, rs117624659 ( $N_{\text{eff}} = 211,214$ ; purple diamond) shows the strongest association with T2D in the region. Two-sided fixed-effect inverse-variance meta-analysis. Variants are coloured based on 1000G Phase 3 East Asian LD with rs117624659. **b**, rs117624659 and an additional candidate variant rs142390274 in high pairwise LD ( $r^2 > 0.80$ ) span a 22-kb region approximately 75 kb upstream of *NKX6-1*. rs117624659 overlaps a region of open chromatin in pancreatic islets and lies within a region that is conserved across vertebrates. **c**, rs117624659-T (T), which is associated with increased risk of T2D, showed greater transcriptional activity in an element cloned in both forward and reverse orientations with respect to *NKX6-1* in MIN6 cells compared to rs117624659-C (C) and an ‘empty vector’ containing a minimal promoter. Data are mean  $\pm$  s.e.m. relative luciferase activity; analyses from two-sided, unpaired *t*-tests using data from  $n=5$  biologically independent samples or independent experiments are shown.

significant heterogeneity in effect size between East Asian and European populations, including 22 that were significant in only one population (Supplementary Table 7) and six with larger effect sizes in one population (for example, *CDKAL1*, *KCNQ1*, and *HNF1B*). Although the overall effect sizes for all 332 variants appear, on average, to be stronger in East Asian individuals than in European individuals, this trend is reduced when each locus is represented only by the lead variant from one population (Extended Data Fig. 8). Specifically, 39 variants identified in the European meta-analysis with imputation using the Haplotype Reference Consortium panel are missing from the comparison because they were rare, monomorphic or poorly imputed in the East Asian meta-analysis, with imputation based on the smaller and more heterogeneous 1000 Genomes reference panel.

Variants that exhibit the largest differences in effect size across ancestries are generally rare ( $\text{MAF} \leq 0.1\%$ ) in European populations but common (for example, *PAX4*, *RANBP3L*) or low-frequency (for example, *ZNF257*, *DGKD*) in East Asian populations. For example, rs142395395 near *ZNF257* ( $\text{RAF} = 96.9\%$ ,  $\text{OR} = 1.24$ , 95% CI 1.19–1.29,  $P = 7.0 \times 10^{-23}$ ) has been reported only twice in 15,414 individuals of non-Finnish European ancestry from the gnomAD database<sup>32</sup>. This variant tags a previously described 415-kb inversion, observed only in East Asian individuals, that disrupts the coding sequence and expression of *ZNF257* as well as lymphoblastoid expression of 81 downstream genes and transcripts<sup>33</sup>. These data suggest that *ZNF257* and/or its downstream target genes influence susceptibility to T2D.

We identified many loci for which the lead variants exhibited similar allele frequencies and effect sizes in both the East Asian and European meta-analyses, but reached genome-wide significance only in the East Asian meta-analysis. Given shared susceptibility across ancestry groups, these loci may be detected in non-East Asian populations when sample sizes increase. Among these variants is rs117624659, located near *NKX6-1* ( $P_{\text{EAS}} = 2.0 \times 10^{-16}$ ,  $P_{\text{EUR}} = 2.2 \times 10^{-4}$ ). This lead variant overlaps a highly

conserved region that shows open chromatin specific to pancreatic islets. We conducted transcriptional reporter assays in MIN6 mouse insulinoma cells and observed that rs117624659 exhibited significant allelic differences in enhancer activity (Fig. 3). In the pancreas, *NKX6.1* is required for the development of insulin-producing  $\beta$  cells and is a potent bifunctional transcriptional regulator<sup>34</sup>. Furthermore, inactivation of *Nkx6.1* in mice induced rapid-onset diabetes resulting from defects in biosynthesis and secretion of insulin<sup>35</sup>. Unexpectedly, the T2D risk allele showed increased transcriptional activity, suggesting either that the variant does not act in isolation or that *NKX6-1* is not the target gene.

At one of the novel T2D-associated loci near *SIX3*, the risk allele of East Asian lead variant rs12712928-C ( $\text{RAF} = 40.2\%$ ,  $\text{OR} = 1.06$ , 95% CI 1.04–1.07,  $P = 1.8 \times 10^{-14}$ ) is common in populations of non-East Asian ancestries, ranging from 16.0% in European individuals to 26.4% in South Asian individuals; however, there was no evidence of association in the other ancestry groups (meta-analysis:  $\text{OR} = 0.98$ , 95% CI 0.96–0.99,  $P = 2.9 \times 10^{-3}$ ; Extended Data Fig. 9a, Supplementary Table 8). Within the East Asian meta-analysis, the direction of effect is consistent across East Asian countries (Extended Data Fig. 9b) and within the contributing cohorts (Extended Data Fig. 9c). The T2D risk allele rs12712928-C is associated with higher fasting glucose levels in East Asian populations, has the strongest association with lower expression of both *SIX3* and *SIX2* in pancreatic islets<sup>19</sup>, and has demonstrated allele-specific binding to the transcription factor GABPA and reduced transcriptional activity<sup>36</sup>. While rs12712928-C is present on only one common haplotype in most populations, it is present on an additional common haplotype (frequency 0.075) in East Asian individuals, suggesting that the effect size attributed to rs12712928 may be influenced by other nearby unknown variants.

## Characterization of T2D-associated loci

To identify potential candidate genes that might underlie the T2D-association signals identified in East Asian individuals, we further characterized 92 known and novel loci for which the lead variant at the primary East Asian association signal was located more than 500 kb from the lead variant of any European T2D association signal<sup>2</sup> (Supplementary Table 9). We characterized loci using prior trait associations, *cis*-regulatory effects on expression (colocalized eQTL), predicted effects on protein sequence, and a literature search (Supplementary Tables 10–13). On the basis of association results from cardiometabolic trait consortia<sup>37</sup>, Biobank Japan<sup>38</sup>, and the UK Biobank<sup>39</sup>, the lead T2D-associated variant at 18 of the 92 loci was associated ( $P < 5 \times 10^{-8}$ ) with at least one additional cardiometabolic trait, most frequently BMI or a fat mass trait (15 loci; Supplementary Tables 10, 12). At 12 of the examined loci, T2D signals were colocalized with *cis*-eQTLs for transcripts in subcutaneous adipose tissue ( $n=5$ ), skeletal muscle ( $n=3$ ), pancreas ( $n=2$ ), islets ( $n=3$ ), or blood ( $n=5$ ; Supplementary Tables 11, 12), generating hypotheses of target genes and directions of effect; further examination of these candidate genes is warranted. At 19 loci, the lead T2D-associated variant or a proxy (East Asian  $r^2 > 0.80$ ) alters the protein sequence (Supplementary Table 12). These variants affect mesenchymal stem cell differentiation and adipogenesis (*GIT2*, *STEAP2* and *JMJD1C*), muscle stem cell biology (*CALCR*), glucose metabolism (*PGM1* and *SCTR*), and insulin secretion (*FGFR4*; Supplementary Table 13). At *SCTR*, which encodes the G-protein coupled secretin receptor, the lead variant encodes Ala122Pro, located in the hormone receptor domain. While mechanistic inference is required, these potential molecular mechanisms suggest new T2D susceptibility genes primarily detected by analyses in East Asian individuals.

T2D loci were also identified at clusters of noncoding RNAs with roles in islet  $\beta$  cell function. One locus includes a set of microRNAs specifically expressed in islet  $\beta$  cells, the maternally expressed noncoding RNA *MEG3*, and the paternally expressed gene *DLK1*. Targets of these microRNAs increase  $\beta$  cell apoptosis<sup>40</sup>, and reduced *Meg3* expression impairs insulin secretion<sup>41</sup>. *DLK1* inhibits adipocyte differentiation, thereby protecting against obesity<sup>3</sup>, and promotes pancreatic ductal cell differentiation into  $\beta$

cells, increasing insulin secretion<sup>42,43</sup>. Other variants near *MEG3* have been associated with type 1 diabetes<sup>44</sup> (EAS and EUR LD<sup>2</sup>=0 with EAS lead variant). The other noncoding RNA locus is the *MIR17HG* cluster of miRNAs, which regulate glucose-stimulated insulin secretion and pancreatic β cell proliferation stress<sup>45</sup>; one of these microRNAs, miR-19a, affects hepatic gluconeogenesis<sup>46</sup>. Yet another T2D locus is located near *TRAFF3*, which is a direct target of the *MIR17HG* microRNA cluster and promotes hyperglycaemia by increasing hepatic glucose production<sup>47,48</sup>. The T2D association results suggest that these noncoding RNAs influence disease susceptibility.

## Discussion

Our T2D GWAS meta-analyses of East Asian individuals have identified 61 novel loci and provided additional insight into the biological basis of T2D. The results emphasize substantial shared T2D susceptibility with European individuals, as shown by the strong correlation of effect sizes among T2D-associated genetic variants with common allele frequencies in populations with either East Asian or European ancestry. Compared to a recent T2D study in individuals of European ancestry<sup>1</sup>, we observed less attenuation of effects on T2D in analyses adjusted for BMI. Loci with a greater effect on T2D after adjusting for BMI include loci with lipodystrophy-like traits that have been identified only in East Asian individuals, adding support to the observation<sup>49,50</sup> that factors beyond overall BMI, such as visceral adiposity or lipodystrophy, may also contribute to T2D in East Asian individuals. The results also reveal novel associations in East Asian individuals that were identified because they have higher allele frequencies in East Asian populations, exhibit larger effect sizes, and/or are influenced by other environmental or behavioural factors such as alcohol consumption.

The identified loci point to multiple plausible molecular mechanisms and many new candidate genes that link T2D susceptibility to diverse biological processes. Following the annotation of loci identified in the East Asian meta-analysis, we speculate that insulin resistance has a substantial role in T2D pathogenesis among East Asian individuals, through skeletal muscle, adipose, and liver development and function. We also provide evidence that multiple distinct association signals in the same region do not necessarily act through the same gene. Conditionally distinct association signals in close proximity can affect different genes that may act in different tissues by different mechanisms, emphasizing the value of identifying functional variants that enable variant-to-gene links to be examined directly. Our results provide a foundation for future biological research into the pathogenesis of T2D and offer potential targets for interventions designed to modify disease risk.

## 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/s41586-020-2263-3>.

- 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).
- Suzuki, K. et al. Identification of 28 new susceptibility loci for type 2 diabetes in the Japanese population. *Nat. Genet.* **51**, 379–386 (2019).
- Moon, Y. S. et al. Mice lacking paternally expressed Pref-1/Dlk1 display growth retardation and accelerated adiposity. *Mol. Cell. Biol.* **22**, 5585–5592 (2002).
- van de Bunt, M. et al. The miRNA profile of human pancreatic islets and beta-cells and relationship to type 2 diabetes pathogenesis. *PLoS One* **8**, e55272 (2013).
- Scott, L. J. et al. The genetic regulatory signature of type 2 diabetes in human skeletal muscle. *Nat. Commun.* **7**, 11764 (2016).
- Civelek, M. et al. Genetic regulation of adipose gene expression and cardio-metabolic traits. *Am. J. Hum. Genet.* **100**, 428–443 (2017).
- Stumvoll, M., Goldstein, B. J. & van Haften, T. W. Type 2 diabetes: principles of pathogenesis and therapy. *Lancet* **365**, 1333–1346 (2005).
- Cho, Y. S. et al. Meta-analysis of genome-wide association studies identifies eight new loci for type 2 diabetes in east Asians. *Nat. Genet.* **44**, 67–72 (2011).
- Huxley, R. et al. Ethnic comparisons of the cross-sectional relationships between measures of body size with diabetes and hypertension. *Obes. Rev.* **9** (Suppl. 1), 53–61 (2008).
- Lassiter, D. G., Sjögren, R. J. O., Gabriel, B. M., Krook, A. & Zierath, J. R. AMPK activation negatively regulates GDAPI, which influences metabolic processes and circadian gene expression in skeletal muscle. *Mol. Metab.* **16**, 12–23 (2018).
- Hoang, C. Q. et al. Transcriptional maintenance of pancreatic acinar identity, differentiation, and homeostasis by PTF1A. *Mol. Cell. Biol.* **36**, 3033–3047 (2016).
- 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).
- Fuchsberger, C. et al. The genetic architecture of type 2 diabetes. *Nature* **536**, 41–47 (2016).
- Kwak, S. H. et al. Nonsynonymous variants in *PAX4* and *GLP1R* are associated with type 2 diabetes in an East Asian population. *Diabetes* **67**, 1892–1902 (2018).
- Klok, M. D., Jakobsdottir, S. & Drent, M. L. The role of leptin and ghrelin in the regulation of food intake and body weight in humans: a review. *Obes. Rev.* **8**, 21–34 (2007).
- Rasmussen-Torvik, L. J. et al. Associations of body mass index and insulin resistance with leptin, adiponectin, and the leptin-to-adiponectin ratio across ethnic groups: the Multi-Ethnic Study of Atherosclerosis (MESA). *Ann. Epidemiol.* **22**, 705–709 (2012).
- Imamura, M. et al. Genome-wide association studies in the Japanese population identify seven novel loci for type 2 diabetes. *Nat. Commun.* **7**, 10531 (2016).
- 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).
- Varshney, A. et al. Genetic regulatory signatures underlying islet gene expression and type 2 diabetes. *Proc. Natl. Acad. Sci. USA* **114**, 2301–2306 (2017).
- Thurner, M. et al. Integration of human pancreatic islet genomic data refines regulatory mechanisms at type 2 diabetes susceptibility loci. *eLife* **7**, e31977 (2018).
- Henselheit, K. D. et al. NKX6 transcription factor activity is required for alpha- and beta-cell development in the pancreas. *Development* **132**, 3139–3149 (2005).
- Yan, R. et al. A novel type 2 diabetes risk allele increases the promoter activity of the muscle-specific small ankyrin 1 gene. *Sci. Rep.* **6**, 25105 (2016).
- Wen, W. et al. Meta-analysis of genome-wide association studies in East Asian-ancestry populations identifies four new loci for body mass index. *Hum. Mol. Genet.* **23**, 5492–5504 (2014).
- Akiyama, M. et al. Genome-wide association study identifies 112 new loci for body mass index in the Japanese population. *Nat. Genet.* **49**, 1458–1467 (2017).
- Prokopenko, I. et al. A central role for GRB10 in regulation of islet function in man. *PLoS Genet.* **10**, e1004235 (2014).
- Hartiala, J. A. et al. Genome-wide association study and targeted metabolomics identifies sex-specific association of CPS1 with coronary artery disease. *Nat. Commun.* **7**, 10558 (2016).
- Okada, Y. et al. Deep whole-genome sequencing reveals recent selection signatures linked to evolution and disease risk of Japanese. *Nat. Commun.* **9**, 1631 (2018).
- Xu, F. et al. ALDH2 genetic polymorphism and the risk of type II diabetes mellitus in CAD patients. *Hypertens. Res.* **33**, 49–55 (2010).
- Kato, N. et al. Meta-analysis of genome-wide association studies identifies common variants associated with blood pressure variation in east Asians. *Nat. Genet.* **43**, 531–538 (2011).
- Takeuchi, F. et al. Confirmation of ALDH2 as a major locus of drinking behavior and of its variants regulating multiple metabolic phenotypes in a Japanese population. *Circ. J.* **75**, 911–918 (2011).
- Schrieks, I. C., Heil, A. L., Hendriks, H. F., Mukama, K. J. & Beulens, J. W. The effect of alcohol consumption on insulin sensitivity and glycemic status: a systematic review and meta-analysis of intervention studies. *Diabetes Care* **38**, 723–732 (2015).
- Lek, M. et al. Analysis of protein-coding genetic variation in 60,706 humans. *Nature* **536**, 285–291 (2016).
- Puig, M. et al. Functional impact and evolution of a novel human polymorphic inversion that disrupts a gene and creates a fusion transcript. *PLoS Genet.* **11**, e1005495 (2015).
- Iype, T. et al. The transcriptional repressor Nkx6.1 also functions as a deoxyribonucleic acid context-dependent transcriptional activator during pancreatic beta-cell differentiation: evidence for feedback activation of the *nkx6.1* gene by Nkx6.1. *Mol. Endocrinol.* **18**, 1363–1375 (2004).
- Taylor, B. L., Liu, F. F. & Sander, M. Nkx6.1 is essential for maintaining the functional state of pancreatic beta cells. *Cell Rep.* **4**, 1262–1275 (2013).
- 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).
- Accelerating Medicines Partnership. Type 2 Diabetes Knowledge Portal <http://www.type2diabetesgenetics.org/home/portalHome> (2019).
- Kanai, M. et al. Genetic analysis of quantitative traits in the Japanese population links cell types to complex human diseases. *Nat. Genet.* **50**, 390–400 (2018).
- Brycroft, C. et al. The UK Biobank resource with deep phenotyping and genomic data. *Nature* **562**, 203–209 (2018).
- Kameswaran, V. et al. Epigenetic regulation of the DLK1-MEG3 microRNA cluster in human type 2 diabetic islets. *Cell Metab.* **19**, 135–145 (2014).
- You, L. et al. Downregulation of long noncoding RNA Meg3 affects insulin synthesis and secretion in mouse pancreatic beta cells. *J. Cell. Physiol.* **231**, 852–862 (2016).
- Wang, Y. et al. Overexpression of Pref-1 in pancreatic islet β-cells in mice causes hyperinsulinemia with increased islet mass and insulin secretion. *Biochem. Biophys. Res. Commun.* **461**, 630–635 (2015).
- Rhee, M. et al. Preadipocyte factor 1 induces pancreatic ductal cell differentiation into insulin-producing cells. *Sci. Rep.* **6**, 23960 (2016).
- Onengut-Gumuscu, S. et al. Fine mapping of type 1 diabetes susceptibility loci and evidence for colocalization of causal variants with lymphoid gene enhancers. *Nat. Genet.* **47**, 381–386 (2015).
- Chen, Y. et al. MicroRNA-17-92 cluster regulates pancreatic beta-cell proliferation and adaptation. *Mol. Cell. Endocrinol.* **437**, 213–223 (2016).
- Dou, L. et al. MiR-19a mediates gluconeogenesis by targeting PTEN in hepatocytes. *Mol. Med. Rep.* **17**, 3967–3971 (2018).
- Chen, Z. et al. Hepatocyte TRAF3 promotes insulin resistance and type 2 diabetes in mice with obesity. *Mol. Metab.* **4**, 951–960 (2015).

48. Liu, F., Cheng, L., Xu, J., Guo, F. & Chen, W. miR-17-92 functions as an oncogene and modulates NF- $\kappa$ B signaling by targeting TRAF3 in MGC-803 human gastric cancer cells. *Int. J. Oncol.* **53**, 2241–2257 (2018).
49. Ma, R. C. & Chan, J. C. Type 2 diabetes in East Asians: similarities and differences with populations in Europe and the United States. *Ann. NY Acad. Sci.* **1281**, 64–91 (2013).
50. Zhu, Y. et al. Racial/ethnic disparities in the prevalence of diabetes and prediabetes by BMI: patient outcomes research to advance learning (PORTAL) multisite cohort of adults in the U.S. *Diabetes Care* **42**, 2211–2219 (2019).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature Limited 2020

- Cassandra N. Spracklen<sup>1,2,89</sup>, Momoko Horikoshi<sup>3,89</sup>, Young Jin Kim<sup>4,89</sup>, Kuang Lin<sup>5,89</sup>, Fiona Bragg<sup>5</sup>, Sanghoon Moon<sup>4</sup>, Ken Suzuki<sup>3,6,7,8</sup>, Claudia H. T. Tam<sup>9,10</sup>, Yasuharu Tabara<sup>11</sup>, Soo-Heon Kwak<sup>12</sup>, Fumihiro Takeuchi<sup>13</sup>, Jirong Long<sup>14</sup>, Victor J. Y. Lim<sup>15</sup>, Jin-Fang Chai<sup>15</sup>, Chien-Hsiun Chen<sup>16</sup>, Masahiro Nakatohchi<sup>17</sup>, Jie Yao<sup>18,19</sup>, Hyek Sun Choi<sup>20</sup>, Apoorva K. Iyengar<sup>1</sup>, Hannah J. Perrin<sup>1</sup>, Sarah M. Brotman<sup>1</sup>, Martijn van de Bunt<sup>21,22</sup>, Anna L. Glyn<sup>21,22,23,87</sup>, Jennifer E. Below<sup>24,25</sup>, Michael Boehnke<sup>26</sup>, Donald W. Bowden<sup>27,28</sup>, John C. Chambers<sup>29,30,31,32,33</sup>, Anubha Mahajan<sup>21,22,88</sup>, Mark I. McCarthy<sup>21,22,23,88</sup>, Maggie C. Y. Ng<sup>24,27</sup>, Lauren E. Petty<sup>24,25</sup>, Weihua Zhang<sup>30,31</sup>, Andrew P. Morris<sup>22,34,35</sup>, Linda S. Adair<sup>36</sup>, Masato Akiyama<sup>6,37,38</sup>, Zheng Bian<sup>39</sup>, Juliana C. N. Chan<sup>9,10,40,41</sup>, Li-Ching Chang<sup>16</sup>, Miao-Li Chee<sup>42</sup>, Yii-Der Ida Chen<sup>18,19</sup>, Yuan-Tsung Chen<sup>16</sup>, Zhengming Chen<sup>5</sup>, Lee-Ming Chuang<sup>43,44</sup>, Shufa Du<sup>36</sup>, Penny Gordon-Larsen<sup>36</sup>, Myron Gross<sup>45</sup>, Xiuqing Guo<sup>18,19</sup>, Yu Guo<sup>3</sup>, Sohee Han<sup>4</sup>, Annie-Green Howard<sup>46</sup>, Wei Huang<sup>47</sup>, Yi-Jen Hung<sup>48,49</sup>, Mi Yeong Hwang<sup>4</sup>, Chii-Min Hwu<sup>50,51</sup>, Sahoko Ichihara<sup>52</sup>, Masato Isono<sup>13</sup>, Hye-Mi Jang<sup>4</sup>, Guozhi Jiang<sup>8,10</sup>, Jost B. Jonas<sup>53</sup>, Yoichiro Kamatani<sup>16,54</sup>, Tomohiro Katsuya<sup>55,56</sup>, Takahisa Kawaguchi<sup>11</sup>, Chea-Chuen Khor<sup>42,57,84</sup>, Katsuhiko Kohara<sup>58</sup>, Myung-Shik Lee<sup>59,60</sup>, Nanette R. Lee<sup>61</sup>, Liming Li<sup>62</sup>, Jianjun Liu<sup>57,63</sup>, Andrea O. Luk<sup>9,10</sup>, Jun Lv<sup>62</sup>, Yukinori Okada<sup>8,64</sup>, Mark A. Pereira<sup>65</sup>, Charumathi Sabanayagam<sup>42,66,67</sup>, Jinxiu Shi<sup>45</sup>, Dong Mun Shin<sup>4</sup>, Wing Yee So<sup>9,40</sup>, Atsushi Takahashi<sup>6,68</sup>, Brian Tomlinson<sup>9,69</sup>, Fuu-Jen Tsai<sup>70</sup>, Rob M. van Dam<sup>15,63</sup>, Yong-Bing Xiang<sup>71,72</sup>, Ken Yamamoto<sup>73</sup>, Toshimasa Yamauchi<sup>7</sup>, Kyunghyun Yoon<sup>74</sup>, Canqing Yu<sup>62</sup>, Jian-Min Yuan<sup>74,75</sup>, Liang Zhang<sup>42</sup>, Wei Zheng<sup>14</sup>, Michiya Igase<sup>76</sup>, Yoon Shin Cho<sup>20</sup>, Jerome I. Rotter<sup>16,19</sup>, Ya-Xing Wang<sup>77</sup>, Wayne H. H. Sheu<sup>49,51,78</sup>, Mitsuhiro Yokota<sup>79</sup>, Jer-Yuarn Wu<sup>16</sup>, Ching-Yu Cheng<sup>42,66,67</sup>, Tien-Yin Wong<sup>42,66,67</sup>, Xiao-Ou Shu<sup>14</sup>, Norihiro Kato<sup>13</sup>, Kyung-Soo Park<sup>12,60,81</sup>, E-Shyong Tai<sup>15,63,82</sup>, Fumihiro Matsuda<sup>11</sup>, Woon-Puay Koh<sup>15,63</sup>, Ronald C. W. Ma<sup>9,10,40,41</sup>, Shiro Maeda<sup>3,84,85</sup>, Iona Y. Millwood<sup>5,86</sup>, Juyoung Lee<sup>4</sup>, Takashi Kadowaki<sup>17</sup>, Robin G. Walters<sup>5,86</sup>, Bong-Jo Kim<sup>4</sup>, Karen L. Mohlke<sup>1</sup> & Xueling Sim<sup>15</sup>

<sup>1</sup>Department of Genetics, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA.

<sup>2</sup>Department of Biostatistics and Epidemiology, School of Public Health and Health Sciences, University of Massachusetts, Amherst, MA, USA. <sup>3</sup>Laboratory for Endocrinology, Metabolism and Kidney Diseases, RIKEN Centre for Integrative Medical Sciences, Yokohama, Japan.

<sup>4</sup>Division of Genome Research, Center for Genome Science, National Institute of Health, Chungcheongbuk-do, Republic of Korea. <sup>5</sup>Nuffield Department of Population Health, University of Oxford, Oxford, UK. <sup>6</sup>Laboratory for Statistical and Translational Genetics, RIKEN Centre for Integrative Medical Sciences, Yokohama, Japan. <sup>7</sup>Department of Diabetes and Metabolic Diseases, Graduate School of Medicine, The University of Tokyo, Tokyo, Japan.

<sup>8</sup>Department of Statistical Genetics, Osaka University Graduate School of Medicine, Osaka, Japan. <sup>9</sup>Department of Medicine and Therapeutics, The Chinese University of Hong Kong, Hong Kong, China. <sup>10</sup>Chinese University of Hong Kong-Shanghai Jiao Tong University Joint Research Centre in Diabetes Genomics and Precision Medicine, The Chinese University of Hong Kong, Hong Kong, China. <sup>11</sup>Center for Genomic Medicine, Kyoto University Graduate School of Medicine, Kyoto, Japan. <sup>12</sup>Department of Internal Medicine, Seoul National University Hospital, Seoul, South Korea. <sup>13</sup>Department of Gene Diagnostics and Therapeutics, Research Institute, National Center for Global Health and Medicine, Tokyo, Japan. <sup>14</sup>Division of Epidemiology, Department of Medicine, Vanderbilt University Medical Center, Nashville, TN, USA. <sup>15</sup>Saw Swee Hock School of Public Health, National University of Singapore and National University Health System, Singapore, Singapore. <sup>16</sup>Institute of Biomedical Sciences, Academia Sinica, Taipei, Taiwan. <sup>17</sup>Department of Integrated Health Sciences, Nagoya University Graduate School of Medicine, Nagoya, Japan. <sup>18</sup>The Institute for Translational Genomics and Population Sciences, The Lundquist Institute for Biomedical Innovation at Harbor-UCLA Medical Center, Torrance, CA, USA. <sup>19</sup>Department of Pediatrics, UCLA School of Medicine, Harbor-UCLA Medical Center, Torrance, CA, USA. <sup>20</sup>Biomedical Science, Hallym University, Chuncheon, South Korea. <sup>21</sup>Oxford Centre for Diabetes, Endocrinology and Metabolism, University of Oxford, Oxford, UK. <sup>22</sup>Wellcome Centre for Human Genetics, University of Oxford, Oxford, UK. <sup>23</sup>Oxford NIHR Biomedical Research Centre, Oxford University Hospitals NHS Foundation Trust, Churchill Hospital, Oxford, UK. <sup>24</sup>Vanderbilt Genetics Institute, Division of Genetic Medicine, Vanderbilt University Medical Center, Nashville, TN, USA. <sup>25</sup>Human Genetics Center, School of Public Health, The University of Texas Health Science Center at Houston, Houston, TX, USA. <sup>26</sup>Department of Biostatistics and Center for Statistical Genetics, University of Michigan, Ann Arbor, MI, USA. <sup>27</sup>Center for Genomics and Personalized Medicine Research, Center for Diabetes Research, Wake Forest School of Medicine, Winston-Salem, NC, USA. <sup>28</sup>Department of Biochemistry, Wake Forest School of Medicine, Winston-Salem, NC, USA. <sup>29</sup>Lee Kong Chian School of Medicine, Nanyang

- Technological University, Singapore, Singapore. <sup>30</sup>Department of Epidemiology and Biostatistics, Imperial College London, London, UK. <sup>31</sup>Department of Cardiology, Ealing Hospital, London North West Healthcare NHS Trust, London, UK. <sup>32</sup>Imperial College Healthcare NHS Trust, Imperial College London, London, UK. <sup>33</sup>MRC-PHE Centre for Environment and Health, Imperial College London, London, UK. <sup>34</sup>Department of Biostatistics, University of Liverpool, Liverpool, UK. <sup>35</sup>School of Biological Sciences, University of Manchester, Manchester, UK. <sup>36</sup>Department of Nutrition, Gillings School of Global Public Health, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA. <sup>37</sup>Laboratory for Statistical Analysis, RIKEN Centre for Integrative Medical Sciences, Yokohama, Japan. <sup>38</sup>Department of Ophthalmology, Graduate School of Medical Sciences, Kyushu University, Fukuoka, Japan. <sup>39</sup>Chinese Academy of Medical Sciences, Beijing, China. <sup>40</sup>Hong Kong Institute of Diabetes and Obesity, The Chinese University of Hong Kong, Hong Kong, China. <sup>41</sup>Li Ka Shing Institute of Health Sciences, The Chinese University of Hong Kong, Hong Kong, China. <sup>42</sup>Singapore Eye Research Institute, Singapore National Eye Centre, Singapore, Singapore. <sup>43</sup>Division of Endocrinology & Metabolism, Department of Internal Medicine, National Taiwan University Hospital, Taipei, Taiwan. <sup>44</sup>Institute of Preventive Medicine, School of Public Health, National Taiwan University, Taipei, Taiwan. <sup>45</sup>Department of Laboratory Medicine and Pathology, University of Minnesota, Minneapolis, MN, USA. <sup>46</sup>Department of Biostatistics, Carolina Population Center, Gillings School of Global Public Health, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA. <sup>47</sup>Department of Genetics, Shanghai-MOST Key Laboratory of Health and Disease Genomics, Chinese National Human Genome Center at Shanghai, Shanghai, China. <sup>48</sup>Division of Endocrine and Metabolism, Tri-Service General Hospital Songshan Branch, Taipei, Taiwan. <sup>49</sup>School of Medicine, National Defense Medical Center, Taipei, Taiwan. <sup>50</sup>Section of Endocrinology and Metabolism, Department of Medicine, Taipei Veterans General Hospital, Taipei, Taiwan. <sup>51</sup>School of Medicine, National Yang-Ming University, Taipei, Taiwan. <sup>52</sup>Department of Environmental and Preventive Medicine, Jichi Medical University School of Medicine, Shimotsuke, Japan. <sup>53</sup>Department of Ophthalmology, Medical Faculty Mannheim of the University of Heidelberg, Mannheim, Germany. <sup>54</sup>Laboratory of Complex Trait Genomics, Department of Computational Biology and Medical Sciences, Graduate School of Frontier Sciences, The University of Tokyo, Tokyo, Japan. <sup>55</sup>Department of Clinical Gene Therapy, Osaka University Graduate School of Medicine, Osaka, Japan. <sup>56</sup>Department of Geriatric and General Medicine, Graduate School of Medicine, Osaka University, Osaka, Japan. <sup>57</sup>Genome Institute of Singapore, Agency for Science, Technology and Research, Singapore, Singapore. <sup>58</sup>Department of Regional Resource Management, Ehime University Faculty of Collaborative Regional Innovation, Ehime, Japan. <sup>59</sup>Severance Biomedical Science Institute and Department of Internal Medicine, Yonsei University College of Medicine, Seoul, South Korea. <sup>60</sup>Department of Medicine, Samsung Medical Center, Sungkyunkwan University School of Medicine, Seoul, South Korea. <sup>61</sup>Department of Anthropology, Sociology and History, University of San Carlos, Cebu City, Philippines. <sup>62</sup>Department of Epidemiology and Biostatistics, Peking University Health Science Centre, Peking University, Beijing, China. <sup>63</sup>Department of Medicine, Yong Loo Lin School of Medicine, National University of Singapore and National University Health System, Singapore, Singapore. <sup>64</sup>Laboratory of Statistical Immunology, Immunology Frontier Research Center (WPI-IFReC), Osaka University, Osaka, Japan. <sup>65</sup>Division of Epidemiology and Community Health, School of Public Health, University of Minnesota, Minneapolis, MN, USA. <sup>66</sup>Ophthalmology & Visual Sciences Academic Clinical Program (Eye ACP), Duke-NUS Medical School, Singapore, Singapore. <sup>67</sup>Department of Ophthalmology, Yong Loo Lin School of Medicine, National University of Singapore and National University Health System, Singapore, Singapore. <sup>68</sup>Department of Genomic Medicine, National Cerebral and Cardiovascular Center, Osaka, Japan. <sup>69</sup>Faculty of Medicine, Macau University of Science and Technology, Macau, China. <sup>70</sup>Department of Medical Genetics and Medical Research, China Medical University Hospital, Taichung, Taiwan. <sup>71</sup>State Key Laboratory of Oncogene and Related Genes, Shanghai Cancer Institute, Renji Hospital, Shanghai Jiaotong University School of Medicine, Shanghai, China. <sup>72</sup>Department of Epidemiology, Shanghai Cancer Institute, Renji Hospital, Shanghai Jiaotong University School of Medicine, Shanghai, China. <sup>73</sup>Department of Medical Biochemistry, Kurume University School of Medicine, Kurume, Japan. <sup>74</sup>Division of Cancer Control and Population Sciences, UPMC Hillman Cancer Center, University of Pittsburgh, Pittsburgh, PA, USA. <sup>75</sup>Department of Epidemiology, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, PA, USA. <sup>76</sup>Department of Anti-aging Medicine, Ehime University Graduate School of Medicine, Ehime, Japan. <sup>77</sup>Beijing Institute of Ophthalmology, Ophthalmology and Visual Sciences Key Laboratory, Beijing Tongren Hospital, Capital Medical University, Beijing, China. <sup>78</sup>Division of Endocrinology and Metabolism, Department of Medicine, Taichung Veterans General Hospital, Taichung, Taiwan. <sup>79</sup>Kurume University School of Medicine, Kurume, Japan. <sup>80</sup>Department of Internal Medicine, Seoul National University College of Medicine, Seoul, South Korea. <sup>81</sup>Department of Molecular Medicine and Biopharmaceutical Sciences, Graduate School of Convergence Science and Technology, Seoul National University, Seoul, South Korea. <sup>82</sup>Duke-NUS Medical School, Singapore, Singapore. <sup>83</sup>Health Services and Systems Research, Duke-NUS Medical School, Singapore, Singapore. <sup>84</sup>Department of Advanced Genomic and Laboratory Medicine, Graduate School of Medicine, University of the Ryukyus, Okinawa, Japan. <sup>85</sup>Division of Clinical Laboratory and Blood Transfusion, University of the Ryukyus Hospital, Okinawa, Japan. <sup>86</sup>Medical Research Council Population Health Research Unit, University of Oxford, Oxford, UK. <sup>87</sup>Present address: Stanford University, Stanford, CA, USA. <sup>88</sup>Present address: Genentech, South San Francisco, CA, USA. <sup>89</sup>These authors contributed equally: Cassandra N. Spracklen, Momoko Horikoshi, Young Jin Kim, Kuang Lin. <sup>✉</sup>e-mail: kadowaki-3im@u-tokyo.ac.jp; robin.walters@ndph.ox.ac.uk; kbj6181@korea.kr; mohlke@med.unc.edu; ephsx@nus.edu.sg

# Article

## Methods

### Ethics statement

All human research was approved by the relevant institutional review boards for each study at their respective sites (Supplementary Table 1) and was conducted according to the Declaration of Helsinki. All participants provided written informed consent.

### Study cohorts and quality control

The East Asian type 2 diabetes (T2D) meta-analyses were performed with studies participating in the Asian Genetic Epidemiology Network (AGEN), a consortium of genetic epidemiology studies of T2D and related traits conducted in individuals of East Asian ancestry, and the Diabetes Meta-analysis of Trans-ethnic Association Studies (DIAMANTE), a consortium examining the genetic contribution to T2D across populations of diverse ancestry including African–American, East Asian, European, Hispanic, and South Asian. The East Asian meta-analysis included 77,418 T2D cases and 356,122 controls from 23 GWAS, including three biobanks, CKB, KBA<sup>51,52</sup>, and BBJ<sup>2</sup> (effective sample size ( $N_{\text{eff}}$ ) = 211,793; Extended Data Fig. 1). A subset of studies for which BMI measurement was available was analysed with and without BMI adjustment in sex-combined and sex-specific models (54,481 cases, 224,231 controls;  $N_{\text{eff}} = 135,780$ ). For each study, T2D case control ascertainment is described in Supplementary Table 1 and summary statistics are provided in Supplementary Table 2. As T2D case definitions across cohorts differ, it is possible that cases of type 1 diabetes and maturity onset diabetes of the young (MODY) are included in these meta-analyses. Included studies were genotyped on either commercially available or customized Affymetrix or Illumina genome-wide genotyping arrays. Array quality control criteria implemented within each study, including variant call rate and Hardy–Weinberg equilibrium, are summarized in Supplementary Table 3. To harmonize the study-level genotype scaffold for imputation to 1000 Genomes (1000G) reference panels, each study adopted a uniform protocol for pre-imputation quality checks. Each study applied the protocol to exclude variants with: (i) mismatched chromosomal positions or alleles not present in the reference panel; (ii) ambiguous alleles (AT/CG) with minor allele frequency (MAF) >40% in the reference panel; or (iii) absolute allele frequency differences >20% compared to East Asian-specific allele frequencies. The genotype scaffold for each study was then imputed to the 1000G Phase 1 or 3 reference panel<sup>53</sup> using minimac3<sup>54</sup> or IMPUTEv2<sup>55</sup>. In BMI-unadjusted analyses, all studies were imputed to 1000G Phase 3. In BMI-adjusted and sex-stratified analyses, all studies were imputed to 1000G Phase 3 except for a subset of Biobank Japan<sup>17</sup>, which was imputed to the 1000G Phase 1 reference panel. No statistical methods were used to predetermine sample size.

### Study-level association analyses

Within each study, all variants were tested for association with T2D assuming an additive model of inheritance within a regression framework, including age, sex, and other study-specific covariates (Supplementary Table 3). To account for population structure and relatedness, association analyses were performed using either FIRTH<sup>56</sup> or mach2dat with additional adjustment for principal components in unrelated individuals or a linear mixed model with kinship matrix implemented in BOLT-LMM<sup>57</sup>. In studies analysed with the linear mixed model, allelic effects and standard errors were converted to the log-odds scale that accounts for case–control imbalance<sup>58</sup>. Within each study, variants were removed if (i) imputation quality score was poor (minimac3  $r^2 < 0.30$ ; IMPUTE2 info score <0.40); (ii) combined case–control minor allele count <5; or (iii) standard error of the log-OR >10. For a subset of the studies, BMI was added as an additional covariate, and association analyses were also performed separately in males and females. For each study and model, association statistics were corrected with genomic

control inflation factor<sup>59</sup> calculated from common variants (MAF  $\geq 5\%$ ) (Supplementary Table 3). For BBJ, we applied the genomic control inflation factor 1.21 as reported<sup>2</sup>.

### Sex-combined meta-analysis

We combined study-level association statistics using fixed-effect meta-analysis with inverse-variance weighting of log-ORs implemented in METAL<sup>60</sup>. Variants with allele frequency differences >20% between 1000G Phase 1 and 3 panels were excluded from the meta-analysis. To assess excess inflation arising from cryptic relatedness and population structure, we applied LD score regression to the meta-analysis summary statistics to estimate residual inflation of summary statistics, using a set of 1,889 unrelated Chinese individuals from the Singapore Chinese Eye Study<sup>61</sup>. The LD score regression intercepts were 0.993 for BMI-unadjusted, and 1.0163 for BMI-adjusted models. As the LD score regression intercepts indicated absence of excess inflation, the meta-analysis results were corrected for inflation using these LD score regression intercepts. For subsequent analyses, we considered only variants that were present in at least 50% of the effective sample size  $N_{\text{eff}}$  (computed as  $4/(1/N_{\text{cases}} + 1/N_{\text{controls}})$ )<sup>60</sup>. Heterogeneity in allelic effect sizes between studies were assessed using fixed-effect inverse-variance weighted meta-analysis  $P_{\text{het}}$ . We further compared the genetic effects from BMI-unadjusted and BMI-adjusted models using fixed-effect inverse-variance weighted meta-analysis  $P_{\text{het}}$ . Loci were defined as novel if the lead variant was: (i) at least 500 kb away and confirmed by GCTA to be conditionally independent from previously reported T2D-associated variants in any ancestry; and (ii) assessed using LocusZoom plots and detailed literature review to be away from known loci with extended LD. Lead variants that mapped to loci already associated with other glycaemic traits were still considered novel for the association with T2D.

### BMI adjustment analyses and effect size comparison

For the subset of studies with both BMI-unadjusted and BMI-adjusted models, we compared the effect sizes and heterogeneity of the lead variants using the standardized mean difference to account for the correlation between the two models<sup>1</sup>. We calculated the Pearson correlation coefficient between effect sizes from the BMI-unadjusted and BMI-adjusted models for all 13.2M variants genome-wide ( $r = 0.98$ ) and for the lead variants at 189 T2D-associated loci ( $r = 0.99$ ).

### Sex-differentiated meta-analysis

The meta-analyses described above were repeated for males and females separately. The male-specific meta-analyses included up to 28,027 cases and 89,312 controls ( $N_{\text{eff}} = 65,660$ ) and the female-specific analyses included up to 27,370 cases and 135,055 controls ( $N_{\text{eff}} = 70,332$ ). LD score regression intercepts were 1.0044 for BMI-unadjusted and 1.0045 for BMI-adjusted models in males and 1.0050 for BMI-unadjusted and 1.0187 for BMI-adjusted models in females. We further performed a test for heterogeneity in allelic effects between males and females as implemented in GWAMA<sup>62,63</sup>.

### Detection of distinct association signals

To detect multiple distinct association signals at each associated locus, we combined overlapping loci when the distance between any pair of lead variants was <1 Mb. We then performed approximate conditional analyses using GCTA<sup>12</sup> with genome-wide meta-analysis summary statistics and LD estimated from 78,000 samples from the Korea Biobank Array<sup>52</sup>. We note the limitations in using a single population reference panel for LD estimation for a meta-analysis of diverse East Asian populations. We present all distinct signals at conditional threshold of  $P < 1 \times 10^{-5}$ , but we suggest that readers exhibit caution and limit inferences from these analyses to signals that show the strongest evidence of association.

## Comparing loci effects between East Asian and European populations

We compared the per-allele effect sizes of lead variants identified from the East Asian BMI-unadjusted sex-combined meta-analysis (183 lead variants) and European BMI-unadjusted sex-combined meta-analysis<sup>1</sup> (231 lead variants; Supplementary Table 7). Across the 414 associated variants from the two ancestries, 12 lead variants overlapped, resulting in 402 unique variants. As the variants in the European analysis were imputed using the Haplotype Reference Consortium reference panel and did not include indel variants, a variant in strong LD (East Asian  $r^2 > 0.90$ ) with the lead East Asian variant was used when the lead variant was an indel, when possible. If the lead East Asian variant or a variant in strong LD (East Asian  $r^2 > 0.90$ ) was not available in the European data from DIAMANTE, we used results from a previous European type 2 diabetes meta-analysis<sup>64</sup>. The effect size comparison plot was restricted to 332 variants for which data were available for both ancestries (Fig. 2a). For loci that were significant in both the East Asian and European meta-analyses, if the lead variants were different, both lead variants were plotted (see Supplementary Table 7). Effect size plots were further restricted to: (i) 278 lead variants with MAF  $\geq 5\%$  in both East Asian and European meta-analyses (Extended Data Fig. 7); (ii) 203 lead variants that were significant in the East Asian meta-analysis (Extended Data Fig. 8a); and (iii) 234 lead variants that were significant in the European meta-analysis (Extended Data Fig. 8b). Differences in effect sizes between the two populations could be due to differences in imputation quality with different reference panels.

## Associations with other metabolic traits and outcomes

We examined publicly available GWAS summary statistics (mostly available through the Type 2 Diabetes Knowledge Portal<sup>37</sup>) to explore associations of the lead variants at the 92 loci for which there are no genome-wide significant European variants within 500 kb (listed in Supplementary Table 9). Association statistics from the following consortia were available for query on the portal (last accessed August 28, 2019): coronary artery disease from CARDIoGRAM<sup>65</sup>, BMI and waist-hip-ratio from GIANT<sup>66,67</sup>, lipid traits from GLGC<sup>68</sup>, and glycaemic traits from MAGIC<sup>69–73</sup>. Additionally, we used available data from the AGEN East Asian meta-analyses for lipids<sup>74</sup>, along with the phenotypic data from the UK Biobank<sup>75</sup>, BioBank Japan<sup>24,38</sup>, and blood pressure data from ICBP<sup>76</sup>. For this analysis, we looked up the effect sizes and *P* values of the East Asian lead variants in the other datasets. If the variant–trait association reached at least nominal significance ( $P < 1 \times 10^{-3}$ ), we included the lookup results in Supplementary Table 10. When the lead East Asian variant was missing in the prior GWAS data, we reported it as NF (not found) in the Table.

## Colocalization with expression quantitative trait loci

We searched publicly available eQTL databases such as GTEx v7<sup>77</sup> and the Parker lab Islet Browser<sup>19</sup> to identify *cis*-eQTLs at the novel loci in adipose (subcutaneous and visceral), blood, pancreas, pancreatic islet, and skeletal muscle tissue. We also searched for *cis*-eQTLs in subcutaneous adipose tissue data from the METSIM study<sup>6</sup>, whole blood<sup>78</sup>, and peripheral blood (BioBankJapan; <http://jenger.riken.jp/en/result>). Colocalized eQTLs were identified if the lead expression level-associated variant and the GWAS lead variant were in high LD ( $r^2 > 0.80$ ) in European individuals to accommodate the predominantly European eQTL data. Reciprocal conditional analyses were also performed using the METSIM data to determine whether the GWAS lead variant and the lead expression-associated variant were part of the same eQTL signal.

## Literature review

We conducted a traditional literature review to identify candidate genes at each novel locus using NCBI Entrez Gene, PubMed and OMIM. We included gene symbols and the following keywords as search terms in PubMed: diabetes, glucose, insulin, islet, adipose, muscle, liver, obesity.

A gene was considered a potential candidate if an apparent link to T2D biology existed, based on prior studies of gene function.

## Functional annotation and experimentation at *NKX6-1*

We used ENCODE<sup>79</sup>, ChromHMM<sup>80</sup>, and Human Epigenome Atlas<sup>81</sup> data available through the UCSC Genome Browser to identify candidate variants at the association signal near *NKX6-1* that overlapped open-chromatin peaks, ChromHMM chromatin states, and chromatin-immunoprecipitation sequencing (ChIP-seq) peaks of histone modifications H4K4me1, H3K4me3, and H3K27ac, and transcription factors in the pancreas and pancreatic islets. MIN6 mouse insulinoma cells (from the ATCC; authenticated by genotyping and tested negative for mycoplasma contamination)<sup>82</sup> were cultured in DMEM (Sigma) supplemented with 10% FBS, 1 mM sodium pyruvate, and 0.1 mM beta-mercaptoethanol. The cell cultures were maintained at 37 °C with 5% CO<sub>2</sub>. To measure variant allelic differences in enhancer activity at the *NKX6-1* locus, we designed oligonucleotide primers (forward: CCCTAGTAATGCCCTTTGTT; reverse: TCAGCCTGAGAAGTCTG TGA) with KpnI and Xhol restriction sites, and amplified the 400-bp DNA region (GRCh37/hg19 -chr4: 85,339,430-85,339,829) around rs117624659. As previously described<sup>80</sup>, we ligated amplified DNA from individuals homozygous for each allele into the multiple cloning site of the pGL4.23 (Promega) minimal promoter luciferase reporter vector in both the forward and reverse orientations with respect to the genome. Clones were isolated and sequenced for genotype and fidelity. MIN6 cells ( $2.1 \times 10^5$  per well) were seeded and grown to 90% confluence in 24-well plates. We co-transfected five independent luciferase constructs and *Renilla* control reporter vector (phRL-TK, Promega) using Lipofectamine 2000 (Life Technologies) and incubated the cells. Forty-eight hours after transfection, the cells were lysed with Passive Lysis Buffer (Promega). Luciferase activity was measured using the Dual-luciferase Reporter Assay System (Promega) according to the manufacturer's instructions and as previously described<sup>83</sup>.

## Reporting summary

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

## Data availability

Summary-level statistics are publicly available on the AGEN consortium website (<https://blog.nus.edu.sg/agen/summary-statistics/t2d-2020>), and the Accelerating Medicines Partnership T2D portal ([http://www.kp4cd.org/dataset\\_downloads/t2d](http://www.kp4cd.org/dataset_downloads/t2d)). A complete list of web resources is available in the Supplementary Information.

51. Kim, Y., Han, B. G. & the KoGES Group. Cohort profile: The Korean Genome and Epidemiology Study (KoGES) Consortium. *Int. J. Epidemiol.* **46**, e20 (2017).
52. Moon, S. et al. The Korea Biobank Array: design and identification of coding variants associated with blood biochemical traits. *Sci. Rep.* **9**, 1382 (2019).
53. Auton, A. et al. A global reference for human genetic variation. *Nature* **526**, 68–74 (2015).
54. Das, S. et al. Next-generation genotype imputation service and methods. *Nat. Genet.* **48**, 1284–1287 (2016).
55. Howie, B., Marchini, J. & Stephens, M. Genotype imputation with thousands of genomes. *G3 (Bethesda)* **1**, 457–470 (2011).
56. Ma, C., Blackwell, T., Boehnke, M. & Scott, L. J. Recommended joint and meta-analysis strategies for case-control association testing of single low-count variants. *Genet. Epidemiol.* **37**, 539–550 (2013).
57. Loh, P. R. et al. Efficient Bayesian mixed-model analysis increases association power in large cohorts. *Nat. Genet.* **47**, 284–290 (2015).
58. Cook, J. P., Mahajan, A. & Morris, A. P. Guidance for the utility of linear models in meta-analysis of genetic association studies of binary phenotypes. *Eur. J. Hum. Genet.* **25**, 240–245 (2017).
59. Devlin, B. & Roeder, K. Genomic control for association studies. *Biometrics* **55**, 997–1004 (1999).
60. Willer, C. J., Li, Y. & Abecasis, G. R. METAL: fast and efficient meta-analysis of genome-wide association scans. *Bioinformatics* **26**, 2190–2191 (2010).
61. Bulik-Sullivan, B. K. et al. LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat. Genet.* **47**, 291–295 (2015).
62. Magi, R., Lindgren, C. M. & Morris, A. P. Meta-analysis of sex-specific genome-wide association studies. *Genet. Epidemiol.* **34**, 846–853 (2010).

63. Mägi, R. & Morris, A. P. GWAMA: software for genome-wide association meta-analysis. *BMC Bioinformatics* **11**, 288 (2010).
64. Scott, R. A. et al. An expanded genome-wide association study of type 2 diabetes in Europeans. *Diabetes* **66**, 2888–2902 (2017).
65. Schunkert, H. et al. Large-scale association analysis identifies 13 new susceptibility loci for coronary artery disease. *Nat. Genet.* **43**, 333–338 (2011).
66. Shungin, D. et al. New genetic loci link adipose and insulin biology to body fat distribution. *Nature* **518**, 187–196 (2015).
67. Yengo, L. et al. Meta-analysis of genome-wide association studies for height and body mass index in ~700000 individuals of European ancestry. *Hum. Mol. Genet.* **27**, 3641–3649 (2018).
68. Willer, C. J. et al. Discovery and refinement of loci associated with lipid levels. *Nat. Genet.* **45**, 1274–1283 (2013).
69. 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).
70. Saxena, R. et al. Genetic variation in GIPR influences the glucose and insulin responses to an oral glucose challenge. *Nat. Genet.* **42**, 142–148 (2010).
71. Strawbridge, R. J. et al. Genome-wide association identifies nine common variants associated with fasting proinsulin levels and provides new insights into the pathophysiology of type 2 diabetes. *Diabetes* **60**, 2624–2634 (2011).
72. 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).
73. 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).
74. Spracklen, C. N. et al. Association analyses of East Asian individuals and trans-ancestry analyses with European individuals reveal new loci associated with cholesterol and triglyceride levels. *Hum. Mol. Genet.* **27**, 1122 (2018).
75. Sudlow, C. et al. UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Med.* **12**, e1001779 (2015).
76. Ehret, G. B. et al. Genetic variants in novel pathways influence blood pressure and cardiovascular disease risk. *Nature* **478**, 103–109 (2011).
77. Gamazon, E. R. et al. Using an atlas of gene regulation across 44 human tissues to inform complex disease- and trait-associated variation. *Nat. Genet.* **50**, 956–967 (2018).
78. Võsa, U. et al. Unraveling the polygenic architecture of complex traits using blood eQTL metaanalysis. Preprint at bioRxiv <https://doi.org/10.1101/447367> (2018).
79. ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74 (2012).
80. Ezzat, S. et al. The cancer-associated FGFR4-G388R polymorphism enhances pancreatic insulin secretion and modifies the risk of diabetes. *Cell Metab.* **17**, 929–940 (2013).
81. Kundaje, A. et al. Integrative analysis of 111 reference human epigenomes. *Nature* **518**, 317–330 (2015).
82. Miyazaki, J. et al. Establishment of a pancreatic beta cell line that retains glucose-inducible insulin secretion: special reference to expression of glucose transporter isoforms. *Endocrinology* **127**, 126–132 (1990).
83. Fogarty, M. P., Cannon, M. E., Vadlamudi, S., Gaulton, K. J. & Mohlke, K. L. Identification of a regulatory variant that binds FOXA1 and FOXA2 at the CDC123/CAMK1D type 2 diabetes GWAS locus. *PLoS Genet.* **10**, e1004633 (2014).

**Acknowledgements** This work was supported by subawards (to X.S. and Y.S.C.) from NIDDK U01DK105554 (J. C. Florez). The authors thank all investigators, staff members and study participants for their contributions to all participating studies, including but not limited to: CAGE (T. Ogihara, Y. Yamori, A. Fujioka, C. Makibayashi, S. Katsuya, K. Sugimoto, K. Kamide and R. Morishita); SBCS (R. Courtney, H. Cai, B. Zhang and J. He); and SMC (D.-H. Kim). A full list of funding, and individual and study acknowledgements are available in the Supplementary Information.

**Author contributions** Project coordination (including supervision of all experiments and analyses): K.L.M. and X.S. Manuscript writing: C.N.S., E.-S.T., M.B., M.H., Y.J.K., K.L., K.L.M., X.S. Core and follow-up analyses including NKX6-1 functional work: C.N.S., M.H., Y.J.K., K.L., A.K.I., H.J.P., S.M.B., K.L.M., X.S. eQTL lookups for ANK1/NKX6-3: M.v.d.B., A.L.G. Core DIAMANTE group: J.E.B. (Hispanic), M.B. (European), D.W.B. (African-American), J.C.C. (South Asian), A.M. (European), M.I.M. (European), M.C.Y.N. (African-American), L.E.P. (Hispanic), W. Zhang (South Asian), A.P.M. (trans-ethnic), K.L.M., X.S. Study-level data analyses: Y.T. (AASC, NAGAHAMA), M.H. (BBJ), K.S. (BBJ), X.S. (BES, DC/SP2, SCHS), F.T. (CAGE-Amagasaki, CAGE-GWAS), M.N. (CAGE-KING), C.N.S. (CHNS, CLHNS), K.L. (CKB), F.B. (CKB), Y.J.K. (KARE, KBA), S. Moon (KBA), C.H.T.T. (HKDR), J.Y. (MESA), X.G. (MESA), J. Long (SBC/SWHS), J.F.C. (SCES), V.J.Y.L. (SiMES), S.-H.K. (SNUH), H.S.C. (SMC), C.-H.C. (TWT2D). Individual cohort study design and/or principal investigators for the individual cohorts: M. Igase (AASC), M.A. (BBJ), S. Maeda (BBJ), T. Kadokawa (BBJ), Y.-X.W. (BES), N.K. (CAGE-Amagasaki, CAGE-GWAS), M.Y. (CAGE-KING), K.L.M. (CHNS, CLHNS), R.G.W. (CKB), E.-S.T. (DC/SP2), J. Lee (KBA), B.-J.K. (KARE, KBA), R.C.W.M. (HKDR), J.I.R. (MESA), F.M. (NAGAHAMA), X.-O.S. (SBC/SWHS), C.-Y.C. (SCES), W.-P.K. (SCHS), T.-Y.W. (SiMES), K.-S.P. (SNUH), Y.S.C. (SMC), W.H.H.S. (TaiChi-G), J.Y.W. (TWT2D) Genotyping and/or phenotyping of the studies: K.K. (AASC), A.T. (BBJ), Y.K. (BBJ), T.Y. (BBJ), Y.O. (BBJ), J.B.J. (BES), T. Katsuya. (CAGE-Amagasaki), M. Isono (CAGE-GWAS), S.I. (CAGE-KING), K. Yamamoto (CAGE-KING), A.-G.H. (CHNS), S.D. (CHNS), W.H. (CHNS), J.S. (CHNS), P.G.-L. (CHNS), C.Y. (CKB), Y.G. (CKB), Z.B. (CKB), J. Lv (CKB), L.L. (CKB), Z.C. (CKB), N.R.L. (CLHNS), L.S.A. (CLHNS), J. Liu (DC/SP2), R.M.V.D. (DC/SP2), S.H. (KBA), K. Yoon (KBA), H.-M.J. (KBA), D.M.S. (KBA), G.J. (HKDR), A.O.L. (HKDR), B.T. (HKDR), W.Y.S. (HKDR), J.C.N.C. (HKDR), M.Y.H. (KARE, KBA), Y.-D.I.C. (MESA), T. Kawaguchi (NAGAHAMA), Y.-B.X. (SBC/SWHS), W. Zheng (SBC/SWHS), L.Z. (SCES), C.-C.K. (SCES), M.A.P. (SCHS), M.G. (SCHS), J.-M.Y. (SCHS), C.S. (SiMES), M.-L.C. (SiMES), M.-S.L. (SMC), C.-M.H. (TaiChi-G), L.-M.C. (TaiChi-G), Y.-J.H. (TaiChi-G), L.-C.C. (TWT2D), Y.-T.C. (TWT2D), F.-J.T. (TWT2D), J.I.R. (MESA).

**Competing interests** M.v.d.B. was supported by a Novo Nordisk postdoctoral fellowship run in partnership with the University of Oxford, and currently a full-time employee of Novo Nordisk. M.I.M and A.M. are currently employees of Genentech. The other authors declare no competing interests.

#### Additional information

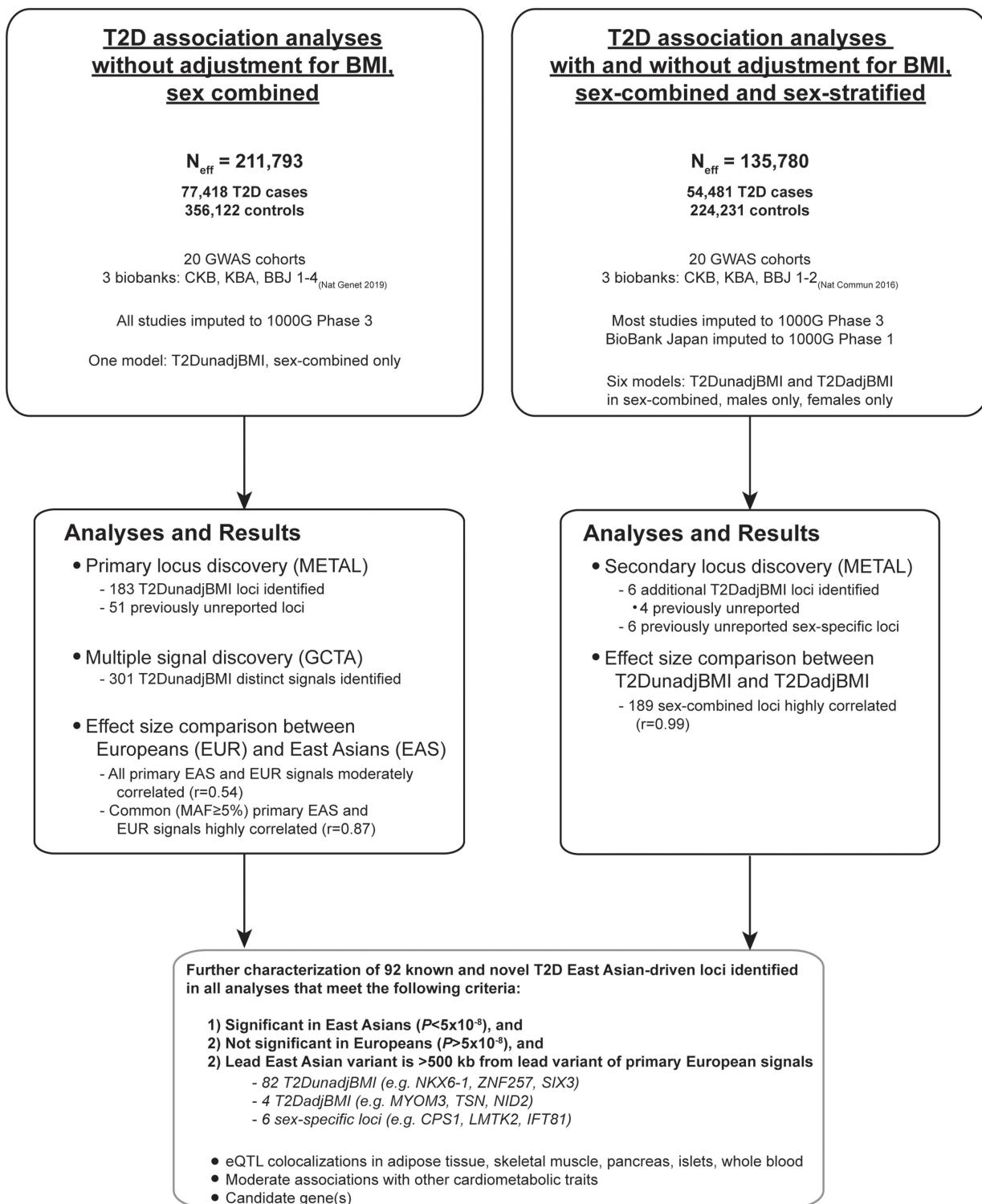
**Supplementary information** is available for this paper at <https://doi.org/10.1038/s41586-020-2263-3>.

**Correspondence and requests for materials** should be addressed to T.K., R.G.W., B.-J.K., K.L.M. or X.S.

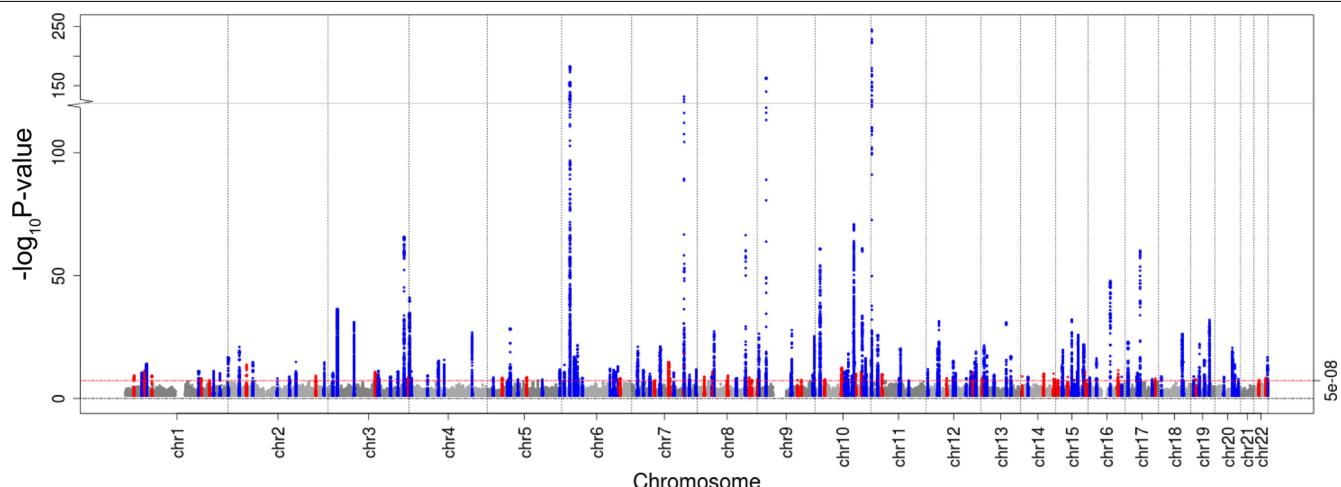
**Peer review information** *Nature* thanks Timothy Frayling, Stephen Rich and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

**Reprints and permissions information** is available at <http://www.nature.com/reprints>.

## East Asian T2D meta-analyses

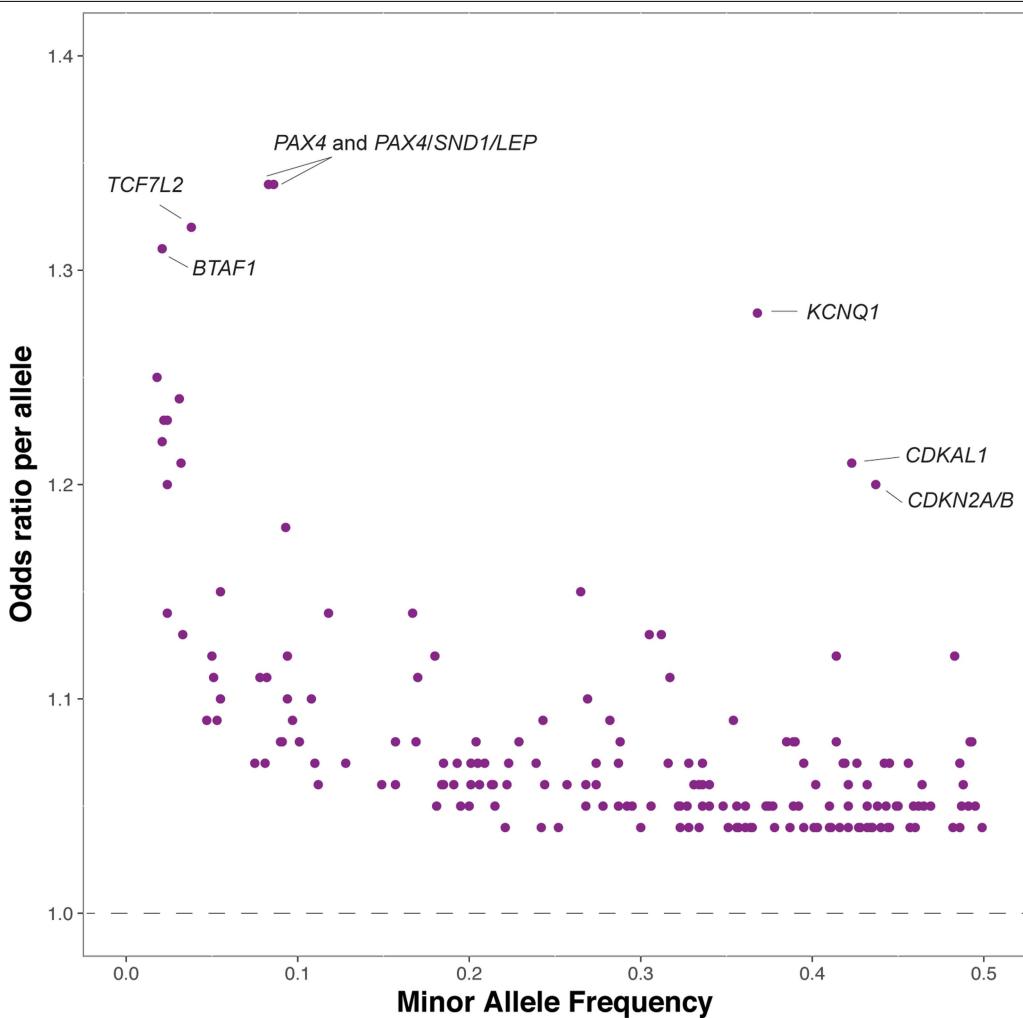


**Extended Data Fig. 1 | Flow chart of study design.** The flow chart shows the different data analyses performed and corresponding summary of association results.



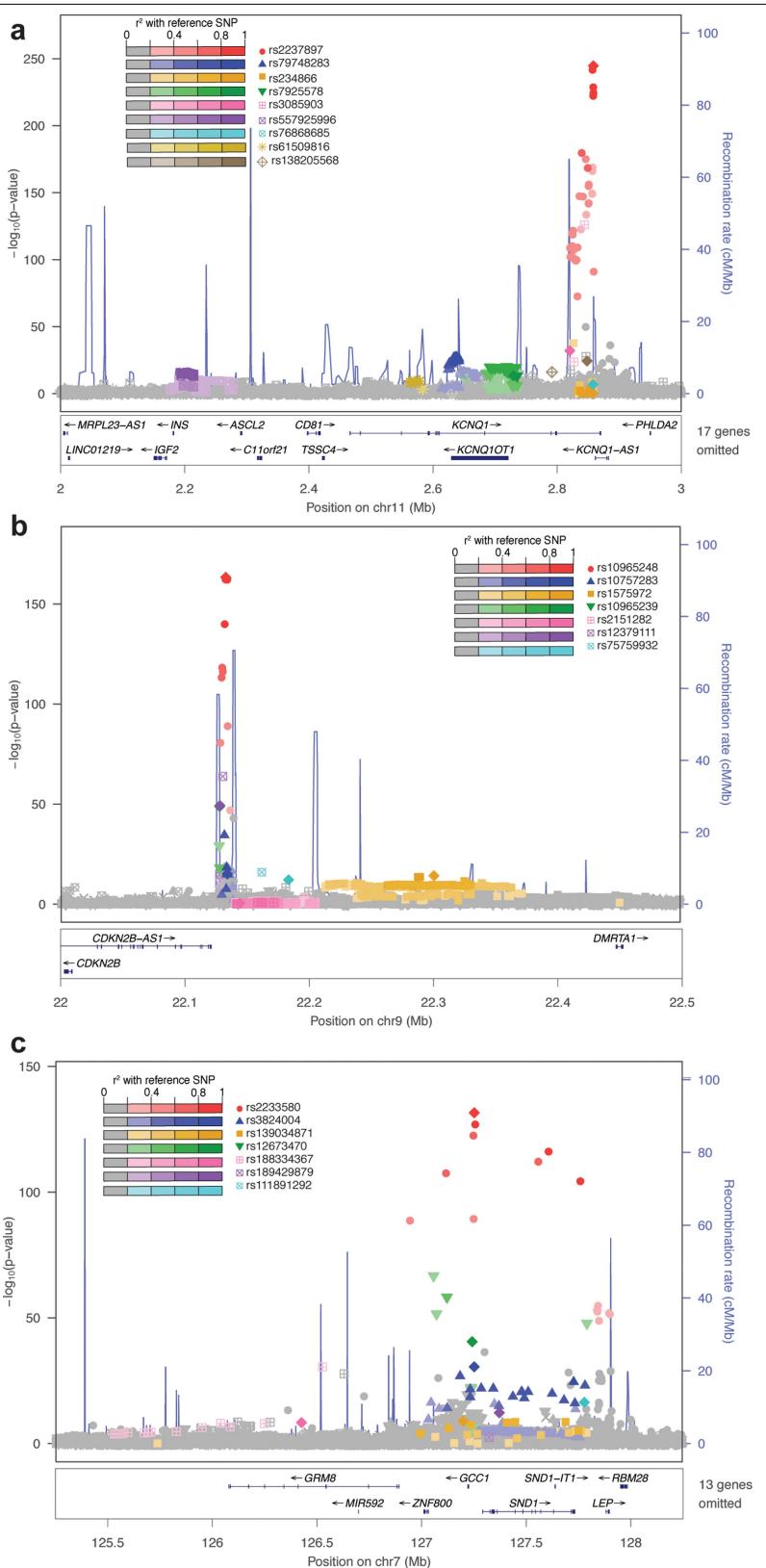
**Extended Data Fig. 2 | Manhattan plot for East Asian T2D meta-analysis association results in model unadjusted for BMI.**  $-\log_{10}P$  values from two-sided fixed-effect inverse-variance genome-wide meta-analysis association results for each variant (yaxis; maximal  $N_{\text{eff}} = 211,793$ ) was plotted

against the genomic position (hg19; xaxis). Known T2D loci achieving genome-wide significance ( $P < 5.0 \times 10^{-8}$ ) in our meta-analysis are shown in blue. Loci achieving genome-wide significance that were previously unreported for T2D association are shown in red.



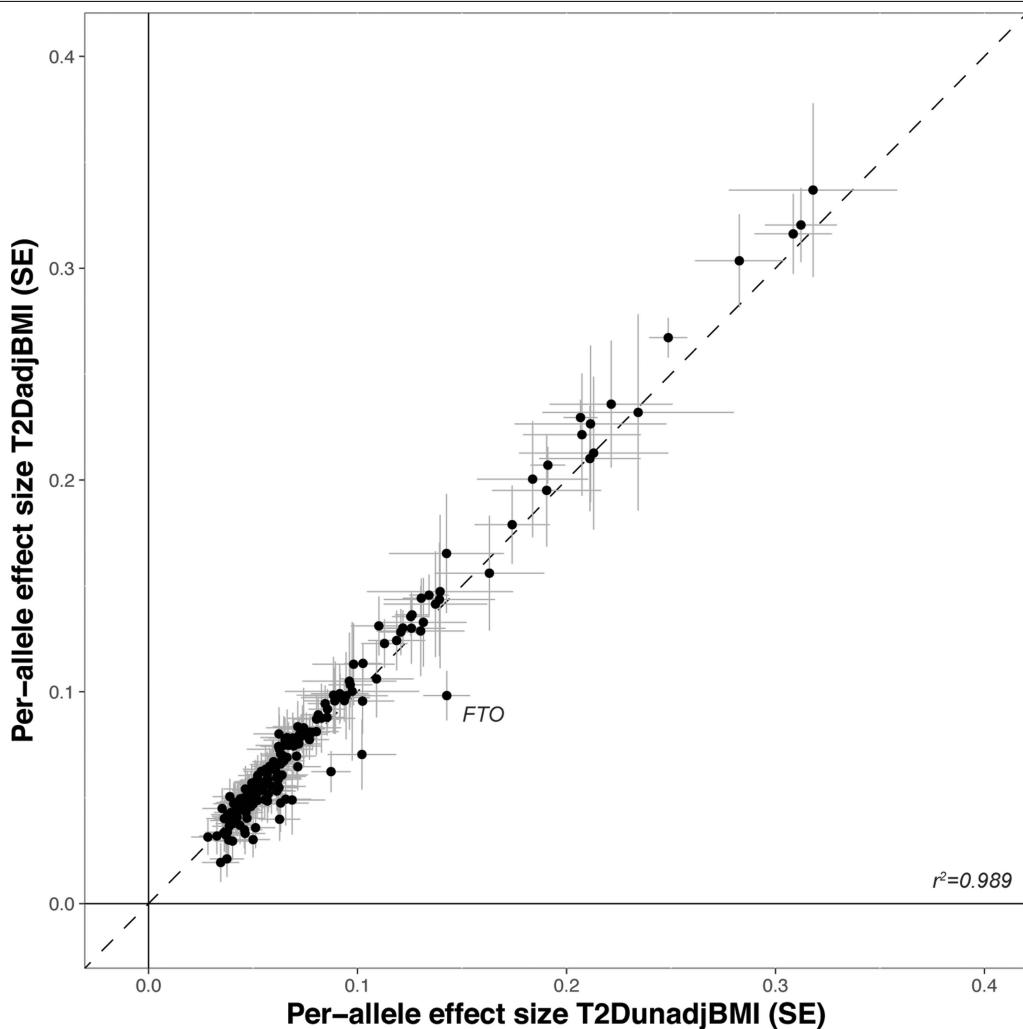
**Extended Data Fig. 3 | The relationship between effect size and minor allele frequency.** Odds ratios (y axis) and minor allele frequencies (x axis) for 189 primary association signals from the T2D BMI-unadjusted models. Odds ratios

are from two-sided fixed-effect inverse-variance meta-analysis on a maximal effective sample size of 211,793.



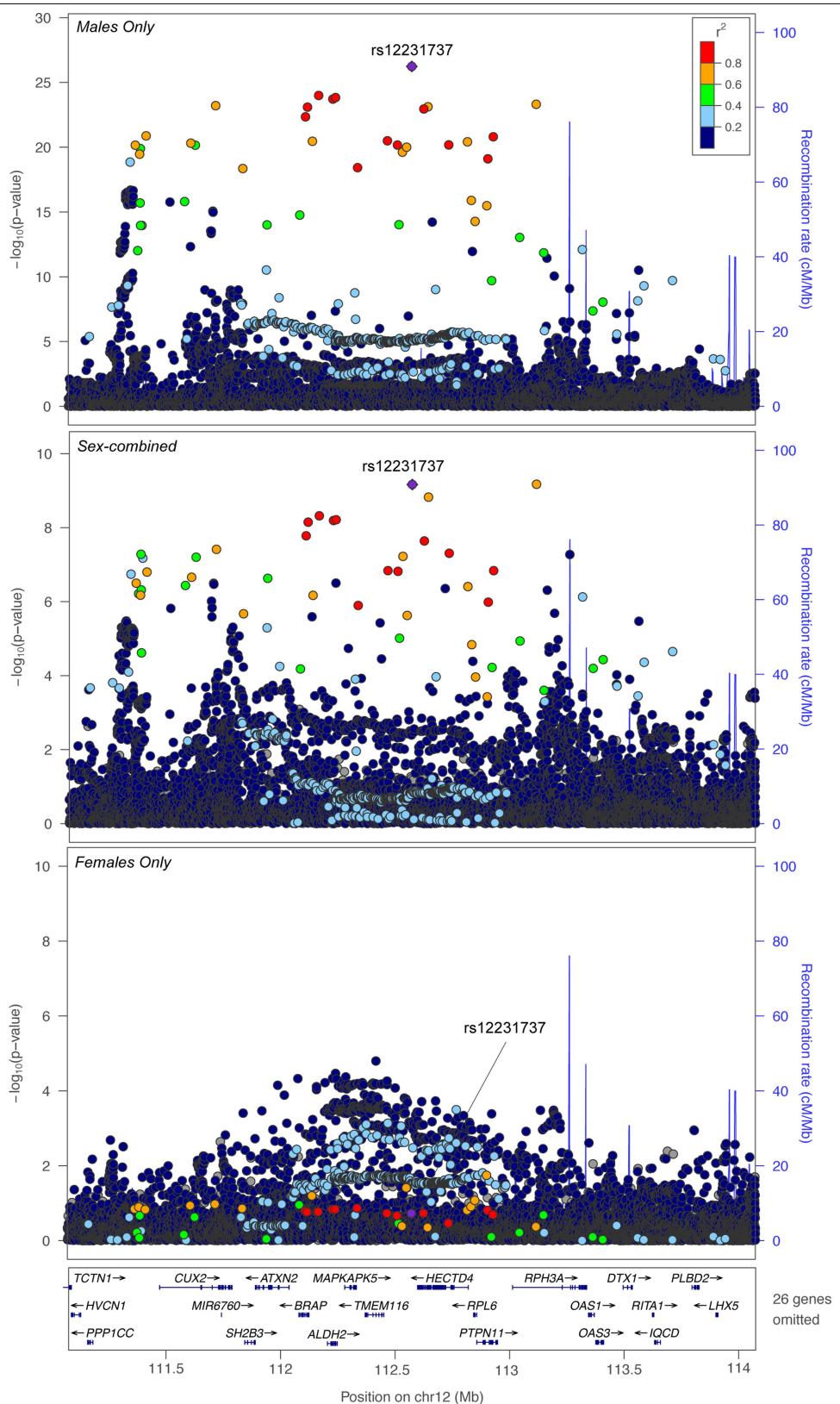
**Extended Data Fig. 4 | Regional association plots at three T2D-associated loci with the strongest association  $P$  values and more than five distinct association signals in East Asian individuals.** **a**, *INS/IGF2/KCNQ1*. **b**, *CDKN2A/B*. **c**, *PAX4/LEP*.  $-\log_{10}P$  values were from the two-sided fixed-effect inverse-variance meta-analysis. Distinct signals ( $P < 1.0 \times 10^{-5}$  from GCTA

conditional analyses) were plotted;  $N_{\text{eff}}$  values for each distinct signal are reported in Supplementary Table 4. Variants are coloured based on East Asian 1000G Phase 3 LD with the lead variants for each association signal shown as diamonds.



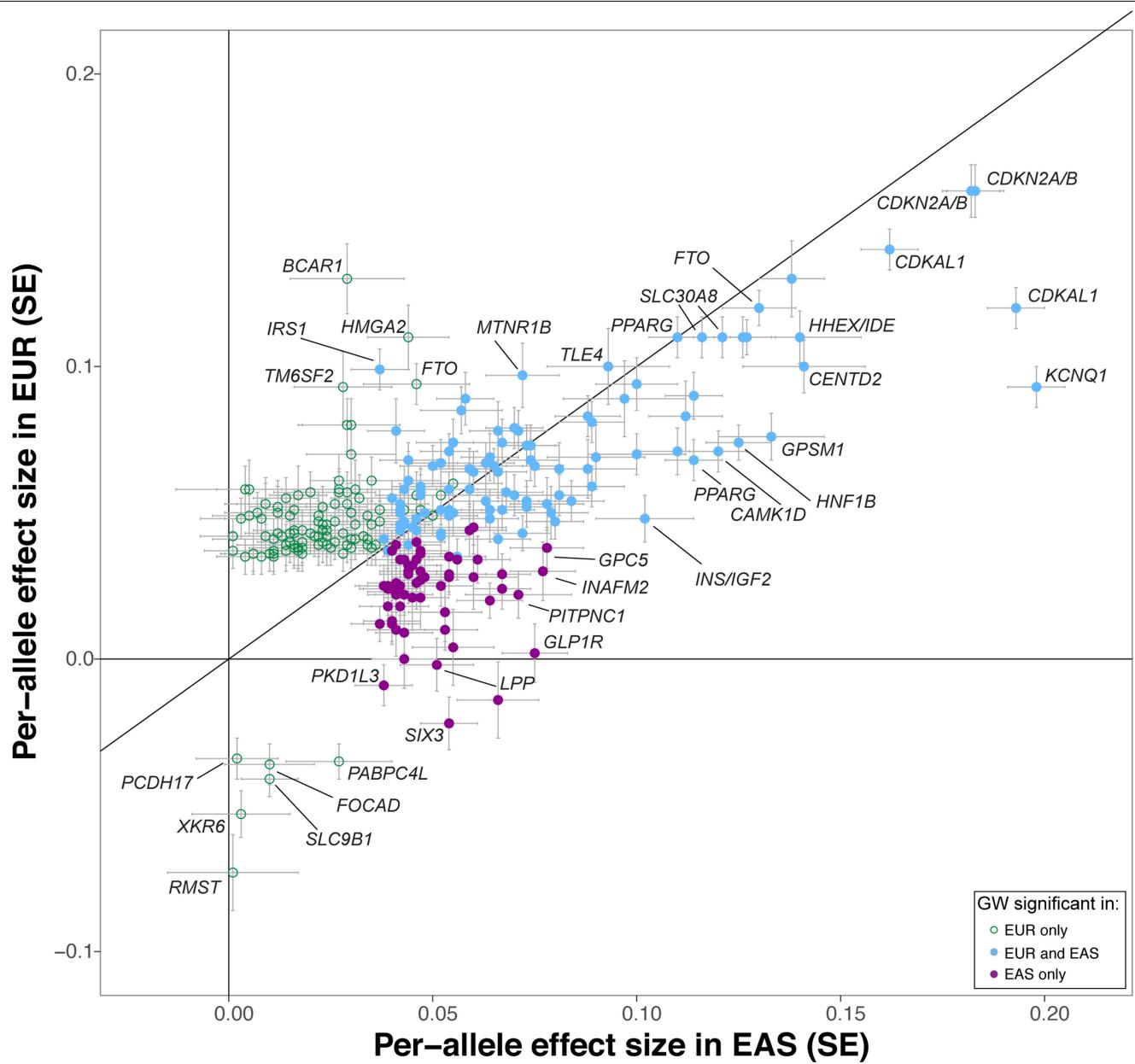
**Extended Data Fig. 5 | Effect size comparison of lead variants in sex-combined models unadjusted and adjusted for BMI.** At 189 lead variants identified in the East Asian BMI-unadjusted sex-combined T2D meta-analysis, per-allele effect sizes ( $\beta$ ) from the BMI-adjusted sex-combined model (T2DadjBMI) were plotted against the BMI-unadjusted sex-combined model

(T2DunadjBMI). Both sex-combined models were from two-sided fixed-effect inverse-variance meta-analyses and included the same set of studies for comparable sample size. Each point denotes the per-allele effect size; grey lines, s.e. Effect sizes between the two models are highly correlated with a Pearson correlation coefficient  $r = 0.99$  (Supplementary Table 4).



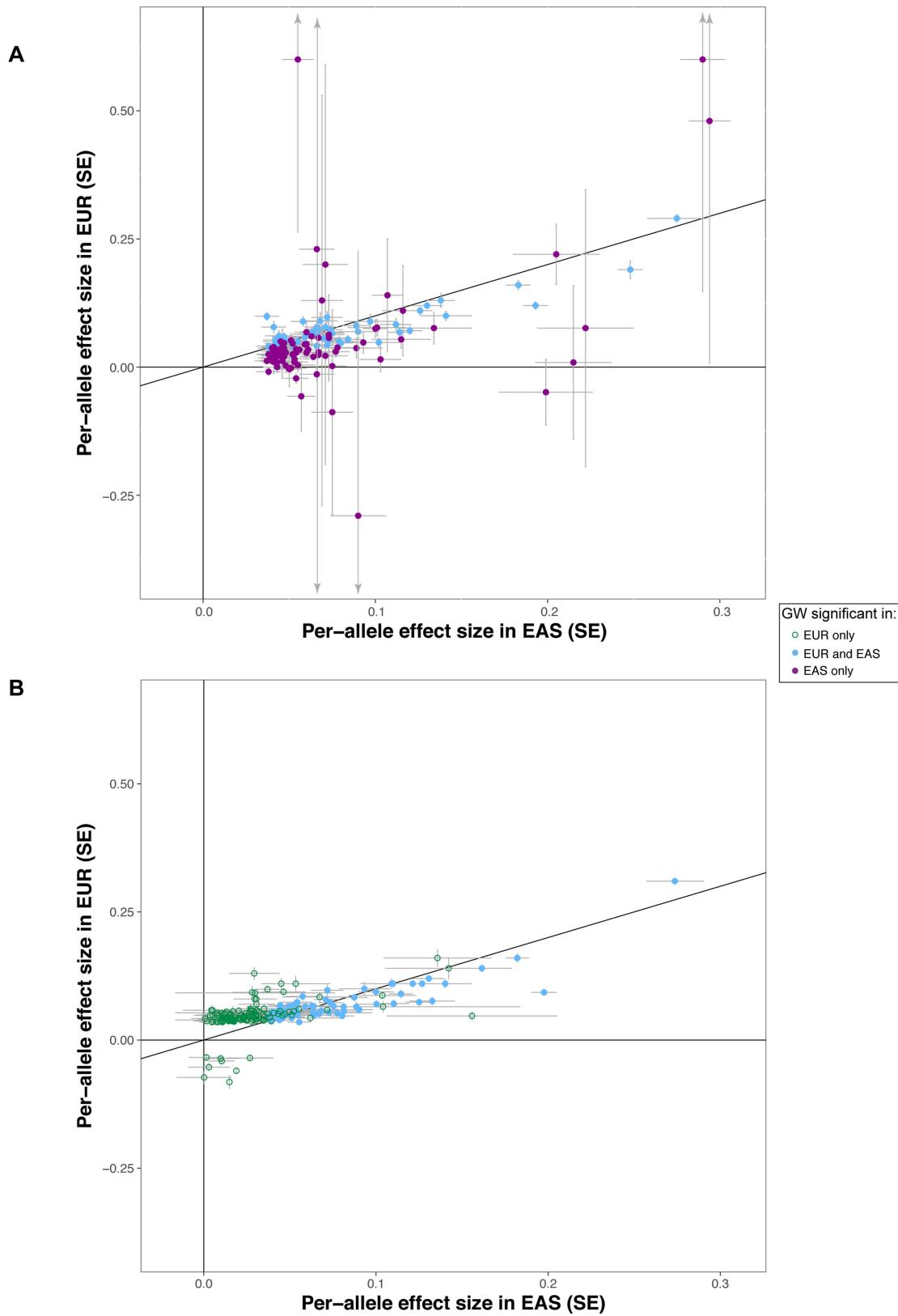
**Extended Data Fig. 6 | Regional plots of male-specific T2D-associated locus, *ALDH2*.** For each plot,  $-\log_{10}P$  values from association results from two-sided fixed-effect inverse-variance meta-analyses for each variant (y axis) were plotted against the genomic position (hg19; x axis). The lead variant rs12231737 plotted is the lead variant from the BMI-unadjusted male-specific meta-analysis ( $N_{eff}=65,202$ ) and also the sex-combined meta-analysis ( $N_{eff}=138,947$ ) from the same subset of individuals included in the

sex-stratified analyses (female-specific  $N_{eff}=70,051$ ). This lead variant rs12231737 is in high LD with rs7768175, identified from the larger BMI-unadjusted sex-combined meta-analysis (East Asian  $r^2=0.80$ ). Top, males only; middle, sex-combined; bottom, females only. Variants are shaded according to East Asian 1000G Phase 3 LD with the lead variant shown as a purple diamond.



**Extended Data Fig. 7 | Effect size comparison of common lead variants (MAF  $\geq 5\%$ ) identified in this East Asian meta-analysis and a previously published European T2D GWAS meta-analysis.** For 278 unique lead variants with MAF  $\geq 5\%$  in both the East Asian and European BMI-unadjusted meta-analyses, per-allele effect sizes ( $\beta$ ) from the European T2D GWAS meta-analysis<sup>2</sup> (y axis) were plotted against per-allele effect sizes from our East Asian meta-analysis (x axis). Effect sizes from both meta-analyses were from

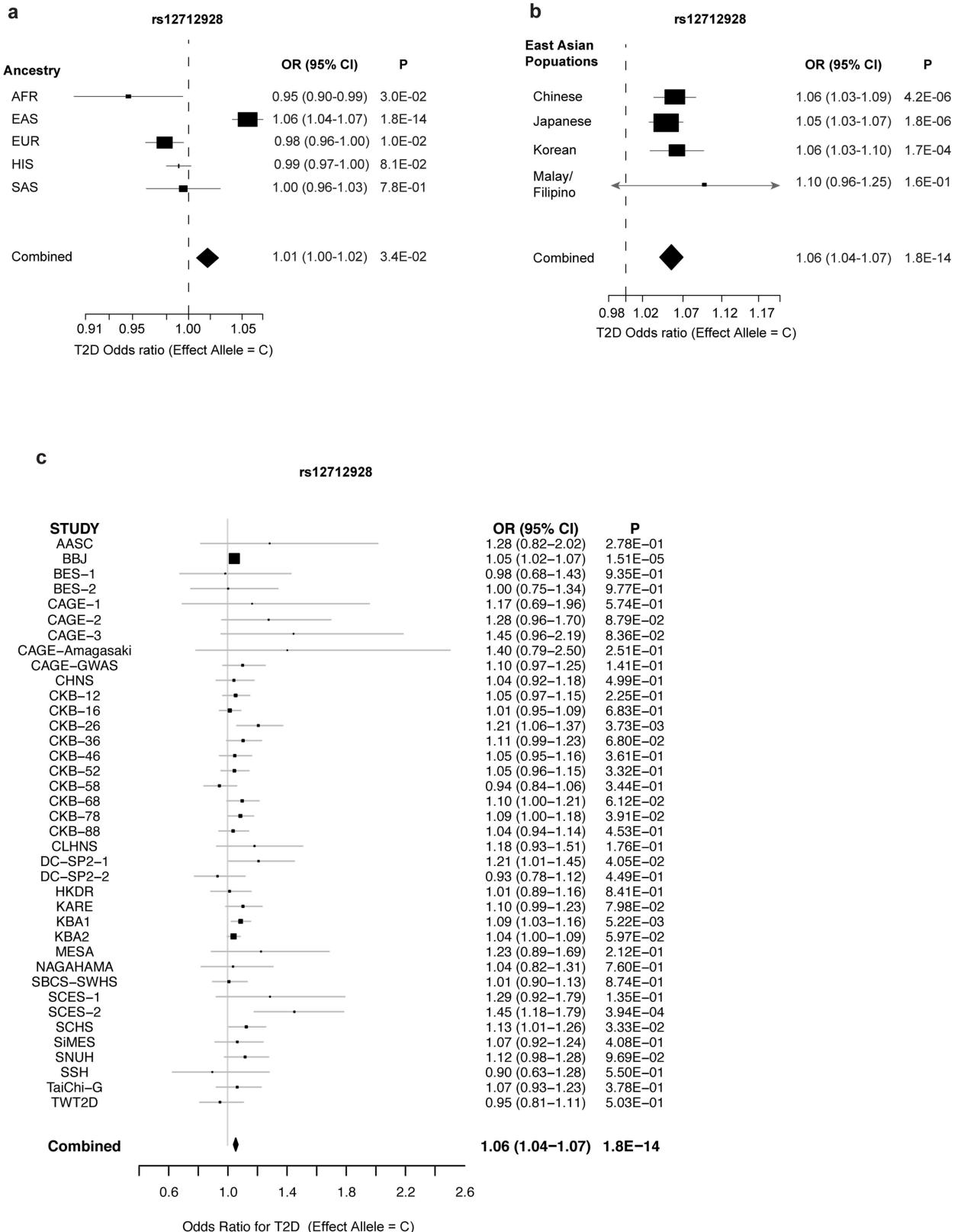
two-sided fixed-effect inverse-variance meta-analyses (maximal  $N_{\text{eff}} = 211,793$  for East Asian and 231,436 for European meta-analyses). Each point denotes the per-allele effect size; grey lines, s.e. Variants are coloured purple if they were significant in the East Asian meta-analysis only, green if they were significant in the European meta-analysis only, and blue if they were significant in both the East Asian and European meta-analyses (see Methods and Supplementary Table 7).



**Extended Data Fig. 8 | Effect size comparison of lead variants identified in East Asian BMI-unadjusted meta-analysis and previously published European T2D GWAS meta-analysis.**

For 332 lead variants identified from the two BMI-unadjusted meta-analyses, per-allele effect sizes ( $\beta$ ) from a European meta-analysis<sup>2</sup> (y axis) were plotted against per-allele effect sizes from our East Asian meta-analysis (x axis). Effect sizes from both meta-analyses were from two-sided fixed-effect inverse-variance meta-analysis (maximal  $N_{\text{eff}} = 211,793$

for East Asian and 231,436 for European meta-analyses). Each point denotes the per-allele effect size; grey lines, s.e. **a**, The 152 lead variants that were significant in the East Asian meta-analysis (purple) or both the East Asian and European meta-analyses (blue). **b**, The 192 lead variants that were significant in the European meta-analysis (green) or both the East Asian and European meta-analyses (blue). These plots include only one variant per locus, in contrast to Fig. 2 and Extended Data Fig. 7.



Extended Data Fig. 9 | See next page for caption.

## Article

**Extended Data Fig. 9 | Forest plots of BMI-unadjusted meta-analysis association results at the *SIX3–SIX2* locus.** Odds ratios (black boxes) and 95% confidence intervals (horizontal lines) for T2D associations at the lead East Asian variant (rs12712928) across ancestries of African-American (AFR), East Asian (EAS), European (EUR)<sup>2</sup>, Hispanic (HIS), and South Asian (SAS) individuals (a); within four major East Asian populations (b; Chinese, Japanese, Korean, and Malay–Filipino, combined due to small sample sizes); and from each contributing cohort (c). Effect sizes from the East Asian study, ancestry, population, and combined meta-analysis were from two-sided fixed-effect

inverse-variance meta-analysis. The size of each box is proportional to the sample size of each contributing study, ancestry or population (Supplementary Table 8). Our East Asian study had >90% power to detect the observed association with a MAF = 0.40, OR = 1.06, and 77,418 T2D cases. Given the number of T2D cases and frequency of rs12712928-C within the other data sets, at 80% power, we can reasonably exclude association OR > 1.07 in EUR and OR > 1.15 in AFR, HIS, and SAS between rs12782928 and T2D. Full study names can be found in Supplementary Table 1 and corresponding sample sizes can be found in Supplementary Table 2.

**Extended Data Table 1 | Novel lead variants associated with type 2 diabetes in East Asian individuals**

Locus	Lead variant	Chr	Pos	RA	NRA	RAF	Cases (N)	Controls (N)	Neff	OR (95% CI)	P	P <sub>het</sub>
VWA5B1	rs60573766	1	20,688,352	C	T	0.64	77,418	356,122	211,793	1.04 (1.03-1.06)	4.3E-10	8.2E-01
MAST2	rs562138031	1	46,244,900	C	CT	0.73	77,221	351,786	211,039	1.06 (1.04-1.07)	4.0E-12	7.5E-01
PGM1	rs2269245	1	64,107,893	G	A	0.82	77,221	351,786	211,039	1.06 (1.04-1.07)	5.4E-10	3.9E-01
TSEN15	rs1327123	1	184,014,593	C	G	0.46	77,418	356,122	211,793	1.04 (1.03-1.05)	7.0E-09	9.2E-01
MDM4	rs201297151	1	204,474,581	CAAAAAAAA	C	0.44	77,221	351,786	211,039	1.04 (1.03-1.05)	3.4E-08	3.7E-01
SIX3	rs12712928	2	45,192,080	C	G	0.40	77,418	356,122	211,793	1.06 (1.04-1.07)	1.8E-14	2.4E-01
IKZF2	rs75179644	2	213,687,103	T	C	0.90	77,418	356,122	211,793	1.08 (1.05-1.10)	5.4E-10	7.6E-01
ZBTB20	rs6806156	3	114,968,018	T	C	0.61	77,418	356,122	211,793	1.05 (1.03-1.06)	1.6E-11	8.7E-01
TFRC	rs9866168	3	195,830,310	T	A	0.64	77,418	356,122	211,793	1.05 (1.03-1.06)	1.6E-09	2.7E-01
RANBP3L	rs16902871	5	36,257,018	G	A	0.15	77,418	356,122	211,793	1.06 (1.04-1.08)	3.3E-09	2.0E-01
PCSK1	rs5556925	5	95,848,503	C	A	0.42	77,418	356,122	211,793	1.04 (1.03-1.05)	3.1E-09	2.2E-01
REPS1	rs9376382	6	139,205,386	C	T	0.60	77,418	356,122	211,793	1.04 (1.03-1.05)	1.5E-08	5.4E-01
HIVEP2	rs9390022	6	143,056,556	T	C	0.80	77,418	356,122	211,793	1.05 (1.03-1.07)	6.4E-09	7.2E-01
ZNF713	rs565050730	7	55,984,953	GA	G	0.33	77,418	356,122	211,793	1.04 (1.03-1.06)	4.4E-08	7.1E-01
STEAP1	rs52469016	7	89,752,238	C	G	0.22	77,418	356,122	211,793	1.07 (1.05-1.08)	1.5E-15	2.4E-01
CALCR	rs2074120	7	93,107,093	A	C	0.32	77,418	356,122	211,793	1.04 (1.03-1.06)	8.4E-09	9.3E-02
GRM8/PAX4	rs117737118	7	126,526,991	G	A	0.09	77,062	350,162	210,460	1.18 (1.14-1.21)	3.3E-31	3.3E-01
ASAH1	rs34642578	8	17,927,609	T	C	0.05	77,418	356,122	211,793	1.09 (1.06-1.13)	1.6E-09	5.2E-02
ZNF703	rs4739515	8	37,391,203	G	C	0.54	77,418	356,122	211,793	1.05 (1.03-1.06)	1.7E-11	7.5E-01
FGFR1	rs328301	8	38,343,012	T	C	0.33	77,418	356,122	211,793	1.04 (1.03-1.06)	4.1E-08	6.9E-01
KCNB2	rs349359	8	73,503,743	C	A	0.24	77,418	356,122	211,793	1.04 (1.03-1.06)	3.1E-08	2.5E-01
GDAP1	rs149265787	8	75,214,398	G	A	0.02	77,392	355,608	211,694	1.14 (1.10-1.19)	5.7E-10	1.1E-01
TRIB1	rs60089934	8	126,471,274	A	G	0.38	77,418	356,122	211,793	1.04 (1.03-1.06)	3.3E-09	1.0E-01
EFR3A	rs73708054	8	132,879,795	C	T	0.25	77,418	356,122	211,793	1.04 (1.03-1.06)	4.4E-08	2.3E-01
DMRT2	rs1016565	9	1,032,567	A	G	0.42	77,418	356,122	211,793	1.04 (1.02-1.05)	2.2E-08	3.4E-01
PTCH1	rs113154802	9	98,278,413	C	T	0.89	77,418	356,122	211,793	1.06 (1.04-1.08)	3.5E-08	7.5E-02
ABCA1	rs201375651	9	107,597,527	CA	C	0.39	77,418	356,122	211,793	1.04 (1.03-1.06)	2.6E-08	2.0E-01
PTF1A	rs77065181	10	23,487,778	A	G	0.05	77,418	356,122	211,793	1.09 (1.06-1.13)	1.6E-08	8.3E-01
ARID5B	rs141583966	10	63,712,602	G	GGTGT	0.91	77,175	351,384	210,874	1.08 (1.05-1.11)	7.7E-10	2.6E-01
JMJD1C	rs148928116	10	64,976,133	T	TA	0.79	77,418	356,122	211,793	1.06 (1.04-1.08)	2.5E-13	8.7E-02
ARHGAP19	rs10736116	10	99,056,921	C	G	0.31	77,418	356,122	211,793	1.05 (1.03-1.06)	9.2E-11	7.7E-01
BBIP1	rs7895872	10	112,678,657	T	G	0.58	77,418	356,122	211,793	1.05 (1.03-1.06)	1.4E-11	5.5E-01
BDNF	rs4922793	11	27,729,505	A	G	0.57	77,418	356,122	211,793	1.04 (1.03-1.06)	1.6E-10	3.8E-01
FAIM2	rs77978149	12	50,269,863	T	C	0.09	77,259	354,498	211,214	1.08 (1.05-1.10)	5.7E-09	2.1E-01
ALDH2	rs149212747	12	111,886,771	A	AC	0.79	77,221	351,786	211,039	1.07 (1.05-1.09)	2.1E-11	1.7E-05
RBM19	rs7307263	12	114,123,722	G	C	0.43	77,418	356,122	211,793	1.04 (1.02-1.05)	3.6E-08	9.6E-01
FGF9	rs9316706	13	22,589,883	A	G	0.35	77,418	356,122	211,793	1.04 (1.03-1.06)	3.3E-09	6.5E-01
NYNRIN	rs12437434	14	24,878,370	C	T	0.71	77,221	351,786	211,039	1.05 (1.03-1.06)	1.0E-09	3.8E-01
LRRK74A	rs58524310	14	77,382,503	G	A	0.33	75,090	353,783	207,126	1.05 (1.03-1.06)	8.4E-11	3.7E-01
DLK1/MEG3/ miRNA cluster	rs73347525	14	101,255,172	A	G	0.76	74,931	352,159	206,547	1.06 (1.04-1.08)	7.5E-11	3.8E-01
TRAF3	rs55700915	14	103,237,952	A	G	0.43	77,221	351,786	211,039	1.04 (1.03-1.06)	1.5E-08	9.9E-01
HERC2	rs76704029	15	28,546,173	T	C	0.73	66,648	265,839	174,183	1.06 (1.04-1.08)	3.4E-08	1.5E-01
MYO5C	rs149336329	15	52,587,740	G	T	0.95	77,221	351,786	211,039	1.11 (1.07-1.14)	1.7E-09	8.3E-01
RGMA	rs61021634	15	93,825,384	A	G	0.44	77,418	356,122	211,793	1.05 (1.03-1.06)	1.4E-11	1.1E-01
IGF1R	rs79826452	15	99,366,409	A	G	0.89	77,418	356,122	211,793	1.07 (1.05-1.10)	3.2E-08	4.1E-01
PKD1L3	rs12600132	16	72,022,534	T	C	0.43	77,418	356,122	211,793	1.04 (1.03-1.05)	6.0E-09	7.9E-01
ZFHX3	rs6416749	16	73,100,308	C	T	0.37	77,062	350,162	210,460	1.05 (1.04-1.07)	3.4E-12	5.9E-01
SUMO2	rs35559984	17	73,187,031	CA	C	0.65	74,931	352,159	206,547	1.05 (1.03-1.07)	7.9E-09	9.8E-01
ZNF799	rs4604181	19	12,509,536	A	C	0.51	77,062	350,162	210,460	1.04 (1.03-1.06)	1.5E-08	5.5E-01
ZNRF3	rs147413364	22	29,380,119	T	TA	0.36	77,175	351,384	210,874	1.04 (1.03-1.06)	3.4E-08	1.4E-01
WNT7B	rs28637892	22	46,313,618	T	G	0.22	63,772	319,376	175,881	1.05 (1.04-1.07)	3.7E-09	4.2E-01

Single variant association results from East Asian fixed-effect inverse-variance meta-analysis (BMI-unadjusted sex-combined model adjusted for age, sex, and study-specific covariates) using METAL. Loci were defined as novel if the lead variant was (i) at least 500 kb away from and confirmed by GCTA to be conditionally independent of previously reported T2D-associated variants in any ancestry, and (ii) assessed using locuszoom plots and biology lookups to be to be away from known loci with extended LD. Four additional variants met the definition for a novel locus but were located within the previously reported major histocompatibility complex (MHC) region; see Supplementary Table 5 for the full list of distinct association signals at the MHC region. rs4804181 was more than 500 kb from primary signal rs3111316 in European meta-analysis, but less than 500 kb from their secondary signal rs755734872. Genome-wide significant association is defined as  $P < 5.0 \times 10^{-8}$ . Physical position based on hg19. Effect alleles are associated with increased risk for T2D. Odds ratios reflect per allele effects of variants on T2D risk.

Chr, chromosome; Pos, position; RAF, risk allele frequency; Neff, effective sample size; OR, odds ratio; CI, confidence interval; P, P value.

# 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 [Authors & Referees](#) 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

The software used have been described in details in Online Methods section and Supplementary Tables 1 and 3.

Softwares and resources included:

Pre-imputation preparation and quality control: <http://www.well.ox.ac.uk/~wrayner/tools/>  
 Pre-phasing: SHAPEITv2/v2.3, [https://mathgen.stats.ox.ac.uk/genetics\\_software/shapeit/shapeit.html](https://mathgen.stats.ox.ac.uk/genetics_software/shapeit/shapeit.html)  
 Pre-phasing: Eagle v2, <https://data.broadinstitute.org/alkesgroup/Eagle/>  
 Imputation: MACHv1.0, minimac3/4, Michigan Imputation Server, <https://imputationserver.sph.umich.edu/index.html>  
 Imputation: IMPUTEv4, <https://jmarchini.org/impute-4/>  
 Association: EPACTSv3.3 (FIRTH, EMMAX), <https://github.com/statgen/EPACTS>  
 Association: RVTESTSv2, <https://github.com/zhanxw/rvtests>  
 Association: SNPTESTv2.5, [https://mathgen.stats.ox.ac.uk/genetics\\_software/snptest/snptest.html](https://mathgen.stats.ox.ac.uk/genetics_software/snptest/snptest.html)  
 Association: mach2datv1.0.24, <http://csg.sph.umich.edu/yli/software.html>  
 Association: BOLT-LMMv2.3, <https://data.broadinstitute.org/alkesgroup/BOLT-LMM/>  
 Meta-analysis: METALv2011-03-25, <http://csg.sph.umich.edu/abecasis/metal/>  
 Meta-analysis: GWAMAv2.2.2, <https://genomics.ut.ee/en/tools/gwama>  
 Analysis of GWAS statistics: GLDSCv1.01, <https://github.com/bulik/lpsc>  
 Conditional analysis: GCTA v1.91.5, <https://cnsgenomics.com/software/gcta/#Overview>  
 Visualization: LocusZoom1.4 <http://locuszoom.org/>  
 Resource: Type 2 Diabetes Knowledge Portal, <http://www.type2diabetesgenetics.org>  
 Resource: GTEx Portal, <https://gtexportal.org/home/>  
 Resource: Parker lab Islet Browser, <http://theparkerlab.org/tools/isletetql/>

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/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

Summary-level statistics will be available at the AGEN consortium website <https://blog.nus.edu.sg/agen/summary-statistics/t2d-2020>, the GWAS catalog <https://www.ebi.ac.uk/gwas/>, and the Accelerating Medicines Partnership (AMP) T2D portal <http://www.type2diabetesgenetics.org/> (scheduled release April 2020).

## 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](https://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 (N>77,000 T2D cases and >356,00 controls of East Asian ancestry) with GWAS data imputed up to 1000 Genomes Phase 3 reference panel to study the role of genetic variants in T2D. This East Asian study had >80% power to detect the observed association with a MAF=0.05, OR=1.12, and 77,418 T2D cases. Our sample size is adequate to observe common known T2D associated regions, and identify 61 novel T2D associated regions.

Data exclusions

Established protocols were used to conduct rigorous data quality control for each GWAS at the study level: variants were first excluded for the following reasons: i) imputation quality score was poor (minimac3 r<sup>2</sup><0.30; IMPUTE2 info score <0.40); ii) combined case control minor allele count <5; or iii) standard error of the log-OR>10 (details in Supplementary Table 3 and Online methods). In addition, to improve the quality of the genotype scaffold in each study, we developed a harmonized protocol in which variants were subsequently removed if: i) mismatched chromosomal positions or alleles not present in the reference panel; ii) ambiguous alleles (AT/GC) with minor allele frequency (MAF) >40% in the reference panel; or iii) absolute allele frequency differences greater than 20% compared to East Asian-specific allele frequencies.

Replication

Because we used all available East Asian data for this analysis, no replication was performed.

Randomization

Not relevant because the study is not experimental.

Blinding

Not relevant because the study is not experimental.

## 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
<input checked="" type="checkbox"/>	<input type="checkbox"/> Antibodies
<input type="checkbox"/>	<input checked="" type="checkbox"/> Eukaryotic cell lines
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology
<input checked="" type="checkbox"/>	<input type="checkbox"/> Animals and other organisms
<input type="checkbox"/>	<input checked="" type="checkbox"/> Human research participants
<input checked="" type="checkbox"/>	<input type="checkbox"/> Clinical data

### Methods

n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> ChIP-seq
<input checked="" type="checkbox"/>	<input type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging

## Eukaryotic cell lines

Policy information about [cell lines](#)

Cell line source(s)

MIN6 cell lines are publicly available through the ATCC.

Authentication

Cell lines were authenticated through genotyping.

Mycoplasma contamination

Cell lines tested negative for mycoplasma contamination.

Commonly misidentified lines  
(See [ICLAC](#) register)

None

## Human research participants

Policy information about [studies involving human research participants](#)

Population characteristics

Analyses were conducted on GWAS summary statistics of 77,418 cases and 356,122 controls of East Asian ancestry from 23 studies. Relevant characteristics include age, sex, and BMI. Study designs include case control, cross-sectional, nested case control, hospital registries and biobanks from China, Hong Kong, Japan, the Philippines, Singapore, South Korea, Taiwan, and the United States. Full description of sample characteristics of each study are provided in Supplementary Tables 1-3.

Recruitment

Participants were recruited originally as a part of numerous studies (case control, cross-sectional, nested case control, hospital registries, and biobanks). For most cohorts, participants were recruited over a long period of time and genotyping was performed on the participants. For T2D case control definition, each study had at least one or more of the following criteria: (1) physician diagnosis and/or on diabetes medication, (2) fasting glucose $\geq$ 126mg/dL, (3) 2h fasting glucose $\geq$ 200mg/dL, (4) random glucose $\geq$ 200mg/dL, or (5) HbA1c $\geq$ 6.5%. Majority of the studies imposed a lower age bound for age at diagnosis, or included a GAD antibodies test. However, as T2D case definitions across cohorts differ, it is possible that cases of type 1 diabetes and maturity onset diabetes of the young (MODY) are included in this study, but these are unlikely to be correlated with genotype or have a significant effect on our analysis. For each study, the case control ascertainment are described in Supplementary Table 1 and summary statistics are provided in Supplementary Table 2.

Ethics oversight

Institutional review boards and/or ethics committees within each of the originating study sites approved the individual study protocols. These include: Kyoto University Graduate School of Medicine, The University of Tokyo, the BBJ Project, The J-MICC Study, the JPHC Study, IMM, ToMMo, RIKEN Yokohama Institute, The Hiroshima Atomic Bomb Casualty Council Health Management Center, The National Center for Global Health and Medicine, Keio University, Hiranuma Clinic, St Marianna University School of Medicine, Toyama University Hospital, Beijing Tongren Hospital, International Medical Center of Japan, Nagoya University School of Medicine, University of North Carolina at Chapel Hill, Chinese National Human Genome Center at Shanghai, Institute of Nutrition and Food Safety at the China Centers for Disease Control, Chinese Center for Disease Control and Prevention, University of Oxford, Singapore General Hospital Ethics Committee, Singapore Eye Research Institute Ethics Committee, National University of Singapore Institutional Review Board, Chinese University of Hong Kong Clinical Research Ethics Committee, Institutional review board at the Korea National Institute of Health, Republic of Korea, Columbia University, Johns Hopkins University, Northwestern University, University of California-Los Angeles, University of Minnesota, Wake Forest University, Vanderbilt University Medical Center, the University of Southern California, the University of Minnesota, the University of Pittsburgh, Biomedical Research Institute at Seoul National University Hospital, Samsung Hospital, The Lundquist Institute (formerly Los Angeles Biomedical Research Institute), China Medical University Hospital, Chia-Yi Christian Hospital, and National Taiwan University Hospital.

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