Identification of 28 new susceptibility loci for type 2 diabetes in the Japanese population

Ken Suzuki1,2,3,4, Masato Akiyama2,5, Kazuyoshi Ishigaki2, Masahiro Kanai6,2, Jun Hosoe1, Nobuhiro Shojima1, Atsushi Hozawa7,8, Aya Kadota9,10, Kiyonori Kuriki11, Mariko Naito12,13, Kozo Tanno14,15, Yasushi Ishigaki16,17, Makoto Hirata18, Koichi Matsuda19, Nakao Iwata20, Masashi Ikeda20, Norie Sawada21, Taiki Yamaji21, Motoki Iwasaki21, Shiro Ikegawa22, Shiro Maeda23,24, Yoshinori Murakami25, Kenji Wakai12, Shoichi Tsugane26, Makoto Sasaki27,28, Masayuki Yamamoto7,8, Yukinori Okada2,29, Michiaki Kubo30, Yoichiro Kamatani1,2,31,34*, Momoko Horikoshi3,34*, Toshimasa Yamauchi1* and Takashi Kadowaki1,32,33*

To understand the genetics of type 2 diabetes in people of Japanese ancestry, we conducted a meta-analysis of four genome-wide association studies (GWAS; 36,614 cases and 155,150 controls of Japanese ancestry). We identified 88 single-ancestry GWAS for type 2 diabetes in the Japanese population4–6. In the most recent study, we conducted a meta-analysis of two GWAS in Japanese subjects (a total of 15,463 cases and 26,183 controls in stage 1, GWAS1 and GWAS2) and reported seven novel type 2 diabetes loci, indicating the utility of single-ancestry GWAS in non-Europeans7. To discover further associations and gain insights into the etiology of type 2 diabetes, we performed two new GWAS for type 2 diabetes in Japanese subjects (GWAS3 and GWAS4), using 1000 Genomes project phase 3 (1KG Phase 3; May 2013) as a reference.

Type 2 diabetes is a metabolic disease caused by multifactorial pathogenesis. To date, >150 susceptibility loci for type 2 diabetes have been identified and reported, mainly from studies of people of European ancestry. However, type 2 diabetes epidemiology is different in people with Japanese versus European ancestry; Japanese people are more prone to type 2 diabetes than Europeans with the same body mass index or waist circumference. We previously performed GWAS for type 2 diabetes in the Japanese population8–10. In the most recent study, we conducted a meta-analysis of two GWAS in Japanese subjects (a total of 15,463 cases and 26,183 controls in stage 1, GWAS1 and GWAS2) and reported seven novel type 2 diabetes loci, indicating the utility of single-ancestry GWAS in non-Europeans7.

To discover further associations and gain insights into the etiology of type 2 diabetes, we performed two new GWAS for type 2 diabetes in Japanese subjects (GWAS3 and GWAS4), using 1000 Genomes project phase 3 (1KG Phase 3; May 2013) as a reference.

*Department of Diabetes and Metabolic Diseases, Graduate School of Medicine, The University of Tokyo, Tokyo, Japan. 1Laboratory for Statistical Analysis, RIKEN Center for Integrative Medical Sciences, Yokohama, Japan. 2Laboratory for Endocrinology, Metabolism and Kidney Diseases, RIKEN Center for Integrative Medical Sciences, Yokohama, Japan. 3Department of Statistical Genetics, Osaka University Graduate School of Medicine, Osaka, Japan. 4Department of Ophthalmology, Graduate School of Medical Sciences, Kyushu University, Fukuoka, Japan. 5Department of Biomedical Informatics, Harvard Medical School, Boston, MA, USA. 6Tohoku Medical Megabank Organization, Sendai, Japan. 7Graduate School of Medicine, Tohoku University, Sendai, Japan. 8Center for Epidemiologic Research in Asia, Shiga University of Medical Science, Otsu, Japan. 9Department of Epidemiology, RIKEN Center for Integrative Medical Sciences, Yokohama, Japan. 10Department of Epidemiology, RIKEN Center for Integrative Medical Sciences, Yokohama, Japan. 11Graduate School of Medical Sciences, Kyushu University, Fukuoka, Japan. 12Department of Public Health, Shiga University of Medical Science, Otsu, Japan. 13Laboratory of Public Health, School of Food and Nutritional Sciences, University of Shizuoka, Shizuoka, Japan. 14Department of Preventive Medicine, Nagoya University Graduate School of Medicine, Nagoya, Japan. 15Department of Oral Epidemiology, Graduate School of Biomedical and Health Sciences, Hiroshima University, Hiroshima, Japan. 16Division of Clinical Research and Epidemiology, Iwate Tohoku Medical Megabank Organization, Disaster Reconstruction Center, Iwate Medical University, Iwate, Japan. 17Department of Hygiene and Preventive Medicine, School of Medicine, Iwate Medical University, Iwate, Japan. 18Division of Innovation & Education, Iwate Tohoku Medical Megabank Organization, Disaster Reconstruction Center, Iwate Medical University, Iwate, Japan. 19Division of Diabetes, Metabolism and Endocrinology, Department of Internal Medicine, Iwate Medical University, Iwate, Japan. 20Institute of Medical Science, The University of Tokyo, Tokyo, Japan. 21Graduate School of Frontier Sciences, The University of Tokyo, Tokyo, Japan. 22Department of Psychiatry, Fujita Health University School of Medicine, Aichi, Japan. 23Division of Epidemiology, Center for Public Health Sciences, National Cancer Center, Tokyo, Japan. 24Laboratory for Bone and Joint Diseases, RIKEN Center for Integrative Medical Sciences, Tokyo, Japan. 25Division of Advanced Genomic and Laboratory Medicine, Graduate School of Medicine, University of the Ryukyus, Nishihara, Japan. 26Division of Clinical Laboratory and Blood Transfusion, University of the Ryukyus Hospital, Nishihara, Japan. 27Division of Molecular Pathology, Institute of Medical Science, The University of Tokyo, Tokyo, Japan. 28Center for Public Health Sciences, National Cancer Center, Tokyo, Japan. 29Iwate Tohoku Medical Megabank Organization, Disaster Reconstruction Center, Iwate Medical University, Iwate, Japan. 30Division of Ultrahigh Field MRI, Institute for Biomedical Sciences, Iwate Medical University, Iwate, Japan. 31Laboratory of Statistical Immunology, Immunology Frontier Research Center (WPI-IFReC), Osaka University, Osaka, Japan. 32RIKEN Center for Integrative Medical Sciences, Yokohama, Japan. 33Kyoto-McGill International Collaborative School in Genomic Medicine, Kyoto University Graduate School of Medicine, Kyoto, Japan. 34Department of Prevention of Diabetes and Lifestyle-Related Diseases, Graduate School of Medicine, The University of Tokyo, Tokyo, Japan. 35Department of Metabolism and Nutrition, Mizonokuchi Hospital, Faculty of Medicine, Teikyo University, Tokyo, Japan. 36These authors contributed equally: M. Horikoshi, Y. Kamatani.

*e-mail: yoichiro.kamatani@riken.jp; momoko.horikoshi@riken.jp; tyamau-tky@umin.net; kadowaki-3im@h.u-tokyo.ac.jp
reference panel for imputation. In addition, we reanalyzed GWAS1 and GWAS2, updating the reference panel for imputation from 1KG Phase 1 to 1KG Phase 3. Subsequently, we conducted fixed-effects meta-analysis combining each of the four GWAS (a total of 36,614 cases and 155,150 controls; Supplementary Figs. 1 and 2, Supplementary Table 1 and Methods). There was no overlap of individuals in the four studies. Although the genomic control inflation factor showed inflation ($\hat{\lambda}_{GC} = 1.21$, Supplementary Fig. 3), linkage disequilibrium (LD)-score regression$^\dagger$ indicated that the inflation was primarily due to polygenic effects (81.4%) rather than biases (estimated mean $\chi^2 = 1.50$ and LD-score intercept $= 1.09$). Because the bias was not large as compared with reported GWAS, we did not apply genomic control correction.

The type 2 diabetes association results at 12,557,761 variants are shown in Fig. 1. Of the 163 previously reported loci$^{11,12}$, lead variants at 139 loci were analyzed, and 95 (68%) of these were replicated with nominal significance ($P < 0.05$ with a consistent direction of effect; Supplementary Table 2).

We detected genome-wide-significant ($P < 5.0 \times 10^{-8}$) (ref. 13) associations at 88 loci, of which 28 were novel ($\geq 1$ Mb away from and not in LD ($r^2 < 0.05$) with reported loci; Table 1 and Supplementary Table 3). When we applied study-level genomic control correction, we identified 75 loci, 23 of which were previously unreported. Had we used a genome-wide-significance threshold that reflected the functional effect of each variant$^{14}$, we would have identified 75 loci, 21 of which were not previously reported (Methods). The novel loci included genes implicated in diabetes and related phenotypes. $ITGA1$ harbors recessively inherited diabetes risk variants$^{15}$, and KSR2 mutation is associated with obesity and insulin resistance$^{16}$.

Of the 88 lead variants, 68 (77%) were common (MAF $\geq 0.05$) in the Japanese (JPN) samples of this study and European (EUR) samples in 1KG Phase 3 (Supplementary Table 3 and Supplementary Fig. 4). When we compared the effect sizes of 69 of the 88 lead variants in Japanese and Europeans that were available in a published European GWAS$^2$ (Supplementary Table 3 and Supplementary Fig. 5), we found a strong positive correlation (Pearson’s $r = 0.87$, $P = 1.4 \times 10^{-13}$) and directional consistency (65 of 69 loci, 94%, sign-test $P = 3.1 \times 10^{-11}$). In addition, when we compared the effect sizes of the 95 of 113 lead variants reported in the European type 2 diabetes GWAS$^3$ that were available in both Japanese and European type 2 diabetes GWAS (Supplementary Table 2 and Supplementary Fig. 6a), we also found a strong positive correlation (Pearson’s $r = 0.74$, $P = 5.9 \times 10^{-14}$) and directional consistency (83 of 95 loci, 87%, sign-test $P = 3.2 \times 10^{-14}$). After this manuscript was submitted, a larger type 2 diabetes GWAS of European ancestry was published$^{17}$. When we repeated the comparison at the lead variants reported in this larger European GWAS, we found a more prominent correlation (Pearson’s $r = 0.83$, $P = 8.7 \times 10^{-13}$) and directional consistency (181 of 192 loci, 94%, sign-test $P = 8.3 \times 10^{-5}$) of the effect sizes (Supplementary Table 4 and Supplementary Fig. 6b). These results indicate that most of the type 2 diabetes susceptibility loci identified in the Japanese or European population had comparable effects on type 2 diabetes in the other population.

We performed stepwise conditional analysis to detect multiple independent signals at a locus. We detected 27 additional signals that reached locus-wide significance ($P < 5.0 \times 10^{-8}$), increasing the total number of signals to 115 (Supplementary Fig. 7 and Supplementary Table 5). Of the 115 independent signals, only six combinations were in LD ($r^2 \geq 0.05$) with each other (rs577402029 and rs463924 in INS–IGF2–KCNQ1; rs2237897 and rs233449 in INS–IGF2–KCNQ1; rs8037894 and rs77820034 in C2CDA4; rs12633613 and rs112332300 in UBE2E2; rs33981001 and rs62508166 in ANK1; and rs10965247 and rs10757282 in CDKN2A and CDKN2B; 0.05 $\leq r^2 < 0.44$). The loci that had the largest number of independent associations included (previously annotated) INS–IGF2 and KCNQ1 loci, which had two and six independent associations, respectively (0.005 $< $ MAF $< 0.38$, 1.06 $< $ odds ratio $< 1.56$).

To generate hypotheses for causal variants and genes mediating the effect at each signal, we searched for missense variants in LD with the identified type 2 diabetes signals. We identified 28 missense variants in 21 genes that were in LD with any of the type 2 diabetes lead variants ($r^2 > 0.6$ in East Asians (EAS) of 1KG phase 3) (Table 2 and Supplementary Table 6). Of these 28, six missense variants coincided with the type 2 diabetes lead variants. At the remaining 22 missense variants, the type 2 diabetes signal association disappeared ($P_{\text{conditional}} > 0.05$) when conditioned on the missense variant, except for p.Pro443Thr in SLC16A11 ($P_{\text{conditional}} = 0.00011$) and p.Gly836Ser in SND1 ($P_{\text{conditional}} = 0.00052$).

Of the 28 missense variants identified, six variants were in novel loci (Table 2). Of these, p.Ala122Pro in SCTR was polymorphic only in East Asians among the populations of the Exome Aggregation Consortium (ExAC) and also only in Japanese individuals among the 1KG Phase 3 populations (Supplementary Table 6). Three of the six variants in novel loci were rare (MAF $< 0.01$) or monomorphic in Europeans but common (MAF $\geq 0.05$) in Japanese individuals (p.Val282Met in GP2) or low frequency (0.01 $\leq $ MAF $< 0.05$) in Japanese (p.Ala122Pro in SCTR and p.Cys517Arg in ZNF257). Of the 28 missense variants identified, nine variants were previously
unreported variants within a 1-Mb distance from reported loci (Table 2). Of these, p.Ala341Thr in CPA1 was polymorphic only in East Asians in ExAC and also only in Japanese individuals in 1KG Phase 3 (Supplementary Table 6). Six of the nine variants were rare (MAF < 0.01) or monomorphic in Europeans but common (MAF ≥ 0.05) in Japanese (p.Ala341Thr in GLP1R and p.Ala414Gly in IRF2BP1) or low-frequency (0.01 ≤ MAF < 0.05) in Japanese individuals (p.Leu225Ser in GRB14, p.Arg192Ser in PAX4, p.Gly836Ser in SN1D1 and p.Ala341Thr in CPA1). Of these 15 previously unreported missense variants, p.Val282Met in GP2, p.Ala341Thr in CPA1 and p.Arg131Gln in GLP1R are notable examples, given their different MAF spectra in Japanese and European individuals and their biological implications for type 2 diabetes.

CPA1 and GP2 have been implicated in the differentiation of exocrine pancreatic acinar cells, which secrete digestive enzymes and have plasticity to convert into insulin-producing beta cells. These genes were specifically expressed in the pancreas in the Genotype Tissue Expression (GTEx) database (Supplementary Fig. 8) and are specifically expressed in acinar cells.

### Table 1 | Association statistics for lead variants in each of the 28 novel type 2 diabetes loci

<table>
<thead>
<tr>
<th>Locus</th>
<th>rsID</th>
<th>Chr</th>
<th>Pos</th>
<th>RA</th>
<th>OA</th>
<th>RAF</th>
<th>OR</th>
<th>95% CI</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>USP4B</td>
<td>rs1825307</td>
<td>1</td>
<td>22068326</td>
<td>A</td>
<td>G</td>
<td>0.25</td>
<td>1.06</td>
<td>1.04–1.09</td>
<td>1.63×10⁻⁹</td>
</tr>
<tr>
<td>FANCL and BCL1A</td>
<td>rs13417036</td>
<td>2</td>
<td>58921777</td>
<td>A</td>
<td>G</td>
<td>0.16</td>
<td>1.07</td>
<td>1.05–1.10</td>
<td>1.67×10⁻⁸</td>
</tr>
<tr>
<td>SCTR</td>
<td>rs3731600</td>
<td>2</td>
<td>120231070</td>
<td>C</td>
<td>G</td>
<td>0.95</td>
<td>1.13</td>
<td>1.08–1.18</td>
<td>1.81×10⁻⁸</td>
</tr>
<tr>
<td>EPC2</td>
<td>rs16829174</td>
<td>2</td>
<td>149455385</td>
<td>A</td>
<td>G</td>
<td>0.59</td>
<td>1.05</td>
<td>1.03–1.07</td>
<td>5.09×10⁻⁹</td>
</tr>
<tr>
<td>DGDK</td>
<td>rs838720</td>
<td>2</td>
<td>234303281</td>
<td>G</td>
<td>C</td>
<td>0.45</td>
<td>1.05</td>
<td>1.04–1.07</td>
<td>1.91×10⁻⁹</td>
</tr>
<tr>
<td>CASR and CD68</td>
<td>rs2134223</td>
<td>3</td>
<td>121956953</td>
<td>A</td>
<td>G</td>
<td>0.47</td>
<td>1.05</td>
<td>1.03–1.07</td>
<td>5.20×10⁻⁹</td>
</tr>
<tr>
<td>GNPD22A and GABRG1</td>
<td>rs10938398</td>
<td>4</td>
<td>45186139</td>
<td>A</td>
<td>C</td>
<td>0.30</td>
<td>1.06</td>
<td>1.04–1.08</td>
<td>2.01×10⁻⁹</td>
</tr>
<tr>
<td>GR5F1 and MOB12</td>
<td>rs12649012</td>
<td>4</td>
<td>71748245</td>
<td>G</td>
<td>A</td>
<td>0.21</td>
<td>1.07</td>
<td>1.04–1.09</td>
<td>3.43×10⁻⁹</td>
</tr>
<tr>
<td>AGPAT9 and NKX6-1</td>
<td>rs201597274</td>
<td>4</td>
<td>85301870</td>
<td>T</td>
<td>C</td>
<td>0.96</td>
<td>1.17</td>
<td>1.12–1.23</td>
<td>2.07×10⁻¹²</td>
</tr>
<tr>
<td>PARP6 and EMB</td>
<td>rs16884025</td>
<td>5</td>
<td>49953459</td>
<td>C</td>
<td>T</td>
<td>0.93</td>
<td>1.14</td>
<td>1.09–1.19</td>
<td>8.48×10⁻⁹</td>
</tr>
<tr>
<td>ITGA1</td>
<td>rs11410487</td>
<td>5</td>
<td>52094244</td>
<td>A</td>
<td>T</td>
<td>0.17</td>
<td>1.07</td>
<td>1.04–1.09</td>
<td>1.75×10⁻⁸</td>
</tr>
<tr>
<td>FGFR4 and ZNF346</td>
<td>rs3135911</td>
<td>5</td>
<td>176513896</td>
<td>A</td>
<td>C</td>
<td>0.44</td>
<td>1.06</td>
<td>1.04–1.08</td>
<td>3.50×10⁻¹⁰</td>
</tr>
<tr>
<td>NUS1</td>
<td>rs80169932</td>
<td>6</td>
<td>117996631</td>
<td>T</td>
<td>C</td>
<td>0.75</td>
<td>1.07</td>
<td>1.05–1.10</td>
<td>8.37×10⁻¹²</td>
</tr>
<tr>
<td>ETV1 and ARL4A</td>
<td>rs56805921</td>
<td>7</td>
<td>13888699</td>
<td>C</td>
<td>G</td>
<td>0.43</td>
<td>1.06</td>
<td>1.04–1.08</td>
<td>2.51×10⁻⁹</td>
</tr>
<tr>
<td>AUTS2</td>
<td>rs6947395</td>
<td>7</td>
<td>69406661</td>
<td>T</td>
<td>A</td>
<td>0.20</td>
<td>1.09</td>
<td>1.07–1.12</td>
<td>4.87×10⁻¹⁰</td>
</tr>
<tr>
<td>C10orf11 and ZNF503</td>
<td>rs7900112</td>
<td>10</td>
<td>77314617</td>
<td>T</td>
<td>A</td>
<td>0.51</td>
<td>1.05</td>
<td>1.04–1.07</td>
<td>2.10×10⁻⁹</td>
</tr>
<tr>
<td>WDR11 and FGFR2</td>
<td>rs10886683</td>
<td>10</td>
<td>122992493</td>
<td>C</td>
<td>T</td>
<td>0.70</td>
<td>1.06</td>
<td>1.04–1.08</td>
<td>3.18×10⁻⁹</td>
</tr>
<tr>
<td>SLC1A2</td>
<td>rs2421897</td>
<td>11</td>
<td>35437863</td>
<td>G</td>
<td>C</td>
<td>0.88</td>
<td>1.08</td>
<td>1.05–1.11</td>
<td>6.58×10⁻¹⁰</td>
</tr>
<tr>
<td>ETS1</td>
<td>rs11819995</td>
<td>11</td>
<td>128389391</td>
<td>T</td>
<td>C</td>
<td>0.22</td>
<td>1.06</td>
<td>1.04–1.08</td>
<td>9.76×10⁻⁹</td>
</tr>
<tr>
<td>KSR2</td>
<td>rs79310463</td>
<td>12</td>
<td>118406696</td>
<td>T</td>
<td>C</td>
<td>0.26</td>
<td>1.06</td>
<td>1.04–1.08</td>
<td>1.83×10⁻⁸</td>
</tr>
<tr>
<td>DLEU7 and KCNNG</td>
<td>rs123378</td>
<td>13</td>
<td>51088809</td>
<td>G</td>
<td>A</td>
<td>0.22</td>
<td>1.06</td>
<td>1.04–1.09</td>
<td>4.93×10⁻⁹</td>
</tr>
<tr>
<td>GP2</td>
<td>rs117267808</td>
<td>16</td>
<td>20332368</td>
<td>A</td>
<td>G</td>
<td>0.08</td>
<td>1.12</td>
<td>1.08–1.16</td>
<td>2.99×10⁻¹²</td>
</tr>
<tr>
<td>ZNF257</td>
<td>rs148316037</td>
<td>19</td>
<td>22257558</td>
<td>T</td>
<td>C</td>
<td>0.97</td>
<td>1.26</td>
<td>1.17–1.34</td>
<td>5.10×10⁻⁹</td>
</tr>
<tr>
<td>NFATC2</td>
<td>rs6021276</td>
<td>20</td>
<td>50153386</td>
<td>T</td>
<td>C</td>
<td>0.43</td>
<td>1.06</td>
<td>1.04–1.07</td>
<td>1.84×10⁻⁹</td>
</tr>
<tr>
<td>CNKSR2</td>
<td>rs6633421</td>
<td>X</td>
<td>21569920</td>
<td>A</td>
<td>G</td>
<td>0.68</td>
<td>1.04</td>
<td>1.03–1.06</td>
<td>2.62×10⁻⁸</td>
</tr>
<tr>
<td>TIMM17B and PCSK1</td>
<td>rs782100977</td>
<td>X</td>
<td>48724648</td>
<td>C</td>
<td>CAA</td>
<td>0.63</td>
<td>1.06</td>
<td>1.04–1.08</td>
<td>4.58×10⁻⁹</td>
</tr>
<tr>
<td>SPIN2A and FAAH2</td>
<td>rs144226500</td>
<td>X</td>
<td>57170781</td>
<td>AT</td>
<td>A</td>
<td>0.30</td>
<td>1.05</td>
<td>1.03–1.06</td>
<td>4.58×10⁻⁹</td>
</tr>
<tr>
<td>IL13RA1</td>
<td>rs11390176</td>
<td>X</td>
<td>117915163</td>
<td>T</td>
<td>TA</td>
<td>0.44</td>
<td>1.06</td>
<td>1.04–1.07</td>
<td>9.82×10⁻¹⁵</td>
</tr>
</tbody>
</table>

Summary statistics of loci that reached genome-wide significance in the meta-analysis (n = 191,764 independent samples). P values (two sided) were derived from meta-analysis with the inverse-variance method under a fixed-effect model. P values were not adjusted for multiple comparisons. Positions (Pos) are based on Human Genome version 19 (hg19), build 37. RA, risk allele; OA, other allele; RAF, risk-allele frequency; OR, odds ratio; CI, confidence interval.

GLP1R encodes a receptor for glucagon-like peptide 1 (GLP1), which induces glucose-dependent insulin secretion. GLP1R agonists are commonly used as drugs for type 2 diabetes. p.Arg131Gln in GLP1R (rs3765467) was in LD with the lead variant rs9394574 (r² = 0.97 in East Asians; Fig. 2a), which resides 245kb away from the neighboring KCNK16 locus (rs1535500, Pjen = 9.56×10⁻¹⁰). rs9394574 was not in LD with rs1535500 (r² < 0.01 in East Asians and Europeans), and conditioning on rs1535500 confirmed independent association of p.Ala341Thr (P conditional = 1.77×10⁻⁸).
exome-wide-significant association of p.Ala316Thr with type 2 diabetes ($P_{\text{GWAS}} = 0.57$ and $P_{\text{ESS}} = 1.59 \times 10^{-11}$, Fig. 2b). Here, we provide what is, to our knowledge, the first report genome-wide significance for GLP1R ($rs3765467$, p.Arg131Gln, $P_{\text{IN}} = 6.10 \times 10^{-14}$; Fig. 2a). p.Arg131Gln was not in LD with p.Ala316Thr ($r^2 = 0$ in Europeans and not defined in East Asians). Whereas p.Arg131Gln was common (MAF$_{\text{JPN}} = 0.18$) in Japanese samples of the present study but rare (MAF$_{\text{JPN}} < 0.01$) in the European samples of 1K Phase 3, p.Ala316Thr was rare in both populations. According to a structure analysis of crystallized GLP1R$_{\text{C}}$, Arg131 is in a highly flexible region (Glu127–Ser136) that links the extracellular domain and transmembrane domain (Fig. 2c,d); Glu127–Lys130, Arg131 and Gly132–Ser136 are colored yellow, red and orange, respectively). In the analysis, Glu127–Arg131 interacts with Glu207–Trp214 in extracellular loop 1 (ECL1; Glu207–Trp214 are shown in pink in Fig. 2c,d). Because arginine has electrically charged side chains, but glutamine has uncharged side chains, Arg131Gln might influence the interactions between Glu127 and Arg131 and between Glu207 and Trp214. The type 2 diabetes–protective allele of p.Arg131Glu substitutes glutamine for arginine and is associated with a $>$100% increase in GLP1-induced insulin secretion$_{10}$. Given the high MAF of p.Arg131Gln in Japanese and East Asians (MAF$_{\text{JPN}} = 0.23$), p.Arg131Gln might be a potential marker for clinical response to GLP1R agonists in Japanese and East Asians. Notably, a study
comparing the change in postprandial plasma glucose area under the curve after single dose injection of lixisenatide in individuals of Japanese and European ancestry has revealed a greater decrease in Japanese subjects with type 2 diabetes.

Next, we searched for lead cis expression quantitative trait locus (cis-eQTL) variants in LD with type 2 diabetes signals by using data from GTEx. The strongest cis-eQTL variants, within a tissue, of 59 transcripts were in LD ($r^2 > 0.6$ in East Asians or Europeans) with type 2 diabetes signals; of these, 16 transcripts including NUS1 were in novel loci (Fig. 3 and Supplementary Table 7).

In the NUS1 locus, the lead type 2 diabetes variant (rs80196932) coincided with the lead cis-eQTL variant of NUS1 in the pancreas and was in LD with those in skeletal muscle and stomach ($r^2 = 1.0$ in East Asians and 0.95 in Europeans, respectively). rs80196932 was common in Japanese and European individuals (MAF$_{JAPANESE} = 0.25$ and MAF$_{EUROPEAN} = 0.20$). In the European type 2 diabetes GWAS, rs80196932 showed nominal association ($P = 9.4 \times 10^{-4}$) and a consistent direction of effect. It is in an evolutionarily conserved region and the 5′ untranslated region of NUS1, and it has been found to overlap with DNase I–hypersensitive sites and active transcription start sites in various tissues including the pancreas. NUS1 is essential for protein glycosylation and intracellular cholesterol trafficking, and it is annotated with ‘congenital disorder of glycosylation’.

We used stratified LD-score regression to quantify the enrichment in heritability in various cell types and tissues in our GWAS for type 2 diabetes. We first evaluated enrichment in ten cell-type groups that were defined by the developers of stratified LD-score.
Fig. 3 | Overlap between type 2 diabetes signals and the lead cis-eQTL variants of the GTEx database. Heat-map representation of the lead eQTL variants that were in LD with type 2 diabetes signals ($r^2 > 0.6$ in East Asians or Europeans). Normalized beta values of lead eQTL alleles that were in phase with type 2 diabetes risk allele are shown. Positive values are in red, and negative values are in blue. Rows show target tissues of eQTL variants. Each column shows the chromosome and base position of type 2 diabetes signals and transcripts whose lead cis-eQTL variants were in LD with the type 2 diabetes signals.

regression as unions of 220 cell-type-specific annotations, reflecting their system- or organ-level structure. We found significant enrichment in adrenal and pancreas annotations ($P=3.9\times10^{-4}$) but not those of other cell-type groups after Bonferroni correction (Supplementary Table 8 and Supplementary Fig. 9a; Methods). Subsequently, we conducted 220 cell-type analyses and observed the most significant enrichment for acetylated histone H3 Lys27 (H3K27ac) sites in pancreatic islets ($P=2.6\times10^{-4}$) (Supplementary Table 9 and Supplementary Fig. 9b; Methods). H3K27ac is a characteristic feature of active enhancers. The heritability of type 2 diabetes was also significantly enriched in H3K4me3 in pancreatic islets, melanocytes, the right ventricle and the fetal brain after Bonferroni correction ($P<2.3\times10^{-4}=0.05/220$). H3K4me3 is a characteristic feature of active promoters.

The functional annotations of the 115 type 2 diabetes signals (missense variants, eQTL and promoter and enhancer marks) are summarized in Supplementary Table 10. Descriptions of the nearest genes in novel loci, genes with newly identified missense variants and eQTL transcripts in novel loci are listed in Supplementary Table 11.

To explore the shared genetics between type 2 diabetes and various traits, we calculated the genetic correlations between type 2 diabetes and 91 complex human traits (32 diseases and 59 quantitative traits) in Japanese subjects by using bivariate LD-score regression (Supplementary Fig. 10 and Supplementary Table 12). As expected, we replicated a positive genetic correlation between type 2 diabetes and several traits related to obesity, cardiovascular disease and hyperlipidemia. In addition, we found previously unreported positive genetic correlation (false discovery rate (FDR) $q<0.01$) with ossification of the posterior longitudinal ligament ($r=0.26$, $P=4.1\times10^{-9}$) and white-blood-cell count ($r=0.17$, $P=4.6\times10^{-5}$). Epidemiological studies have reported positive phenotypic correlations between type 2 diabetes and ossification of the posterior longitudinal ligament or white-blood-cell count, in agreement with the observations from our genetic analysis.

To gain biological insights, we performed a transethnic comparison of the molecular pathways suggested by the type 2 diabetes GWAS of Japanese and Europeans. Of the 1,077 pathways, 17 and 13 pathways were significantly associated with type 2 diabetes in Japanese and European individuals, respectively (FDR $q<0.05$) (Table 3 and Supplementary Tables 13 and 14). We found that the pathway of mature-onset diabetes in young individuals was most significantly associated with type 2 diabetes in both populations (FDR $q_{JPN}=6.3\times10^{-8}$ and FDR $q_{EUR}=1.2\times10^{-4}$). Mature-onset diabetes of the young is monogenic diabetes characterized by impaired beta cell function. Pathways of beta cells, development, prostate cancer and G1 phase were also significantly associated with type 2 diabetes in both populations. However, several pathways, including NOTCH signaling and insulin secretion, showed stronger associations in one of the two populations than in the other population (12 and 8 pathways with FDR $q<0.05$ in only one of the populations). Because the difference in the power of the GWAS may lead to different associations of pathways with type 2 diabetes in Japanese and European individuals, these pathways should be assessed in larger GWAS in the future.

In summary, we performed expanded meta-analysis of type 2 diabetes GWAS in Japanese subjects and identified 115 signals across 88 type 2 diabetes loci, of which 28 loci with 30 signals were
novel. Most of the lead variants across 88 type 2 diabetes loci were rare or monomorphic in Europeans. Transethnic molecular pathway analysis elucidated ethnically shared or heterogeneous effects of pathways on type 2 diabetes, highlighting the role of beta cell dysfunction in type 2 diabetes in Japanese and Europeans.

Together, these findings provide insights into the etiology of type 2 diabetes in Japanese and Europeans.

Note added in proof: While our paper was under revision, two additional reports\textsuperscript{45,46} were published that mapped T2D associations to regions reported here at genome-wide significance.

### URLs
- 1000 Genomes Project, http://www.1000genomes.org/
- CADD, https://cadd.gs.washington.edu/
- Chimera, https://www.cgl.ucsf.edu/chimera/
- DIAGRAM, http://www.diagramp consortium.org/
- GPCRdb, http://gpcrdb.org/
- GTEx, http://www.gtexportal.org/home/
- HaploReg, https://pubs.broadinstitute.org/mammals/haploreg/haploreg.pl
- mach2dat, https://genome.sph.umich.edu/wiki/Mach2dat_A nalysis_with_MACH_output/
- METAL, https://genome.sph.umich.edu/wiki/METAL/
- PLINK, https://www.cog-genomics.org/plink2/;
- Polyphen2, http://genetics.bwh.harvard.edu/pph2/;
- R, https://www.r-project.org/;
- SHAPEIT, https://mathgen.stats.ox.ac.uk/software/shapeit/shapeit.html;
- SIFT, http://sift.bii.a-star.edu.sg/;
- ToMMo, http://www.megabank.tohoku.ac.jp/english/;

### Online content
Any methods, additional references, Nature Research reporting summaries, source data, statements of data availability and associated accession codes are available at https://doi.org/10.1038/s41588-018-0332-4.

Received: 14 June 2018; Accepted: 10 December 2018;
Published online: 04 February 2019

### References

Acknowledgements
We acknowledge the staff of the BBJ Project, IMM, ToMMo, the JPHC Study and the J-MICC Study for their outstanding assistance in collecting samples and clinical information. We also acknowledge the members of the Genetic Study Group of the Investigation Committee on the Ossification of Spinal Ligaments for recruiting subjects to this study, which was supported by the Japan Agency for Medical Research and Development (AMED (grants JP17ek0109221n0011) (S.I.). The study of psychiatric disorders was supported by AMED under grants JP18hm0107907 (N.I., M. Ikeda and K.Y.), JP18km0405201 (N.I.) and JP18km0405208 (M. Ikeda). We are grateful to members of The Rotary Club of Osaka-Midousui District 2660 Rotary International in Japan for supporting our study. This research was funded by the Tailor-Made Medical Treatment Program (the BMS Project) of the Ministry of Health, Labour and Welfare of Japan from 1989 to 2010 (S.T.). The study of psychiatric disorders was supported by the Japan Agency for Medical Research and Development (AMED) (grants JP18km0105003 and JP18km0105004, M.S.) ToMMo is supported by MEET and AMED (grants JP18km0405202 and JP18km0405001, M.Y.). The JPIC study has been supported by the National Cancer Center Research and Development Fund since 2011 and was supported by a Grant-in-Aid for Cancer Research from the Ministry of Health, Labour and Welfare of Japan from 1989 to 2010 (S.T.). The J-MICC Study was supported by Grants-in-Aid for Scientific Research for Priority Areas of Cancer (17015018) and Innovative Areas (22100001) and the JSPS KAKENHI Grant (16H06277) from MEET (A.K., K.K., M.N. and K.W.).

Author contributions
K.S., M.A., M. Horikoshi and Y.K. designed the study and wrote the manuscript. K.S., M.A., M. Horikoshi, K.I. and M. Kanai performed statistical analysis. J.H., N. Shojima, A.H., A.K., K.K., M.N. and K.W. contributed to data acquisition. M. Horikoshi, K.I. and M. Kanai performed statistical analysis. J.H., N. Shojima, A.H., A.K., K.K., M.N. and K.W. contributed to data acquisition. M. Horikoshi, K.Y., M. Kubo, T. Yamauuchi and T.K. supervised the study. All authors contributed to and approved the final version of the manuscript.

Competing interests
The authors declare no competing interests.

Additional information
Supplementary information is available for this paper at https://doi.org/10.1038/s41588-018-0332-4.
Reprints and permissions information is available at www.nature.com/reprints.
Correspondence and requests for materials should be addressed to Y.K., M.H., T.Y. or T.K.
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 America, Inc. 2019.
Methods

Study participants. In GWAS1 and GWAS2, we used case and control individuals registered in the Biobank Japan (BBJ) project,35 Osaka-Midousuji Rotary Club and Pharma SNP consortium, as previously analyzed and reported.48,49 Forty-eight samples in GWAS2 were removed to withdraw consent of other reasons (Life Sciences Reporting Summary). For the case samples in GWAS3 and GWAS4, we selected participants 40 years of age or older from the BBJ project who had type 2 diabetes but did not have type 1 diabetes, mitochondriod diabetes, mature-onset diabetes of the young or any types of diabetes other than type 2 diabetes. The subtype of diabetes was extracted from medical records in which diagnosis was made by physicians at the participating hospitals. Definitions of subtypes of diabetes in Japan have been described elsewhere.11 Participants with concentrations of antibodies to glutamic acid decarboxylase >5 IU ml−1 were excluded from the type 2 diabetes cases. For the control samples in GWAS3 and GWAS4, we selected participants registered in the BBJ Project, who did not have diabetes but did not have type 1 diabetes or participants from the following population-based cohorts: Tohoku Medical Megabank Organization (ToMMo),40 Iwate Tohoku Medical Megabank Organization (IMM),41 the Japan Public Health Center–based Prospective (JPHC) Study42 and the Japan Multi-Institutional Collaborative Cohort (J-MICC) Study.43 All control individuals were 20 years of age or older. The characteristics of the cohorts are described in detail in the Supplementary Note and Life Sciences Reporting Summary. For GWAS3 and GWAS4 sample quality control, we removed samples with a call rate <0.98. We removed related individuals with PI_HAT>0.1. PI_HAT is an index of relatedness between two individuals based on identity by descent (IBD) defined in PLINK.44 PI_HAT is defined by the following equation: $\text{PI\_HAT} = \text{P(IBD = 2)} + 0.5\times \text{P(IBD = 1)}$. To identify population stratification, we conducted principal component analysis for genotype by using FastPCA. We excluded outliers from the East Asian cluster and divided samples into Hondo (GWAS3) and Ryukyu (GWAS4) clusters.45 All participating studies obtained informed consent from all participants by following the protocols approved by their institutional ethical committees before enrollment. The ethical committees at all collaborating facilities (The University of Tokyo, the BBJ Project, The J-MICC Study, the JPHC Study, IMM and ToMMo) approved this research project. All relevant ethical regulations were followed in this research project. There was no overlap of individuals in the four studies.

Genome-wide quality-control procedures. All samples in GWAS1, GWAS3 and GWAS4 were genotyped with Illumina HumanOmniExpressExome BeadChips or a combination of Illumina HumanOmniExpress and HumanExome BeadChips. GWAS2 samples were genotyped with the Illumina Human 610K SNP array. Before imputation, genotyped variants were excluded according to the following conditions. In GWAS1, GWAS3, GWAS2 and GWAS4, we excluded variants with (i) heterozygosity count <5, (ii) Hardy–Weinberg equilibrium $P$ (calculated in all samples for autosomes and in females for $X$ chromosome) $<1.0\times10^{-4}$ in each chip or (iii) concordance rate $<0.99$ with in-house whole-genome-sequence data using overlapping samples. All extracted variants had a call rate $>0.99$. In GWAS2, we also excluded variants with (i) call rate $<0.99$, (ii) MAF $<0.01$, (iii) differential missingness $P<1.0\times10^{-4}$ or (iv) Hardy–Weinberg equilibrium $P$ (calculated in all samples for autosomes and in females for $X$ chromosome) $<1.0\times10^{-4}$.

Prephasing and genotyping imputation. Prephasing of autosomes and $X$ chromosomes was performed with either EAGLE v2.3 (for GWAS1, GWAS3 and GWAS4) or SHAPEIT v2.r837 (for GWAS2) in males and females together. Prephased autosomes were imputed up to the 1KG Phase 3 reference panel (phase 3, May 2013 release, all samples) with minimacl3 (2.0.1). Similarly, imputation of $X$ chromosomes was performed with the 1KG Phase 3 reference panel and minimacl3, but in males and females separately. The pseudautosomal region (2.6–154.9 Mb) was excluded from the reference panel before imputation. Allele dosage data was imputed from 0 to 2 in males under the assumption of full dosage compensation.

SNP association analyses and meta-analyses. Association analysis of autosomes of GWAS1–GWAS4 was performed independently with mach2dat (1.0.24) with adjustment for age and sex and principied components, by assuming an additive model. After filtering of variants with minimacl3 imputation quality score $r^2>0.3$, the four GWAS were meta-analyzed with the inverse-variance method under a fixed-effect model with METAL. Variants that had MAF $>0.001$ and were reported in at least two of the four GWAS were selected after the meta-analysis. For $X$ chromosomes, we performed association analyses in males and females separately in each GWAS, using mach2dat with adjustment for age and ten principal components. We selected SNPs with minimacl3 imputation quality score $r^2>0.3$. We integrated the results of males and females in each GWAS through allele dosage imputation. Allele dosage information was matched to the reference panels used for imputation in the Japanese and European type 2 diabetes meta-analysis. We used SUM statistics and

Evaluation of confounding biases with LD-score regression. To estimate confounding biases derived from cryptic relatedness and population stratification, we conducted LD-score regression $r^2(1.0.0)$. We set the population and sample prevalence of type 2 diabetes as 0.075 and 0.191, respectively. The population prevalence of type 2 diabetes was estimated from the National Health and Nutrition Survey and Population Estimates of Japan (see URLs). We used LD scores for the East Asian population provided with the software.

Conditional association analyses. For the purpose of finding multiple type 2 diabetes–associated signals within 88 type 2 diabetes loci (with a 500-kb margin on both sides of the most up- and downstream genome-wide-significant variants at each locus), we conducted a stratified conditional association analysis. We performed conditional analyses of GWAS1–GWAS4 independently, conditioning on the index variants of the 88 type 2 diabetes loci in the meta-analysis. We then combined the results of conditional analyses of GWAS1–GWAS4 by using the fixed-effects inverse-variance method. We repeated this process until the index variants fell below the significance threshold of $5.0\times10^{-8}$, on the basis of the approximate average number of multiple tests in each locus. This threshold is more stringent than the threshold used in previous European type 2 diabetes GWAS ($1.0\times10^{-8}$).

Functional annotation and eQTL analyses. To characterize associated variants, we performed eQNOVAR on functionally annotated variants of the type 2 diabetes GWAS in East Asian 1KG Phase 3 with the 115 type 2 diabetes lead variants identified by conditional analysis. We annotated missense variants with SIFT, Polyphen2-HDIV and CADD scores, using ANNOVAR to assess their possible functional effects. Supplementary figures showing gene expression for CPA1 and GPP2 were downloaded from GTEx3 portal (release v7). The three-dimensional ribbon model of GLP1R, Fig. 2d was based on Protein Data Bank ID 5NX2, which was prepared with University of California San Francisco Chimera version 1.12 referring to Jazayeri et al.46 The snake plot of GLP1R in Fig. 2d was created with CPGCRdb revision 9656102 (see URLs). We searched for overlaps between the 115 type 2 diabetes lead variants and promoters and enhancer marks by using HaploReg v4.1, setting the source for epigenetic annotation as ChromHMM core 15-state model. HaploReg v4.1 was also used to inspect the lead eQTL variant of NUSI (rs80196932).

Genetic correlation. In our previous work48,49, bivariate LD-score regression was conducted to estimate genetic correlations among complex diseases and clinical laboratory measurements in the Japanese population. We updated the previous type 2 diabetes GWAS50 to the current type 2 diabetes GWAS (GWAS1–GWAS4) and tested for genetic correlation between type 2 diabetes and 91 human complex traits (32 diseases and 59 quantitative traits). We removed traits whose sample size was $<5000$. We included subjects with type 2 diabetes and dyslipidemia for the GWAS of bipolar disorder. East Asian LD score and summary statistics of SNPs provided by the software developer were used for the regression. SNPs in the major histocompatibility complex region (chromosome 6, 26–34 Mb) were excluded from the analysis. We computed FDR $q$ with the Benjamini–Hochberg method.

Linkage disequilibrium score (LDSC) partitioning heritability. To assess the enrichment of heritability in tissue and cell types for type 2 diabetes in the Japanese population, we used stratified LD-score regression10 of ten cell-type group annotations and 220 cell-type specific annotations provided by the authors. We used the East Asian LD score of SNPs provided with the software and the summary statistics of Japanese type 2 diabetes GWAS. Variants were filtered by removing SNPs with MAF $>0.3$, by adapting the LD structure of East Asians and Europeans. Custom reference genotype information was matched to the reference panels used for imputation in the Japanese and European type 2 diabetes meta-analysis. We used SUM statistics and
predefined pathway libraries from KEGG, REACTOME and BIOCARTA with default parameters.

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

**Data availability**


**References**

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.

Statistical parameters

When statistical analyses are reported, confirm that the following items are present in the relevant location (e.g. figure legend, table legend, main text, or Methods section).

<table>
<thead>
<tr>
<th>n/a</th>
<th>Confirmed</th>
</tr>
</thead>
<tbody>
<tr>
<td>❌</td>
<td>❌</td>
</tr>
<tr>
<td>❌</td>
<td>❌</td>
</tr>
<tr>
<td>❌</td>
<td>❌</td>
</tr>
<tr>
<td>❌</td>
<td>❌</td>
</tr>
<tr>
<td>❌</td>
<td>❌</td>
</tr>
<tr>
<td>❌</td>
<td>❌</td>
</tr>
<tr>
<td>❌</td>
<td>❌</td>
</tr>
<tr>
<td>❌</td>
<td>❌</td>
</tr>
<tr>
<td>❌</td>
<td>❌</td>
</tr>
<tr>
<td>❌</td>
<td>❌</td>
</tr>
<tr>
<td>❌</td>
<td>❌</td>
</tr>
</tbody>
</table>

- The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
- An indication of 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 statistics 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
- Clearly defined error bars
- State explicitly what error bars represent (e.g. SD, SE, CI)

Our web collection on statistics for biologists may be useful.

Software and code

Policy information about availability of computer code

Data collection

We used publicly available databases for the analyses. The used databases are listed below. Further details are described in the Method section in our manuscript.

- GPCRdb (revision 96fd5102), http://gpcrdb.org/
- 1000 Genomes Project (phase 3), http://www.1000genomes.org/
- The Exome Aggregation Consortium (ExAC, release 1), http://exac.broadinstitute.org/
- GTEx (release v6), http://www.gtexportal.org/home/

Data analysis

We used publicly available softwares for the analyses. The used softwares are listed below. Further details are described in the Method section in our manuscript.

- PLINK 1.9, https://www.cog-genomics.org/plink2/
- Minimac3 (2.0.1), https://genome.sph.umich.edu/wiki/Minimac3/
- mach2dat (1.0.24), https://genome.sph.umich.edu/wiki/Mach2dat:_Association_with_MACH_output
- Chimera (1.12), https://www.cgl.ucsf.edu/chimera/
- SHAPEIT (v2.r837), https://mathgen.stats.ox.ac.uk/genetics_software/shapeit/shapeit.html
- R (3.3.2), https://www.r-project.org/
Locuszoom (v1.2), http://locuszoom.sph.umich.edu/locuszoom/;
LDSC (v1.0.0), https://github.com/bulik/ldsc/;
LDpair (3.0), https://analysistools.nci.nih.gov/LDlink/;
PASCAL, https://www2.unil.ch/cbg/index.php?title=Pascal;
METAL (2011-03-25), https://genome.sph.umich.edu/wiki/METAL;

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 upon request. 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


Field-specific reporting

Please select 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/authors/policies/ReportingSummary-flat.pdf

Life sciences study design

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

Sample size

We aimed to create the largest possible sample size in the Japanese population for the meta-analysis to gain the statistical power. We included as many case and control subjects as possible in our analysis.

Data exclusions

In GWAS1 and GWAS2, we used case and control individuals that were previously analyzed and reported. Forty-eight samples in GWAS1 and 20 samples in GWAS2 were removed due to withdrawal of consent or other reasons. For the case samples in GWAS3 and GWAS4, we selected participants aged 40 or above from the BBJ project who had T2D but did not have type 1 diabetes (T1D), mitochondrial diabetes, maturity onset diabetes of the young or any types of diabetes other than T2D. Participants whose anti-glutamic acid decarboxylase (GAD) antibody was more than 5 IU/ml were excluded from T2D cases. All the control individuals were aged 20 or above. For GWAS3 and GWAS4 sample quality control, we removed samples with a call rate < 0.98. We removed related individuals with PI_HAT > 0.1. PI_HAT is an index of relatedness between two individuals based on identity by descent (IBD) implemented in PLINK. Exclusion criteria were established before conducting association studies.

Replication

Association study was conducted in four different sets of Japanese subjects and the results were combined using fixed-effects meta-analysis. Using summary statistics of published European GWAS, we checked the effect sizes of the T2D loci identified in this study. We found strong positive correlation (Pearson’s r = 0.87, P = 1.4 × 10^-22) and directional consistency (65 loci, 94%, sign test P = 3.1 × 10^-15) in the two populations. Because our participating cohorts are the largest cohorts in Japan, to the best of our knowledge, there are no Japanese or east Asian replication cohort with enough sample sizes.

Randomization

Since we conducted case-control genetic study, we did not randomly allocated subjects into groups.

Blinding

Since we conducted case-control genetic study, blinding is not relevant to our study.

Reporting for specific materials, systems and methods
Policy information about studies involving human research participants

Population characteristics

The number of the male and female subjects in the meta analysis were 100,257 (52.3%) and 91,507 (47.7%), respectively. The mean age of the male and female subjects in the meta analysis were 62.7 and 61.2, respectively.

Recruitment

1. Descriptions of the participating cohorts
The Biobank Japan (BBI) Project
The BBJ Project (http://biobankjp.org) began at the Institute of Medical Science, the University of Tokyo in 2003. As of today, the BBJ Project has recruited about 200,000 participants with any of the 47 types of diseases. These participants were recruited from 12 medical institutions around Japan. Participating institutions are the Cancer Institute Hospital of Japanese Foundation for Cancer Research, Tokushukai Hospitals, Osaka Medical Center for Cancer and Cardiovascular Diseases, Nihon University School of Medicine, Tokyo Metropolitan Geriatric Hospital, Iwate Medical University, Shiga University of Medical Science, Juntendo University, Fukujjui Hospital, Izuka Hospital, Nippon Medical School, and National Hospital Organization Osaka National Hospital.

The Japan Multi-Institutional Collaborative Cohort (J-MICC) Study
In the J-MICC Study, individuals 35 to 69 years of age (40,892 men and 51,750 women) answered medical questionnaires and provided blood samples at the baseline survey during 2004-2014 for the present study. Participants were recruited from 14 study areas in Japan from community dwellers, health checkup examinees, or first-visit patients in a cancer hospital. For the current analyses, around 500 to 2,000 subjects were selected from each study area, considering the number of respondents from each area and the geographical distribution of participants. Written informed consent was obtained from all the participants. The ethics committees of Nagoya University, the affiliation of the principal investigator, and all the other institutions approved the protocol for the J-MICC Study. The J-MICC Study was funded by Grants-in-Aid for Scientific Research for Priority Areas of Cancer (No. 17015018) and Innovative Areas (No. 22150001) and ISPS KAKENHI Grant (No. 16H06277) from the Japanese Ministry of Education, Science, Sports, Culture and Technology (MEXT). Participating institutions are Chiba Cancer Center, Shiga University of Medical Science, Tsuruga City College of Nursing, University of Shizuoka, Kyushu University, Nagoya City University, Kyoto Prefectural University of Medicine, Aichi Cancer Center, University of Tokushima, Nagoya University, Saga University, and Kagoshima University.

The Japan Public Health Center-based Prospective Study (JPHC)
The samples of the JPHC Study were derived from a cohort of 33,736 individuals in 9 public health center (PHC) areas who answered a self-administered questionnaire and provided 10 mL of venous blood at the baseline survey. Before the enrollment, informed consent was obtained from all participants by obeying the study protocol approved by the institutional ethics committees of the National Cancer Center, Tokyo, Japan. At the first step of sample selection, we stratified the cohort by sex, 5-year age categories, and 9 PHC areas, and then performed a random sampling. In the random sampling process, a similar proportion of subjects was selected from each stratum. Before conducting genetic study using the JPHC Study samples, we acquired an approval from the institutional review board of the National Cancer Center (approval number: 2011-044), Tokyo, Japan, and provided all eligible participants with the opportunity to deny participation in the study.

The Tohoku Medical Megabank (TMM) Project
The TMM Project is a ten-year reconstruction project from the Great East Japan Earthquake, 2011 performed by Tohoku University (http://www.megabank.tohoku.ac.jp/english/) and Iwate Medical University (http://iwate-megabank.org/en/). The TMM Project performs two prospective cohort studies in Miyagi and Iwate Prefectures, Japan; the TMM Community-Based Cohort Study (TMM CommCohort Study) and the TMM Birth and Three-Generation Cohort Study (TMM BirThree Cohort Study). The TMM CommCohort Study is a population-based cohort study of adults. The TMM CommCohort Study recruited around 84,000 participants greater than or equal to 20 years of age between 2013 and 2016. The TMM BirThree Cohort Study has recruited approximately 60,000 participants, which includes fetuses and their parents, siblings, grandparents, and extended family members as of July 2016. The TMM BirThree Cohort Study will recruit 70,000 or more participants by March 2017. Written informed consent to genetic studies was obtained from all participants in the TMM Project at baseline examination, medical data (questionnaires, blood and urine tests, and physiological measurements) and biological specimens (blood and urine) have been collected. These medical information and samples are stored in the integrated biobank of the TMM Project. DNA samples of the subjects of the TMM CommCohort Study recruited in 2013 have been analyzed by using the Illumina OmniExpressExome array (N=10,000). Information about age and sex has been obtained by using self-administered questionnaires and by reviewing municipal basic resident register.