nature genetics

LETTERS

https://doi.org/10.1038/s41588-018-0332-4

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

Ken Suzuki[1,2,3,4], Masato Akiyama[2,5], Kazuyoshi Ishigaki[2], Masahiro Kanai[6,2], Jun Hosoe[1], Nobuhiro Shojima[1], Atsushi Hozawa[7,8], Aya Kadota[9,10], Kiyonori Kuriki[11], Mariko Naito[12,13], Kozo Tanno[14,15], Yasushi Ishigaki[16,17], Makoto Hirata[18], Koichi Matsuda[19], Nakao Iwata[20], Masashi Ikeda[20], Norie Sawada[21], Taiki Yamaji[21], Motoki Iwasaki[21], Shiro Ikegawa[22], Shiro Maeda[3,23,24], Yoshinori Murakami[25], Kenji Wakai[12], Shoichiro Tsugane[26], Makoto Sasaki[27,28], Masayuki Yamamoto[7,8], Yukinori Okada[2,4,29], Michiaki Kubo[30], Yoichiro Kamatani[2,31,34*], Momoko Horikoshi[3,34*], Toshimasa Yamauchi[1*] and Takashi Kadowaki[1,32,33*]

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 type 2 diabetes–associated loci ($P < 5.0 \times 10^{-8}$) with 115 independent signals ($P < 5.0 \times 10^{-6}$), of which 28 loci with 30 signals were novel. Twenty-eight missense variants were in linkage disequilibrium ($r^2 > 0.6$) with the lead variants. Among the 28 missense variants, three previously unreported variants had distinct minor allele frequency (MAF) spectra between people of Japanese and European ancestry ($MAF_{JPN} > 0.05$ versus $MAF_{EUR} < 0.01$), including missense variants in genes related to pancreatic acinar cells (*GP2*) and insulin secretion (*GLP1R*). Transethnic comparisons of the molecular pathways identified from the GWAS results highlight both ethnically shared and heterogeneous effects of a series of pathways on type 2 diabetes (for example, monogenic diabetes and beta cells).

Type 2 diabetes is a metabolic disease caused by multifactorial pathogenesis[1]. To date, >150 susceptibility loci for type 2 diabetes have been identified and reported, mainly from studies of people of European ancestry[2]. 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[3]. We previously performed GWAS for type 2 diabetes in the Japanese population[4–6]. 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-Europeans[7].

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)[8] as a

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

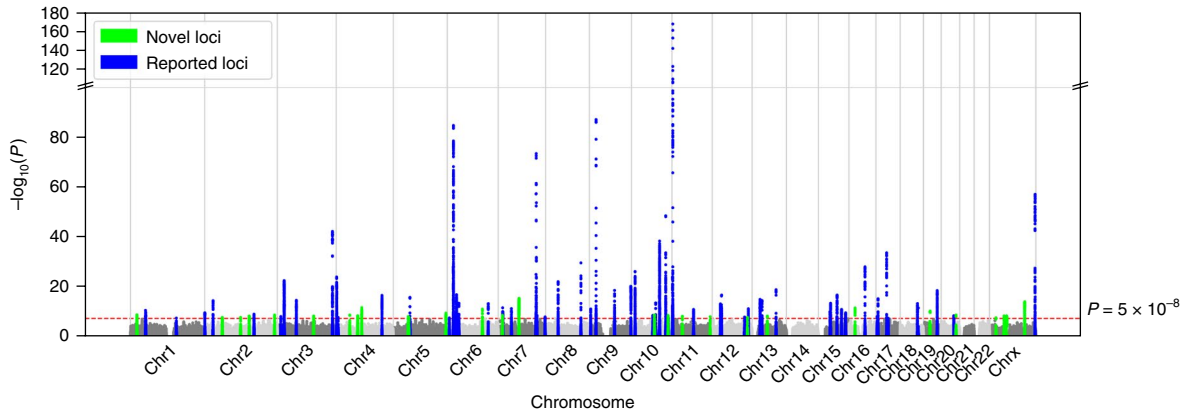

Fig. 1 | Manhattan plot of meta-analysis. Manhattan plot of genome-wide-association results of the meta-analysis in 36,614 cases and 155,150 controls. *P* values (two sided) were derived from the meta-analysis by using the inverse-variance method under a fixed-effect model. *P* values were not adjusted for multiple comparisons. $-\log_{10}(P)$ for each of up to 12,557,761 variants (*y* axis) was plotted against genomic position (NCBI Build 37; *x* axis). Association signals that reached genome-wide significance ($P<5.0\times10^{-8}$) are shown in green if novel and in blue if previously reported. Chr, chromosome.

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 ($\lambda_{GC}=1.21$, Supplementary Fig. 3), linkage disequilibrium (LD)-score regression[9] 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[9], 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[2,10–12], 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 ($P<5.0\times10^{-8}$) (ref. [13]) associations at 88 loci, of which 28 were novel (>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 ≥ 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 $r=0.87$, $P=1.4\times10^{-22}$) and directional consistency (65 of 69 loci, 94%, sign-test $P=3.1\times10^{-15}$). In addition, when we compared the effect sizes of the 95 of 113 lead variants reported in the European type 2 diabetes GWAS[2] 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 $r=0.74$, $P=5.9\times10^{-18}$) and directional consistency (83 of 95 loci, 87%, sign-test $P=3.2\times10^{-14}$). 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 $r=0.83$, $P=8.7\times10^{-51}$) and directional consistency (181 of 192 loci, 94%, sign-test $P=8.3\times10^{-41}$) 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 ($P<5.0\times10^{-6}$), 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 ($r^2\geq0.05$) with each other (rs77402029 and rs463924 in *INS–IGF2–KCNQ1*; rs2237897 and rs233449 in *INS–IGF2–KCNQ1*; rs8037894 and rs77820034 in *C2CD4A*; 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 ($0.005<\text{MAF}<0.38$, $1.06<\text{odds ratio}<1.56$).

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 ($P_{conditional}>0.05$) when conditioned on the missense variant, except for p.Pro443Thr in *SLC16A11* ($P_{conditional}=0.00011$) and p.Gly836Ser in *SND1* ($P_{conditional}=0.0052$).

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 ≥ 0.05) in Japanese individuals (p.Val282Met in *GP2*) or low frequency ($0.01\leq\text{MAF}<0.05$) in Japanese (p.Ala122Pro in *SCTR* and p.Cys517Arg in *ZNF257*). Of the 28 missense variants identified, nine variants were previously

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

Locus	rsID	Chr	Pos	RA	OA	RAF	OR	95% CI	P
USP48	rs1825307	1	22068326	A	G	0.25	1.06	1.04–1.09	1.63×10^{-9}
FANCL and *BCL11A*	rs13417036	2	58921777	A	G	0.16	1.07	1.05–1.10	1.67×10^{-8}
SCTR	rs3731600	2	120231070	C	G	0.95	1.13	1.08–1.18	1.81×10^{-8}
EPC2	rs16829174	2	149455385	A	G	0.59	1.05	1.03–1.07	5.09×10^{-9}
DGKD	rs838720	2	234303281	G	C	0.45	1.05	1.04–1.07	1.91×10^{-9}
CASR and *CD86*	rs2134223	3	121956953	A	G	0.47	1.05	1.03–1.07	5.20×10^{-9}
GNPDA2 and *GABRG1*	rs10938398	4	45186139	A	G	0.30	1.06	1.04–1.08	2.01×10^{-9}
GRSF1 and *MOB1B*	rs12649012	4	71748245	G	A	0.21	1.07	1.04–1.09	3.43×10^{-9}
AGPAT9 and *NKX6-1*	rs201597274	4	85301870	T	C	0.96	1.17	1.12–1.23	2.07×10^{-12}
PARP8 and *EMB*	rs16884025	5	49953459	C	T	0.93	1.14	1.09–1.19	8.48×10^{-9}
ITGA1	rs11410487	5	52094244	A	AT	0.17	1.07	1.04–1.09	1.75×10^{-8}
FGFR4 and *ZNF346*	rs3135911	5	176513896	A	C	0.44	1.06	1.04–1.08	3.50×10^{-10}
NUS1	rs80196932	6	117996631	T	C	0.75	1.07	1.05–1.10	8.37×10^{-12}
ETV1 and *ARL4A*	rs56805921	7	13888699	C	G	0.43	1.06	1.04–1.08	2.51×10^{-9}
AUTS2	rs6947395	7	69406661	T	A	0.20	1.09	1.07–1.12	4.87×10^{-16}
C10orf11 and *ZNF503*	rs7900112	10	77314617	T	A	0.51	1.05	1.04–1.07	2.10×10^{-9}
WDR11 and *FGFR2*	rs10886863	10	122929493	C	T	0.70	1.06	1.04–1.08	3.18×10^{-9}
SLC1A2	rs2421897	11	35437863	G	C	0.88	1.08	1.05–1.11	6.58×10^{-9}
ETS1	rs11819995	11	128389391	T	C	0.22	1.06	1.04–1.08	9.76×10^{-9}
KSR2	rs79310463	12	118406696	T	C	0.26	1.06	1.04–1.08	1.83×10^{-8}
DLEU7 and *KCNRG*	rs123378	13	51088809	G	A	0.22	1.06	1.04–1.09	4.93×10^{-9}
GP2	rs117267808	16	20323168	A	G	0.08	1.12	1.08–1.16	2.99×10^{-12}
ZNF257	rs148316037	19	22257558	T	C	0.97	1.26	1.17–1.34	5.10×10^{-11}
NFATC2	rs6021276	20	50155386	T	C	0.43	1.06	1.04–1.07	1.84×10^{-9}
CNKSR2	rs6633421	X	21569920	A	G	0.68	1.04	1.03–1.06	2.62×10^{-8}
TIMM17B and *PCSK1N*	rs782100977	X	48724648	C	CAA	0.63	1.06	1.04–1.08	4.58×10^{-9}
SPIN2A and *FAAH2*	rs144226500	X	57170781	AT	A	0.30	1.05	1.03–1.06	4.58×10^{-9}
IL13RA1	rs11390176	X	117915163	T	TA	0.44	1.06	1.04–1.07	9.82×10^{-15}

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.

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.Arg131Gln 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 *SND1* 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[18,19], which secrete digestive enzymes[20] and have plasticity to convert into insulin-producing beta cells[21]. These genes were specifically expressed in the pancreas in the Genotype Tissue Expression (GTEx) database[22] (Supplementary Fig. 8) and are specifically expressed in acinar cells[23]. p.Val282Met in *GP2* (rs78193826, $P_{JPN}=4.57\times10^{-12}$, $MAF_{JPN}=0.08$ and $MAF_{EUR}=0.001$) was in high LD ($r^2=0.98$ in East Asians) with the lead variant rs117267808 ($P_{JPN}=2.99\times10^{-12}$). The association of the lead variant at this locus disappeared completely when conditioned on p.Val282Met ($P_{conditional}=0.35$). p.Ala341Thr in *CPA1* (rs77792157, $P_{JPN}=3.07\times10^{-9}$, $MAF_{JPN}=0.05$ and monomorphic in Europeans) was the lead variant of this locus and was 441 kb away from the neighboring *KLF14* locus[24] (rs972283, $P_{JPN}=1.36\times10^{-4}$). The *CPA1* lead missense variant, p.Ala341Thr, was not in LD with *KLF14* rs972283 ($r^2=0.003$ in East Asians), and conditioning on rs972283 confirmed independent association of p.Ala341Thr ($P_{conditional}=1.77\times10^{-8}$).

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 ($r^2=0.97$ in East Asians; Fig. 2a), which resides 245 kb away from the neighboring *KCNK16* locus[25] (rs1535500, $P_{JPN}=9.56\times10^{-10}$). rs9394574 was not in LD with rs1535500 ($r^2<0.01$ in East Asians and Europeans), and conditioning on rs1535500 confirmed independent association of rs9394574 ($P_{conditional=}8.59\times10^{-14}$). Previous candidate gene and exome-wide association studies (ExWAS) in the European population have reported that p.Ala316Thr in *GLP1R* (rs10305492) was associated primarily with fasting glucose and with type 2 diabetes ($P_{EUR}=0.01$–9.4×10^{-5}) in follow-up analyses[26–28]. Recent large-scale European GWAS[2] and ExWAS[11] for type 2 diabetes have not reported genome-wide- or

Table 2 | Missense variants in LD with type 2 diabetes signals

Variant	Chr	Pos	Gene	Amino acid change	RA	OA	OR	95% CI	*P*	MAF_{JPN}[a]	MAF_{EUR}[b]
Coding-variant associations outside established GWAS regions											
rs3731600	2	120231070	*SCTR*	p.Ala122Pro	C	G	1.13	1.08–1.18	1.81×10^{-8}	0.046	0.000
rs1499280	5	52096889	*PELO*	p.Leu221Met	C	A	1.07	1.04–1.09	2.79×10^{-8}	0.16	0.074
rs1966265	5	176516631	*FGFR4*	p.Val10Ile	G	A	1.06	1.04–1.07	1.33×10^{-9}	0.45	0.18
rs351855	5	176520243	*FGFR4*	p.Gly388Arg	A	G	1.05	1.04–1.07	1.23×10^{-9}	0.43	0.29
rs78193826	16	20328666	*GP2*	p.Val282Met p.Val429Met p.Val285Met p.Val432Met	T	C	1.12	1.08–1.15	4.57×10^{-12}	0.078	0.001
rs148612115	19	22272101	*ZNF257*	p.Cys517Arg	T	C	1.25	1.17–1.34	1.95×10^{-10}	0.026	0.000
Previously unreported coding-variant associations within 1 Mb from established GWAS regions											
rs75536691	2	165381518	*GRB14*	p.Leu225Ser	A	G	1.21	1.14–1.29	9.46×10^{-10}	0.022	0.000
rs3765467	6	39033595	*GLP1R*	p.Arg131Gln	G	A	1.09	1.07–1.12	6.10×10^{-14}	0.18	0.001
rs35452727	7	44266184	*CAMK2B*	p.Glu510Lys	T	C	1.08	1.06–1.11	2.35×10^{-10}	0.18	0.054
rs3824004	7	127253551	*PAX4*	p.Arg192Ser	T	G	1.18	1.12–1.24	1.12×10^{-10}	0.029	0.000
rs150524780	7	127729628	*SND1*	p.Gly836Ser	A	G	1.18	1.11–1.24	7.19×10^{-9}	0.026	0.000
rs77792157	7	130025713	*CPA1*	p.Ala341Thr	G	A	1.15	1.10–1.20	3.07×10^{-9}	0.047	0.000
rs60158447	19	46387792	*IRF2BP1*	p.Ala414Gly	G	C	1.09	1.05–1.13	2.14×10^{-6}	0.072	0.010
rs56863346	X	57936433	*ZXDA*	p.His141Arg	C	T	1.07	1.04–1.09	3.27×10^{-8}	0.20	0.14
rs774062078	X	57936451	*ZXDA*	p.Cys135Ser	G	C	1.07	1.04–1.10	4.43×10^{-8}	0.19	0.12
Established coding-variant associations											
rs1260326	2	27730940	*GCKR*	p.Leu446Pro	C	T	1.07	1.05–1.09	5.38×10^{-15}	0.44	0.41
rs1801282	3	12393125	*PPARG*	p.Pro12Ala	C	G	1.16	1.10–1.22	4.98×10^{-8}	0.031	0.12
rs10947804	6	39282036	*KCNK17*	p.Ser21Gly	C	T	1.06	1.04–1.08	1.90×10^{-9}	0.37	0.49
rs11756091	6	39282806	*KCNK16*	p.Pro254His p.Pro301His	T	G	1.06	1.04–1.08	1.08×10^{-9}	0.37	0.49
rs1535500	6	39284050	*KCNK16*	p.Ala277Glu	T	G	1.06	1.04–1.08	9.56×10^{-10}	0.37	0.49
rs2233580	7	127253550	*PAX4*	p.Arg192His	T	C	1.32	1.28–1.36	4.10×10^{-74}	0.092	0.000
rs13266634	8	118184783	*SLC30A8*	p.Arg276Trp p.Arg325Trp	C	T	1.11	1.09–1.13	2.52×10^{-30}	0.41	0.28
rs60980157	9	139235415	*GPSM1*	p.Ser391Leu	C	T	1.14	1.11–1.17	1.51×10^{-20}	0.13	0.24
rs3764002	12	108618630	*WSCD2*	p.Thr266Ile	C	T	1.05	1.03–1.07	2.92×10^{-8}	0.49	0.26
rs75493593	17	6945087	*SLC16A11*	p.Pro443Thr	T	G	1.12	1.09–1.16	3.80×10^{-13}	0.084	0.017
rs75418188	17	6945483	*SLC16A11*	p.Gly340Ser	T	C	1.14	1.11–1.18	6.70×10^{-16}	0.078	0.017
rs13342692	17	6946287	*SLC16A11*	p.Asp127Gly	C	T	1.14	1.10–1.18	7.14×10^{-16}	0.078	0.023
rs117767867	17	6946330	*SLC16A11*	p.Val113Ile	T	C	1.14	1.10–1.18	8.28×10^{-16}	0.078	0.017

Positions are based on Human Genome version 19 (hg19), build 37. Established coding-variant associations include previously reported coding variants or coding variants that were in high LD ($r^2>0.9$ in East Asians) with reported coding variants. *P* values (two sided) were derived from meta-analysis ($n=191{,}764$ independent samples) using the inverse-variance method under a fixed-effect model. *P* values were not adjusted for multiple comparisons. All variants in the list reached genome-wide significance ($P<5.0\times10^{-8}$) except for rs60158447 in *IRF2BP1*. [a]MAF_{JPN}, MAF in Japanese samples of the present study; [b]MAF_{EUR}, MAF in European samples of 1000 Genomes Project.

exome-wide-significant association of p.Ala316Thr with type 2 diabetes ($P_{GWAS}=0.57$ and $P_{ExWAS}=1.59\times10^{-3}$, Fig. 2b). Here, we provide what is, to our knowledge, the first report genome-wide significance for *GLP1R* (rs3765467, p.Arg131Gln, $P_{JPN}=6.10\times10^{-14}$; 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 ($MAF_{JPN}=0.18$) in Japanese samples of the present study but rare ($MAF_{EUR}<0.01$) in the European samples of 1KG Phase 3, p.Ala316Thr was rare in both populations. According to a structure analysis of crystallized GLP1R[29], 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[30]. Given the high MAF of p.Arg131Gln in Japanese and East Asians ($MAF_{EAS}=0.23$), p.Arg131Gln might be a potential marker for clinical response to GLP1R agonists in Japanese and East Asians. Notably, a study

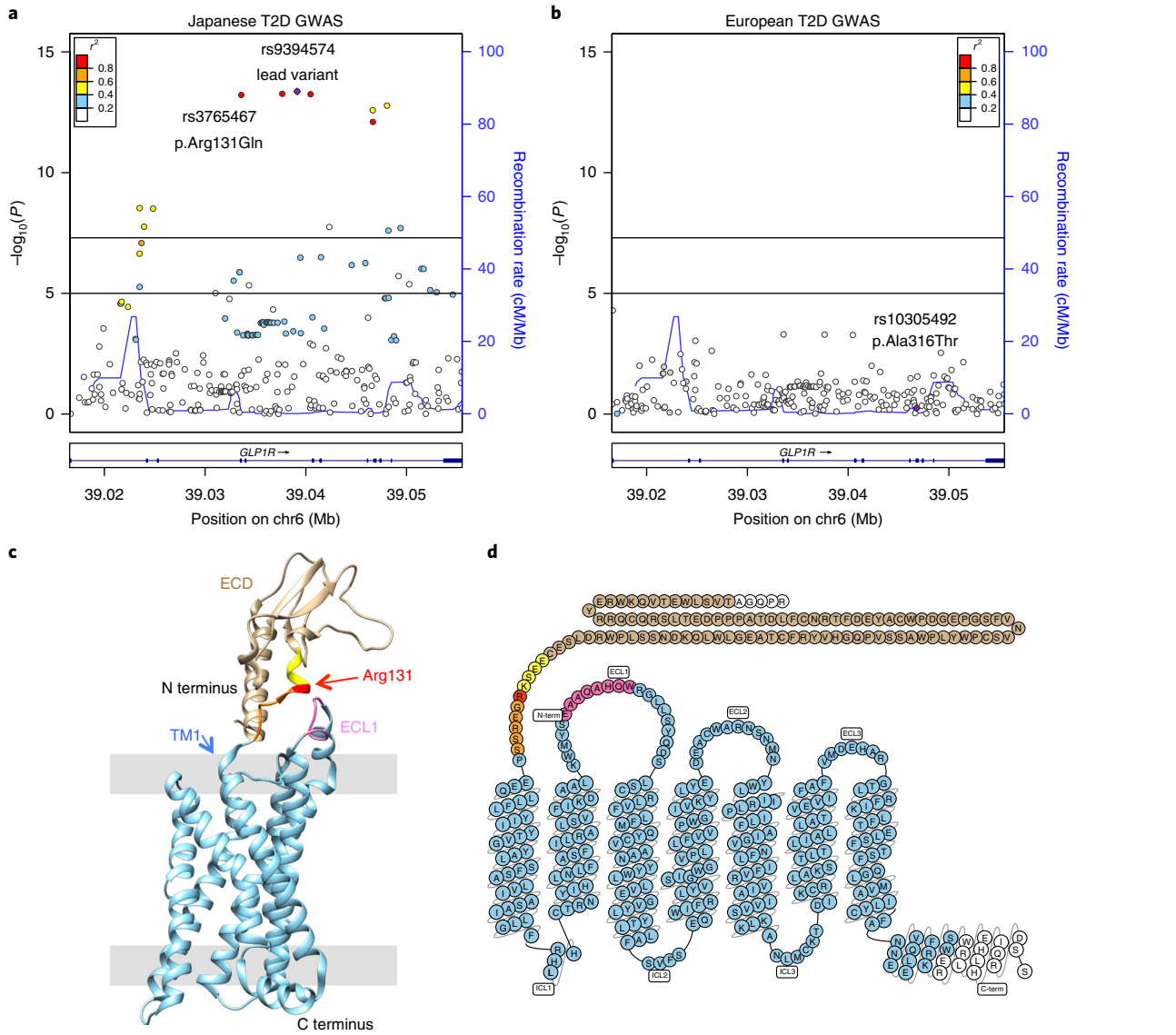

Fig. 2 | Regional association plots and the structure of GLP1R. a,**b**, Regional association plots of *GLP1R* region in Japanese ($n=191{,}764$, **a**) and European ($n=159{,}208$, **b**) type 2 diabetes (T2D) GWAS. *P* values (two sided) were derived from the meta-analyses by using an inverse-variance method under a fixed-effect model. *P* values were not adjusted for multiple comparisons. $-\log_{10}(P)$ is on the *y* axis, and variant position is on the *x* axis. Each variant is colored according to the strength of LD (r^2) with the reference variant (purple). The reference variant is the lead variant (rs9394574) in Japanese GWAS and p.Ala316Thr of *GLP1R* (rs10305492) in European GWAS. **c**, Three-dimensional ribbon model of GLP1R, as viewed parallel to the cell membrane. Approximate membrane boundaries are represented by gray boxes. The model is color-coded as follows: cyan, transmembrane domain; brown, extracellular domain (ECD); yellow, residues Glu127–Lys130; red, Arg131; orange, Gly132–Ser136; pink, Glu207–Trp214. ECL, extracellular loop; ICL, intracellular loop; TM, transmembrane helix. **d**, Snake plot of GLP1R. Each amino acid is color coded in the same way as in the ribbon model. N-term, N terminus; C-term, C terminus.

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[31].

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 ($MAF_{JPN}=0.25$ and $MAF_{EUR}=0.20$). In the European type 2 diabetes GWAS[2], rs80196932 showed nominal association ($P=9.4\times10^{-4}$) 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[32]. *NUS1* is essential for protein glycosylation[33] and intracellular cholesterol trafficking[34], and it is annotated with 'congenital disorder of glycosylation' (Online Mendelian Inheritance in Man, MIM 617082)[35].

We used stratified LD-score regression[36] 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

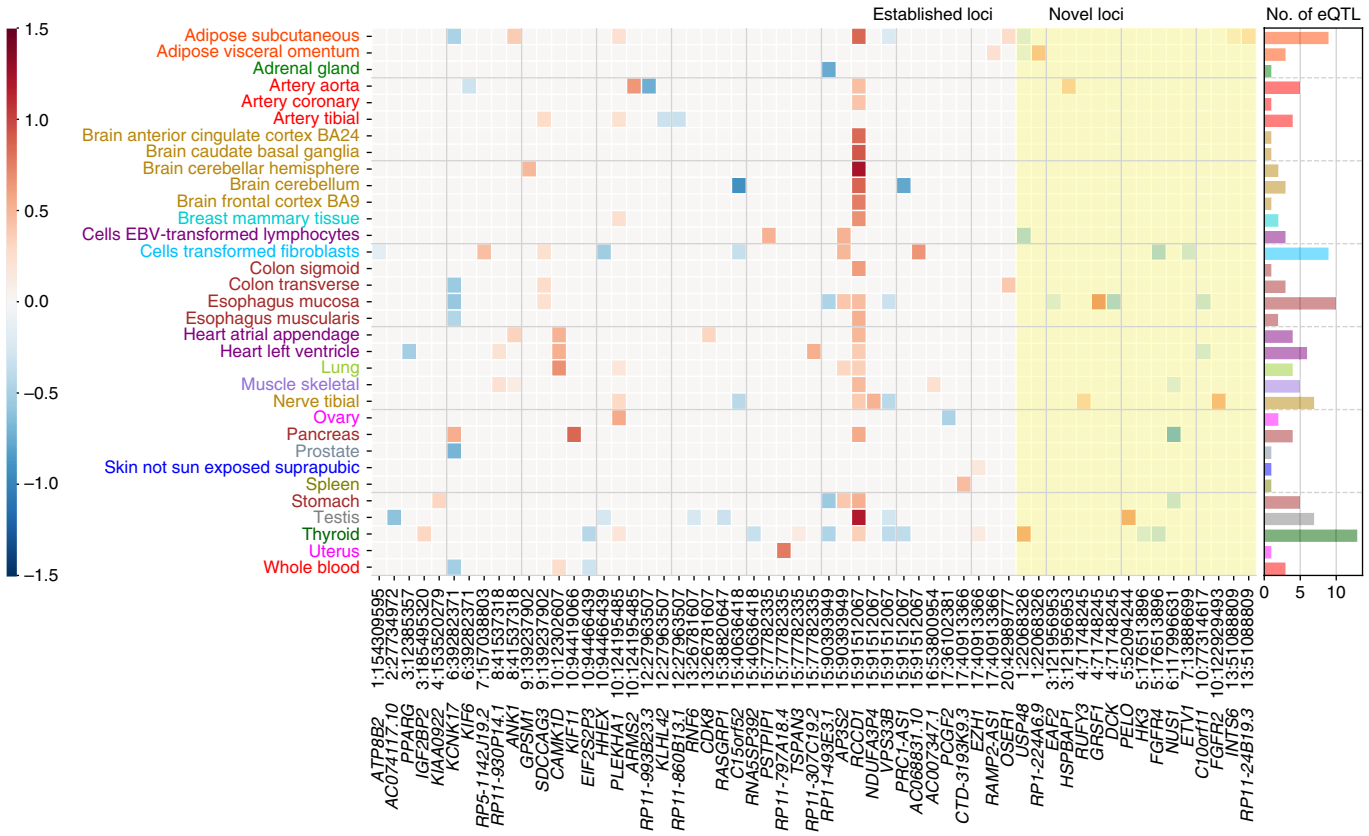

Fig. 3 | Overlap between type 2 diabetes signals and the lead cis-eQTL variants of the GTEx database. Heat-map representation of the lead eQTL variants that were in LD with type 2 diabetes signals ($r^2 > 0.6$ in East Asians or Europeans). Normalized beta values of lead eQTL alleles that were in phase 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 ($P = 3.9 \times 10^{-8}$) 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 ($P = 2.6 \times 10^{-7}$) (Supplementary Table 9 and Supplementary Fig. 9b; Methods). H3K27ac is a characteristic feature of active enhancers[37]. 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 ($P < 2.3 \times 10^{-4} = 0.05/220$). H3K4me3 is a characteristic feature of active promoters[38].

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[39,40] by using bivariate LD-score regression[41] (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 ($r_g = 0.26$, $P = 4.1 \times 10^{-5}$) and white-blood-cell count ($r_g = 0.17$, $P = 4.6 \times 10^{-5}$). Epidemiological studies have reported positive phenotypic correlations between type 2 diabetes and ossification of the posterior longitudinal ligament[42] or white-blood-cell count[43], 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[2]. 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 $q_{JPN} = 6.3 \times 10^{-10}$ and FDR $q_{EUR} = 1.2 \times 10^{-3}$). Mature-onset diabetes of the young is monogenic diabetes characterized by impaired beta cell function[44]. 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

Table 3 | Pathway analysis of type 2 diabetes GWAS results for Japanese and European individuals

		FDR *q* of pathways	
Database	**Pathway name**	**Japanese**	**Europeans**
Pathways with FDR *q* < 0.05 in both Japanese and Europeans			
KEGG	Mature-onset diabetes of the young	6.3×10^{-10}	1.2×10^{-3}
REACTOME	Regulation of beta cell development	6.1×10^{-6}	0.028
REACTOME	Developmental biology	8.6×10^{-3}	0.016
KEGG	Prostate cancer	0.016	0.031
REACTOME	G1 phase	0.046	0.043
Pathways with FDR *q* < 0.05 only in Japanese			
REACTOME	Regulation of gene expression in beta cells	4.0×10^{-4}	0.084
REACTOME	Signaling by NOTCH	7.6×10^{-3}	0.17
REACTOME	Integration of energy metabolism	8.6×10^{-3}	0.41
REACTOME	NOTCH1 intracellular domain regulates transcription	8.6×10^{-3}	0.41
REACTOME	Regulation of insulin secretion	8.6×10^{-3}	0.29
REACTOME	Negative regulation of FGFR signaling	0.010	0.48
KEGG	Phosphatidylinositol signaling system	0.011	0.54
KEGG	Chronic myeloid leukemia	0.022	0.18
REACTOME	Signaling by NOTCH1	0.022	0.61
KEGG	Pathways in cancer	0.046	0.16
REACTOME	Cell cycle mitotic	0.046	0.10
KEGG	NOTCH signaling pathway	0.047	0.27
Pathways with FDR *q* < 0.05 only in Europeans			
KEGG	Type II diabetes mellitus	0.11	0.011
KEGG	ABC transporters	0.70	0.028
REACTOME	Pre NOTCH transcription and translation	0.20	0.028
REACTOME	Fatty acid triacylglycerol and ketone body metabolism	0.26	0.031
REACTOME	Pre NOTCH expression and processing	0.17	0.031
REACTOME	Diabetes pathways	0.24	0.037
REACTOME	PPARA activates gene expression	0.30	0.043
KEGG	Tight junction	0.72	0.050

Results of pathway analyses using Japanese ($n=191,764$) and European ($n=159,208$) type 2 diabetes meta-analyses are indicated. Chi-square tests were conducted, and FDR *q* values were calculated by the Benjamini–Hochberg method. Pathways with FDR $q<0.05$ in both Japanese and Europeans, only in Japanese and only in Europeans are indicated.

novel. Most of the lead variants across 88 type 2 diabetes loci were common in both Japanese and European individuals and showed highly correlated effects on type 2 diabetes. Investigation of missense and eQTL variants at each signal suggested potential causal genes. A substantial number of the newly identified missense variants 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[45,46] *were published that mapped T2D associations to regions reported here at genome-wide significance.*

URLs. 1000 Genomes Project, http://www.1000genomes.org/; ANNOVAR, http://annovar.openbioinformatics.org/en/latest/; CADD, https://cadd.gs.washington.edu/; Chimera, https://www.cgl.ucsf.edu/chimera/; DIAGRAM, http://www.diagram-consortium.org/; FastPCA, https://data.broadinstitute.org/alkesgroup/EIGENSOFT/; GPCRdb, http://gpcrdb.org/; GTEx, http://www.gtexportal.org/home/; HaploReg, https://pubs.broadinstitute.org/mammals/haploreg/haploreg.php; IMM, http://iwate-megabank.org/en/; JENGER, http://jenger.riken.jp/en/; LDpair, https://analysistools.nci.nih.gov/LDlink/; LDSC, https://github.com/bulik/ldsc/; Locuszoom, http://locuszoom.sph.umich.edu/locuszoom/; mach-2dat, https://genome.sph.umich.edu/wiki/Mach2dat:_Association_with_MACH_output; METAL, https://genome.sph.umich.edu/wiki/METAL; Minimac3, https://genome.sph.umich.edu/wiki/Minimac3; National Health and Nutrition Survey, http://www.nibiohn.go.jp/eiken/english/research/pdf/nhns2012.pdf; NBDC Human Database, https://humandbs.biosciencedbc.jp/en/; PASCAL, https://www2.unil.ch/cbg/index.php?title=Pascal; PLINK, https://www.cog-genomics.org/plink2/; Polyphen2, http://genetics.bwh.harvard.edu/pph2/; Population Estimates of Japan, http://www.stat.go.jp/english/data/jinsui/2012np/index.htm; R, https://www.r-project.org/; SHAPEIT, https://mathgen.stats.ox.ac.uk/genetics_software/shapeit/shapeit.html; SIFT, http://sift.bii.a-star.edu.sg/; The BioBank Japan Project, https://biobankjp.org/english/index.html/; ExAC, http://exac.broadinstitute.org/; The J-MICC Study, http://www.jmicc.com/; The JPHC Study, http://epi.ncc.go.jp/en/jphc/index.html; ToMMo, http://www.megabank.tohoku.ac.jp/english/; VEP, http://www.ensembl.org/info/docs/tools/vep/index.html/.

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

1. Stumvoll, M., Goldstein, B. J. & van Haeften, T. W. Type 2 diabetes: pathogenesis and treatment. *Lancet* **371**, 2153–2156 (2008).
2. Scott, R. A. et al. An expanded genome-wide association study of type 2 diabetes in Europeans. *Diabetes* **66**, 2888–2902 (2017).
3. 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).
4. Unoki, H. et al. SNPs in *KCNQ1* are associated with susceptibility to type 2 diabetes in East Asian and European populations. *Nat. Genet.* **40**, 1098–1102 (2008).
5. Yamauchi, T. et al. A genome-wide association study in the Japanese population identifies susceptibility loci for type 2 diabetes at *UBE2E2* and *C2CD4A-C2CD4B*. *Nat. Genet.* **42**, 864–868 (2010).
6. Hara, K. et al. Genome-wide association study identifies three novel loci for type 2 diabetes. *Hum. Mol. Genet.* **23**, 239–246 (2014).
7. 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).
8. Auton, A. et al. A global reference for human genetic variation. *Nature* **526**, 68–74 (2015).
9. Bulik-Sullivan, B. K. et al. LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat. Genet.* **47**, 291–295 (2015).
10. Zhao, W. et al. Identification of new susceptibility loci for type 2 diabetes and shared etiological pathways with coronary heart disease. *Nat. Genet.* **49**, 1450–1457 (2017).

11. Mahajan, A. et al. Refining the accuracy of validated target identification through coding variant fine-mapping in type 2 diabetes. *Nat. Genet.* **50**, 559–571 (2018).
12. Bonàs-Guarch, S. et al. Re-analysis of public genetic data reveals a rare X-chromosomal variant associated with type 2 diabetes. *Nat. Commun.* **9**, 321 (2018).
13. Kanai, M., Tanaka, T. & Okada, Y. Empirical estimation of genome-wide significance thresholds based on the 1000 Genomes Project data set. *J. Hum. Genet.* **61**, 861–866 (2016).
14. Sveinbjornsson, G. et al. Weighting sequence variants based on their annotation increases power of whole-genome association studies. *Nat. Genet.* **48**, 314–317 (2016).
15. Grarup, N. et al. Identification of novel high-impact recessively inherited type 2 diabetes risk variants in the Greenlandic population. *Diabetologia* **61**, 2005–2015 (2018).
16. Pearce, L. R. et al. KSR2 mutations are associated with obesity, insulin resistance, and impaired cellular fuel oxidation. *Cell* **155**, 765–777 (2013).
17. 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).
18. Zhou, Q. et al. A multipotent progenitor domain guides pancreatic organogenesis. *Dev. Cell* **13**, 103–114 (2007).
19. Cogger, K. F. et al. Glycoprotein 2 is a specific cell surface marker of human pancreatic progenitors. *Nat. Commun.* **8**, 331 (2017).
20. Logsdon, C. D. & Ji, B. The role of protein synthesis and digestive enzymes in acinar cell injury. *Nat. Rev. Gastroenterol. Hepatol.* **10**, 362–370 (2013).
21. Zhou, Q., Brown, J., Kanarek, A., Rajagopal, J. & Melton, D. A. In vivo reprogramming of adult pancreatic exocrine cells to β-cells. *Nature* **455**, 627–632 (2008).
22. The GTEx Consortium. et al. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science* **348**, 648–660 (2015).
23. Muraro, M. J. et al. A single-cell transcriptome atlas of the human pancreas. *Cell Syst.* **3**, 385–394.e3 (2016).
24. Voight, B. F. et al. Twelve type 2 diabetes susceptibility loci identified through large-scale association analysis. *Nat. Genet.* **42**, 579–589 (2010).
25. Morris, A. P. et al. Large-scale association analysis provides insights into the genetic architecture and pathophysiology of type 2 diabetes. *Nat. Genet.* **44**, 981–990 (2012).
26. Wessel, J. et al. Low-frequency and rare exome chip variants associate with fasting glucose and type 2 diabetes susceptibility. *Nat. Commun.* **6**, 5897 (2015).
27. Mahajan, A. et al. Identification and functional characterization of G6PC2 coding variants influencing glycemic traits define an effector transcript at the G6PC2-ABCB11 locus. *PLoS Genet.* **11**, e1004876 (2015).
28. Scott, R. A. et al. A genomic approach to therapeutic target validation identifies a glucose-lowering GLP1R variant protective for coronary heart disease. *Sci. Transl. Med.* **8**, 341ra76 (2016).
29. Jazayeri, A. et al. Crystal structure of the GLP-1 receptor bound to a peptide agonist. *Nature* **546**, 254–258 (2017).
30. Sathananthan, A. et al. Common genetic variation in GLP1R and insulin secretion in response to exogenous GLP-1 in nondiabetic subjects: a pilot study. *Diabetes Care* **33**, 2074–2076 (2010).
31. Seino, Y. et al. Pharmacodynamics of the glucagon-like peptide-1 receptor agonist lixisenatide in Japanese and Caucasian patients with type 2 diabetes mellitus poorly controlled on sulphonylureas with/without metformin. *Diabetes Obes. Metab.* **16**, 739–747 (2014).
32. Ward, L. D. & Kellis, M. HaploRegv4: systematic mining of putative causal variants, cell types, regulators and target genes for human complex traits and disease. *Nucleic Acids Res.* **44**, 877–881 (2016).
33. Harrison, K. D. et al. Nogo-B receptor is necessary for cellular dolichol biosynthesis and protein N-glycosylation. *EMBO J.* **30**, 2490–2500 (2011).
34. Harrison, K. D. et al. Nogo-B receptor stabilizes Niemann-Pick type C2 protein and regulates intracellular cholesterol trafficking. *Cell Metab.* **10**, 208–218 (2009).
35. Park, E. J. et al. Mutation of Nogo-B receptor, a subunit of cis-prenyltransferase, causes a congenital disorder of glycosylation. *Cell Metab.* **20**, 448–457 (2014).
36. Finucane, H. K. et al. Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nat. Genet.* **47**, 1228–1235 (2015).
37. Creyghton, M. P. et al. Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proc. Natl Acad. Sci. USA* **107**, 21931–21936 (2010).
38. Heintzman, N. D. et al. Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nat. Genet.* **39**, 311–318 (2007).
39. 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).
40. 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).
41. Bulik-Sullivan, B. et al. An atlas of genetic correlations across human diseases and traits. *Nat. Genet.* **47**, 1236–1241 (2015).
42. Kobashi, G. et al. High body mass index after age 20 and diabetes mellitus are independent risk factors for ossification of the posterior longitudinal ligament of the spine in Japanese subjects: a case-control study in multiple hospitals. *Spine* **29**, 1006–1010 (2004).
43. Nakanishi, N., Yoshida, H., Matsuo, Y., Suzuki, K. & Tatara, K. White blood-cell count and the risk of impaired fasting glucose or type II diabetes in middle-aged Japanese men. *Diabetologia* **45**, 42–48 (2002).
44. American Diabetes Association. Classification and diagnosis of diabetes: standards of medical care in diabetes. *Diabetes Care* **41** (Suppl. 1), S13–S27 (2018).
45. Xue, A. et al. Genome-wide association analyses identify 143 risk variants and putative regulatory mechanisms for type 2 diabetes. *Nat. Commun.* **9**, 2941 (2018).
46. 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).

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 the ossification of the posterior longitudinal ligament GWAS used in this study, which was supported by the Japan Agency for Medical Research and Development (AMED) (17ek0109223h0001) (S.I.). The study of psychiatric disorders was supported by AMED under grants JP18dm0107097 (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-Midosuji District 2660 Rotary International in Japan for supporting our study. This research was funded by the Tailor-Made Medical Treatment Program (the BBJ Project) of the Ministry of Education, Culture, Sports, Science, and Technology (MEXT) and AMED (M. Kubo and Y.M.). This research was also supported by the Advanced Genome Research and Bioinformatics Study to Facilitate Medical Innovation (GRIFIN) in the Platform Program for Promotion of Genome Medicine (P3GM) of AMED (T.K.), JP18km0405202. IMM is supported by MEXT and AMED (grants JP18km0105003 and JP18km0105004, M.S.) ToMMo is supported by MEXT and AMED (grants JP18km0105001, JP18km0105002, JP18km0405203 and JP18km0405001, M.Y.). The JPHC 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 (221S0001) and the JSPS KAKENHI Grant (16H06277) from MEXT (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., K.T., Y.I., M. Hirata, K.M., N.I., M. Ikeda, N. Sawada, T. Yamaji, M. Iwasaki, S.I., S.M., Y.M., K.W., S.T., M.S., M.Y. and Y.O. contributed to data acquisition. M. Horikoshi, Y.K., M. Kubo, T. Yamauchi 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.

Methods

Study participants. In GWAS1 and GWAS2, we used case and control individuals registered in the BioBank Japan (BBJ) project[47], Osaka-Midousuji Rotary Club and Pharma SNP consortium, as previously analyzed and reported[7]. Forty-eight samples in GWAS1 and 20 samples in GWAS2 were removed, owing to withdrawal of consent or 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, mitochondrial 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[48]. 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 had diseases other than diabetes or participants from the following population-based cohorts: Tohoku Medical Megabank Organization (ToMMo)[49], Iwate Tohoku Medical Megabank Organization (IMM)[49], the Japan Public Health Center–based Prospective (JPHC) Study[50] and the Japan Multi-Institutional Collaborative Cohort (J-MICC) Study[51]. 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) implemented in PLINK[52]. PI_HAT is defined by the following equation: PI_HAT = P(IBD = 2) + 0.5 × 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[53]. 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 610 K SNP array. Before imputation, genotyped variants were excluded according to the following conditions. In GWAS1, GWAS3 and GWAS4, we excluded variants with (i) hterozygosity count <5, (ii) Hardy–Weinberg-equilibrium P (calculated in all samples for autosomes and in females for X chromosome) < 1.0×10^{-6} 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 excluded variants with (i) call rate <0.99, (ii) MAF < 0.01, (iii) differential missingness $P < 1.0 \times 10^{-6}$ or (iv) Hardy–Weinberg-equilibrium P (calculated in all samples for autosomes and in females for X chromosome) < 1.0×10^{-6}.

Prephasing and genotype 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)[8] with minimac3 (2.0.1). Similarly, imputation of X chromosomes was performed with the 1KG Phase 3 reference panel and minimac3, but in males and females separately. The pseudoautosomal region (2.6–154.9 Mb) was excluded from the reference panel before imputation. Allelic dosages were 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, sex and ten principal components, by assuming an additive model. After filtering of variants with minimac3 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 minimac3 imputation quality score (r^2) > 0.3. We integrated the results of males and females in each GWAS through the inverse-variance method under the assumption of full dosage compensation and fixed-effect model. We followed the same procedure as that in autosomes for subsequent steps. Regional association plots were produced with Locuszoom (v1.2). Adjacent genome-wide-significant ($P < 5.0 \times 10^{-8}$) variants were grouped in one locus if they were located within 1 Mb from each other. We used $P = 5.0 \times 10^{-8}$ as the primary threshold for genome-wide significance. In the Supplementary Note, the empirical threshold reflecting the prior probability of the effect of each variant[14] is discussed. We calculated allele frequency by using PLINK on the basis of the 1KG Phase 3 reference panel (May 2013 release) or ExAC release 1.

Evaluation of confounding biases with LD-score regression. To estimate confounding biases derived from cryptic relatedness and population stratification, we conducted LD-score regression[9] (v1.0.0). 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 stepwise conditional meta-analysis. We first 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 5.0×10^{-6}, 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[2] (1.0×10^{-5}).

Functional annotation and eQTL analyses. To characterize associated variants, we used ANNOVAR to gain functional annotation of variants in LD ($r^2 \geq 0.6$ 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 *GP2* were downloaded from GTEx[22] portal (release v7). The three-dimensional ribbon model of *GLP1R* in Fig. 2c 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.[29]. The snake plot of *GLP1R* in Fig. 2d was created with GPCRdb revision 96fd5102 (see URLs). We searched for overlaps between the 115 type 2 diabetes lead variants (and proxies with $r^2 \geq 0.6$ in East Asians or Europeans in 1KG Phase 3) and lead cis-eQTL variants in GTEx[22] (release v6). GTEx is a large-scale collaborative work in which genomes and transcriptomes of multiple tissues from individuals of various ancestries have been analyzed. We considered only eQTLs with an FDR q <0.05 as listed by the providers, and we chose variants showing the most significant association for each transcript in each tissue. We searched for overlaps between the lead variants of 115 type 2 diabetes signals and promoter 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 *NUS1* (rs80196932).

Genetic correlation. In our previous work[39,40], bivariate LD-score regression[41] was conducted to estimate genetic correlations among complex diseases[54–56] and clinical laboratory measurements in the Japanese population. We updated the previous type 2 diabetes GWAS[39] 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 <5,000. We excluded subjects with type 2 diabetes and dyslipidemia from controls in 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 regression[36] 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 with low imputation quality scores ($r^2 < 0.7$ in at least one of the four GWAS) and variants within major histocompatibility complex region (chromosome 6, 26–34 Mb) were excluded from the regression analysis.

Transethnic pathway analysis. We conducted transethnic comparison of the associations of molecular pathways[57] estimated from the summary statistics of Japanese and European type 2 diabetes GWAS in autosomes, by using PASCAL. Summary statistics of the Japanese GWAS were adjusted for study-level genomic control in accordance with the correction method of the European GWASs[2]. PASCAL computes gene-based scores (SUM and MAX scores) by aggregating SNP P values and calculates pathway scores by merging the scores of genes belonging to the same pathway. We prepared custom reference genotype information from East Asian samples in 1KG Phase 3 and European samples in 1KG Phase 1 v3 to adapt 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

GWAS summary statistics of type 2 diabetes will be publicly available at our website (JENGER, http://jenger.riken.jp/en/) and the National Bioscience Database Center (NBDC, https://humandbs.biosciencedbc.jp/en/) Human Database. Genotype data of case samples are available at NBDC under research ID hum0014.

References

47. Nagai, A. et al. Overview of the BioBank Japan Project: study design and profile. *J. Epidemiol.* **27**, S2–S8 (2017).
48. Seino, Y. et al. Report of the committee on the classification and diagnostic criteria of diabetes mellitus. *Diabetol. Int.* **1**, 212–228 (2010).
49. Kuriyama, S. et al. The Tohoku Medical Megabank Project: design and mission. *J. Epidemiol.* **26**, 493–511 (2016).
50. Tsugane, S. et al. The JPHC study: design and some findings on the typical Japanese diet. *Jpn. J. Clin. Oncol.* **44**, 777–782 (2014).
51. Hamajima, N. et al. The Japan Multi-Institutional Collaborative Cohort Study (J-MICC Study) to detect gene-environment interactions for cancer. *Asian Pac. J. Cancer Prev.* **8**, 317–323 (2007).
52. Purcell, S. et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).
53. 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).
54. Ogura, Y. et al. A functional SNP in BNC2 is associated with adolescent idiopathic scoliosis. *Am. J. Hum. Genet.* **97**, 337–342 (2015).
55. Hirota, T. et al. Genome-wide association study identifies eight new susceptibility loci for atopic dermatitis in the Japanese population. *Nat. Genet.* **44**, 1222–1226 (2012).
56. Nakajima et al. A genome-wide association study identifies susceptibility loci for ossification of the posterior longitudinal ligament of the spine. *Nat. Genet.* **46**, 1012–1016 (2014).
57. Okada, Y., Raj, T. & Yamamoto, K. Ethnically shared and heterogeneous impacts of molecular pathways suggested by the genome-wide meta-analysis of rheumatoid arthritis. *Rheumatology* **55**, 186–189 (2015).

nature research

Corresponding author(s): Takashi Kadowaki, Toshimasa Yamauchi, Momoko Horikoshi, and Yoichiro Kamatani

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).

n/a	Confirmed	
☐	☒	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/; DIAGRAM, http://www.diagram-consortium.org;
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; EIGENSOFT (6.1), https://data.broadinstitute.org/alkesgroup/EIG6.1/; 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; ANNOVAR, http://annovar. openbioinformatics.org/en/latest/; Chimera (1.12), https://www.cgl.ucsf.edu/chimera/; SHAPEIT (v2.r837), https://mathgen.stats. ox.ac.uk/genetics_software/shapeit/shapeit.html; HaploReg (v4.1), http:// www.broadinstitute.org/mammals/haploreg/haploreg.php; 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

GWAS summary statistics of type 2 diabetes will be publically available at our website (JENGER, http://jenger.riken.jp/en/) and the National Bioscience Database Center (NBDC, https://humandbs.biosciencedbc.jp/en/) Human Database. Genotype data of case samples are available at NBDC under research ID hum0014.

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

Materials & experimental systems

n/a	Involved in the study
☒	☐ Unique biological materials
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology
☒	☐ Animals and other organisms
☐	☒ Human research participants

Methods

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

Human research participants

Policy information about studies involving human research participants

Population characteristics	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 (BBJ) Project The BBJ Project (http://biobankjp.org) began at the Institute of Medical Science, the University of Tokyo in 20031. 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, Fukujuji Hospital, Iizuka 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 study2. 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. 221S0001) and JSPS 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) areas3 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 ethical 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 - Iwate Tohoku Medical Megabank Organization (IMM) - Tohoku Medical Megabank Organization (ToMMo) 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/)4. 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.