Genome-wide association study identifies 112 new loci for body mass index in the Japanese population

Masato Akiyama[1], Yukinori Okada[1–3], Masahiro Kanai[1], Atsushi Takahashi[1,4], Yukihide Momozawa[5], Masashi Ikeda[6], Nakao Iwata[6], Shiro Ikegawa[7], Makoto Hirata[8], Koichi Matsuda[9], Motoki Iwasaki[10], Taiki Yamaji[10], Norie Sawada[10], Tsuyoshi Hachiya[11], Kozo Tanno[11,12], Atsushi Shimizu[11], Atsushi Hozawa[13,14], Naoko Minegishi[13,14], Shoichiro Tsugane[15], Masayuki Yamamoto[13,14], Michiaki Kubo[16] & Yoichiro Kamatani[1,17]

Obesity is a risk factor for a wide variety of health problems. In a genome-wide association study (GWAS) of body mass index (BMI) in Japanese people ($n = 173,430$), we found 85 loci significantly associated with obesity ($P < 5.0 \times 10^{-8}$), of which 51 were previously unknown. We conducted trans-ancestral meta-analyses by integrating these results with the results from a GWAS of Europeans and identified 61 additional new loci. In total, this study identifies 112 novel loci, doubling the number of previously known BMI-associated loci. By annotating associated variants with cell-type-specific regulatory marks, we found enrichment of variants in CD19+ cells. We also found significant genetic correlations between BMI and lymphocyte count ($P = 6.46 \times 10^{-5}$, $r_g = 0.18$) and between BMI and multiple complex diseases. These findings provide genetic evidence that lymphocytes are relevant to body weight regulation and offer insights into the pathogenesis of obesity.

Obesity, which is heritable, is a risk factor for various diseases[1–3]. Thus far, GWAS have identified more than 100 loci associated with BMI[4–7], the most commonly used measurement for obesity. However, these loci explained only a small fraction (~3%) of heritability in Europeans[6]. Recent polygenic analyses have suggested that genetic variants could explain more than 20% of the heritability[6,8]; therefore, most of the genetic components of BMI have not yet been discovered. Whereas most previous studies were in individuals of European ancestry, GWAS in non-Europeans have identified different loci[5]. Furthermore, epidemiological surveys reported differences in BMI across ethnicities; for example, the prevalence of obesity was lower in Asians than in Europeans and North Americans[1]. In addition, Asians tend to develop diabetes with a lower BMI than Europeans[9]. These differences suggest that analyzing different populations could yield further insight into the etiology of obesity.

To identify genetic loci associated with obesity and gain more insight into body weight regulation, we conducted a GWAS that included >170,000 Japanese volunteers (study design is shown in

Supplementary Fig. 1). We then conducted a comprehensive integration with the previous GWAS of Europeans, functional annotations of the associated loci and genetic correlation analyses between BMI and other complex diseases in humans.

RESULTS
Association signals at 85 loci found in Japanese population
This GWAS included 158,284 Japanese subjects who participated in the BioBank Japan (BBJ) project[10–12]. After whole-genome imputation using East Asian (EAS) samples of the 1000 Genomes Project (1KGP)[13] as a reference, we performed a GWAS using 6,108,953 single-nucleotide variants. Although we applied stringent quality controls (QCs), the genomic inflation factor showed notable inflation ($\lambda_{GC} = 1.44$) (Supplementary Fig. 2a). To dissect observed inflation into the polygenic effects and potential biases, we performed linkage disequilibrium (LD) score regression[14]. The estimated mean χ^2 and intercept (i.e., the index of biases) values were 1.63 and 1.07, respectively (Supplementary Fig. 2b). These values suggested that

[1]Laboratory for Statistical Analysis, RIKEN Center for Integrative Medical Sciences, Yokohama, Japan. [2]Department of Statistical Genetics, Osaka University Graduate School of Medicine, Osaka, Japan. [3]Laboratory of Statistical Immunology, Immunology Frontier Research Center (WPI-IFReC), Osaka University, Suita, Japan. [4]Laboratory for Omics Informatics, Omics Research Center, National Cerebral and Cardiovascular Center, Osaka, Japan. [5]Laboratory for Genotyping Development, RIKEN Center for Integrative Medical Sciences, Yokohama, Japan. [6]Department of Psychiatry, Fujita Health University School of Medicine, Aichi, Japan. [7]Laboratory for Bone and Joint Diseases, RIKEN Center for Integrative Medical Sciences, Tokyo, Japan. [8]Institute of Medical Science, the University of Tokyo, Tokyo, Japan. [9]Graduate school of Frontier Sciences, the University of Tokyo, Tokyo, Japan. [10]Division of Epidemiology, Center for Public Health Sciences, National Cancer Center, Tokyo, Japan. [11]Iwate Tohoku Medical Megabank Organization, Iwate Medical University, Iwate, Japan. [12]Department of Hygiene and Preventive Medicine, School of Medicine, Iwate Medical University, Iwate, Japan. [13]Tohoku Medical Megabank Organization, Tohoku University, Sendai, Japan. [14]Graduate School of Medicine, Tohoku University, Sendai, Japan. [15]Center for Public Health Sciences, National Cancer Center, Tokyo, Japan. [16]RIKEN Center for Integrative Medical Sciences, Yokohama, Japan. [17]Center for Genomic Medicine, Kyoto University Graduate School of Medicine, Kyoto, Japan. Correspondence should be addressed to Y.K. (yoichiro.kamatani@riken.jp).

Table 1 Results of the condition analysis

SNP	Positional candidate gene	Chr.	Position[a]	Allele REF/ALT	Crude			Condition		
					β[b]	s.e.m.	P value	β[b]	s.e.m.	P value
rs939584	*ALKAL2, TMEM18*	2	621558	C/T	0.054	0.006	1.39×10^{-19}	0.053	0.006	1.55×10^{-18}
rs4430979	*FAM110C*	2	47208	T/G	0.031	0.005	5.40×10^{-10}	0.029	0.005	6.32×10^{-9}
rs2495707	*HIF1AN, PAX2*	10	102425949	A/G	−0.025	0.004	6.21×10^{-9}	−0.023	0.004	3.80×10^{-8}
rs11191021	*BTRC*	10	103231641	A/G	0.021	0.004	8.12×10^{-9}	0.020	0.004	4.97×10^{-8}
rs2540034	*ADCY9*	16	4022694	C/T	0.028	0.004	3.49×10^{-11}	0.029	0.004	1.11×10^{-11}
rs7199766	*SLX4*	16	3637429	G/A	0.020	0.004	1.25×10^{-7}	0.021	0.004	3.95×10^{-8}
rs12597682	*GPR139, GP2*	16	20258432	C/A	−0.030	0.005	6.80×10^{-11}	−0.031	0.005	3.90×10^{-11}
rs1011939	*GPRC5B, GPR139*	16	19992996	G/A	−0.025	0.004	8.26×10^{-11}	−0.025	0.004	4.73×10^{-11}

Chr., chromosome; REF, reference allele; ALT, alternative allele.

[a]Positions are based on Human Genome version 19 (hg19), build 37. [b]Alternative alleles were treated as effect alleles.

polygenic effects induced a majority (89%) of the inflation, and the influence of biases was minimal in comparison with previously reported meta-GWAS[14]. Therefore, we did not apply GC correction to our GWAS.

We observed genome-wide significant (GWS; $P < 5.0 \times 10^{-8}$) association signals at 72 loci (at least 1 Mb apart) (**Supplementary Fig. 2c**). We further evaluated 134 lead variants ($P < 1.0 \times 10^{-6}$) using replication sets composed of 15,146 subjects from two independent Japanese population-based cohort studies (the Japan Public Health Center–based Prospective Study (JPHC) and the Tohoku Medical Megabank Project (TMM)). We observed 83% of directional consistency (111 of 134 loci; P for sign test = 5.0×10^{-14}), and all the nominal associations ($P < 0.05$ in the replication sets) were in the same direction (P for sign test = 2.2×10^{-31}). When we combined the results of the GWAS with replication sets, six loci fell below the GWS threshold, whereas 17 additional loci surpassed the GWS level (**Supplementary Tables 1** and **2**). Accordingly, 83 loci were significantly associated with BMI (**Supplementary Table 3a** and **Supplementary Data Sets 1** and **2**).

To identify the BMI-associated loci with sex-dependent effects, we performed a GWAS stratified by sex (**Supplementary Fig. 3**). We identified two male-specific loci that had not been previously reported (rs77511173 and rs111612372) (**Supplementary Tables 3b** and **4** and **Supplementary Fig. 3c**). We compared the effect sizes of identified variants between sexes; we found that only rs77511173 showed a significant difference ($P_{difference} = 5.88 \times 10^{-5}$, $\alpha = 0.05/85$) (**Supplementary Table 5** and **Supplementary Fig. 4**). The effect sizes of the variants on the X chromosome were not different under the assumption of the full dosage compensation ($P_{difference} > 0.05$). We observed nominal differences at two reported sex-dependent loci, *SEC16B*[6] (stronger in females, $P_{difference} = 3.60 \times 10^{-3}$) and *KCNQ1* (ref. 5) (stronger in males, $P_{difference} = 3.49 \times 10^{-3}$), but not at other loci (*PCSK1*, *CDKAL1* and *ALDH2*)[5]. The effect sizes of the associated alleles were strongly correlated between sexes (Pearson's $r = 0.92$, $P = 5.21 \times 10^{-38}$). We also evaluated the shared heritability between sexes. After applying bivariate LD score regression[15], we observed a significant genetic correlation ($r_g = 0.93$, s.e. = 0.04, $P = 8.3 \times 10^{-151}$), indicating that polygenicity was shared. In total, 85 loci, including five loci on the X chromosome and two male-specific loci, were significantly associated with BMI (**Supplementary Table 3**), and 51 were novel.

To assess the robustness, we evaluated the associations of the significant variants in the replication sets. Among them, 49 of 85 (57.6%) were associated with nominal significance ($P < 0.05$). Furthermore, the directions of the effects were consistent at 83 loci (97.6%; P for sign test = 1.89×10^{-22}). We did not find any statistical differences

in the number of directionally consistent and nominally associated variants between the previously reported loci and the newly identified loci ($P = 1$ for Fisher's exact test), suggesting that most of the newly identified loci were not false positives. We observed strong correlation in the effect sizes of associated alleles between the GWAS and the meta-analysis of replication sets ($r = 0.91$, $P = 3.81 \times 10^{-34}$).

Next, we performed a conditional analysis to explore BMI-associated variants independent of the lead variants within the loci. We found four significant second signals that satisfied GWS after conditioning by the lead variants within each region (in a range of ±1 Mb of each lead variant) (**Table 1** and **Supplementary Fig. 5**), and two of these had already been found in Europeans[6] (*GPRC5B–GP2* and *NLRC3–ADCY9*). Another two signals were novel secondary associations (*FAM110C* and *BTRC*). Therefore, we determined 89 independent susceptibility variants within 85 loci.

To distinguish association signals in the Japanese population from those previously reported in Europeans, we compared the association signals in our GWAS and Europeans by visual inspection and found three distinct signals (**Supplementary Fig. 6**). When we calculated LD of 1KGP between lead variants in each GWAS, these variants were in weak LD ($r^2 < 0.02$) in both European (EUR) and EAS. Furthermore, by conditioning on lead variants of Europeans, we confirmed that these associations were independent of previously reported ones ($P_{condition} < 5.0 \times 10^{-8}$) (**Supplementary Table 6**).

Among the 89 independently associated variants, rs148546399 was low frequency (minor allele frequency (MAF) ≤ 5%) in East Asian and monomorphic in Europeans; 13 variants had a MAF ≤5% in Europeans but were common in East Asian (MAF > 5%); and two variants (rs10208649 and rs7903146) were low frequency in East Asian and common in Europeans according to 1KGP (**Supplementary Tables 1** and **4**). When we annotated the variants in LD ($r^2 \geq 0.8$) with the 89 variants that reached GWS, we found 29 missense variants (**Supplementary Table 7**), including a missense variant in *GPR101* (rs1190736 (C > A);p.Val124Leu; GenPept NP_473362.1), in which mutations lead to acromegaly[16]. We additionally evaluated variants in LD with rarer variant rs10208649 (MAF = 1.2% in EAS), and found the rare coding variant of *GPR75* in moderate LD (rs80328470 (T > C); p.Thr27Ala; GenPept NP_006785.1; MAF = 0.8% in EAS; $r^2 = 0.66$), which is monomorphic in Europeans according to 1KGP and ExAC[17].

We examined the pleiotropy of the 89 identified variants using the GWAS catalog database. We found 32 loci (36%) that had reported associations for other traits (**Supplementary Tables 8** and **9**). The trait that most frequently overlapped with BMI was type 2 diabetes (T2D; 11.2%), followed by age at menarche (9.0%) and height (6.7%). These overlaps were also observed in Europeans[6].

Table 2 New loci identified by trans-ethnic meta-analysis

SNP	Chr.[a]	Position[a] (bp)	Alleles REF/ALT	1KG phase 1 ALT freq EAS	1KG phase 1 ALT freq EUR	Candidate gene(s)	BBJ GWAS β[b]	BBJ GWAS s.e.m.	BBJ GWAS P value	European GWAS β[b]	European GWAS s.e.m.	European GWAS P value	N all	MANTRA Posterior mean allelic effect	MANTRA s.e.m.	MANTRA Log-10BF	MANTRA Log10BF for heterogeneity	Fixed-effect β[b]	Fixed-effect s.e.m.	Fixed-effect P value	P_{het}[c]
rs12044597	1	1,708,801	A/G	0.47	0.51	NADK	0.013	0.004	3.16×10^{-4}	0.015	0.003	1.35×10^{-6}	474,968	0.014	0.002	7.49	0.05	0.014	0.002	1.54×10^{-9}	0.64
rs11185092	1	107,886,278	A/G	0.16	0.22	NTNG1	0.020	0.005	7.46×10^{-5}	0.018	0.005	9.97×10^{-5}	392,297	0.019	0.003	6.29	0.07	0.019	0.003	2.89×10^{-8}	0.80
rs10923724	1	119,546,842	C/T	0.58	0.57	TBX15,WARS2	−0.015	0.004	2.89×10^{-5}	−0.012	0.003	1.33×10^{-4}	480,391	−0.013	0.002	6.28	−0.09	−0.013	0.002	2.34×10^{-8}	0.46
rs10754220	1	197,244,290	G/A	0.19	0.30	CRB1	−0.025	0.005	2.86×10^{-7}	−0.013	0.004	1.28×10^{-3}	392,262	−0.018	0.003	6.85	−0.02	−0.018	0.003	7.90×10^{-9}	0.07
rs823114	1	205,719,532	G/A	0.50	0.53	NUCKS1	−0.014	0.004	4.51×10^{-5}	−0.012	0.003	2.02×10^{-4}	473,158	−0.013	0.002	6.12	0.04	−0.013	0.002	4.29×10^{-8}	0.53
rs9786986	1	235,656,632	G/T	0.49	0.11	B3GALNT2	0.016	0.004	4.80×10^{-6}	0.020	0.005	7.43×10^{-5}	480,088	0.017	0.003	7.28	−0.15	0.017	0.003	1.55×10^{-9}	0.60
rs4596023	2	48,955,683	A/G	0.66	0.73	LHCGR, STON1-GTF2A1L	0.016	0.004	1.14×10^{-5}	0.013	0.004	2.84×10^{-4}	480,125	0.014	0.003	6.45	−0.14	0.014	0.003	1.50×10^{-8}	0.54
rs7569376	2	205,385,322	T/C	0.47	0.79	ICOS,PARD3B	−0.017	0.004	3.70×10^{-6}	−0.015	0.005	2.52×10^{-3}	391,110	−0.016	0.003	6.02	−0.04	−0.016	0.003	3.50×10^{-08}	0.73
rs972540	2	207,244,783	A/G	0.19	0.24	ZDBF2,ADAM23	0.013	0.005	7.66×10^{-3}	0.018	0.003	2.28×10^{-7}	480,378	0.016	0.003	6.77	0.10	0.016	0.003	6.39×10^{-9}	0.37
rs7613875	3	49,971,514	C/A	0.14	0.52	MON1A,RBM6	0.015	0.005	1.50×10^{-3}	0.016	0.003	7.36×10^{-7}	464,114	0.016	0.003	7.00	−0.08	0.016	0.003	4.38×10^{-9}	0.95
rs1225051	3	131,645,972	G/A	0.46	0.44	CPNE4	0.012	0.004	6.28×10^{-4}	0.016	0.004	9.32×10^{-6}	392,295	0.014	0.003	6.29	0.13	0.014	0.003	3.11×10^{-8}	0.41
rs7621025	3	136,272,246	T/C	0.84	0.73	STAG1	0.015	0.005	1.77×10^{-3}	0.017	0.004	2.80×10^{-6}	480,294	0.016	0.003	6.47	0.10	0.016	0.003	2.07×10^{-8}	0.72
rs4834272	4	113,313,986	T/C	0.53	0.31	ALPK1	0.014	0.004	5.33×10^{-5}	0.013	0.003	3.71×10^{-5}	480,262	0.014	0.002	6.77	0.26	0.014	0.002	1.04×10^{-8}	0.83
rs7720894	5	60,733,933	G/C	0.39	0.48	ZSWIM6	−0.020	0.004	6.12×10^{-7}	−0.011	0.004	3.22×10^{-3}	392,090	−0.015	0.003	6.20	0.00	−0.015	0.003	2.89×10^{-8}	0.10
rs3849724	5	173,290,977	G/T	0.25	0.50	LINC01485, CPEB4	−0.018	0.004	5.85×10^{-6}	−0.014	0.003	9.76×10^{-6}	480,165	−0.015	0.003	7.94	−0.02	−0.015	0.002	4.95×10^{-10}	0.37
rs2228213	6	12,124,855	G/A	0.23	0.33	HIVEP1	−0.017	0.004	1.60×10^{-5}	−0.016	0.003	1.04×10^{-6}	480,386	−0.016	0.003	8.65	0.03	−0.016	0.003	1.03×10^{-10}	0.81
rs7779181	7	32,345,283	T/C	0.23	0.20	PDE1C, LOC100130673	0.015	0.004	2.52×10^{-4}	0.019	0.004	2.37×10^{-5}	392,261	0.017	0.003	6.32	0.22	0.017	0.003	2.63×10^{-8}	0.58
rs10269783	7	49,616,203	G/A	0.60	0.37	CDC14C,VWC2	0.015	0.004	3.42×10^{-5}	0.014	0.003	4.14×10^{-6}	477,965	0.015	0.002	7.78	−0.20	0.015	0.002	4.87×10^{-10}	0.92
rs3779273	7	77,828,940	G/A	0.44	0.40	MAGI2	−0.014	0.004	1.18×10^{-4}	−0.015	0.003	1.52×10^{-6}	480,399	−0.014	0.002	7.66	−0.17	−0.014	0.002	7.55×10^{-10}	0.82
rs12156392	8	28,140,103	C/A	0.35	0.41	ELP3,PNOC	0.019	0.004	6.59×10^{-7}	0.011	0.004	5.28×10^{-3}	392,039	0.015	0.003	6.16	0.06	0.015	0.003	4.42×10^{-8}	0.11
rs10091344	8	34,132,075	G/A	0.44	0.34	DUSP26, LINC01288	−0.015	0.004	2.21×10^{-5}	−0.016	0.004	9.62×10^{-5}	392,217	−0.015	0.003	6.80	0.05	−0.015	0.003	8.33×10^{-9}	0.93
rs10811901	9	23,356,935	G/A	0.86	0.62	LINC01239, LOC101929563	0.023	0.004	7.57×10^{-8}	0.010	0.003	1.89×10^{-3}	461,004	0.015	0.003	6.54	0.11	0.015	0.003	1.26×10^{-8}	0.02
rs580809	9	101,518,442	C/T	0.56	0.75	ANKS6	−0.016	0.004	7.74×10^{-6}	−0.012	0.004	6.75×10^{-4}	480,376	−0.014	0.002	6.31	0.02	−0.014	0.002	2.70×10^{-8}	0.42
rs7024334	9	109,072,075	T/G	0.42	0.77	TMEM38B, MIR8081	−0.014	0.004	1.55×10^{-4}	−0.016	0.004	1.41×10^{-5}	473,417	−0.015	0.003	6.53	−0.07	−0.015	0.003	1.14×10^{-8}	0.66
rs2270204	9	131,042,734	T/G	0.58	0.27	SW15	0.016	0.004	5.47×10^{-6}	0.019	0.004	2.60×10^{-7}	475,247	0.018	0.003	9.56	−0.12	0.018	0.003	9.02×10^{-12}	0.63
rs2163188	10	65,314,711	G/C	0.33	0.50	REEP3	0.015	0.004	8.63×10^{-5}	0.015	0.004	5.03×10^{-5}	388,579	0.015	0.003	6.50	0.72	0.015	0.003	1.67×10^{-8}	0.99
rs5215	11	17,408,630	C/T	0.62	0.66	KCNJ11	0.018	0.004	5.33×10^{-7}	0.014	0.003	1.13×10^{-5}	474,995	0.016	0.002	9.10	0.14	0.016	0.002	3.22×10^{-11}	0.34
rs3026401	11	31,807,524	C/T	0.52	0.81	PAX6	0.016	0.004	3.59×10^{-6}	0.016	0.005	8.37×10^{-4}	392,284	0.016	0.003	6.56	0.20	0.016	0.003	1.13×10^{-8}	0.90
rs506338	11	64,440,920	T/C	0.23	0.71	NRXN2	0.017	0.005	2.30×10^{-4}	0.015	0.003	1.30×10^{-5}	480,264	0.016	0.003	6.65	−0.04	0.016	0.003	9.72×10^{-9}	0.62
rs11607976	11	69,279,111	T/C	0.07	0.30	MYEOV, LINC01488	−0.027	0.008	6.54×10^{-4}	−0.017	0.003	7.72×10^{-7}	479,024	−0.018	0.003	7.01	0.18	−0.018	0.003	3.85×10^{-9}	0.25
rs10899469	11	78,018,313	T/C	0.41	0.19	GAB2	−0.015	0.004	2.51×10^{-5}	−0.016	0.004	1.66×10^{-4}	477,329	−0.015	0.003	6.46	−0.02	−0.015	0.003	1.49×10^{-8}	0.89

(continued)

Table 2 New loci identified by trans-ethnic meta-analysis (continued)

SNP	Chr.[a]	Position[a] (bp)	Alleles REF/ALT	1KG phase 1 ALT freq EAS	EUR	Candidate gene(s)	BBJ GWAS β[b]	s.e.m.	P value	European GWAS β[b]	s.e.m.	P value	N all	Trans-ancestral meta-analysis MANTRA Posterior mean allelic effect	s.e.m.	Log10BF	Log10BF for heterogeneity	Fixed-effect meta-analysis β[b]	s.e.m.	P value	P_{het}[c]
rs10772983	12	17,141,582	C/T	0.68	0.57	SKP1P2	−0.014	0.004	2.42×10^{-4}	−0.012	0.003	4.62×10^{-5}	478,531	−0.013	0.002	6.18	0.34	−0.013	0.002	3.45×10^{-8}	0.73
rs11105839	12	91,237,920	T/A	0.48	0.36	LINC00936, LINC00615	−0.013	0.004	2.84×10^{-4}	−0.013	0.003	2.97×10^{-5}	474,917	−0.013	0.002	6.28	0.16	−0.013	0.002	2.54×10^{-8}	0.96
rs2321882	13	59,451,989	G/C	0.09	0.24	LOC101926897, DIAPH3	0.026	0.006	3.60×10^{-6}	0.016	0.005	3.64×10^{-4}	392,005	0.020	0.004	6.52	−0.07	0.020	0.004	1.32×10^{-8}	0.17
rs1927790	13	96,922,191	T/C	0.36	0.45	HS6ST3	0.010	0.004	5.29×10^{-3}	0.015	0.003	7.87×10^{-7}	480,335	0.013	0.002	6.15	0.05	0.013	0.002	2.70×10^{-8}	0.27
rs12895330	14	33,305,343	G/C	0.24	0.49	AKAP6, NPAS3	0.020	0.004	6.17×10^{-6}	0.016	0.004	1.07×10^{-4}	385,665	0.018	0.003	7.18	0.18	0.018	0.003	3.54×10^{-9}	0.44
rs709400	14	104,149,475	A/G	0.10	0.40	KLC1	−0.010	0.005	5.35×10^{-2}	−0.017	0.003	1.02×10^{-7}	480,397	−0.015	0.003	6.05	−0.40	−0.015	0.003	1.80×10^{-8}	0.31
rs12440086	15	27,038,492	C/A	0.59	0.54	GABRB3, GABRA5	0.017	0.004	3.02×10^{-6}	0.012	0.004	1.74×10^{-3}	392,170	0.015	0.003	6.21	−0.05	0.015	0.003	3.06×10^{-8}	0.33
rs1559677	15	47,738,063	A/G	0.43	0.36	SEMA6D	0.016	0.004	1.06×10^{-5}	0.011	0.003	5.77×10^{-4}	480,342	0.013	0.002	6.01	−0.23	0.013	0.002	3.37×10^{-8}	0.30
rs9302817	16	6,163,936	T/G	0.43	0.18	RBFOX1	−0.015	0.004	6.15×10^{-5}	−0.019	0.005	1.54×10^{-4}	392,179	−0.016	0.003	6.02	0.06	−0.016	0.003	4.56×10^{-8}	0.49
rs7200543	16	15,129,970	A/G	0.38	0.32	PDXDC1	−0.015	0.004	4.16×10^{-5}	−0.015	0.003	1.03×10^{-5}	480,268	−0.015	0.002	7.44	−0.14	−0.015	0.002	1.24×10^{-9}	0.97
rs7195386	16	24,578,458	T/C	0.46	0.53	RBBP6	−0.016	0.004	6.15×10^{-6}	−0.013	0.004	4.89×10^{-4}	392,282	−0.015	0.003	6.58	0.38	−0.015	0.003	1.42×10^{-8}	0.51
rs2307022	16	68,381,978	A/G	0.88	0.65	PRMT7	−0.015	0.006	8.15×10^{-3}	−0.016	0.003	9.86×10^{-7}	480,262	−0.016	0.003	6.26	0.28	−0.016	0.003	3.41×10^{-7}	0.83
rs4925114	17	17,711,270	A/G	0.07	0.60	RAI1	0.029	0.006	1.86×10^{-6}	0.014	0.004	5.11×10^{-4}	392,169	0.019	0.003	6.26	0.18	0.018	0.003	3.04×10^{-8}	0.04
rs4986044	17	21,261,560	C/T	0.58	0.49	MAP2K3, KCNJ12	−0.013	0.004	3.65×10^{-4}	−0.016	0.003	1.42×10^{-7}	476,195	−0.015	0.002	8.32	1.22	−0.015	0.002	2.17×10^{-10}	0.53
rs12150665	17	34,914,787	T/C	0.39	0.40	GGNBP2	−0.015	0.004	5.06×10^{-5}	−0.016	0.003	2.44×10^{-7}	480,289	−0.016	0.002	8.98	0.12	−0.016	0.002	4.67×10^{-11}	0.78
rs6504108	17	46,292,923	C/T	0.78	0.70	SKAP1	−0.019	0.004	1.04×10^{-5}	−0.017	0.003	3.58×10^{-7}	480,392	−0.018	0.003	9.46	0.06	−0.018	0.003	1.51×10^{-11}	0.74
rs9304204	18	36,896,003	A/G	0.83	0.81	LINC00669	0.017	0.004	4.56×10^{-5}	0.018	0.005	2.67×10^{-4}	392,282	0.017	0.003	6.03	0.01	0.017	0.003	4.53×10^{-8}	0.92
rs4129322	18	50,605,642	G/A	0.31	0.10	DCC	−0.020	0.004	6.21×10^{-8}	−0.011	0.007	1.23×10^{-1}	392,300	−0.018	0.003	6.07	0.11	−0.018	0.003	4.23×10^{-8}	0.20
rs7243785	18	52,475,162	A/G	0.38	0.23	DYNAP, RAB22B	0.018	0.004	2.70×10^{-6}	0.018	0.004	6.97×10^{-5}	382,050	0.018	0.003	7.85	0.01	0.018	0.003	7.79×10^{-10}	0.91
rs12454712	18	60,845,884	T/C	0.45	0.40	BCL2	0.019	0.004	4.79×10^{-7}	0.017	0.004	1.46×10^{-5}	401,822	0.018	0.003	9.06	−0.09	0.018	0.003	3.26×10^{-11}	0.77
rs17710386	18	63,461,201	T/C	0.31	0.34	CDH7	0.017	0.004	1.94×10^{-5}	0.012	0.003	2.26×10^{-4}	474,561	0.014	0.003	6.10	−0.13	0.014	0.003	2.77×10^{-8}	0.41
rs4805566	19	30,918,316	T/A	0.69	0.67	ZNF536	0.018	0.004	2.23×10^{-5}	0.018	0.004	8.59×10^{-6}	389,755	0.018	0.003	7.74	0.01	0.018	0.003	7.88×10^{-10}	0.98
rs1884897	20	6,612,832	A/G	0.89	0.60	CASC20, BMP2	0.014	0.006	2.42×10^{-2}	0.016	0.003	6.83×10^{-7}	480,315	0.015	0.003	6.12	0.02	0.015	0.003	3.64×10^{-8}	0.79

Loci that reached genome-wide significance after meta-analysis with European results are shown. Chr, chromosome; REF, reference allele; ALT, alternative allele; BF, Bayes' factor. [a]Positions are based on Human Genome version 19 (hg19), build 37. [b]Alternative alleles were treated as effect alleles. [c]P for heterogeneity between studies estimated by Cochran's Q test.
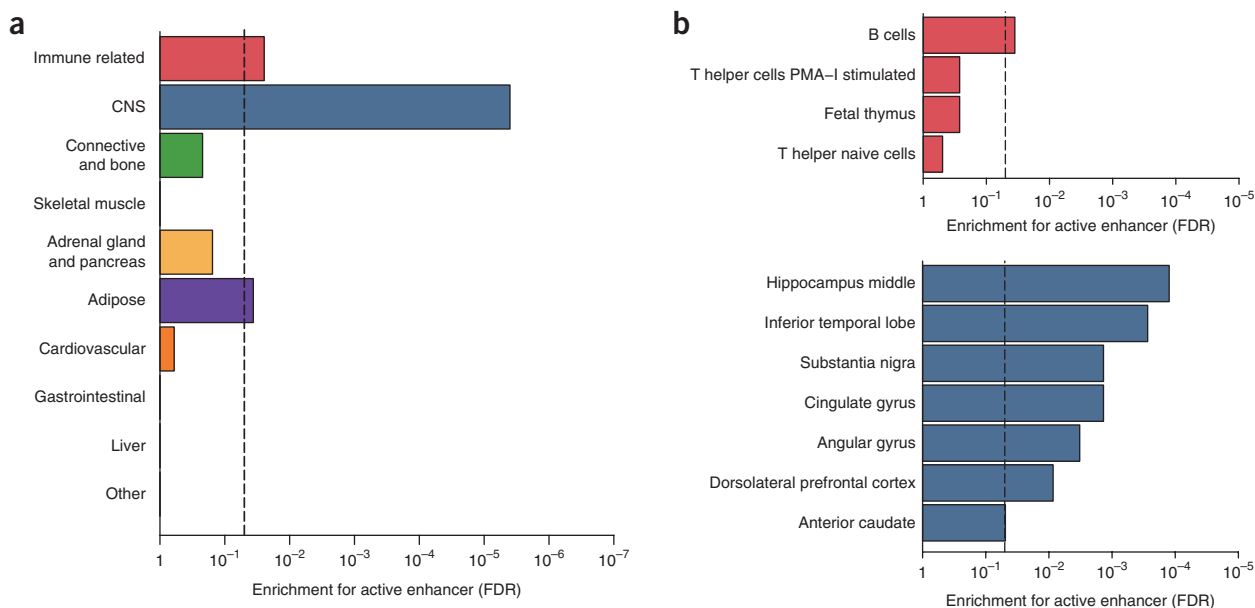
Figure 1 Enrichment of identified variants in active enhancers. (**a**) Enrichment of the variants included in the 99% credible sets for active enhancer in 10 cell groups (**a**) and immune-related cell and CNS groups (**b**). Shown are cell types with $P < 0.05$ in **b**. P values were calculated by 1×10^7 permutations. FDR was estimated using Benjamini–Hochberg method. Vertical dashed lines denote FDR = 0.05.

61 new loci identified from trans-ethnic meta-analyses

To further identify loci associated with BMI, we conducted trans-ethnic meta-analyses using our GWAS and publicly available results of Europeans[6] by MANTRA[18]. We first analyzed data sets without sex stratification (N_{max} = 480,438) and observed significant association signals ($\log_{10}$ Bayes' factor >6) (ref. 19) across 163 loci (**Supplementary Table 10**, **Supplementary Fig. 7** and **Supplementary Data Set 3**). Among them, 54 loci have not been reported (**Table 2**). We note that the previously reported sex-specific loci (*ZBTB10* and *LOC646736*)[6] and an age-dependent locus (*SLC22A3*)[7] reached significance. We observed 66 of 78 (84.6%) autosomal loci that reached the threshold for significance in Japanese individuals, and 70 of 91 (76.9%) previously reported loci by the GIANT consortium[6] remained significant (variants found by sex-stratified GWAS were excluded). This proportion was not statistically different between studies (P for χ^2 test = 0.72). We found that 163 significant variants showed correlation ($r = 0.82$, $P = 7.52 \times 10^{-42}$) and directional consistency in the effect sizes between studies (98.8%, $P = 2.28 \times 10^{-45}$) (**Supplementary Fig. 8**), indicating that BMI-associated loci were generally shared across ethnicities. However, we also observed differences in the effect size at 14 loci ($P_{het} < 0.05/163$).

We also performed meta-analysis of sex-stratified GWAS and found six male-specific loci and one female-specific locus that had not been reported (**Supplementary Table 11** and **Supplementary Fig. 9**). We found distinct but closely located signals in 16q12, where the distances between male-specific rs17795934 and female-specific rs1564981 was less than 1 Mb but in low LD ($r^2 < 0.01$ in both EUR and EAS samples of 1KGP; **Supplementary Fig. 9c**). Comparison of the effect sizes of the significant variants in this trans-ethnic analysis showed strong correlation between sexes ($r = 0.94$, $P = 1.59 \times 10^{-77}$) with perfect directional consistency (**Supplementary Table 12**), again indicating shared genetic components between sexes.

Using the MANTRA result, we constructed 99% credible sets in each identified locus (**Supplementary Table 13**). When we compared the intervals of the genomic region including variants in the 99%

credible set, we observed significant reduction in comparison to the previous trans-ancestral fine-mapping result[6] ($P = 0.04$, Wilcoxon signed rank test) (**Supplementary Table 14**). Annotation of the variants included in the 99% credible sets identified 38 missense variants, including 15 variants found in the Japanese GWAS (**Supplementary Table 15**). Although the majority of the missense variants showed lower posterior probability (PP), we found two likely causal variants with relatively higher PP (*KCNJ11*, p.Val337Ile (GenPept NP_000516.3), PP = 0.75, and *HIVEP1*, p.Met1609Ile (GenPept NP_002105.3), PP = 0.62).

Cell types related to obesity

To investigate specific tissues genetically associated with obesity, we evaluated enrichment of the identified variants included in the 99% credible sets for the cell-type-specific active enhancer constructed by the Roadmap project[20] (**Supplementary Note**). After grouping 64 cell types into 10 categories, we tested for enrichment of the variants. We observed significant enrichment in the active enhancers of the three cell groups (false discovery rate (FDR) < 5%; immune related, central nervous system (CNS) and adipose) (**Fig. 1a** and **Supplementary Tables 16** and **17**). Next, we analyzed each cell type belonging to immune-related and CNS cell-type groups (we only included one cell type in adipose group). We found that all cell types belonging to CNS showed significant enrichment (FDR < 5%, 1.23- to 1.68-fold enrichment), whereas B cells were enriched only in immune-related cells (FDR < 5%, 1.60-fold enrichment) (**Fig. 1b** and **Supplementary Table 18**).

Next, we conducted an enrichment analysis, based on cell-type-specific trimethylation of histone H3 at Lys4 (H3K4me3), in 34 cell types using epiGWAS software[21]. As this analysis requires LD information, we restricted it to the 84 autosomal variants found in the Japanese GWAS. We found three enriched cell types with nominal significance (**Fig. 2** and **Supplementary Table 19**): pancreatic islets ($P = 0.0013$), B cells (CD19+ primary cells, $P = 0.0038$) and brain cells (inferior temporal lobe, $P = 0.0499$). We also applied this method to 97 variants
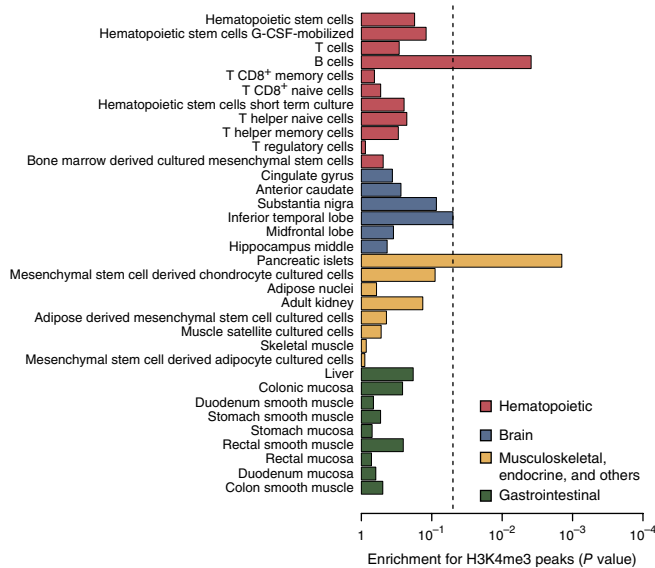
Figure 2 Enrichment of identified variants in H3K4me3 peaks. We analyzed enrichment of identified variants in H3K4me3 peaks of 34 cell types using epi-GWAS software with 1×10^6 permutations. The scale of the x axis corresponds to the P value in the $\log_{10}$ scale for each cell type. The dashed line indicates $P = 0.05$.
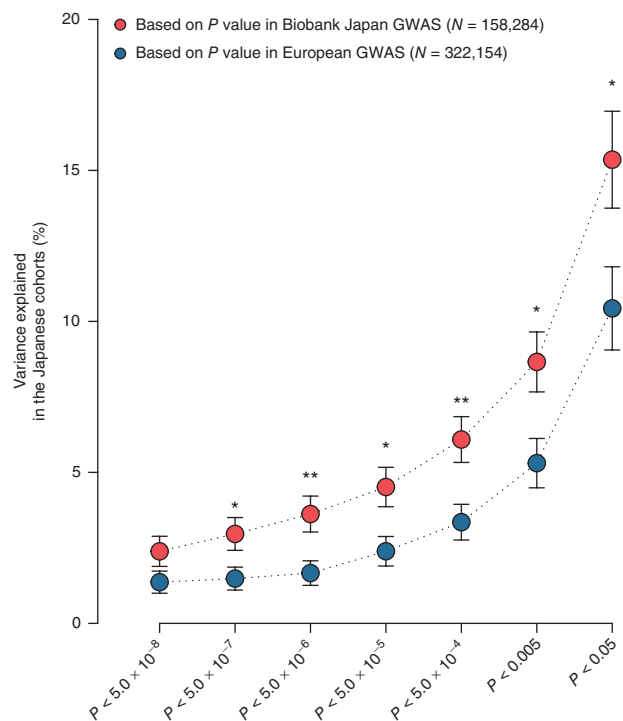

Figure 3 Variance explained by subsets of associated variants in Japanese cohorts. We plotted the estimated variance explained by subsets of variants below the P value threshold from the current Japanese GWAS (red circles) and the GWAS of Europeans (blue circles). Explained variance was estimated from two Japanese population-based cohorts ($N = 5,612$ and $6,434$ in the JPHC and the TMM, respectively). Weighted average from two cohorts is shown. Error bars denote the standard error of estimates. Single asterisks indicate significant differences between the explained variance estimated from each study ($P < 0.05$). Double asterisks denote the Bonferroni-corrected level of significance ($P < 0.05/7$).

found in the GIANT consortium[6] and found that only CD34+ primary cells showed significant enrichment ($P = 0.02$).

Together, our cell-type specificity analyses using the different sets of variants and regulatory markers consistently suggested that B cells and cells belonging to CNS were relevant to body weight regulation. Although the previous studies highlighted the tissue in the CNS[6,22], our results suggest that gene regulation in immune-related cells (B cells), adipose and pancreas may also have a role in the pathogenesis of obesity. We annotated the lead variants of the identified loci with expression quantitative trait loci (eQTL)[23,24] in candidate cell types and found 69 overlaps (**Supplementary Table 20**).

Variance explained by variants in the GWAS

The lead variants of the 83 identified loci found in Japanese GWAS explained 2.8% of the phenotypic variance (**Supplementary Table 3**), which is similar to that found in Europeans (2.7% by 97 variants)[6]. Although we included five variants on the X chromosome that were not analyzed in the GWAS of Europeans, they explained only 0.09% of the phenotypic difference.

To assess the polygenic effects, we estimated the variance explained by all the GWAS variants in the Japanese population-based cohorts using the genomic restricted maximum likelihood (GREML) method implemented in GCTA[25]. The estimated proportion of the phenotypic variance explained by autosomal variants was $29.8 \pm 3.4\%$ (s.e.m.). This fraction was similar to that found in the European population ($27.4 \pm 2.5\%$)[8]. When we estimated the explained variance of each chromosome[26], a weak linear relationship with chromosome lengths was observed ($P = 7.31 \times 10^{-3}$, $r^2 = 0.30$) (**Supplementary Fig. 10**). The X chromosome explained $2.3 \pm 0.8\%$ of the phenotypic variance.

We further examined the variance explained in the Japanese cohorts by examining a subset of the variants under a certain P value threshold according to our GWAS results of Japanese and those of Europeans[6] using variants analyzed in both studies. As expected, the explained variance could be increased by lowering the threshold of the P value (**Fig. 3** and **Supplementary Fig. 11**). When we compared our GWAS

with the GWAS of Europeans, estimates based on our GWAS better explained the phenotypic variance in Japanese cohorts (on average, 1.81-fold higher; range: 1.47–2.17).

Pathway analysis

Using the results of GWAS in the Japanese individuals, we performed MAGENTA[27] to investigate biological pathways associated with obesity and found three significant pathways (ethanol oxidation, glycolysis gluconeogenesis and maturity-onset diabetes of the young; FDR: $q < 0.05$) (**Supplementary Tables 21** and **22**). We found that the neurotrophin signaling pathway, which was previously reported in Europeans[6], showed nominal enrichment in our study ($P = 0.0052$, $q = 0.25$).

Stratified analysis by T2D

As previous reports suggested that the effect sizes of BMI at a few loci differed between individuals with T2D and individuals who did not develop diabetes (*TCF7L2* (ref. 6) and *CDKAL1* (ref. 28)), we evaluated 193 variants by separating GWAS samples into a group with T2D ($n = 31,609$) and a group without diabetes ($n = 125,816$). We observed significant directional consistency (96.4%, $P = 1.29 \times 10^{-26}$, sign test) and a significant correlation between the effects ($r = 0.85$, $P = 8.69 \times 10^{-55}$). Nevertheless, we found evidence of heterogeneity not only at two reported loci but also at five other loci (*HHEX*, *IGFBP2*, *KCNQ1*, *CDKN2B* and *DUSP9*; $\alpha = 0.05/193$) (**Supplementary Figs. 12** and **13**
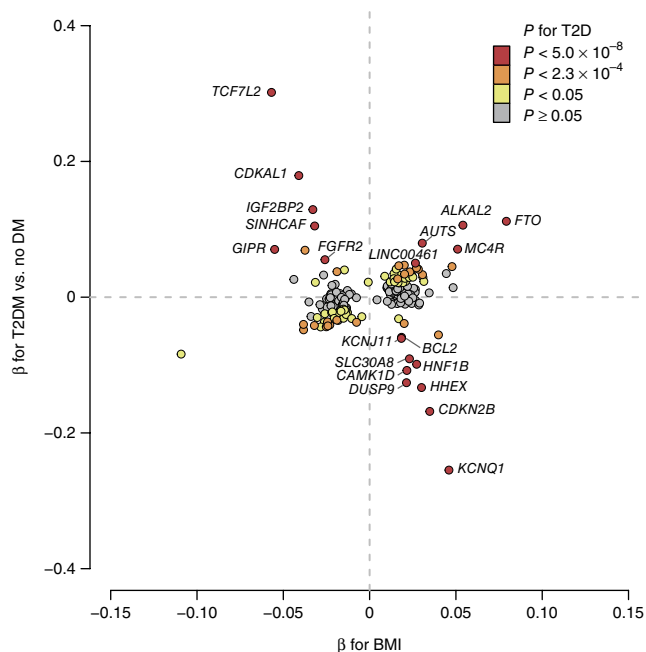
Figure 4 Scatter plot of the effect sizes for BMI and T2D. Effect sizes of 193 identified variants for BMI (x axis) and T2D (y axis). Plots were colored according to the significance level for T2D.

and **Supplementary Table 23**). The effects of these variants were stronger in individuals with T2D. Considering that observed effect size of *TCF7L2* in Japanese individuals with T2D (β = 0.12, s.e.m. = 0.02) was similar to that in Europeans[6] (β = 0.11, s.e.m. = 0.01; $P_{difference}$ = 0.52), some variants may have different impact according to diabetes status.

We also evaluated the effects of BMI-associated variants on T2D susceptibility. Of the 193 variants, we found 20 that showed GWS levels of association for T2D (**Supplementary Table 24**). When we compared the effects of both traits, we observed a bidirectional relationship (**Fig. 4**). A positive relationship was observed at five loci. A negative relationship was found at 15 loci. Although epidemiological studies reported that a higher BMI was a risk for T2D[1], our results indicated that the relationship between genetic variants and T2D susceptibility is more complicated. Of the variants that showed negative relationships, many were reportedly associated with glycemic and/or insulin regulation[29,30] and insulin processing[30]. Although some loci (*SINHCAF* and *FGFR2*) have not been reported to be associated with these traits, this observation implied that they might also share biological pathways.

To assess the possibility of biased estimation in our BMI GWAS due to index event bias, we compared the effect sizes for BMI at these loci between GWAS and Japanese population-based cohorts. Among 16 variants evaluated in both data sets, we observed a perfect directional consistency and a significant correlation in effect sizes (r = 0.94, P = 6.63 × 10^{-8}) (**Supplementary Fig. 14**) without significant heterogeneity (P_{het} > 0.05/16), suggesting a low possibility of false positives due to case-mix samples in our GWAS.

Genetic correlation analysis

We evaluated the genetic correlations between the results of this study and 33 other GWAS of Asians using bivariate LD score regression[15] (**Fig. 5**, **Supplementary Table 25** and **Supplementary Note**).

We found six significant positive correlations (FDR q < 0.05), including correlations with T2D, cardiovascular diseases (ischemic stroke, myocardial infarction and peripheral arterial diseases), asthma and ossification of posterior longitudinal ligament of the spine (OPLL). Significant negative correlations were observed in adolescent idiopathic scoliosis (AIS), schizophrenia and rheumatoid arthritis (RA). Genetic correlations with T2D and coronary artery disease have also been reported in Europeans[15]. Other genetic correlations have not been previously reported. To compare our results with those in Europeans, we searched LD Hub[31] and found seven results of genetic correlations that were also evaluated in the current study (**Supplementary Table 26**). Similar r_g values were observed between our study and those in Europeans in the traits reported as significant in our analysis (cardiovascular diseases, RA, schizophrenia and T2D).

Our cell-type-specific analyses suggested involvement of lymphocytes in the regulation of body weight. Although the genetic involvement of lymphocytes in obesity has not been elucidated, previous clinical studies have shown positive correlations between BMI and the counts of leukocytes and their subtypes in peripheral blood[32,33]. We hypothesized that the counts of these cells might be influenced by genetic components that also regulate BMI. We analyzed the genetic correlation between BMI and three major hematological traits (N_{GWAS} = 110,397–111,268) (**Supplementary Note**) and found a significant positive correlation with white blood cell (WBC) counts (P = 8.50 × 10^{-5}, r_g = 0.14) (**Fig. 5b** and **Supplementary Table 27**). Next, we evaluated the genetic correlations between BMI and the counts of five WBC subtypes (N_{GWAS} = 63,197). Of these correlations, only lymphocytes showed a significant genetic correlation with BMI (P = 6.46 × 10^{-5}, r_g = 0.18). These results suggested that BMI and lymphocytes shared genetic components that regulate both traits.

DISCUSSION

Through a large GWAS of non-European subjects, we found 85 BMI-associated loci, which accounted for ~3% of the phenotypic differences in Japanese volunteers. Of the 51 newly identified loci, five were found on the X chromosome, and two had sex-dependent effects. Conditional analysis identified four additional signals within the loci. Furthermore, combining the results with GWAS of Europeans identified 61 more novel loci, bringing the total number of known susceptibility loci to >200.

We showed evidence of genetic differences between Europeans and East Asians. First, comparisons of the effect sizes suggested that susceptibility loci were generally shared between Europeans and East Asians, which was comparable to trans-ethnic studies for other traits[34,35]. However, variants that were grouped on the basis of the associations in our GWAS were able to explain larger phenotypic differences in Japanese cohorts compared with the variants selected from the GWAS of Europeans. We considered four possible explanations: (i) different causal variants between populations, (ii) differences in the effect sizes of causal variants across ethnicities, (iii) difference in LD structure between causal variants and GWAS SNPs and (iv) differences in allele frequencies of the causal variants. Second, we observed association signals, including second signals, that were independent of previously reported associations. This could be the result of independent causal variants or distinct LD structures reflecting the same causal variants across ethnicities. Our results emphasize the importance of investigating multiple populations to discover novel genetic components. Third, pathway analysis indicated that the neurotrophin signaling pathway influenced BMI across ethnicities, but we also found three pathways that were not reported in Europeans. Considering selection at *ADH1B* and *ALDH2* in East Asians[36,37], the
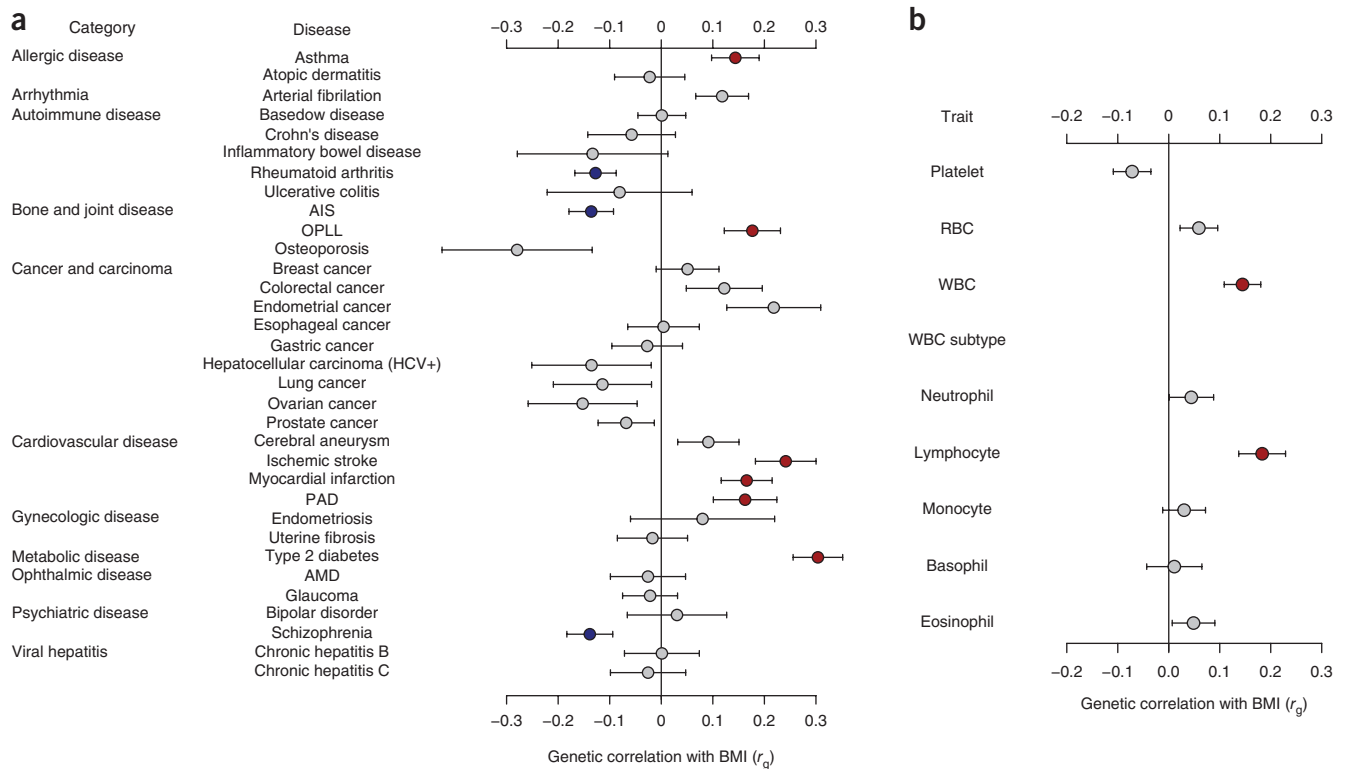
Figure 5 Genetic correlations between BMI and analyzed traits. (**a**,**b**) Genetic correlations between BMI and diseases (**a**) or hematological traits (**b**) estimated from cross-trait LD score regression. Error bars indicate the standard error of estimates. Significant (FDR $q < 0.05$) genetic correlations are shown in red and blue for positive and negative relationships, respectively. PAD, peripheral artery disease; AMD, age-related macular degeneration; AIS, adolescent idiopathic scoliosis; RBC, red blood cell; WBC, white blood cell; HCV, hepatitis C virus.

ethanol oxidation and glycolysis gluconeogenesis pathways—which include both *ADH* and *ALDH* family genes—may have a larger impact on the BMI of East Asians. These differences highlight the need for the use of diverse populations to better understand the genetic underpinnings of BMI.

Our results suggest two possible ways that lymphocytes are involved in the regulation of body weight. First, gene regulation in B cells might influence body weight regulation, because the identified variants were enriched in CD19+ cells-specific chromatin marks. The previous investigation of Europeans using DEPICT suggested that candidate genes underlying BMI-associated loci were highly expressed in precursor cells B lymphoid and lymphoid progenitor cells[6]. Substantial biological evidence has shown the involvement of immune cells in obesity via inflammation and insulin resistance[38]. Recent studies have also indicated that B cells have an important role in obesity through activation of proinflammatory macrophages and T cells and the production of IgG antibodies in insulin resistance[39]. Changes in gene regulation in B cells might affect BMI through these mechanisms. Second, the increased number of leukocytes in people with obesity[32,33] might be influenced by genetic variants, as we found a significant positive genetic correlation between BMI and counts of WBCs, especially lymphocytes. Our observations indicate that the same genetic variants were associated with increases (or decreases) in BMI and counts of these cells. Because genetic correlations could not pinpoint the shared causal variants, further study is needed to elucidate the genetic components of both traits and clarify their underlying biological pathways. Our results provide genetic evidence for the involvement of lymphocytes in the etiology of obesity.

Our cell-type-specificity analysis also suggested that cells in pancreatic islets and adipose tissue are relevant to obesity. When we looked up the existing genetic correlations in LD Hub[31], we found significant positive genetic correlations with BMI in measures related to insulin secretion fasting insulin level and homeostatic model assessment of beta cell function (HOMA-β) and insulin resistance homeostatic model assessment of insulin resistance (HOMA-IR). These facts indicated that BMI shares heritability with insulin modulation. A previous investigation indicated the enrichment of the BMI heritability in pancreatic islets[22]. Visceral adipose tissue is insulin responsive in individuals with obesity and is a major driver of insulin resistance via its chronic inflammation[40]. Furthermore, a recent study indicated that the variant at *FTO* region leads to *IRX3* and *IRX5* disruption during adipocyte differentiation and causes body weight gain through a cell-autonomous shift from white adipose browning and thermogenesis to lipid storage[41]. Our results may link this biological knowledge and suggest that gene regulation in these tissues may influence body weight.

We investigated the genetic correlations between BMI and other traits and found significant correlations with nine diseases. As we were able to replicate the correlations reported in Europeans, genetic correlation may be reproducible across ethnicities. Indeed, a genetic correlation between BMI and schizophrenia was also suggested in Europeans ($P = 0.0002$, $r_g = −0.10$)[15]. In observational studies, consistent relationships were found between BMI and T2D, cardiovascular diseases, asthma, OPLL and AIS[1,42–44]; however, to our knowledge, the relationships between BMI and RA and between BMI and schizophrenia have been controversial[45–48]. Besides the positive genetic correlation between BMI and T2D, we found that some BMI-increasing variants reduce the risk of T2D. Considering that T2D confers

risk of multiple diseases, BMI-associated variants may not uniformly influence complex diseases. Our results refine the list of diseases that share heritability with obesity and provide insight into the contribution of genetic components in complex diseases.

In conclusion, we performed a large-scale GWAS of BMI and found 112 novel loci. These results highlight that investigation of diverse populations helps to identify new loci and narrow down the candidate genomic region. Our findings provide insight into the links between BMI and complex diseases and provide genetic evidence that lymphocytes have a role in the pathogenesis of obesity.

URLs. BioBank Japan, https://biobankjp.org/english/index.html; JPHC, http://epi.ncc.go.jp/en/jphc/index.html; TMM, http://www.megabank.tohoku.ac.jp/english/; PLINK, https://www.cog-genomics.org/plink2; MACH, http://csg.sph.umich.edu//abecasis/MaCH/; EpiGWAS, https://www.broadinstitute.org/mpg/epigwas/; ANNOVAR, http://annovar.openbioinformatics.org/en/latest/; SHAPEIT, https://mathgen.stats.ox.ac.uk/genetics_software/shapeit/shapeit.html; HaploReg, http://www.broadinstitute.org/mammals/haploreg/haploreg.php; GTEx, http://www.gtexportal.org/home/; R, https://www.r-project.org/; MAGENTA, https://www.broadinstitute.org/mpg/magenta/; GIANT consortium; https://www.broadinstitute.org/collaboration/giant/index.php/GIANT_consortium; Locuszoom, http://locuszoom.sph.umich.edu/locuszoom/; 1000 Genomes Project, http://www.1000genomes.org/; HapMap project, http://hapmap.ncbi.nlm.nih.gov/; GCTA, http://cnsgenomics.com/software/gcta/; LDSC, https://github.com/bulik/ldsc/; GWAS catalog, https://www.ebi.ac.uk/gwas/; ROADMAP Epigenomics Project, http://www.roadmapepigenomics.org/; LD Hub, http://ldsc.broadinstitute.org/; bedtools, http://bedtools.readthedocs.io/en/latest/; JGA, http://trace.ddbj.nig.ac.jp/jga/index_e.html; (NBDC) Human Database, https://humandbs.biosciencedbc.jp/en/.

METHODS
Methods, including statements of data availability and any associated accession codes and references, are available in the online version of the paper.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

AUTHOR CONTRIBUTIONS
M.A., Y.K. and M. Kubo conceived and designed the study. K.M., M.H. and M. Kubo collected and managed the BBJ sample. M. Iwasaki, T.Y., N.S. and S.T. collected and managed JPHC sample and information. T.H., K.T., A.S., A.H., N.M. and M.Y. collected and managed the TMM sample. Y.M. and M. Kubo performed genotyping. M.A., M. Kanai, Y.K. and A.T. performed statistical analysis. S.I., M. Ikeda and N.I. contributed to data acquisition. Y.O., A.T., Y.K. and M. Kubo supervised the study. M.A., Y.O., Y.K. and M. Kubo wrote the manuscript.

1. Haslam, D.W. & James, W.P.T. Obesity. *Lancet* **366**, 1197–1209 (2005).
2. Wilson, P.W.F., D'Agostino, R.B., Sullivan, L., Parise, H. & Kannel, W.B. Overweight and obesity as determinants of cardiovascular risk: the Framingham experience. *Arch. Intern. Med.* **162**, 1867–1872 (2002).
3. Renehan, A.G., Tyson, M., Egger, M., Heller, R.F. & Zwahlen, M. Body-mass index and incidence of cancer: a systematic review and meta-analysis of prospective observational studies. *Lancet* **371**, 569–578 (2008).
4. Speliotes, E.K. *et al.* Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index. *Nat. Genet.* **42**, 937–948 (2010).
5. Wen, W. *et al.* Meta-analysis of genome-wide association studies in East Asian–ancestry populations identifies four new loci for body mass index. *Hum. Mol. Genet.* **23**, 5492–5504 (2014).
6. Locke, A.E. *et al.* Genetic studies of body mass index yield new insights for obesity biology. *Nature* **518**, 197–206 (2015).
7. Winkler, T.W. *et al.* The influence of age and sex on genetic associations with adult body size and shape: a large-scale genome-wide interaction study. *PLoS Genet.* **11**, e1005378 (2015).
8. Yang, J. *et al.* Genetic variance estimation with imputed variants finds negligible missing heritability for human height and body mass index. *Nat. Genet.* **47**, 1114–1120 (2015).
9. Yoon, K.H. *et al.* Epidemic obesity and type 2 diabetes in Asia. *Lancet* **368**, 1681–1688 (2006).
10. Nakamura, Y. The BioBank Japan Project. *Clin. Adv. Hematol. Oncol.* **5**, 696–697 (2007).
11. Nagai, A. *et al.* Overview of the BioBank Japan Project: study design and profile. *J. Epidemiol.* **27**, S2–S8 (2017).
12. Hirata, M. *et al.* Cross-sectional analysis of BioBank Japan clinical data: a large cohort of 200,000 patients with 47 common diseases. *J. Epidemiol.* **27**, S9–S21 (2017).
13. Abecasis, G.R. *et al.* A map of human genome variation from population-scale sequencing. *Nature* **467**, 1061–1073 (2010).
14. Bulik-Sullivan, B.K. *et al.* LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat. Genet.* **47**, 291–295 (2015).
15. Bulik-Sullivan, B. *et al.* An atlas of genetic correlations across human diseases and traits. *Nat. Genet.* **47**, 1236–1241 (2015).
16. Trivellin, G. *et al.* Gigantism and acromegaly due to Xq26 microduplications and *GPR101* mutation. *N. Engl. J. Med.* **371**, 2363–2374 (2014).
17. Lek, M. *et al.* Analysis of protein-coding genetic variation in 60,706 humans. *Nature* **536**, 285–291 (2016).
18. Morris, A.P. Transethnic meta-analysis of genomewide association studies. *Genet. Epidemiol.* **35**, 809–822 (2011).
19. Wang, X. *et al.* Comparing methods for performing trans-ethnic meta-analysis of genome-wide association studies. *Hum. Mol. Genet.* **22**, 2303–2311 (2013).
20. Roadmap Epigenomics Consortium. Integrative analysis of 111 reference human epigenomes. *Nature* **518**, 317–330 (2015).
21. Trynka, G. *et al.* Chromatin marks identify critical cell types for fine mapping complex trait variants. *Nat. Genet.* **45**, 124–130 (2013).
22. Finucane, H.K. *et al.* Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nat. Genet.* **47**, 1228–1235 (2015).
23. GTEx Consortium. The Genotype-Tissue Expression (GTEx) project. *Nat. Genet.* **45**, 580–585 (2013).
24. Westra, H.-J. *et al.* Systematic identification of *trans* eQTLs as putative drivers of known disease associations. *Nat. Genet.* **45**, 1238–1243 (2013).
25. Yang, J., Lee, S.H., Goddard, M.E. & Visscher, P.M. GCTA: a tool for genome-wide complex trait analysis. *Am. J. Hum. Genet.* **88**, 76–82 (2011).
26. Yang, J. *et al.* Genome partitioning of genetic variation for complex traits using common SNPs. *Nat. Genet.* **43**, 519–525 (2011).
27. Segrè, A.V., Groop, L., Mootha, V.K., Daly, M.J. & Altshuler, D. Common inherited variation in mitochondrial genes is not enriched for associations with type 2 diabetes or related glycemic traits. *PLoS Genet.* **6**, e1001058 (2010).
28. Okada, Y. *et al.* Common variants at *CDKAL1* and *KLF9* are associated with body mass index in east Asian populations. *Nat. Genet.* **44**, 302–306 (2012).
29. Scott, R.A. *et al.* Large-scale association analyses identify new loci influencing glycemic traits and provide insight into the underlying biological pathways. *Nat. Genet.* **44**, 991–1005 (2012).
30. Dimas, A.S. *et al.* Impact of type 2 diabetes susceptibility variants on quantitative glycemic traits reveals mechanistic heterogeneity. *Diabetes* **63**, 2158–2171 (2014).
31. Zheng, J. *et al.* LD Hub: a centralized database and web interface to perform LD score regression that maximizes the potential of summary level GWAS data for SNP heritability and genetic correlation analysis. *Bioinformatics* **33**, 272–279 (2017).
32. Julius, S., Egan, B.M., Kaciroti, N.A., Nesbitt, S.D. & Chen, A.K. In prehypertension leukocytosis is associated with body mass index but not with blood pressure or incident hypertension. *J. Hypertens.* **32**, 251–259 (2014).
33. Ilavská, S. *et al.* Association between the human immune response and body mass index. *Hum. Immunol.* **73**, 480–485 (2012).

34. Okada, Y. *et al.* Genetics of rheumatoid arthritis contributes to biology and drug discovery. *Nature* **506**, 376–381 (2014).

35. Liu, J.Z. *et al.* Association analyses identify 38 susceptibility loci for inflammatory bowel disease and highlight shared genetic risk across populations. *Nat. Genet.* **47**, 979–986 (2015).

36. Oota, H. *et al.* The evolution and population genetics of the *ALDH2* locus: random genetic drift, selection, and low levels of recombination. *Ann. Hum. Genet.* **68**, 93–109 (2004).

37. Han, Y. *et al.* Evidence of positive selection on a class I ADH locus. *Am. J. Hum. Genet.* **80**, 441–456 (2007).

38. Johnson, A.M.F. & Olefsky, J.M. The origins and drivers of insulin resistance. *Cell* **152**, 673–684 (2013).

39. Winer, D.A. *et al.* B cells promote insulin resistance through modulation of T cells and production of pathogenic IgG antibodies. *Nat. Med.* **17**, 610–617 (2011).

40. Winer, D.A., Winer, S., Chng, M.H., Shen, L. & Engleman, E.G. B lymphocytes in obesity-related adipose tissue inflammation and insulin resistance. *Cell. Mol. Life Sci.* **71**, 1033–1043 (2014).

41. Claussnitzer, M. *et al.* FTO obesity variant circuitry and adipocyte browning in humans. *N. Engl. J. Med.* **373**, 895–907 (2015).

42. Hjellvik, V., Tverdal, A. & Furu, K. Body mass index as predictor for asthma: a cohort study of 118,723 males and females. *Eur. Respir. J.* **35**, 1235–1242 (2010).

43. Inamasu, J., Guiot, B.H. & Sachs, D.C. Ossification of the posterior longitudinal ligament: an update on its biology, epidemiology, and natural history. *Neurosurgery* **58**, 1027–1039, discussion 1027–1039 (2006).

44. Tam, E.M. *et al.* Lower muscle mass and body fat in adolescent idiopathic scoliosis are associated with abnormal leptin bioavailability. *Spine* **41**, 940–946 (2016).

45. Turesson, C., Bergström, U., Pikwer, M., Nilsson, J.-Å. & Jacobsson, L.T.H. A high body mass index is associated with reduced risk of rheumatoid arthritis in men, but not in women. *Rheumatology* **55**, 307–314 (2016).

46. Qin, B. *et al.* Body mass index and the risk of rheumatoid arthritis: a systematic review and dose-response meta-analysis. *Arthritis Res. Ther.* **17**, 86 (2015).

47. Zammit, S. *et al.* Height and body mass index in young adulthood and risk of schizophrenia: a longitudinal study of 1 347 520 Swedish men. *Acta Psychiatr. Scand.* **116**, 378–385 (2007).

48. Wyatt, R.J., Henter, I.D., Mojtabai, R. & Bartko, J.J. Height, weight and body mass index (BMI) in psychiatrically ill US Armed Forces personnel. *Psychol. Med.* **33**, 363–368 (2003).

ONLINE METHODS

Subjects. We used BMI and genome-wide variant data obtained from the BBJ[10,11], which has enrolled approximately 200,000 individuals and consists of 47 diseases[12]. For the GWAS, we set the eligibility criteria as follows: (i) age ≥18, (ii) height and weight registered, and (iii) height within threefold the interquartile range (IQR). For the quality control of GWAS, we excluded samples with a call rate ≤0.98. Closely related samples, which were estimated using identity by state (IBS), were excluded by visual inspection. We performed principal component analysis (PCA) for genotype using an in-house program based on the algorithm implemented by smartpca[49], and we excluded outliers from the East Asian cluster. Finally, we calculated the Z-score for height by linear regression using age, sex, status of 47 diseases, and the top 10 principal components (PCs) and excluded individuals outside of ±4 s.d. for the purpose of quality control of the phenotype data.

Subjects of the population-based prospective cohorts were independently recruited in the JPHC and the TMM. The inclusion and exclusion criteria for the analysis were the same as for the GWAS. After exclusion of individuals who lacked clinical information including age, sex, height and body weight, we excluded related individuals with PI_HAT > 0.125, which was calculated using PLINK[50]. To calculate the Z-score for height, we used age, sex, and the top 10 PCs as covariates and excluded outliers using the same criterion used in the GWAS. We obtained approval from ethics committees at all collaborating facilities. Informed consent was obtained from all participants before enrollment. Characteristics of the cohorts are shown in the **Supplementary Note**.

Phenotype. The BMI information obtained from medical records was standardized using a rank-based inverse-normal transformation. We first calculated residuals for the log-transformed BMI by linear regression analysis using age, age-squared, sex, status of 47 registered diseases, and the top 10 PCs as covariates. Then, we transformed residuals using a rank-based inverse-normal transformation. In the cohort studies, linear regression was performed using age, age[2], sex, and the top 10 PCs for the log-transformed BMI in each cohort, and we transformed residuals using a rank-based inverse-normal transformation.

Genotyping and imputation. The subjects of the GWAS were genotyped using the Illumina HumanOmniExpressExome BeadChip or in combination of Illumina HumanOmniExpress and HumanExome BeadChips. We aligned the probe sequence in the manifest files of genotyping array to GRCh37.3 reference using BLAST to convert genotypes into forward strand, and phased haplotypes using MACH[51]. Imputation was performed by minimac (v0.1.1). We used EAS samples of 1KGP (phase1v3) as a reference. We performed quality control on the reference panel as follows: we excluded 11 closely related individuals estimated from IBS and excluded variants with a minor allele frequency (MAF) < 1.0% and a P value for Hardy–Weinberg equilibrium ≤ 1.0 × 10[−6] from the reference panel. After the imputation, we used SNPs with an imputation quality of Rsq ≥ 0.7 for the GWAS.

In the population-based cohorts, DNA samples were genotyped using the Illumina HumanOmniExpressExome BeadChip. We phased haplotypes using SHAPIT[52] (v2.r778) and performed imputation using Minimac3 (1.0.13)[51] with the same reference panel used for the GWAS. For replication purposes, we allowed Rsq ≥ 0.5 throughout the study.

To investigate the X chromosome, we called genotypes using BeadStudio. First, we generated genoplots using only female samples. After that, we added male samples to genoplots and called genotypes. Genotypes called as heterozygote were treated as missing in male samples. For the genotype QC, we excluded variants with a MAF < 0.05% and a variant call rate ≤ 0.99 in either male or female samples. We also excluded variants with P value for Hardy–Weinberg equilibrium ≤ 1.0 × 10[−6] in female samples. Haplotype phasing and imputation were performed separately for males and females in the GWAS. Allelic dosages were imputed from 0 to 2 in male samples under the assumption of full dosage compensation. In the population based-cohorts, pre-phasing was performed for males and females together using SHAPEIT, and the data were imputed separately. The pseudo-autosomal region was excluded from the reference before imputation.

Association analyses. The GWAS was performed using imputed allele dosages and fitted to an additive genetic model by mach2qtl (ref. 51). We performed a stepwise conditional analysis to determine independent association signals near associated loci (lead variants ±1 Mb) using samples included in the GWAS until the top associated variants fell below the GWS level in each stepwise procedure. Meta-analysis was performed using the inverse-variance method under the assumption of the fixed effect model. Heterogeneity between studies and P for difference in the stratified analyses were calculated by Cochran's Q test. To investigate associations of the variants on X chromosome, we performed association analyses using mach2qtl in male and female, respectively. After that, we integrated the results of each sex by inverse-variance method under the assumption of the full dosage compensation.

We estimated the variance explained by each identified variant using the following formula using the allele frequency (f) estimated in the GWAS and estimates of the additive effect (β) in meta-analysis: Explained variance = β^2 $(1 − f)$ $2f$. To estimate the additive explained variance of 83 independent lead variants of identified loci, the explained variance of each individual variant was summed.

R (v3.1.3) was used for these analyses. Regional association plots were generated by Locuszoom[53] (v1.3) and R.

Estimation of biases in the GWAS using LD score regression. To evaluate biases resulting from population stratification and cryptic relatedness, we performed LD score regression[14] with the default settings. LD scores for the East Asian population provided with software were used.

Sex-stratified GWAS. For the sex-stratified GWAS, the samples analyzed in the GWAS were divided by sex. We recalculated phenotypes in each sex stratum using residuals calculated by linear regression analysis using age, age-squared, status of diseases, and the top 10 PCs as covariates. For the evaluation in the replication sets (population-based cohorts), we selected variants at each locus that showed $P < 1.0 \times 10^{-6}$ unless loci showed this level of association in the primary GWAS. As a result, we selected 12 variants for males and 3 variants for females.

T2D stratified analysis. We selected 31,609 individuals with T2D from samples analyzed in the GWAS. Individuals with type 1 diabetes, maturity-onset diabetes of the young, mitochondrial diabetes, or gestational diabetes were excluded from this analysis. We analyzed 193 lead variants of the loci that reached genome-wide significance level. In case the lead variants were different between Japanese-only analysis and trans-ethnic analysis, we choose lead variants found in the Japanese GWAS. The BMI phenotype was re-evaluated from the residuals calculated by a linear regression analysis using age, age[2], sex, status of 46 diseases, and the top 10 PCs after stratification. We tested association for T2D susceptibility using logistic regression analysis with 10 PCs as covariates.

Trans-ethnic meta-analysis. To carry out a meta-GWAS using our GWAS and publicly available results from Europeans[6] conducted by the GIANT consortium, we downloaded the summary statistics from their website. After adding the positional information using the HapMap project phase2 to the summary statistics of the GIANT consortium, chromosomal position was converted to hg19 by the lift-over tool. We excluded the variants which were duplicated, and without allele frequency information, and included the variants which were analyzed in the both studies. Finally, we analyzed 2,078,865 autosomal variants in this study. We also conducted the meta-analyses of the sex-stratified GWAS in the same way.

For the meta-analyses, we used MANTRA[18] (v2) software that was designed for trans-ethnic meta-analysis with allowing heterogeneity in allelic effects. Prior model of the relatedness between the studies was estimated by dmatcal script in the software using the allele frequency of the analyzed variants. We considered $\log_{10}BF > 6$ as a significant threshold according to the previous simulation study[19]. To compare with result of the fixed effect model, we also performed meta-analysis using inverse variance method, and estimated heterogeneity using Cochran's Q test only for the significant loci.

Credible sets. Using the results from the trans-ethnic meta-analyses, we constructed the set of the variants that are likely to be causal based on the MANTRA[18] analysis. For the each locus, we considered the window which was

determined by position of the lead variants ±500 kb, and calculated the posterior probability (PP) of the causal variant for the jth variant under the assumption of a single causal variant in each locus using the following formula:

$$PP_j = \frac{BF_j}{\Sigma_k BF_k}$$

where k denotes the variants in the window. After that, we selected the smallest set of variants to account for more than or equal to 99% of the PP in each locus.

Genetic correlation. To estimate the genetic correlations, bivariate LD score regression[15] was conducted using the results from the current GWAS of all samples, the GWAS of 33 complex diseases, and the hematological traits with the LD scores for the East Asian population (**Supplementary Note**). All the GWAS that were analyzed only included Asians. We calculated the FDR using the Benjamini–Hochberg procedure.

GREML analysis. To estimate the variance explained by autosomal variants, we selected imputed variants with Rsq ≥ 0.7 and generated a genetic relatedness matrix (GRM) for the Japanese cohorts by GCTA (v1.25)[25]. For the GREML analysis, we used samples whose genetic relationship was less than 0.05. Then, 5,612 and 6,434 samples in the JPHC and the TMM, respectively, were included in these analyses. The log-transformed BMI values were then standardized to be Z-scores adjusted for age, age-squared, and sex in each cohort. Explained variance was estimated using the reml function implemented in GCTA software with default settings. We also estimated the variance explained by each chromosome using the reml function and the "–reml-no-constrain" option. The X chromosome was analyzed under the assumption that there was full dosage compensation in females. We calculated weighted averages from results of two cohorts using inverse-variance weighting. In the regression analysis for chromosome length, negative values were constrained at zero after calculation of weighted averages.

Pleiotropy. We downloaded the NHGRI GWAS catalog on 19 February 2016. We converted chromosomal positions from hg38 to hg19 using the liftOver tool in the UCSC genome browser. To prioritize other reported traits, we manually selected variants that satisfied one of the following criterion for each identified variant from the GWAS catalog: (i) variants with $r^2 \geq 0.5$ in East Asians if the publication included Asians, (ii) variants with $r^2 \geq 0.5$ in both East Asians and other population if the original report not studied East Asian, and (iii) the physical distance from lead variants in the present GWAS to variants reported in the GWAS catalog was within 25 kb. LD was calculated according to 1KGP.

Tissue-specific active enhancer enrichment. To investigate cell types and cell groups that are relevant to obesity, we conducted an enhancer enrichment analysis using variants included in the 99% credible sets constructed by trans-ethnic meta-analysis. As there are several types of enhancer information, we annotated the variants included in the trans-ethnic GWAS and compared the overlaps of three different types of enhancers (15-state model state 7; 18-state model state 9; and DNaseI-accessible enhancer) in each cell type. We observed that active enhancers in the ChromHMM 18-state model (state 9) show the lowest overlap across cell types among analyzed enhancers (**Supplementary Note**). Therefore, we used the active enhancer of ChromHMM 18-state model for this analysis.

For analysis of cell-type and cell-group specificity, we excluded the cell types derived from cell lines, labeled as cultured cells, or belonging to embryonic stem cell or induced pluripotent stem cell groups. To test statistical significance, we compared the overlaps between variants in the 99% credible sets and the same number of randomly selected ones from the background variants located at the position of the lead variants ±500 kb and not selected in

the credible sets. To avoid bias resulting from allele frequency differences, we controlled MAF in EAS and EUR for the background variants with variants in the credible sets by dividing MAF into four categories for each population: MAF < 5%; 5% < MAF ≤ 15%; 15% < MAF ≤ 30%; MAF > 30%) in each permutation procedure. We calculated empirical P value by 10^7 permutations using R. FDR was estimated by Benjamini–Hochberg method.

Tissue-specific H3K4me3 enrichment. To evaluate overlap of tissue-specific H3K4me3 at 85 identified autosomal variants, we used epiGWAS software[23]. We calculated LD (r^2 cutoff = 0.8) between the lead variant and variants within a 500-kb window using computeLD.py script in the software with LD information from the EAS samples in 1KGP. We analyzed the H3K4me3 peaks of 34 tissues obtained from the Roadmap project, which were included in the software. To estimate the statistical significance of cell-type-specific overlap, 10^6 permutations were performed. The H3K4me3 peaks are shown in a regional association plot (**Supplementary Data Sets 1** and **2**). We also analyzed 97 variants reported by the GIANT consortium in the same way using LD information of the EUR samples in 1KGP.

Functional annotation and eQTL analysis. For the annotation of nonsynonymous variants, ANNOVAR[54] was used. We annotated the variants with enhancer information using bedtools (v2.17.0).

To evaluate the overlaps between GWAS signal and *cis*-eQTL, we annotated the lead variants found in the trans-ethnic meta-analysis. We targeted the tissues which were genetically suggested to be involved by our cell-type specificity analyses. To prioritize eQTLs, we used GTEx[23] (release v6) and whole-blood eQTL information[24]. We set the significance threshold for eQTL as FDR < 0.05. In each data set, we considered only the variants that were analyzed in our trans-ethnic analysis, and selected the variants that showed the strongest association for each transcript. We considered the overlap to be significant if the GWAS lead variant was in LD ($r^2 \geq 0.8$) in EUR samples of 1KGP with the lead variant of the eQTL.

Pathway analysis. We used MAGENTA[27] for pathway analysis. Genes in the HLA region were excluded from the analysis. To adapt the LD structure of East Asians, we pruned the variants in EAS samples of 1KGP using the '–indep-pairwise' option of PLINK[50]. Default parameters were applied for other settings. We regarded the pathways that showed a FDR $q < 5\%$ as significant.

49. Price, A.L. *et al.* Principal components analysis corrects for stratification in genome-wide association studies. *Nat. Genet.* **38**, 904–909 (2006).
50. 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).
51. Li, Y., Willer, C.J., Ding, J., Scheet, P. & Abecasis, G.R. MaCH: using sequence and genotype data to estimate haplotypes and unobserved genotypes. *Genet. Epidemiol.* **34**, 816–834 (2010).
52. Delaneau, O., Marchini, J. & Zagury, J.-F. A linear complexity phasing method for thousands of genomes. *Nat. Methods* **9**, 179–181 (2011).
53. Pruim, R.J. *et al.* LocusZoom: regional visualization of genome-wide association scan results. *Bioinformatics* **26**, 2336–2337 (2010).
54. Wang, K., Li, M. & Hakonarson, H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* **38**, e164 (2010).

nature research

Corresponding Author: Yoichiro Kamatani

Date: Aug 3, 2017

Life Sciences Reporting Summary

Nature Research wishes to improve the reproducibility of the work we publish. This form is published with all life science papers and is intended to promote consistency and transparency in reporting. All life sciences submissions use this form; while some list items might not apply to an individual manuscript, all fields must be completed for clarity.

For further information on the points included in this form, see Reporting Life Sciences Research. For further information on Nature Research policies, including our data availability policy, see Authors & Referees and the Editorial Policy Checklist.

▶ Experimental design

1. **Sample size**

   Describe how sample size was determined.

   > As I commented to the editors by e-mail, we answered as "Not applicable" since it is expected to detect new loci after exceeding sample size of millions; we haven't achieved it, and it will not be realized soon in the complex disease GWAS setting. Current practice we understand is to use as much subjects as possible, and interpret the GWAS results; under this situation, I would say that the description about the sample size determination is "Not applicable".

2. **Data exclusions**

   Describe any data exclusions.

   > We excluded the samples and variants based on the standard quality control procedure in GWAS. Detailed information on quality controls were sufficiently described in our manuscript.

3. **Replication**

   Describe whether the experimental findings were reliably reproduced.

   > Variants associated in GWAS were investigated in two independent set of the Japanese population based cohorts. By comparing the estimates between GWAS and replication sets, we interpreted that most of the newly identified loci were not false-positives.
   >
   > We also performed the meta-analysis of the GWASs using publicly available GWAS summary statistics of Europeans. We confirmed the effect sizes of newly discovered variants in this analysis were not statistically differed between studies by estimating heterogeneity between them.

4. **Randomization**

   Describe how samples/organisms/participants were allocated into experimental groups.

   > We allocated the samples according to the study groups who recruited the participants.

5. **Blinding**

   Describe whether the investigators were blinded to group allocation during data collection and/or analysis.

   > In the replication studies, we analyzed two Japanese population-based cohorts separately. After that, we combined these results using the inverse variance method. Results of GWAS and replication sets were also combined using the same statistical method.

Note: all studies involving animals and/or human research participants must disclose whether blinding and randomization were used.

6. Statistical parameters

For all figures and tables that use statistical methods, confirm that the following items are present in relevant figure legends (or the Methods section if additional space is needed).

n/a	Confirmed	
☐	☒	The <u>exact</u> sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement (animals, litters, cultures, etc.)
☐	☒	A description of how samples were collected, noting whether measurements were taken from distinct samples or whether the same sample was measured repeatedly.
☐	☒	A statement indicating how many times each experiment was replicated
☐	☒	The statistical test(s) used and whether they are one- or two-sided (note: only common tests should be described solely by name; more complex techniques should be described in the Methods section)
☐	☒	A description of any assumptions or corrections, such as an adjustment for multiple comparisons
☐	☒	The test results (e.g. *p* values) given as exact values whenever possible and with confidence intervals noted
☐	☒	A summary of the descriptive statistics, including central tendency (e.g. median, mean) and variation (e.g. standard deviation, interquartile range)
☐	☒	Clearly defined error bars

See the web collection on statistics for biologists for further resources and guidance.

▶ Software

Policy information about availability of computer code

7. Software

Describe the software used to analyze the data in this study.	We used publicly available softwares for the analysis. The used softwares were listed in the URLs section, and clarified in the method section in our manuscript.

For all studies, we encourage code deposition in a community repository (e.g. GitHub). Authors must make computer code available to editors and reviewers upon request. The *Nature Methods* guidance for providing algorithms and software for publication may be useful for any submission.

▶ Materials and reagents

Policy information about availability of materials

8. Materials availability

Indicate whether there are restrictions on availability of unique materials or if these materials are only available for distribution by a for-profit company.	No unique materials were used.

9. Antibodies

Describe the antibodies used and how they were validated for use in the system under study (i.e. assay and species).	Not applicable.

10. Eukaryotic cell lines

a. State the source of each eukaryotic cell line used.	Not applicable.
b. Describe the method of cell line authentication used.	Not applicable.
c. Report whether the cell lines were tested for mycoplasma contamination.	Not applicable.
d. If any of the cell lines used in the paper are listed in the database of commonly misidentified cell lines maintained by ICLAC, provide a scientific rationale for their use.	Not applicable.

▶ Animals and human research participants

Policy information about studies involving animals; when reporting animal research, follow the ARRIVE guidelines

11. Description of research animals

Provide details on animals and/or animal-derived materials used in the study.	Not applicable.

12. Description of human research participants

Describe the covariate-relevant population characteristics of the human research participants.

We used three independent sets of participants in this study. As shown in Supplementary Note 2, these were 1) BBJ (N = 158,284, N for male = 85,894, mean age: 62.6), 2) JPHC (N = 7,379, N for male = 2,475, mean age: 53.6), and 3) TMM (N = 7,76, N for male = 2,623, mean age: 61.1). These characteristics were adjusted in the association analysie as described in the methods of our manuscript.