

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

秋山, 雅人

<https://hdl.handle.net/2324/1931996>

出版情報 : Kyushu University, 2017, 博士 (医学), 論文博士

バージョン :

権利関係 : Public access to the fulltext file is restricted for unavoidable reason (2)



The contents of this article is now published by *Nature Genetics*:
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).

<https://www.nature.com/articles/ng.3951>

DOI number: 10.1038/ng.3951

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

Authors: Masato Akiyama¹, Yukinori Okada^{1,2,3}, Masahiro Kanai¹, Atsushi Takahashi^{1,4}, Yukihide Momozawa⁵, Masashi Ikeda⁶, Nakao Iwata⁶, Shiro Ikegawa⁷, Makoto Hirata⁸, Koichi Matsuda⁹, Motoki Iwasaki¹⁰, Taiki Yamaji¹⁰, Norie Sawada¹⁰, Tsuyoshi Hachiya¹¹, Kozo Tanno^{11,12}, Atsushi Shimizu¹¹, Atsushi Hozawa^{13,14}, Naoko Minegishi^{13,14}, Shoichiro Tsugane¹⁵, Masayuki Yamamoto^{13,14}, Michiaki Kubo¹⁶ and Yoichiro Kamatani^{1, 17}.

Affiliations:

1. Laboratory for Statistical Analysis, RIKEN Center for Integrative Medical Sciences, Yokohama 230-0045, Japan.
2. Department of Statistical Genetics, Osaka University Graduate School of Medicine, Osaka 565-0871, 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 565-8565, Japan.
5. Laboratory for Genotyping Development, RIKEN Center for Integrative Medical Sciences, Yokohama 230-0045, Japan.
6. Department of Psychiatry, Fujita Health University School of Medicine, Aichi 470-1192,

Japan.

7. Laboratory for Bone and Joint Diseases, RIKEN Center for Integrative Medical Sciences,
Tokyo 108-8639, Japan.

8. Institute of Medical Science, The University of Tokyo, Tokyo 108-8639, Japan.

9. Graduate school of Frontier Sciences, The University of Tokyo, Tokyo 108-8639, Japan.

10. Division of Epidemiology, Center for Public Health Sciences, National Cancer Center,
Tokyo 104-0045, Japan.

11. Iwate Tohoku Medical Megabank Organization, Iwate Medical University, Iwate 028-3694,
Japan.

12. Department of Hygiene and Preventive Medicine, School of Medicine, Iwate Medical
University, Iwate 028-3694, Japan.

13. Tohoku Medical Megabank Organization, Tohoku University, Sendai 980-8573, Japan.

14. Graduate School of Medicine, Tohoku University, Sendai 980-8575, Japan.

15. Center for Public Health Sciences, National Cancer Center, Tokyo 104-0045, Japan.

16. RIKEN Center for Integrative Medical Sciences, Yokohama 230-0045, Japan.

17. Center for Genomic Medicine, Kyoto University Graduate School of Medicine, Kyoto
606-8507, Japan

Correspondence to: Yoichiro Kamatani (yoichiro.kamatani@riken.jp)

44

45 **Abstract**

46 Obesity is a risk factor for a wide variety of health problems. By carrying out a genome-wide
47 association study (GWAS) of body mass index (BMI) in Japanese ($n = 173,430$), we found
48 85 loci significantly associated ($P < 5.0 \times 10^{-8}$), of which 51 were novel. We conducted
49 trans-ancestral meta-analyses by integrating these results with the results from a GWAS of
50 Europeans, and revealed 61 additional new loci. In total, this study identifies 112 novel loci
51 which double the known BMI loci. By annotating associated variants with cell-type specific
52 regulatory marks, we found the enrichment of variants in CD19+ cells. We also found a
53 significant genetic correlation between BMI and lymphocyte count ($P = 6.46 \times 10^{-5}$, $r_g = 0.18$),
54 and between BMI and multiple complex diseases. These findings provide genetic evidence
55 that lymphocytes were relevant to body weight regulation, and insights into the pathogenesis
56 of obesity.

57

58

59 **Introduction**

60 Obesity, which is heritable, is a risk factor for various diseases¹⁻³. Thus far, GWASs have
61 identified more than 100 loci associated with BMI⁴⁻⁷, which is the most commonly used
62 measurement for obesity. However, these loci explained only a small fraction (~3%) of
63 heritability in Europeans⁶. Recent polygenic analyses have suggested that the genetic
64 variants could explain more than 20% of the heritability^{6, 8}. Therefore, the majority of the
65 genetic components of BMI have not yet been discovered. Whereas most previous studies
66 were performed using individuals with European ancestry, GWASs in non-Europeans have
67 revealed different loci⁵. Furthermore, epidemiological surveys reported differences in BMI
68 across ethnicities, such as the prevalence of the obese in Asians was lower than that of
69 Caucasians⁹. In addition, Asians tend to develop diabetes with a lesser BMI in comparison
70 with Europeans¹⁰. These differences imply that investigations in different populations could
71 yield further insights into the etiology of obesity.

72 To identify genetic loci associated with obesity and to expand the knowledge on
73 body weight regulation, we conducted a GWAS that included >170,000 Japanese subjects
74 (study design is shown in **Supplementary Fig.1**). We then conduct a comprehensive
75 integration with the previous GWAS of Europeans, functional annotations of the associated
76 loci and genetic correlation analyses between BMI and other complex diseases in humans.

77

78

79 **RESULTS**80 **Association signals at 85 loci found in Japanese population**

81 For this GWAS, 158,284 Japanese subjects who participated in the BioBank Japan (BBJ)
 82 project¹¹⁻¹³ were included. After whole-genome imputation using East Asian (EAS) samples
 83 of the 1000 genome project (1KGP)¹⁴ as a reference, we performed a GWAS using
 84 6,108,953 single-nucleotide variants. Although we applied stringent quality controls (QCs),
 85 the genomic inflation factor showed remarkable inflation ($\lambda_{GC} = 1.44$; **Supplementary Fig.**
 86 **2a**). To dissect observed inflation into the polygenic effects and potential biases, we
 87 performed LD score regression¹⁵. The estimated mean chi-square and intercept (i.e., the
 88 index of biases) values were 1.63 and 1.07 (**Supplementary Fig. 2b**), respectively. These
 89 values suggested that polygenic effects induced a majority (89%) of the inflation, and the
 90 influence of biases was minimal in comparison with previously reported meta-GWASs¹⁵.
 91 Therefore, we did not apply GC correction to our GWAS.

92 We observed genome-wide significant (GWS; $P < 5.0 \times 10^{-8}$) levels of association
 93 signals at 72 loci (at least 1 Mb apart from each other; **Supplementary Fig. 2c**). We further
 94 evaluated 134 lead variants with $P < 1.0 \times 10^{-6}$ using replication sets composed of 15,146
 95 subjects from two independent Japanese population-based cohort studies (the Japan Public
 96 Health Center-based Prospective Study (JPHC) and the Tohoku Medical Megabank Project
 97 (TMM)). We observed 83% of directional consistency (111 of 134 loci; P for sign test = $5.0 \times$
 98 10^{-14}), and all the nominal associations ($P < 0.05$ in the replication sets) were in the same
 99 direction (P for sign test = 2.2×10^{-31}). When we combined the results of the GWAS with
 100 replication sets, six loci fell below the GWS threshold, whereas 17 additional loci surpassed
 101 the GWS level (**Supplementary Table 1 and 2**). Accordingly, 83 loci were significantly
 102 associated with BMI (**Supplementary Table 3a, Supplementary Data Set 1 and 2**).

103 To identify the BMI-associated loci with sex-dependent effects, we performed a

104 GWAS stratified by sex (**Supplementary Fig. 3**). We identified two male-specific loci that
 105 had not been previously reported (rs77511173 and rs111612372, **Supplementary Table 3b**,
 106 **4** and **Supplementary Fig. 3c**). We compared the effect sizes of identified variants between
 107 sexes; we found that only rs77511173 showed a significant difference ($P_{\text{difference}} = 5.88 \times 10^{-5}$,
 108 $\alpha = 0.05/85$, **Supplementary Table 5** and **Supplementary Fig. 4**). The effect sizes of the
 109 variants on the X chromosome were not differed under the assumption of the full dosage
 110 compensation ($P_{\text{difference}} > 0.05$). We observed nominal differences at two reported
 111 sex-dependent loci, *SEC16B*⁶ (stronger in females, $P_{\text{difference}} = 3.60 \times 10^{-3}$) and *KCNQ1*⁵
 112 (stronger in males, $P_{\text{difference}} = 3.49 \times 10^{-3}$), but not at other loci (*PCSK1*, *CDKAL1* and
 113 *ALDH2*)⁵. The effect sizes of the associated alleles were strongly correlated between sexes
 114 (Pearson's correlation coefficients [r] = 0.92, $P = 5.21 \times 10^{-38}$). We also evaluated the shared
 115 heritability between sexes. By applying bivariate LD score regression¹⁶, we observed a
 116 significant genetic correlation ($r_g = 0.93$, standard error [SE] = 0.04, $P = 8.3 \times 10^{-151}$),
 117 indicating that polygenicity was shared. Taken together, 85 loci, including five loci on the X
 118 chromosome and two male-specific loci, were significantly associated with BMI
 119 (**Supplementary Table 3**), and 51 were novel.

120 To assess the robustness, we evaluated the associations of the significant variants
 121 in the replication sets. Among them, 49 of 85 (57.6%) were associated with nominal
 122 significance ($P < 0.05$). Furthermore, the directions of the effects were consistent at 83 loci
 123 (97.6%; P for sign test = 1.89×10^{-22}). We did not find any statistical differences in the
 124 number of directionally consistent and nominally associated variants between the previously
 125 reported loci and the newly identified loci ($P = 1$ for Fisher's exact test), suggesting that most
 126 of the newly identified loci were not false-positives. We observed strong correlation in the
 127 effect sizes of associated alleles between the GWAS and the meta-analysis of replication
 128 sets ($r = 0.91$, $P = 3.81 \times 10^{-34}$).

129 Next, we performed a conditional analysis to explore BMI-associated variants
130 independent of the lead variants within the loci. We found four significant second signals that
131 satisfied GWS after conditioning by the lead variants within each region (in a range of ± 1 Mb
132 of each lead variant; **Table 1** and **Supplementary Fig. 5**), and two of these had already
133 been found in Europeans⁶ (*GPRC5B/GP2* and *NLRC3/ADCY9*). Another two signals were
134 novel secondary associations (*FAM110C/TMEM18* and *HIF1AN-PAX2/BTRC*). Therefore,
135 we determined 89 independent susceptibility variants within 85 loci.

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

144 Among the 89 independently associated variants, rs148546399 was low-frequency
145 (minor allele frequency [MAF] $\leq 5\%$) in East Asian and monomorphic in Europeans, 13
146 variants had a MAF $\leq 5\%$ in Europeans but common in East Asian (MAF $> 5\%$), and two
147 variants (rs10208649 and rs7903146) were low-frequency in East Asian and common in
148 Europeans according to 1KGP (**Supplementary Table 1** and **4**). When we annotated the
149 variants in LD ($r^2 \geq 0.8$) with the 89 variants reached GWS, we found 29 missense variants
150 (**Supplementary Table 7**), including a missense variant in *GPR101* (rs1190736 [C > A];
151 NP_473362.1:p.Val124Leu), the causative gene of acromegaly¹⁷. We additionally evaluated
152 variants in LD with rarer variant rs10208649 (MAF = 1.2 % in EAS), and found the rare
153 coding variant of *GPR75* in moderate LD (rs80328470 [T > C]; NP_006785.1: p.Thr27Ala,

MAF = 0.8 % in EAS; $r^2 = 0.66$), which is monomorphic in Europeans according to 1KGP and ExAC¹⁸.

We searched 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 Table 8** and **Supplementary Table 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 Europeans⁶.

Identification of 61 new loci through trans-ethnic meta-analyses

To further identify loci associated with BMI, we conducted trans-ethnic meta-analyses using our GWAS and publically available results of Europeans⁶ by MANTRA¹⁹. We first analyzed data sets without sex stratification ($N_{\max} = 480,438$), and observed significant level of the association signals ($\log_{10}(\text{Bayes' factor}) > 6$)²⁰ 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 significant level. We observed 66 of 78 (84.6%) autosomal loci which reached significant level in Japanese and 70 of 91 (76.9%) previously reported loci by the GIANT consortium⁶ remained to be significant (variants found by sex-stratified GWAS were excluded). This proportion was not statistically differed between studies (P for chi-square test = 0.72). Significant 163 variants showed a significant 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 BMI associated loci were generally shared across ethnicities. On the other hand, we observed differences in the effect size at 14 loci ($P_{\text{het}} < 0.05/163$).

We also performed meta-analysis of sex-stratified GWASs and found six male

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

187 Using the MANTRA result, we constructed 99% credible sets in each identified locus
188 (**Supplementary Table 13**). When we compared the intervals of the genomic region
189 including variants in the 99% credible set, we observed significant reduction in comparison
190 with the previous trans-ancestral fine-mapping result⁶ (P for Wilcoxon signed rank test =
191 0.04; **Supplementary Table 14**). Annotation of the variants included in the 99% credible sets
192 revealed 38 missense variants, including 15 variants found in the Japanese GWAS
193 (**Supplementary Table 15**). Although the majority of the missense variants showed lower
194 posterior probability (PP), we found two likely causal variants with relatively higher PP
195 (KCNJ11, NP_000516.3: p.Val337Ile, PP = 0.75, and HIVEP1, NP_002105.3: p.Met1609Ile,
196 PP = 0.62).

197

198 **Cell types related to obesity**

199 To investigate specific tissues genetically associated with obesity, we evaluated enrichment
200 of the identified variants included in the 99% credible sets for the cell-type specific active
201 enhancer constructed by the Roadmap project²¹ (**Supplementary Note**). After grouping 64
202 cell-types into 10 categories, we tested for enrichment of the variants, and observed
203 significant enrichment in the active enhancer of the three cell-groups (false discovery rate

[FDR] < 5%; immune related, central nervous system (CNS), and adipose; **Fig.1a, Supplementary Table16** and **Supplementary Table17**). Next, we analyzed each cell-types 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 – 1.68 fold enrichment), whereas B cells was only enriched in immune-related cells (FDR < 5%, 1.60 fold enrichment; **Fig.1b, Supplementary Table18**).

Next, we conducted another enrichment analysis using the cell-type-specific trimethylation of histone H3 at lysine 4 (H3K4me3) peaks of 34 cell types using epiGWAS software²². Since this analysis requires LD information, we restricted the 84 autosomal variants found in the Japanese GWAS for the analysis. We found three enriched cell types with nominal significance (**Fig. 2** and **Supplementary Table 19**). The cell types were pancreatic islets ($P = 0.0013$), B cells (CD19+ primary cells, $P = 0.0038$), and brain (inferior temporal lobe, $P = 0.0499$). We also applied this method to 97 variants found in the GIANT consortium⁶, and found only T cells (CD34+ cells) showed significant enrichment ($P = 0.02$).

Considering 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, 23}, our results imply that gene regulation in immune related cells (B cells and T cells), adipose and pancreas may also have a role in the pathogenesis of obesity. We annotated the lead variants of the identified loci with eQTL^{24, 25} in candidate cell-types, and found 69 overlaps (**Supplementary Table 20**).

225

226 Variance explained by variants in the GWAS

227 The lead variants of the 83 identified loci found in Japanese GWAS explained 2.8% of the
228 phenotypic variance (**Supplementary Table 3**), which is similar to that found in Europeans

229 (2.7% by 97 variants)⁶. Although we included five variants on the X chromosome, which
230 were not analyzed in the GWAS of Europeans, they explained only 0.09% of the phenotypic
231 difference.

232 To assess the polygenic effects, we estimated the variance explained by all the
233 GWAS variants in the Japanese population-based cohorts using the genomic restricted
234 maximum likelihood (GREML) method implemented in GCTA²⁶. The estimated proportion of
235 the phenotypic variance explained by autosomal variants was $29.8 \pm 3.4\%$ (mean \pm SE).
236 This fraction was similar to that found in the European population ($27.4 \pm 2.5\%$)⁸. When we
237 estimated the explained variance of each chromosome²⁷, a weak linear relationship with
238 chromosome lengths was observed ($P = 7.31 \times 10^{-3}$, $r^2 = 0.30$; **Supplementary Fig. 10**).
239 The X chromosome explained $2.3 \pm 0.8\%$ of the phenotypic variance.

240 We further examined the variance explained in the Japanese cohorts by examining a
241 subset of the variants under a certain *P*-value threshold according to the GWAS results of
242 current Japanese and those of previous Europeans⁶ using variants analyzed in both studies.
243 As expected, the explained variance was increased by lowering the threshold of the *P*-value
244 (**Fig. 3** and **Supplementary Fig. 11**). When we compared our GWAS with the GWAS of
245 Europeans, estimates based on our GWAS better explained the phenotypic variance in
246 Japanese cohorts (on average, 1.81 fold higher; range: 1.47–2.17).

247

248 **Pathway analysis**

249 Using the result of GWAS in the Japanese individuals, we performed MAGENTA²⁸ to
250 investigate biological pathways associated with obesity and found three significant pathways
251 (ethanol oxidation, glycolysis gluconeogenesis, and maturity-onset diabetes of the young;
252 FDR: $q < 0.05$; **Supplementary Table 21** and **22**). We found that the neurotrophin signaling
253 pathway, which was previously reported in Europeans⁶, showed nominal enrichment in our

254 study ($P = 0.0052$, $q = 0.25$).

255

256 **Stratified analysis by type 2 diabetes**

257 As previous reports suggested that the effect sizes of BMI at a few loci differed between
258 individuals with T2D and individuals who did not develop diabetes (*TCF7L2* (ref. 6) and
259 *CDKAL1* (ref.29)), we evaluated 193 variants by separating GWAS samples into a group of
260 individuals with T2D ($n = 31,609$) and a group of individuals without diabetes ($n = 125,816$).
261 We observed significant directional consistency (96.4%, P for sign test = 1.29×10^{-26}) and a
262 significant correlation between the effects ($r = 0.85$, $P = 8.69 \times 10^{-55}$). Nevertheless, we
263 found evidence of heterogeneity not only at two reported loci but also at five other loci
264 (*HHEX*, *IGFBP2*, *KCNQ1*, *CDKN2B*, and *DUSP9*; $\alpha = 0.05/193$, **Supplementary Fig. 12, 13**
265 and **Supplementary Table 23**). The effects of these variants were stronger in individuals
266 with T2D. Considering that observed effect size of *TCF7L2* in Japanese individuals with T2D
267 ($\beta = 0.12$, $SE = 0.02$) was similar to that in the result of Europeans⁶ ($\beta = 0.11$, $SE = 0.01$;
268 $P_{\text{difference}} = 0.52$), some variants may have different impact according to diabetes status.

269 We also evaluated the effects of BMI-associated variants on T2D susceptibility. Of
270 the 193 variants, we found 20 that showed GWS levels of association for T2D
271 (**Supplementary Table 24**). When we compared the effects of both traits, a bidirectional
272 relationship was observed (**Fig. 4**). A positive relationship was observed at five loci. A
273 negative relationship was found at 15 loci. Although epidemiological studies reported that a
274 higher BMI was a risk for T2D¹, our results indicated that the relationship between genetic
275 variants and T2D susceptibility is more complicated. Of the variants that showed negative
276 relationships, many were reportedly associated with glycemic and/or insulin regulation^{30, 31},
277 and insulin processing³¹. Although some loci (*FAM60A* and *FGFR2*) have not been reported
278 to be associated with these traits, this observation implied that they might also share

279 biological pathways.

280 To assess the possible biased estimation in our BMI GWAS due to index event bias,
281 we compared the effect sizes for BMI at these loci between GWAS and Japanese
282 population-based cohorts. Among 16 variants which were evaluated in both data sets, we
283 observed a perfect directional consistency and a significant correlation in effect sizes ($r =$
284 0.94 , $P = 6.63 \times 10^{-8}$; **Supplementary Fig.14**) without significant heterogeneity ($P_{\text{het}} >$
285 $0.05/16$), suggesting a low possibility of false positive due to case-mix samples in our
286 GWAS.

287

288 Genetic correlation analysis

289 We evaluated the genetic correlation between the results of this study and 33 other GWASs
290 of Asians using bivariate LD score regression¹⁶ (**Fig. 5a, Supplementary Table 25** and
291 **Supplementary note**). We found six significant positive correlations (FDR $q < 0.05$),
292 including correlations with T2D, cardiovascular diseases (ischemic stroke, myocardial
293 infarction, and peripheral arterial diseases), asthma, and ossification of posterior longitudinal
294 ligament of the spine (OPLL). Significant negative correlations were observed in adolescent
295 idiopathic scoliosis (AIS), schizophrenia, and rheumatoid arthritis (RA). Genetic correlations
296 with T2D and coronary artery disease have also been reported in Europeans¹⁶. Other
297 genetic correlations have not been previously reported. To compare our results with that in
298 Europeans, we searched LD hub³² and found seven results of genetic correlations that were
299 also evaluated in the current study (**Supplementary Table 26**). Similar r_g values were
300 observed between ours and Europeans in the traits reported as significant in our analysis
301 (cardiovascular diseases, RA, schizophrenia, and T2D).

302 Our cell-type-specific analyses suggested involvement of lymphocytes in the
303 regulation of body weight. Although the genetic involvement of lymphocytes in obesity has

304 not been elucidated, previous clinical studies have shown positive correlations between BMI
305 and the counts of leukocytes and its subtypes in peripheral blood^{33, 34}. We hypothesized that
306 the counts of these cells might be influenced by genetic components that also regulate BMI.
307 We analyzed the genetic correlation between BMI and three major hematological traits
308 ($N_{\text{GWAS}} = 110,397 - 111,268$; **Supplementary Note**) and found a significant positive
309 correlation with white blood cell (WBC) counts ($P = 8.50 \times 10^{-5}$, $r_g = 0.14$; **Fig. 5b** and
310 **Supplementary Table 27**). Next, we evaluated the genetic correlations between BMI and
311 the counts of five WBC subtypes ($N_{\text{GWAS}} = 63,197$). Of these correlations, only lymphocytes
312 showed a significant genetic correlation with BMI ($P = 6.46 \times 10^{-5}$, $r_g = 0.18$). These results
313 implied that BMI and lymphocytes shared genetic components that regulate both traits.
314

315

316 **DISCUSSION**

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

323 We showed evidence of genetic differences between Europeans and East Asians.
324 First, comparisons of the effect sizes suggested that susceptibility loci were generally shared
325 between Europeans and East Asians, which was compatible with trans-ethnic studies for
326 other traits^{35, 36}. On the other hand, variants that were grouped based on the associations of
327 our GWAS were able to explain larger phenotypic differences in Japanese cohorts compared
328 with the variants selected from the GWAS of Europeans. We considered four possible
329 explanation: 1) different causal variants underlie between populations, 2) differences in the
330 effect sizes of causal variants across ethnicities, 3) difference in LD structure between
331 causal variants and GWAS SNPs, 4) differences in allele frequencies of the causal variants.
332 Second, we observed association signals, including second signals, independent of
333 previously reported associations. This could be the result of independent causal variants or
334 distinct LD structures reflecting the same causal variants across ethnicities. Our results
335 emphasize the importance of investigating multiple population to reveal uncovered genetic
336 components. Third, pathway analysis indicated that the neurotrophin signaling pathway
337 influenced BMI across ethnicities, but we also found three pathways that were not reported
338 in Europeans. Considering selection at *ADH1B* and *ALDH2* genes in East Asians^{37, 38}, the
339 role of the ethanol oxidation pathway and the glycolysis gluconeogenesis pathway, which
340 include both *ADH* and *ALDH* family genes, may have a larger impact on the BMI of East

341 Asians. These differences highlight the need for the use of diverse populations to better
342 understand the genetic underpinnings of BMI.

343 Our results suggest two possible ways that lymphocytes are involved in the
344 regulation of body weight. First, gene regulation in B cells might influence body weight
345 regulation because the identified variants were enriched in CD19+ cells-specific chromatin
346 marks. The previous investigation of Europeans using DEPICT suggested candidate genes
347 underlying BMI associated loci were highly expressed in precursor cells B lymphoid and
348 lymphoid progenitor cells⁶. Substantial biological evidence has shown the involvement of
349 immune cells in obesity via inflammation and insulin resistance³⁹. Recent studies have also
350 indicated that B cells play an important role in obesity through activation of pro-inflammatory
351 macrophages and T cells and the production of IgG antibodies in insulin resistance⁴⁰.
352 Changes in gene regulation in B cells might affect BMI through these biological mechanisms.
353 Second, the increased number of leukocytes in subjects with obesity^{33, 34} might be
354 influenced by genetic variants because we found a significant positive genetic correlation
355 between BMI and counts of WBCs, especially lymphocytes. Our observations indicate that
356 the same genetic variants were increasing (or decreasing) both BMI and counts of these
357 cells. Because genetic correlations could not pinpoint the shared causal variants, further
358 study is needed to elucidate the genetic components of both traits and clarify their underlying
359 biological pathways. Our results provide genetic evidence for the involvement of
360 lymphocytes in the etiology of obesity.

361 Our cell-type specificity analysis also suggested that cells in pancreatic islets and
362 adipose tissue are relevant to obesity. When we looked up the existing genetic correlations
363 in LD hub³², significant positive genetic correlations with BMI were found in insulin secretion
364 related measurements (fasting insulin level and HOMA- β) and insulin resistance related trait
365 (HOMA-IR). These facts indicated that BMI shares heritability with insulin modulation. The

366 previous investigation indicated the enrichment of the BMI heritability in pancreatic islets²³.
367 Visceral adipose tissue is insulin responsive during obesity, and a major driver of insulin
368 resistance via its chronic inflammation⁴¹. Furthermore, recent study indicated that the variant
369 at *FTO* region lead *IRX3* and *IRX5* disruption during adipocyte differentiation, and causes
370 body weight gain through cell-autonomous shift from white adipose browning and
371 thermogenesis to lipid storage⁴². Our results may link to these biological knowledge, and
372 suggest that gene regulation in these tissues may influence body weight.

373 We investigated the genetic correlations between BMI and several other traits and
374 found significant correlations with nine diseases. As we were able to replicate the
375 correlations that were reported in Europeans, genetic correlation may be reproducible
376 across ethnicities. Indeed, a genetic correlation between BMI and schizophrenia was also
377 suggested in Europeans ($P = 0.0002$, $r_g = -0.10$)¹⁶. In observational studies, consistent
378 relationships were found between BMI and T2D, cardiovascular diseases, asthma, OPLL,
379 and AIS^{1, 43-45}; however, to our knowledge, the relationships between BMI and RA, and BMI
380 and schizophrenia have been controversial⁴⁶⁻⁴⁹. Besides the positive genetic correlation
381 between BMI and T2D, we found some BMI-increasing variants reduce the risk of T2D.
382 Considering T2D confers risk of the multiple diseases, BMI-associated variants may not
383 uniformly influence complex diseases. Our results refine the list of diseases which shares
384 heritability with obesity, and provide insight into the contribution of genetic components in
385 complex diseases.

386 In conclusion, we performed a large-scale GWAS of BMI and found 112 novel loci.
387 These results highlight that investigating diverse population contributes to identify uncovered
388 loci and to narrow down the candidate genomic region. Our findings provide insight into the
389 links between BMI and complex diseases and provide genetic evidence that lymphocytes
390 play a role in the pathogenesis of obesity.

391

392

393 **URLs**

394 **BioBank Japan:** <https://biobankjp.org/english/index.html>,

395 **JPHC:** <http://epi.ncc.go.jp/en/jphc/index.html>,

396 **TMM:** <http://www.megabank.tohoku.ac.jp/english/>,

397 **PLINK:** <https://www.cog-genomics.org/plink2>,

398 **MACH:** <http://csg.sph.umich.edu/abecasis/MaCH/>,

399 **EpiGWAS:** <https://www.broadinstitute.org/mpg/epigwas/>,

400 **ANNOVAR:** <http://annovar.openbioinformatics.org/en/latest/>,

401 **SHAPEIT:** https://mathgen.stats.ox.ac.uk/genetics_software/shapeit/shapeit.html,

402 **HaploReg:** <http://www.broadinstitute.org/mammals/haploreg/haploreg.php>,

403 **GTEEx:** <http://www.gtexportal.org/home/>,

404 **R:** <https://www.r-project.org/>,

405 **MAGENTA:** <https://www.broadinstitute.org/mpg/magenta/>,

406 **GIANT consortium:**

407 https://www.broadinstitute.org/collaboration/giant/index.php/GIANT_consortium,

408 **Locuszoom:** <http://locuszoom.sph.umich.edu/locuszoom/>,

409 **1000 genome project:** <http://www.1000genomes.org/>,

410 **HapMap project:** <http://hapmap.ncbi.nlm.nih.gov/>,

411 **GCTA:** <http://cnsgenomics.com/software/gcta/>,

412 **LDSC:** <https://github.com/bulik/ldsc/>,

413 **GWAS catalog:** <https://www.ebi.ac.uk/gwas/>,

414 **ROADMAP epigenomics PROJECT:** <http://www.roadmapepigenomics.org/>,

415 **LD hub:** <http://ldsc.broadinstitute.org/>

416 **bedtools:** <http://bedtools.readthedocs.io/en/latest/>

417

418

419 **Acknowledgments**

420 We would like to acknowledge the staff of the TMM, the JPHC, and the BBJ for collecting
421 samples and clinical information. We are grateful to the staff of the RIKEN Center for
422 Integrative Medical Sciences for genotyping and data management. We thank S. K. Low, K.
423 Suzuki and M.Horikoshi for advice on statistical analyses, and Professor Morris for providing
424 us with the MANTRA software. This study was funded by the BioBank Japan project (M.A.,
425 Y.O., M. Kanai, A.T., Y.M., M.H., K.M., M. Kubo and Y.K.) and Tohoku Medical Megabank
426 project (T.H., K.T., A.S., A.H., N.M., and M.Y.), which is supported by the Ministry of
427 Education, Culture, Sports, Sciences and Technology of Japanese government and the
428 Japan Agency for Medical Research and Development. The JPHC Study is supported by the
429 National Cancer Research and Development Fund since 2010, and was by a Grant-in-Aid
430 for Cancer Research from the Ministry of Health, Labour and Welfare of Japan from 1989 to
431 2010 (M.Iwasaki., T.Y., N.S., and S.T.). GWASs of psychiatric disorders were the results of
432 the Strategic Research Program for Brain Sciences (SRPBS) from the Japan Agency for
433 Medical Research and Development (A.T., M.Ikeda., N.I., M. Kubo and Y.K.).

434

435 **Author contributions**

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

443

444 **Competing financial interests**

445 The authors declare no competing financial interests.

446

Reference for main text

- 1) Haslam, D. W. & James, W. P. T. Obesity. *Lancet* **366**, 1197–209 (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**, 1867–72 (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**, 569–78 (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**, 937–48 (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**, 5492–504 (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 Genetics* **11**, (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**, 1114–20 (2015).
- 9) Haslam, D. W. & James, W. P. T. Obesity. *Lancet* **366**, 1197–209 (2005).
- 10) Yoon, K. H. et al. Epidemic obesity and type 2 diabetes in Asia. *Lancet* **368**, 1681–1688 (2006).
- 11) Nakamura, Y. The BioBank Japan Project. *Clin. Adv. Hematol. Oncol.* **5**, 696–7 (2007).
- 12) Nagai, A. et al. Overview of the BioBank Japan Project: Study design and profile. *J.*

472 *Epidemiol.* **27**, 2–8 (2017).

473 13) Hirata, M. et al. Cross-sectional analysis of BioBank Japan clinical data: A large cohort
474 of 200,000 patients with 47 common diseases. *J. Epidemiol.* **27**, 9–21 (2017).

475 14) Abecasis, G. R. et al. A map of human genome variation from population-scale
476 sequencing. *Nature* **467**, 1061–73 (2010).

477 15) Bulik-Sullivan, B. K. et al. LD Score regression distinguishes confounding from
478 polygenicity in genome-wide association studies. *Nat. Genet.* **47**, 291–5 (2015).

479 16) Bulik-Sullivan, B. et al. An atlas of genetic correlations across human diseases and
480 traits. *Nat. Genet.* **47**, 1236–41 (2015).

481 17) Trivellin, G. et al. Gigantism and acromegaly due to Xq26 microduplications and
482 GPR101 mutation. *N. Engl. J. Med.* **371**, 2363–74 (2014).

483 18) Lek, M. et al. Analysis of protein-coding genetic variation in 60,706 humans. *Nature* **536**,
484 285–291 (2016).

485 19) Morris, A. P. Transethnic meta-analysis of genomewide association studies. *Genet.*
486 *Epidemiol.* **35**, 809–822 (2011).

487 20) Wang, X. et al. Comparing methods for performing trans-ethnic meta-analysis of
488 genome-wide association studies. *Hum. Mol. Genet.* **22**, 2303–2311 (2013).

489 21) Roadmap Epigenomics Consortium. Integrative analysis of 111 reference human
490 epigenomes. *Nature* **518**, 317–330 (2015).

491 22) Trynka, G. et al. Chromatin marks identify critical cell types for fine mapping complex
492 trait variants. *Nat. Genet.* **45**, 124–30 (2013).

493 23) Finucane, H. K. et al. Partitioning heritability by functional annotation using
494 genome-wide association summary statistics. *Nat. Genet.* **47**, 1228–35 (2015).

495 24) Lonsdale J, et al. The Genotype-Tissue Expression (GTEx) project. *Nat. Genet.* **45**,
496 580–5 (2013).

- 497 25) Westra, H.-J. et al. Systematic identification of trans eQTLs as putative drivers of known
498 disease associations. *Nat. Genet.* **45**, 1238–43 (2013).
- 499 26) Yang, J., Lee, S. H., Goddard, M. E. & Visscher, P. M. GCTA: a tool for genome-wide
500 complex trait analysis. *Am. J. Hum. Genet.* **88**, 76–82 (2011).
- 501 27) Yang, J. et al. Genome partitioning of genetic variation for complex traits using common
502 SNPs. *Nat. Genet.* **43**, 519–25 (2011).
- 503 28) Segrè, A. V, Groop, L., Mootha, V. K., Daly, M. J. & Altshuler, D. Common inherited
504 variation in mitochondrial genes is not enriched for associations with type 2 diabetes or
505 related glycemc traits. *PLoS Genet.* **6**, (2010).
- 506 29) Okada, Y. et al. Common variants at CDKAL1 and KLF9 are associated with body mass
507 index in east Asian populations. *Nat. Genet.* **44**, 302–306 (2012).
- 508 30) Scott, R. A. et al. Large-scale association analyses identify new loci influencing
509 glycemc traits and provide insight into the underlying biological pathways. *Nat. Genet.* **44**,
510 991–1005 (2012).
- 511 31) Dimas, A. S. et al. Impact of type 2 diabetes susceptibility variants on quantitative
512 glycemc traits reveals mechanistic heterogeneity. *Diabetes* **63**, 2158–71 (2014).
- 513 32) Zheng, J. et al. LD Hub: a centralized database and web interface to perform LD score
514 regression that maximizes the potential of summary level GWAS data for SNP heritability
515 and genetic correlation analysis. *Bioinformatics* **33**, 272–279 (2017).
- 516 33) Julius, S., Egan, B. M., Kaciroti, N. A., Nesbitt, S. D. & Chen, A. K. In prehypertension
517 leukocytosis is associated with body mass index but not with blood pressure or incident
518 hypertension. *J Hypertens.* **32**, 251–9 (2014).
- 519 34) Ilavská, S. et al. Association between the human immune response and body mass
520 index. *Hum. Immunol.* **73**, 480–5 (2012).

- 521 35) Okada, Y. et al. Genetics of rheumatoid arthritis contributes to biology and drug
522 discovery. *Nature* **506**, 376–81 (2014).
- 523 36) Liu, J. Z. et al. Association analyses identify 38 susceptibility loci for inflammatory bowel
524 disease and highlight shared genetic risk across populations. *Nat. Genet.* **47**, 979–86
525 (2015).
- 526 37) Oota, H. et al. The evolution and population genetics of the ALDH2 locus: Random
527 genetic drift, selection, and low levels of recombination. *Ann. Hum. Genet.* **68**, 93–109
528 (2004).
- 529 38) Han, Y. et al. Evidence of positive selection on a class I ADH locus. *Am. J. Hum. Genet.*
530 **80**, 441–56 (2007).
- 531 39) Johnson, A. M. F. & Olefsky, J. M. The origins and drivers of insulin resistance. *Cell* **152**,
532 673–84 (2013).
- 533 40) Winer, D. A. et al. B cells promote insulin resistance through modulation of T cells and
534 production of pathogenic IgG antibodies. *Nat. Med.* **17**, 610–7 (2011).
- 535 41) Winer, D. a, Winer, S., Chng, M. H. Y., Shen, L. & Engleman, E. G. B Lymphocytes in
536 obesity-related adipose tissue inflammation and insulin resistance. *Cell. Mol. Life Sci.* **71**,
537 1033–1043 (2013).
- 538 42) Claussnitzer, M. et al. FTO Obesity Variant Circuitry and Adipocyte Browning in Humans.
539 *N. Engl. J. Med.* **373**, 895–907 (2015).
- 540 43) Hjellevik, V., Tverdal, A & Furu, K. Body mass index as predictor for asthma: a cohort
541 study of 118,723 males and females. *Eur. Respir. J.* **35**, 1235–42 (2010).
- 542 44) Inamasu, J., Guiot, B. H. & Sachs, D. C. Ossification of the posterior longitudinal
543 ligament: an update on its biology, epidemiology, and natural history. *Neurosurgery* **58**,
544 1027–39 (2006).

- 545 45) Tam, E. M. et al. Lower Muscle Mass and Body Fat in Adolescent Idiopathic Scoliosis
546 are Associated With Abnormal Leptin Bioavailability. *Spine* (Phila. Pa. 1976). **41**, 940–6
547 (2015).
- 548 46) Turesson, C., Bergström, U., Pikwer, M., Nilsson, J.-Å. & Jacobsson, L. T. H. A high
549 body mass index is associated with reduced risk of rheumatoid arthritis in men, but not in
550 women. *Rheumatology* (Oxford). **55**, 307–14 (2016).
- 551 47) Qin, B. et al. Body mass index and the risk of rheumatoid arthritis: a systematic review
552 and dose-response meta-analysis. *Arthritis Res. Ther.* **17**, 86 (2015).
- 553 48) Zammit, S. et al. Height and body mass index in young adulthood and risk of
554 schizophrenia: a longitudinal study of 1 347 520 Swedish men. *Acta Psychiatr. Scand.*
555 **116**, 378–85 (2007).
- 556 49) Wyatt, R. J., Henter, I. D., Mojtabai, R. & Bartko, J. J. Height, weight and body mass
557 index (BMI) in psychiatrically ill US Armed Forces personnel. *Psychol. Med.* **33**, 363–8
558 (2003).
- 559

560

561 **Figure legends for main text**

562

563 **Figure 1. Enrichment of identified variants in active enhancer.**

564 We evaluated enrichment of the variants included in the 99% credible sets for active
565 enhancer. We calculated empirical P -values by 1×10^7 permutations, and estimated FDR
566 using Benjamini–Hochberg method. Vertical dashed lines denote $\text{FDR} = 0.05$. (a)
567 Enrichment for active enhancer in 10 cell-groups. (b) Enrichment for active enhancer in
568 immune related cell group and CNS group. We showed the cell-types which showed
569 P -values < 0.05 . Each bar is colored according to the cell-groups as shown in (a).

570

571 **Figure 2. Enrichment of identified variants in H3K4me3 peaks.**

572 We analyzed enrichment of identified variants in H3K4me3 peaks of 34 cell types using
573 epi-GWAS software with 1×10^6 permutations. The scale of the X axis corresponds to the
574 P -value in the \log_{10} scale for each cell type. The dashed line indicates $P = 0.05$.

575

576 **Figure 3. Variance explained by subsets of associated variants in Japanese cohorts.**

577 We plotted the estimated variance explained by subsets of variants below the P -value
578 threshold from the current Japanese GWAS (red circle) and the GWAS of Europeans (blue
579 circle). Explained variance was estimated from two Japanese population-based cohorts ($N =$
580 5,612 and 6,434 in the JPHC and the TMM, respectively). Weighted average from two
581 cohorts is shown. Error bars denote the standard error of estimates. Single asterisks indicate
582 significant differences between the explained variance estimated from each study ($P < 0.05$).
583 Double asterisks denote the Bonferroni-corrected level of significance ($P < 0.05/7$).

584

585 **Figure 4. Scatter plot of the effect sizes for BMI and T2D**

586 Effect sizes of 193 identified variants for BMI (X axis) and T2D (y axis). Plots were colored
587 according to the significance level for T2D.

588

589 **Figure 5. Genetic correlation**

590 Each plot shows the genetic correlation (r_g) between BMI and analyzed traits (a: diseases,
591 and b: hematological traits) estimated from cross-trait LD score regression. Error bars
592 indicate the standard error of estimates. Significant (FDR $q < 0.05$) genetic correlations are
593 colored red and blue for positive and negative relationships, respectively. OPLL: ossification
594 of posterior longitudinal ligament of the spine, PAD: peripheral artery disease, AMD:
595 age-related macular degeneration, AIS: adolescent idiopathic scoliosis, RBC: red blood cell,
596 WBC: white blood cell.

Table 1. Result of the condition analysis

SNP	Positional Candidate gene	Chr.	Position ^a	Allele REF/ALT	Crude			Condition		
					beta ^b	s.e.m	P value	beta ^b	s.e.m	P value
rs939584	<i>FAM150B, TMEM18</i>	2	621,558	C/T	0.054	0.006	1.39E-19	0.053	0.006	1.55E-18
rs4430979	<i>FAM110C</i>	2	47,208	T/G	0.031	0.005	5.40E-10	0.029	0.005	6.32E-09
rs2495707	<i>HIF1AN, PAX2</i>	10	102,425,949	A/G	-0.025	0.004	6.21E-09	-0.023	0.004	3.80E-08
rs11191021	<i>BTRC</i>	10	103,231,641	A/G	0.021	0.004	8.12E-09	0.020	0.004	4.97E-08
rs2540034	<i>ADCY9</i>	16	4,022,694	C/T	0.028	0.004	3.49E-11	-0.029	0.004	1.11E-11
rs7199766	<i>SLX4</i>	16	3,637,429	G/A	0.020	0.004	1.25E-07	0.021	0.004	3.95E-08
rs1259768	<i>GPR139, GP2</i>	16	20,258,432	C/A	-0.030	0.005	6.80E-11	-0.031	0.005	3.90E-11
rs1011939	<i>GPRC5B, GPR139</i>	16	19,992,996	G/A	-0.025	0.004	8.26E-11	-0.025	0.004	4.73E-11

Result of the conditional analysis.

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

^aPositions are based on Build37 (hg19).

^bAlternative alleles were treated as effective allele.

Table. 2 New loci identified by trans-ethnic meta-analyses

SNP	Analysis	Chr. ^a	1KG phase 1					BBJ GWAS			European GWAS			Trans-ancestral meta-analysis								
			Position ^a (bp)	Alleles REF/ALT	ALT freq		Candidate gene(s)							N all	MANTRA				Fixed-effect meta-analysis			
					EAS	EUR		beta ^b	s.e.m	P value	beta ^b	s.e.m	P value		Posterior mean allelic effect	s.e.m	Log ₁₀ BF	Log ₁₀ BF for heterogeneity	beta ^b	s.e.m	P value	Phet ^c
rs12044597	All	1	1,708,801	A/G	0.47	0.51	NADK	0.013	0.004	3.16E-04	0.015	0.003	1.35E-06	474,968	0.014	0.002	7.49	0.05	0.014	0.002	1.54E-09	0.64
rs11185092	All	1	107,886,278	A/G	0.16	0.22	NTNG1	0.020	0.005	7.46E-05	0.018	0.005	9.97E-05	392,297	0.019	0.003	6.29	0.07	0.019	0.003	2.89E-08	0.80
rs10923724	All	1	119,546,842	C/T	0.58	0.57	TBX15,WARS2	-0.015	0.004	2.89E-05	-0.012	0.003	1.33E-04	480,391	-0.013	0.002	6.28	-0.09	-0.013	0.002	2.34E-08	0.46
rs10754220	All	1	197,244,290	G/A	0.19	0.30	CRB1	-0.025	0.005	2.86E-07	-0.013	0.004	1.28E-03	392,262	-0.018	0.003	6.85	-0.02	-0.018	0.003	7.90E-09	0.07
rs823114	All	1	205,719,532	G/A	0.50	0.53	NUCKS1	-0.014	0.004	4.51E-05	-0.012	0.003	2.02E-04	473,158	-0.013	0.002	6.12	0.04	-0.013	0.002	4.29E-08	0.53
rs9786986	All	1	235,656,632	G/T	0.49	0.11	B3GALNT2	0.016	0.004	4.80E-06	0.020	0.005	7.43E-05	480,088	0.017	0.003	7.28	-0.15	0.017	0.003	1.55E-09	0.60
rs4596023	All	2	48,955,683	A/G	0.66	0.73	LHCGR,STON1-GTF2A1L	0.016	0.004	1.14E-05	0.013	0.004	2.84E-04	480,125	0.014	0.003	6.45	-0.14	0.014	0.003	1.50E-08	0.54
rs7569376	All	2	205,385,322	T/C	0.47	0.79	ICOS,PARD3B	-0.017	0.004	3.70E-06	-0.015	0.005	2.52E-03	391,110	-0.016	0.003	6.02	-0.04	-0.016	0.003	3.50E-08	0.73
rs972540	All	2	207,244,783	A/G	0.19	0.24	ZDBF2,ADAM23	0.013	0.005	7.66E-03	0.018	0.003	2.28E-07	480,378	0.016	0.003	6.77	0.10	0.016	0.003	6.39E-09	0.37
rs7613875	All	3	49,971,514	C/A	0.14	0.52	MON1A,RBM6	0.015	0.005	1.50E-03	0.016	0.003	7.36E-07	464,114	0.016	0.003	7.00	-0.08	0.016	0.003	4.38E-09	0.95
rs1225051	All	3	131,645,972	G/A	0.46	0.44	CPNE4	0.012	0.004	6.28E-04	0.016	0.004	9.32E-06	392,295	0.014	0.003	6.29	0.13	0.014	0.003	3.11E-08	0.41
rs7621025	All	3	136,272,246	T/C	0.84	0.73	STAG1	0.015	0.005	1.77E-03	0.017	0.004	2.80E-06	480,294	0.016	0.003	6.47	0.10	0.016	0.003	2.07E-08	0.72
rs4834272	All	4	113,313,986	T/C	0.53	0.31	ALPK1	0.014	0.004	5.33E-05	0.013	0.003	3.71E-05	480,262	0.014	0.002	6.77	0.26	0.014	0.002	1.04E-08	0.83
rs7720894	All	5	60,733,933	G/C	0.39	0.48	ZSWIM6	-0.020	0.004	6.12E-07	-0.011	0.004	3.22E-03	392,090	-0.015	0.003	6.20	0.00	-0.015	0.003	2.89E-08	0.10
rs3849724	All	5	173,290,977	G/T	0.25	0.50	LINC01485,CPEB4	-0.018	0.004	5.85E-06	-0.014	0.003	9.76E-06	480,165	-0.015	0.003	7.94	-0.02	-0.015	0.002	4.95E-10	0.37
rs2228213	All	6	12,124,855	G/A	0.23	0.33	HIVEP1	-0.017	0.004	1.60E-05	-0.016	0.003	1.04E-06	480,386	-0.016	0.003	8.65	0.03	-0.016	0.003	1.03E-10	0.81
rs7779181	All	7	32,345,283	T/C	0.23	0.20	PDE1C,LOC100130673	0.015	0.004	2.52E-04	0.019	0.004	2.37E-05	392,261	0.017	0.003	6.32	0.22	0.017	0.003	2.63E-08	0.58
rs10269783	All	7	49,616,203	G/A	0.60	0.37	CDC14C,VWC2	0.015	0.004	3.42E-05	0.014	0.003	4.14E-06	477,965	0.015	0.002	7.78	-0.20	0.015	0.002	4.87E-10	0.92

rs3779273	All	7	77,828,940	G/A	0.44	0.40	<i>MAGI2</i>	-0.014	0.004	1.18E-04	-0.015	0.003	1.52E-06	480,399	-0.014	0.002	7.66	-0.17	-0.014	0.002	7.55E-10	0.82
rs12156392	All	8	28,140,103	C/A	0.35	0.41	<i>ELP3,PNOC</i>	0.019	0.004	6.59E-07	0.011	0.004	5.28E-03	392,039	0.015	0.003	6.16	0.06	0.015	0.003	4.42E-08	0.11
rs10091344	All	8	34,132,075	G/A	0.44	0.34	<i>DUSP26,LINC01288</i>	-0.015	0.004	2.21E-05	-0.016	0.004	9.62E-05	392,217	-0.015	0.003	6.80	0.05	-0.015	0.003	8.33E-09	0.93
rs10811901	All	9	23,356,935	G/A	0.86	0.62	<i>LINC01239,LOC101929563</i>	0.023	0.004	7.57E-08	0.010	0.003	1.89E-03	461,004	0.015	0.003	6.54	0.11	0.015	0.003	1.26E-08	0.02
rs580809	All	9	101,518,442	C/T	0.56	0.75	<i>ANKS6</i>	-0.016	0.004	7.74E-06	-0.012	0.004	6.75E-04	480,376	-0.014	0.002	6.31	0.02	-0.014	0.002	2.70E-08	0.42
rs7024334	All	9	109,072,075	T/G	0.42	0.77	<i>TMEM38B,MIR8081</i>	-0.014	0.004	1.55E-04	-0.016	0.004	1.41E-05	473,417	-0.015	0.003	6.53	-0.07	-0.015	0.003	1.14E-08	0.66
rs2270204	All	9	131,042,734	T/G	0.58	0.27	<i>SWI5</i>	0.016	0.004	5.47E-06	0.019	0.004	2.60E-07	475,247	0.018	0.003	9.56	-0.12	0.018	0.003	9.02E-12	0.63
rs2163188	All	10	65,314,711	G/C	0.33	0.50	<i>REEP3</i>	0.015	0.004	8.63E-05	0.015	0.004	5.03E-05	388,579	0.015	0.003	6.50	0.72	0.015	0.003	1.67E-08	0.99
rs5215	All	11	17,408,630	C/T	0.62	0.66	<i>KCNJ11</i>	0.018	0.004	5.33E-07	0.014	0.003	1.13E-05	474,995	0.016	0.002	9.10	0.14	0.016	0.002	3.22E-11	0.34
rs3026401	All	11	31,807,524	C/T	0.52	0.81	<i>PAX6</i>	0.016	0.004	3.59E-06	0.016	0.005	8.37E-04	392,284	0.016	0.003	6.56	0.20	0.016	0.003	1.13E-08	0.90
rs506338	All	11	64,440,920	T/C	0.23	0.71	<i>NRXN2</i>	0.017	0.005	2.30E-04	0.015	0.003	1.30E-05	480,264	0.016	0.003	6.65	-0.04	0.016	0.003	9.72E-09	0.62
rs11607976	All	11	69,279,111	C/T	0.07	0.30	<i>MYEOV,LINC01488</i>	-0.027	0.008	6.54E-04	-0.017	0.003	7.72E-07	479,024	-0.018	0.003	7.01	0.18	-0.018	0.003	3.85E-09	0.25
rs10899469	All	11	78,018,313	T/C	0.41	0.19	<i>GAB2</i>	-0.015	0.004	2.51E-05	-0.016	0.004	1.66E-04	477,329	-0.015	0.003	6.46	-0.02	-0.015	0.003	1.49E-08	0.89
rs10772983	All	12	17,141,582	C/T	0.68	0.57	<i>SKP1P2</i>	-0.014	0.004	2.42E-04	-0.012	0.003	4.62E-05	478,531	-0.013	0.002	6.18	0.34	-0.013	0.002	3.45E-08	0.73
rs11105839	All	12	91,237,920	T/A	0.48	0.36	<i>LINC00936,LINC00615</i>	-0.013	0.004	2.84E-04	-0.013	0.003	2.97E-05	474,917	-0.013	0.002	6.28	0.16	-0.013	0.002	2.54E-08	0.96
rs2321882	All	13	59,451,989	G/C	0.09	0.24	<i>LOC101926897,DIAPH3</i>	0.026	0.006	3.60E-06	0.016	0.005	3.64E-04	392,005	0.020	0.004	6.52	-0.07	0.020	0.004	1.32E-08	0.17
rs1927790	All	13	96,922,191	T/C	0.36	0.45	<i>HS6ST3</i>	0.010	0.004	5.29E-03	0.015	0.003	7.87E-07	480,335	0.013	0.002	6.15	0.05	0.013	0.002	2.70E-08	0.27
rs12895330	All	14	33,305,343	G/C	0.24	0.49	<i>AKAP6,NPAS3</i>	0.020	0.004	6.17E-06	0.016	0.004	1.07E-04	385,665	0.018	0.003	7.18	0.18	0.018	0.003	3.54E-09	0.44
rs709400	All	14	104,149,475	A/G	0.10	0.40	<i>KLC1</i>	-0.010	0.005	5.35E-02	-0.017	0.003	1.02E-07	480,397	-0.015	0.003	6.05	-0.40	-0.015	0.003	1.80E-08	0.31
rs12440086	All	15	27,038,492	C/A	0.59	0.54	<i>GABRB3,GABRA5</i>	0.017	0.004	3.02E-06	0.012	0.004	1.74E-03	392,170	0.015	0.003	6.21	-0.05	0.015	0.003	3.06E-08	0.33
rs1559677	All	15	47,738,063	A/G	0.43	0.36	<i>SEMA6D</i>	0.016	0.004	1.06E-05	0.011	0.003	5.77E-04	480,342	0.013	0.002	6.01	-0.23	0.013	0.002	3.37E-08	0.30
rs9302817	All	16	6,163,936	T/G	0.43	0.18	<i>RBFOX1</i>	-0.015	0.004	6.15E-05	-0.019	0.005	1.54E-04	392,179	-0.016	0.003	6.02	0.06	-0.016	0.003	4.56E-08	0.49
rs7200543	All	16	15,129,970	A/G	0.38	0.32	<i>PDXDC1</i>	-0.015	0.004	4.16E-05	-0.015	0.003	1.03E-05	480,268	-0.015	0.002	7.44	-0.14	-0.015	0.002	1.24E-09	0.97
rs7195386	All	16	24,578,458	T/C	0.46	0.53	<i>RBBP6</i>	-0.016	0.004	6.15E-06	-0.013	0.004	4.89E-04	392,282	-0.015	0.003	6.58	0.38	-0.015	0.003	1.42E-08	0.51
rs2307022	All	16	68,381,978	A/G	0.88	0.65	<i>PRMT7</i>	-0.015	0.006	8.15E-03	-0.016	0.003	9.86E-07	480,262	-0.016	0.003	6.26	0.28	-0.016	0.003	3.41E-08	0.83

rs4925114	All	17	17,711,270	A/G	0.07	0.60	<i>RAI1</i>	0.029	0.006	1.86E-06	0.014	0.004	5.11E-04	392,169	0.019	0.003	6.26	0.18	0.018	0.003	3.04E-08	0.04
rs4986044	All	17	21,261,560	C/T	0.58	0.49	<i>MAP2K3,KCNJ12</i>	-0.013	0.004	3.65E-04	-0.016	0.003	1.42E-07	476,195	-0.015	0.002	8.32	1.22	-0.015	0.002	2.17E-10	0.53
rs12150665	All	17	34,914,787	T/C	0.39	0.40	<i>GGNBP2</i>	-0.015	0.004	5.06E-05	-0.016	0.003	2.44E-07	480,289	-0.016	0.002	8.98	0.12	-0.016	0.002	4.67E-11	0.78
rs6504108	All	17	46,292,923	C/T	0.78	0.70	<i>SKAP1</i>	-0.019	0.004	1.04E-05	-0.017	0.003	3.58E-07	480,392	-0.018	0.003	9.46	0.06	-0.018	0.003	1.51E-11	0.74
rs9304204	All	18	36,896,003	A/G	0.83	0.81	<i>LINC00669</i>	0.017	0.004	4.56E-05	0.018	0.005	2.67E-04	392,282	0.017	0.003	6.03	0.01	0.017	0.003	4.53E-08	0.92
rs4129322	All	18	50,605,642	G/A	0.31	0.10	<i>DCC</i>	-0.020	0.004	6.21E-08	-0.011	0.007	1.23E-01	392,300	-0.018	0.003	6.07	0.11	-0.018	0.003	4.23E-08	0.20
rs7243785	All	18	52,475,162	A/G	0.38	0.23	<i>DYNAP,RAB27B</i>	0.018	0.004	2.70E-06	0.018	0.004	6.97E-05	382,050	0.018	0.003	7.85	0.01	0.018	0.003	7.79E-10	0.91
rs12454712	All	18	60,845,884	T/C	0.45	0.40	<i>BCL2</i>	0.019	0.004	4.79E-07	0.017	0.004	1.46E-05	401,822	0.018	0.003	9.06	-0.09	0.018	0.003	3.26E-11	0.77
rs17710386	All	18	63,461,201	T/C	0.31	0.34	<i>CDH7</i>	0.017	0.004	1.94E-05	0.012	0.003	2.26E-04	474,561	0.014	0.003	6.10	-0.13	0.014	0.003	2.77E-08	0.41
rs4805566	All	19	30,918,316	T/A	0.69	0.67	<i>ZNF536</i>	0.018	0.004	2.23E-05	0.018	0.004	8.59E-06	389,755	0.018	0.003	7.74	0.01	0.018	0.003	7.88E-10	0.98
rs1884897	All	20	6,612,832	A/G	0.89	0.60	<i>CASC20,BMP2</i>	0.014	0.006	2.42E-02	0.016	0.003	6.83E-07	480,315	0.015	0.003	6.12	0.02	0.015	0.003	3.64E-08	0.79

Loci reached genome-wide significance after meta-analysis to European results were shown.

Chr., chromosome; REF, reference allele; ALT, alternative allele; BF, Bayes' factor.

^aPositions are based on Build37 (hg19).

^bAlternative alleles were treated as effective allele.

^cP for heterogeneity between studies estimated by Cochran's Q test.

METHOD

Subjects

We used BMI and genome-wide variant data obtained from the BBJ¹¹⁻¹², which has enrolled approximately 200,000 individuals consisting of 47 various disease¹³. For the GWAS, we set the eligibility criteria as follows: 1) aged ≥ 18 , 2) registered height and weight, and 3) individuals whose height was within 3 times of interquartile range (IQR) of upper/lower quartile. 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 smartpca⁵⁰, 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 out of ± 4 standard deviations (SD) 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 PLINK⁵¹. 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, age², 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, age², 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 MACH⁵². 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 $\leq 1.0 \times 10^{-6}$ from the reference panel. After the imputation, we used SNPs with an imputation quality of $R_{sq} \geq 0.7$ for the GWAS.

In the population-based cohorts, DNA samples were genotyped using the Illumina HumanOmniExpressExome BeadChip. We phased haplotypes using SHAPIT⁵³ (v2.r778) and performed imputation using Minimac3⁵² (1.0.13) with the same reference panel used for the GWAS. For replication purposes, we allowed $R_{sq} \geq 0.5$ throughout the study.

To investigate the X chromosome, we called genotype using BeadStudio. First, we

generated genoplot using only female. After that, we added male samples to genoplot and called genotypes. Genotypes called as heterozygote were treated as missing in male. For the genotype QC, we excluded variants with a MAF < 0.05 % and a variant call rate ≤ 0.99 in either male or female. We also excluded variants with P -value for Hardy–Weinberg equilibrium $\leq 1.0 \times 10^{-6}$ in female. Haplotype phasing and imputation were performed separately for males and females in the GWAS. Allelic dosage were imputed from 0 to 2 in male 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 mach2qtl⁵². 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 step-wise 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** = $\beta^2(1 - f)2f$. To estimate the additive

explained variance of 83 independent lead variants of identified loci, the explained variance of each individual variant was summed.

R (ver. 3.1.3) was used for these analyses. Regional association plots were generated by Locuszoom⁵⁴ (ver.1.3) 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 regression¹⁵ with the default settings. LD scores for the East Asian population provided with software was 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, age², 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, and gestational diabetes were excluded from this analysis. We analyzed 193 lead variants of the loci reached genome-wide significant level. In case the lead variants were differed 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, age², 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 result of Europeans⁶ which was conducted by 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 GWASs in the same way.

For the meta-analyses, we used MANTRA¹⁹ (v2) software which 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 study²⁰. 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 which are likely to be causal based on the MANTRA¹⁹ 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 j_{th} variant under the assumption of a single causal variant in each locus using the following formula:

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

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 regression¹⁶ was conducted using the results from the current GWAS of all samples, the GWASs of 33 complex diseases, and the hematological traits with the LD scores for the East Asian population (**Supplementary note**). All the GWASs that were analyzed only included Asians. We calculated the false discovery rate using the Benjamini–Hochberg procedure.

GREML analysis

To estimate the variance explained by autosomal variants, we selected imputed variants with $R_{sq} \geq 0.7$ and generated a genetic relatedness matrix (GRM) for the Japanese cohorts by GCTA (ver. 1.25)²⁶. 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, age², 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 Feb 19, 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: 1) variants with $r^2 \geq 0.5$ in East Asians if the publication included Asians, 2) variants with $r^2 \geq 0.5$ in both East Asians and other population if the original report not studied East Asian, and 3) 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 which are relevant to obesity, we conducted an enhancer enrichment analysis using variants included in the 99% credible sets constructed by trans-ethnic meta-analysis. Since 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-types. We observed that active enhancer in the ChromHMM 18 state model (state 9) shows lowest overlaps across cell-types among analyzed enhancers (**Supplementary Note**). Therefore, we determined to use active enhancer of ChromHMM 18 state model for this analysis.

For the cell-type and cell-group specificity analysis, we excluded the cell-types which are derived from cell-lines, labeled as cultured cell, and belonging to ESC or iPSC

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 which are located at 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 (1: $MAF < 5\%$, 2: $5\% < MAF \leq 15\%$, 3: $15\% < MAF \leq 30\%$, 4: $MAF > 30\%$) 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 overlaps of tissue-specific H3K4me3 at 85 identified autosomal variants, we used epiGWAS software²⁴. We calculated LD (r^2 cutoff = 0.8) between the lead variant and variants within 500kb 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 was included in the software. To estimate the statistical significance of cell-type-specific overlaps, 10^6 permutations were performed. The H3K4me3 peaks were shown in a regional association plot (**Supplementary Data Set 1 and 2**). We also analyzed 97 variants reported by GIANT consortium in the same way using LD information of the EUR samples in 1KGP.

Functional annotation and eQTL analysis

For the annotation of the nonsynonymous variant, ANNOVAR⁵⁵ 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 its involvement by our cell-type specificity analyses. To prioritize eQTLs, we used GTEx²⁴ (Release v6) and whole-blood eQTL information²⁵. These data were downloaded from their website. We set significant threshold for eQTL as $FDR < 0.05$. In each datasets, we only considered the variants which were analyzed in our trans-ethnic analysis, and selected the variants which showed strongest association for each transcript. We regarded significant overlap if the GWAS lead variant was in LD ($r^2 \geq 0.8$) in EUR samples of 1KGP with lead variant of the eQTL.

Pathway analysis

We used MAGENTA²⁸ 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 PLINK⁵¹. Default parameters were applied for other settings. We regarded the pathways that showed a $FDR\ q < 5\%$ as significant.

Data availability

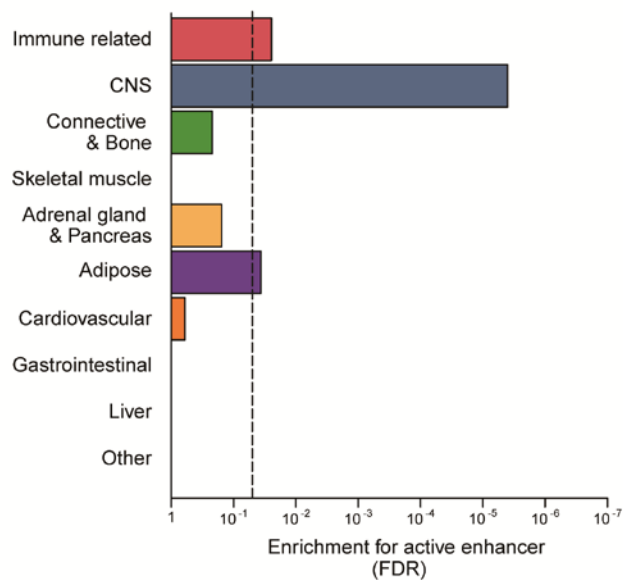
The genotype data, BMI measurements, and related phenotype information that support the findings of this study are available in Japanese Genotype-phenotype Archive (JGA; http://trace.ddbj.nig.ac.jp/jga/index_e.html) with the accession codes: JGAS00000000114 for the study, JGAD00000000123 for the genotype data, and JGAD00000000124 for the BMI measurements, respectively. The summary statistics of the GWAS have been deposited in the National Bioscience Database Center (NBDC) Human Database (<https://humandbs.biosciencedbc.jp/en/>) with the Data Set ID (hum0014.v6.158k.v1).

Methods-only references

- 50) Price, A. L. et al. Principal components analysis corrects for stratification in genome-wide association studies. *Nat. Genet.* **38**, 904–9 (2006).
- 51) Purcell, S. et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–75 (2007).
- 52) 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**, 816–34 (2010).
- 53) Delaneau, O., Marchini, J. & Zagury, J.-F. A linear complexity phasing method for thousands of genomes. *Nat. Methods* **9**, 179–81 (2012).
- 54) Pruim, R. J. et al. LocusZoom: regional visualization of genome-wide association scan results. *Bioinformatics* **26**, 2336–7 (2010).
- 55) Wang, K., Li, M. & Hakonarson, H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* **38**, e164 (2010).

Figure 1

a



b

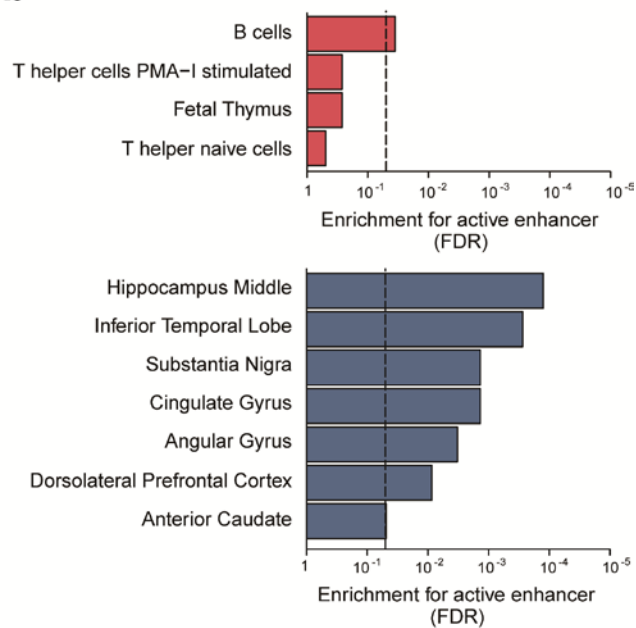


Figure 2

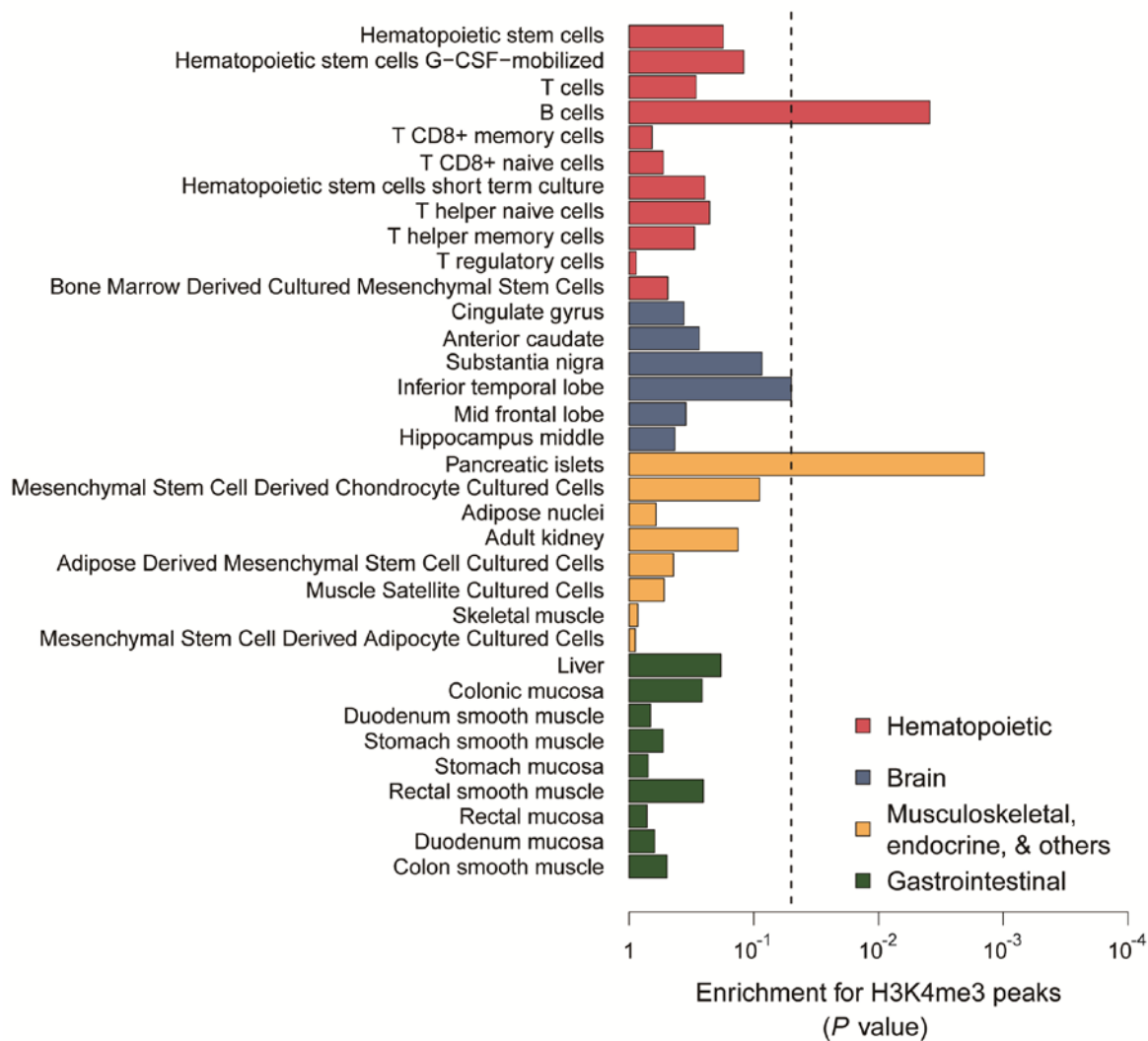


Figure 3

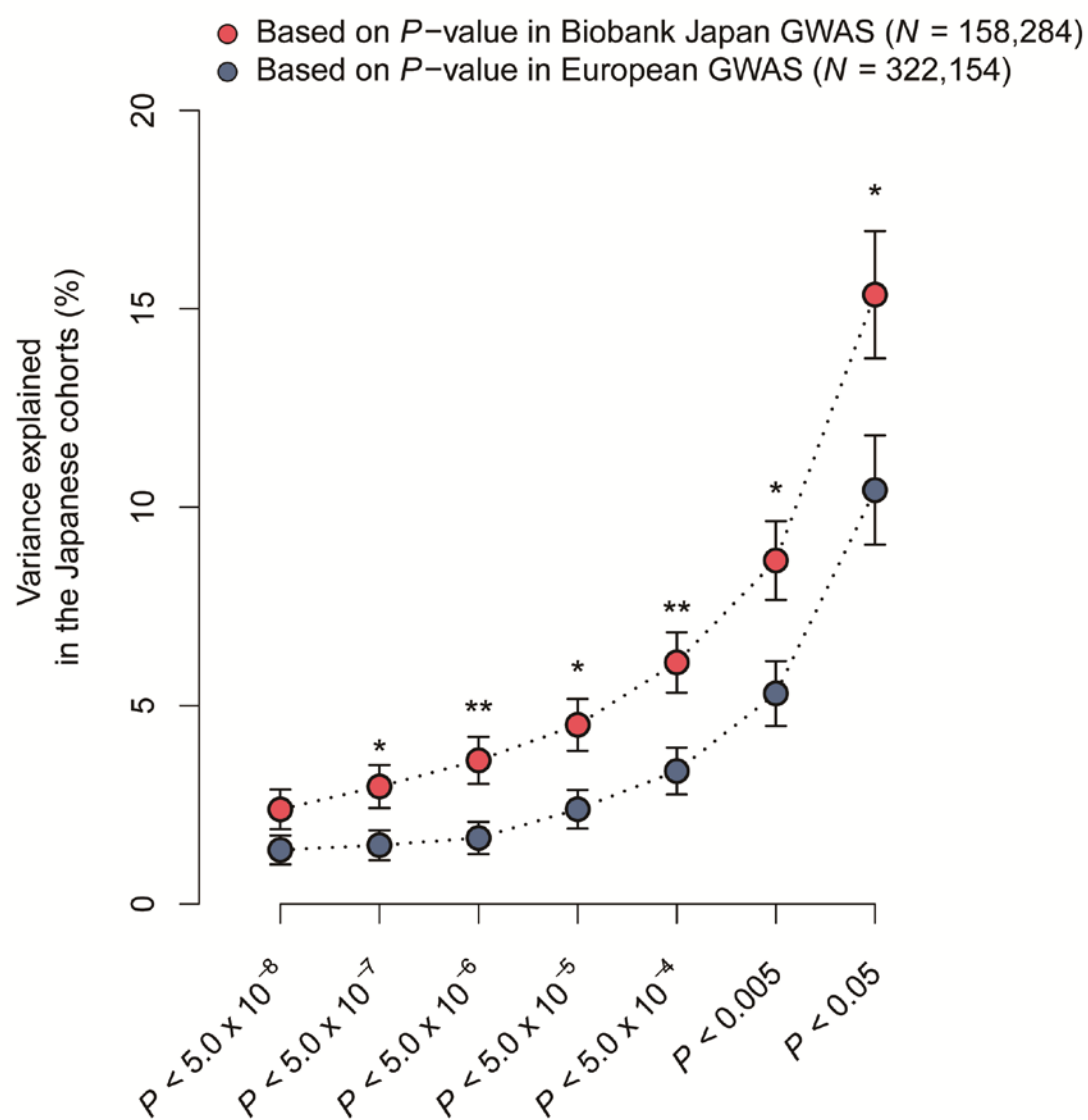


Figure 4

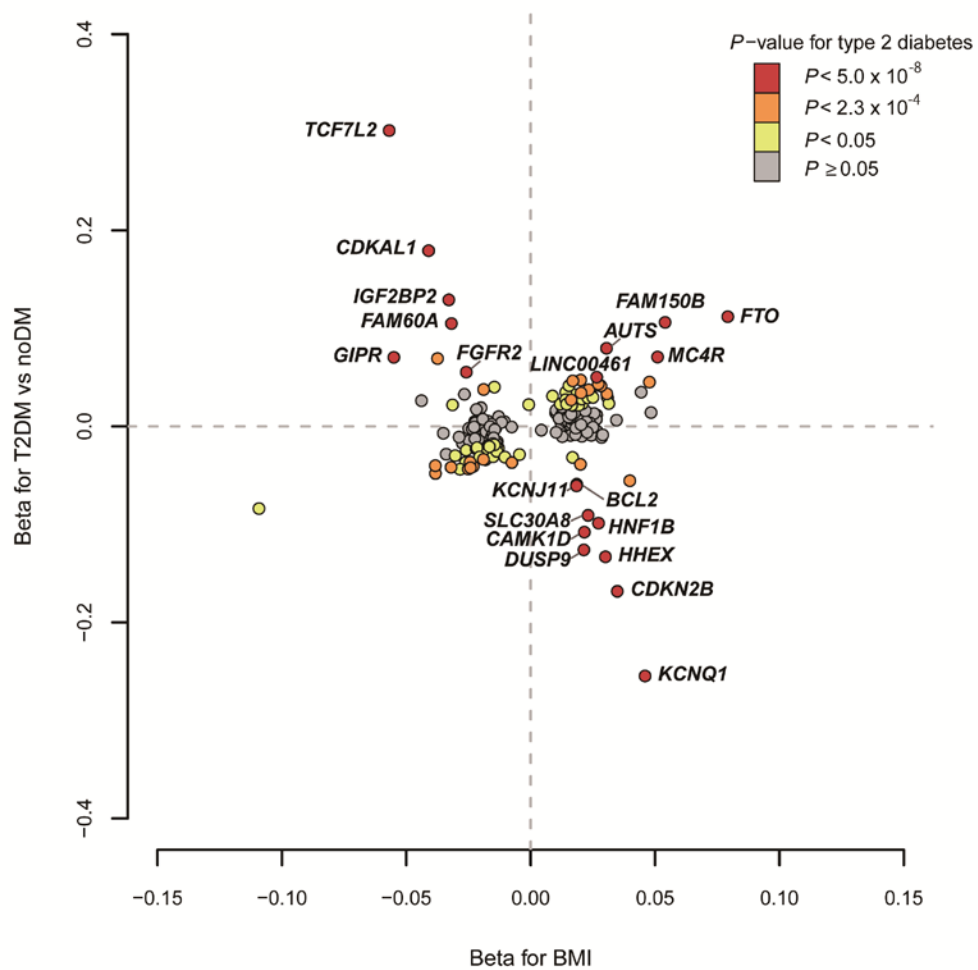
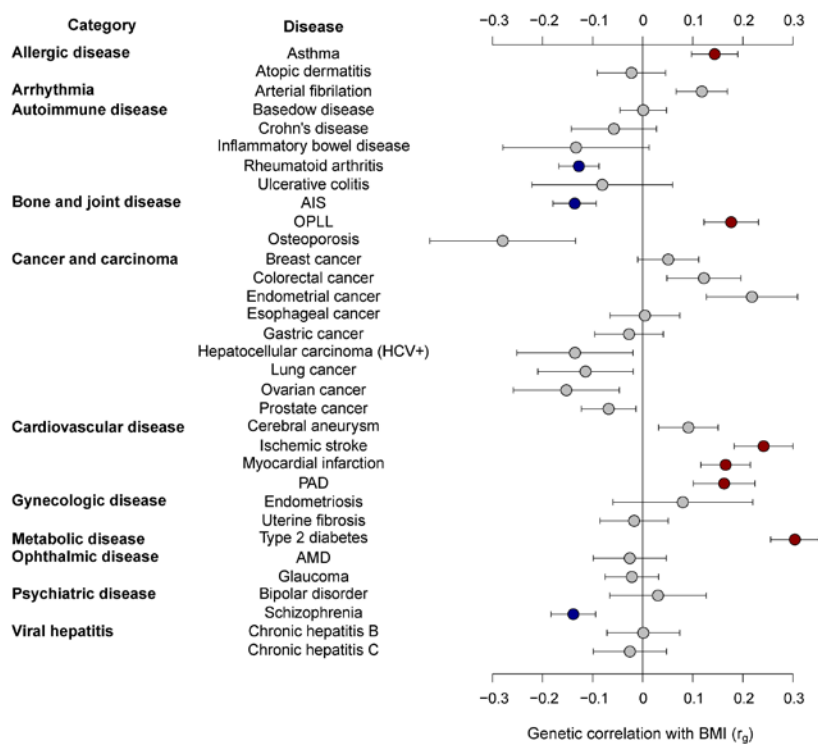


Figure 5

a.



b.

