Meta-analysis identifies common variants associated with body mass index in east Asians

Wanqing Wen1–3,53, Yoon-Shin Cho4,5,53, Wei Zheng1–3,53, Rajkumar Dorajoo6,7,53, Norihiro Kato8,53, Lu Qi9,53, Chien-Hsiun Chen10,11,53, Ryan J Delahanty1–3, Yukinori Okada12,13, Yasuharu Tabara14, Dongfeng Gu15, Dingliang Zhu16–19, Christopher A Haiman20, Zengnan Mo21, Yu-Tang Gao22, Seang-Mei Saw23, Min-Jin Go4, Fumihiko Takeuchi8, Li-Ching Chang10, Yoshihiro Kokubo24, Jun Liang25, Mei Hao26, Loïc Le Marchand27, Yi Zhang16–18, Yanling Hu28, Tien-Yin Wong29–31, Jirong Long1–3, Bok-Ghee Han4, Michiaki Kubo32, Ken Yamamoto33, Mei-Hsin Su10, Tetsuro Miki34, Brian E Henderson20, Huaidong Song35, Aihua Tan36, Jiang He26, Daniel P-K Ng23, Qiuyin Cai1–3, Tatsuhiko Tsunoda37, Fuu-Jen Tsai11, Naoharu Iwai38, Gary K Chen20, Jiajun Shi1–3, Jianfeng Xu39, Xueling Sim40, Yong-Bing Xiang22, Shiro Maeda41, Rick T H Ong6,42, Chun Li43, Yusuke Nakamura44, Tin Aung29,30, Naoyuki Kamatani12, Jian-Jun Liu6, Wei Lu45, Mitsuhiro Yokota46, Mark Seielstad6,47, Cathy S J Fann10, The Genetic Investigation of ANthropometric Traits (GIANT) Consortium48, Jer-Yuarn Wu10,11,54, Jong-Young Lee4,54, Frank B Hu9,49,54, Toshihiro Tanaka50,54, E Shyong Tai23,51,52,54 & Xiao-Ou Shu1–3,53

Multiple genetic loci associated with obesity or body mass index (BMI) have been identified through genome-wide association studies conducted predominantly in populations of European ancestry. We performed a meta-analysis of associations between BMI and approximately 2.4 million SNPs in 27,715 east Asians, which was followed by in silico and de novo replication studies in 37,691 and 17,642 additional east Asians, respectively. We identified ten BMI-associated loci at genome-wide significance (\(P < 5.0 \times 10^{-8}\)), including seven previously identified loci (FTO, SEC16B, MC4R, GIPR-QPCTL, ADCY3-DNAJC27, BDNF and MAP2K5) and three novel loci in or near the CDKAL1, PCSK1 and GP2 genes. Three additional loci nearly reached the genome-wide significance threshold, including two previously identified loci in the GNPDA2 and TFPAP2B genes and a newly identified signal near PAX6, all of which were associated with BMI with \(P < 5.0 \times 10^{-7}\). Findings from this study may shed light on new pathways involved in obesity and demonstrate the value of conducting genetic studies in non-European populations.

Genome-wide association studies (GWAS) have thus far identified 37 genetic loci associated with obesity or BMI1–11. Virtually all of these studies were conducted in populations of European ancestry and included limited data from Asian populations8,11. Asians, who account for over 60% of the world’s population, have higher percentages of body fat and increased metabolic disease risk than individuals of European ancestry with the same BMI12. Therefore, studies conducted in Asian populations, in addition to allowing an evaluation of the extent to which genetic markers of obesity identified in North American and European populations can be generalized, also facilitate the dissection of the genetic architecture of obesity and the identification of genetic variants of particular importance in Asians.

The initial genome-wide association meta-analysis of BMI included approximately 2.4 million genotyped or imputed SNPs generated from eight GWAS including 27,715 east Asians (stage 1). This was followed by an in silico replication analysis conducted among 37,691 east Asians from an additional seven GWAS (stage 2) and a subsequent de novo replication study conducted among 17,642 east Asians from three studies (stage 3). Details of the study designs are provided in Supplementary Figure 1, Supplementary Tables 1–3 and the Supplementary Note.

The stage 1 meta-analysis was performed using the METAL program, and study-specific genomic control adjustment was applied (see Online Methods). The stage 1 analysis revealed that three well-established loci (FTO, SEC16B and MC4R) were associated with BMI at or near the level of genome-wide significance \((P < 5 \times 10^{-8})\) (Fig. 1 and Table 1).

In stage 2, we analyzed 798 SNPs that were associated with BMI at \(P < 1.0 \times 10^{-4}\) in stage 1 and 50 additional SNPs that were previously reported to be associated with BMI but which did not reach \(P < 1.0 \times 10^{-4}\) in stage 1. Seven additional GWAS conducted in east Asian populations participated in the stage 2 study. We combined data from stage 2 with the stage 1 meta-analysis results in meta-analyses with adjustment for both study-specific genomic control inflation and estimated significance.
The reported effect sizes for BMI-related SNPs in studies of populations of European ancestry are usually greater than 3% of the s.d. of BMI. Given the sample sizes of our study, we had adequate statistical power (>0.8) to detect a SNP with such an effect size and with a MAF of >0.2 in stage 1 or a MAF of >0.08 in the combined stage 1 and 2 data at a significance of $P < 0.05$. The index SNPs in the 19 previously identified loci that were not replicated in our study at $P < 0.05$ had either very small effect sizes or very low MAFs in east Asians (Supplementary Table 4).

One representative SNP from the four newly identified loci at or near the CDKL1, PCSK1, PAX6 and GP2 genes and the three loci at the GIPR-QPCTL, ADCY3-DNAJC27 and MAP2K5 genes that were reported by the GIANT consortium (Supplementary Table 4) were selected for further replication in stage 3 using de novo genotyping in three studies that included a total of 17,642 subjects (Supplementary Tables 1 and 2). In stage 3 analyses, the directions of the associations between BMI and the seven SNPs were consistent with the corresponding associations in stages 1 and 2. The final results derived from combined data from all three stages showed that six SNPs at or near GIPR-QPCTL, ADCY3-DNAJC27, MAP2K5, DNAJC27, PCSK1 and GP2 were associated with BMI with genome-wide significance ($P = 1.02 \times 10^{-8}$ to $9.39 \times 10^{-14}$) (Table 1), and rs652722 near the PAX6 gene was found to be associated with BMI with $P = 7.65 \times 10^{-8}$ (Supplementary Table 6). The variance in BMI explained by these SNPs is presented in Table 1.

We also evaluated the association of BMI with these seven SNPs in the GIANT consortium data. Four of these SNPs (rs65481, rs4776970 and rs1167166) at or near the AGCY3-DNAJC27, PCSK1, MAP2K5 and GIPR-QPCTL loci, respectively, showed a significant association with BMI at $P < 0.007$ ($P = 0.05/7$, to account for seven tests) (Supplementary Table 7). Although the effect sizes of these seven loci were smaller than those of the well-established variants in the FTO, MC4R and SEC16B loci (2.55–4.22% of s.d. versus 5.51–7.92%; Table 1), their effect sizes were larger and the explained variances were larger in east Asians than in Europeans (Supplementary Table 7), with the exception of rs4776970 in the MAP2K5 gene, which was independently identified by both our study and the GIANT

![Manhattan plot showing the significance of associations between BMI and SNPs in the stage 1 data. The SNPs in previously reported genes that show significant associations with BMI are highlighted in red. The SNPs in newly identified loci that are significantly associated with BMI are highlighted in blue.](image)

**Figure 1**

- Table 1: Identified loci associated with BMI variation in east Asian populations
- Supplementary Table 4
- Supplementary Table 5
- Supplementary Table 7
- Supplementary Tables 1, 4, 5

| Gene | Chr. | SNP | Genotype | EAF | $\beta$ (s.e.m.) | $P$ value by stage$^b$ | Final$^b$ | Explained variance$^c$ |
|------|------|-----|----------|-----|-----------------|-------------------|-----------|---------------------|
|      |      |     |          |     |     | 1      | 2      | 3      |                     |
| FTO  | 16   | rs17817449 | G/T | 0.17 | 7.92 (1.06) | 6.13 $\times 10^{-12}$ | 8.18 $\times 10^{-14}$ | 4.60 $\times 10^{-27}$ | 0.18% |
| SEC16B | 1  | rs574367 | T/G | 0.20 | 5.93 (0.92) | 2.38 $\times 10^{-11}$ | 1.28 $\times 10^{-10}$ | 9.47 $\times 10^{-20}$ | 0.11% |
| MC4R | 18   | rs6567160 | G/T | 0.21 | 5.51 (0.93) | 6.32 $\times 10^{-10}$ | 3.35 $\times 10^{-9}$ | 2.13 $\times 10^{-15}$ | 0.10% |
| GIPR-QPCTL | 19 | rs11671664 | A/G | 0.50 | 4.22 (0.76) | 1.29 $\times 10^{-5}$ | 2.57 $\times 10^{-8}$ | 3.57 $\times 10^{-3}$ | 0.09% |
| ADCY3-DNAJC27 | 2 | rs6545814 | A/G | 0.45 | 3.26 (0.76) | 1.20 $\times 10^{-5}$ | 1.62 $\times 10^{-5}$ | 1.05 $\times 10^{-5}$ | 0.03% |
| BDNF | 11   | rs2626 | G/T | 0.44 | 4.97 (0.83) | 1.18 $\times 10^{-5}$ | 2.72 $\times 10^{-9}$ | 3.56 $\times 10^{-3}$ | 0.12% |
| MAP2K5 | 15 | rs4776970 | A/T | 0.22 | 2.55 (0.90) | 1.10 $\times 10^{-6}$ | 4.63 $\times 10^{-3}$ | 2.90 $\times 10^{-3}$ | 0.02% |
| GPNDA2 | 4  | rs10938397 | G/T | 0.29 | 3.72 (0.85) | 1.60 $\times 10^{-3}$ | 1.30 $\times 10^{-5}$ | 9.69 $\times 10^{-8}$ | 0.06% |
| TFAP2B | 6  | rs4715210 | C/T | 0.21 | 3.05 (0.91) | 1.12 $\times 10^{-5}$ | 7.64 $\times 10^{-4}$ | 1.61 $\times 10^{-7}$ | 0.03% |

| Gene | Chr. | SNP | Genotype | EAF | $\beta$ (s.e.m.) | $P$ value by stage$^b$ | Final$^b$ | Explained variance$^c$ |
|------|------|-----|----------|-----|-----------------|-------------------|-----------|---------------------|
|      |      |     |          |     |     | 1      | 2      | 3      |                     |
| CDDK1 | 6   | rs9356744 | G/T | 0.58 | 3.39 (0.76) | 3.21 $\times 10^{-5}$ | 7.67 $\times 10^{-6}$ | 3.02 $\times 10^{-3}$ | 0.06% |
| PCSK1 | 5   | rs261967 | C/A | 0.41 | 3.77 (0.77) | 1.22 $\times 10^{-5}$ | 9.36 $\times 10^{-7}$ | 8.46 $\times 10^{-1}$ | 0.07% |
| GP2  | 16   | rs12597579 | C/T | 0.80 | 4.09 (0.96) | 7.13 $\times 10^{-5}$ | 2.07 $\times 10^{-5}$ | 1.45 $\times 10^{-1}$ | 0.05% |
| PAX6 | 11   | rs652722 | C/T | 0.61 | 2.75 (0.77) | 2.84 $\times 10^{-5}$ | 3.70 $\times 10^{-4}$ | 1.89 $\times 10^{-1}$ | 0.04% |

Chr., chromosome.  
$^a$Effect allele/or other allele.  
$^b$Effect allele frequency in Asians, estimated from stages 1 and 2.  
$^c$Per allele effect of SNPs (in percentage) on BMI, obtained from stage 2 data only.  
$^d$Derived from meta-analysis.  
$^e$Values for the combined data were adjusted for both study-specific inflation factors and the estimated inflation factor for the stage 1 meta-analysis statistic.  
$^f$Combination of all available data from the three stages.  
$^g$The effect sizes obtained from stage 2 data were used to estimate the explained variance.
The explained variance of this SNP is 0.03% in Europeans (Supplementary Table 7) and 0.02% in Asians (Table 1).

As shown in Table 1, the SNP in FTO had the greatest effect on BMI and accounted for the largest proportion of the variance (0.18%) in our study population. Together, the 10 loci associated with BMI that reached genome-wide significance explained 0.87% of the interindividual variation in BMI, and all 22 loci that were associated with BMI at P < 0.05 explained 1.18% (Online Methods and Supplementary Table 4). These explained variance values are lower than those reported by the GIANT consortium (1.45% for the SNPs overall and 0.34% for FTO)\(^8\). After exclusion of SNPs within these 22 loci that were associated with BMI at P < 0.05, the number of SNPs with small observed P values for an association with BMI still exceeded the expected number (Fig. 2), suggesting that additional BMI-related loci remain to be uncovered in east Asians.

The newly identified associations of four SNPs at or near the CDKAL1, PCSK1, PAX6 and GP2 genes with BMI were consistent across studies, gender and ancestry and remained little changed after subjects with chronic diseases (cancer or diabetes) were excluded (Supplementary Table 6). Meta-analyses of obesity as a dichotomous outcome (BMI $\geq 27.5$)\(^{14}\) also showed similar associations, with odds ratios per allele ranging from 1.05 to 1.10 (Supplementary Table 8). Of the studies participating in our analyses, one stage 2 study, the Singapore Cohort study Of The Risk factors for Myopia (SCORM), only included children (9 year olds). In data from the SCROM study, all four loci had an association with BMI consistent with the meta-analysis, and the rs652722 SNP near PAX6 was associated with nominal significance (P = 0.0335) (Supplementary Table 6). Excluding the SCORM study from the analysis had little effect on the results.

The consistency of the findings across studies and populations suggests that population structure alone did not account for the significant associations we identified. In addition, multiple SNPs in LD with each other showed similar associations in the combined stage 1 and 2 data at each locus (Fig. 3 and Supplementary Table 9). This finding, together with the identification of similar associations in the de novo replication stage, suggests that our results are unlikely to have been caused by genotyping or imputation errors.

The locus represented by the rs9356744 SNP (6p22.3) contains the CDKAL1 gene, which has been reported to affect type 2 diabetes risk in a number of studies\(^{15-17}\). A recent study identified an association between a SNP in CDKAL1, rs4712526, and BMI in 8-year-old children\(^{18}\). rs4712526 was not included in stage 2, but stage 1 data for this SNP were consistent with the previous report (Supplementary Table 10). This SNP is in strong LD with rs9356744 ($r^2 = 0.87$) in Asians. Excluding participants with type 2 diabetes resulted in a similar association (P = 4.01 × 10^{-8}) (Supplementary Table 6), indicating that the association of rs9356744 with BMI was not driven by inclusion of individuals with diabetes in the study. Additionally, two SNPs in the CDKAL1 gene (rs9356744 and rs9368222; Supplementary Table 9) are cis expression quantitative trait loci (eQTLs) for the nearby E2F3 gene, a transcription factor and tumor suppressor\(^{19}\). In an accompanying paper, Okada et al.\(^{20}\) identified another SNP (rs2206734) in the
CDKAL1 gene that is associated with BMI. rs9356744 and rs2206734 are in strong LD in Asians ($r^2 = 0.932$) and in weaker LD in Europeans ($r^2 = 0.396$). In data from the GIANT consortium, rs9356744 was not associated with BMI ($P = 0.186$; Supplementary Table 7), but rs2206734 was ($P = 0.0049$). These findings suggest that the functional SNP encoding risk for obesity is in LD with both rs9356744 and rs2206734 in east Asians but only with rs2206734 in populations of European ancestry.

At the chromosome 5 locus (5q15), the top SNP, rs261967, along with 13 other SNPs that are in strong LD ($r^2 = 1.0$) with it, reached genome-wise significance in the combined stage 1 and 2 data (Supplementary Table 9). The nearest gene to this locus is PCSK1 (81.3 kb upstream). A candidate-gene study reported two common nonsynonymous coding variants (rs62334 and rs62335) in the PCSK1 gene that were associated with obesity. However, these two SNPs showed no association with BMI in our study (Supplementary Table 10), nor were they in LD with the 14 SNPs we identified ($r^2 = 0$) according to HapMap Asian data. rs261967 showed an association with BMI ($P = 0.00158$; Supplementary Table 7) in the data provided by the GIANT consortium.

At the chromosome 16 locus (16p12.3), the identified SNP, rs12597579, is near the GPR139 and GP2 genes. Multiple SNPs in this region showed an association with BMI in stage 1 that nearly met the significance threshold that was required for them to be evaluated in stage 2 (Fig. 3d). One of these SNPs, rs12598578 ($P = 1.63 \times 10^{-4}$; Supplementary Table 10), which is in LD ($r^2 = 0.968$ in Asians) with the identified rs12597579 SNP, is highly conserved across species according to the TRANSFAC database, and the common G allele creates a Ying-Yang transcription factor binding site.

The top SNP at the chromosome 11 locus (11p13), rs652722, is approximately 66.0 kb from the nearest gene, PAx6. This SNP is in LD with several SNPs that, according to the SCAN database, are predicted to be eQTLs for a number of genes potentially important in the regulation of body weight. Among the potentially regulated genes is MIF, according to HapMap lymphoblastoid cell lines. According to data from these cell lines, two SNPs (rs621611 and rs679887) in LD with rs652722 are significantly associated with MIF expression in SCAN. High plasma levels of the MIF protein are related to higher BMI. Another gene whose expression is associated with rs652722 is PFKP, which, along with FTO, has been associated with increased BMI, hip circumference and weight.

We identified multiple signals at the ADCY3-DNAJC27 locus (Supplementary Table 4). The rs11676272 SNP ($P = 5.88 \times 10^{-10}$) encodes a predicted missense mutation in the ADCY3 gene that causes a p.Ser107Pro alteration in the protein, and this change is predicted to be potentially deleterious according to PolyPhen. This SNP is also associated with expression of the POMC gene, which regulates energy balance. In addition, rs11676272 and rs6545814 at this locus ($r^2 = 0.98$ for LD between the two SNPs in Asians) are both eQTLs for the ADCY3 gene.

In conclusion, our study identified ten loci that are associated with BMI at the genome-wide significance level ($P < 5.0 \times 10^{-8}$), including seven loci previously identified in populations of European ancestry (FTO, SEC16B, MC4R, GIPR–QPCRTL, ADCY3–DNAJC27, BDNF and MAP2K3) and three newly identified loci in or near the CDKAL1, PCSK1 and GP2 genes. Three additional loci nearly reached genome-wide significance, including two previously identified SNPs in the GNPD2A and TTPAP2B genes and a newly identified marker near PAx6, all having $P < 5.0 \times 10^{-7}$. Of the three previously reported loci at GIPR–QPCRTL, ADCY3–DNAJC27 and MAP2K5, conditional analyses showed that only the SNPs identified by our study were associated with BMI in east Asian populations. The representative SNP (rs261967) near the newly identified association with PCSK1 showed a significant association ($P = 0.00158$) with BMI in a European population. As expected, the explained variance of the previously reported loci was generally lower in east Asians than in Europeans, whereas the explained variance for the newly identified loci from this study was generally higher in east Asians than in Europeans. The identification of new loci may shed light on new pathways involved in obesity. Future fine mapping of mixed-ancestry populations could lead to the identification of causal links.

**METHODS**

Methods and any associated references are available in the online version of the paper at http://www.nature.com/naturegenetics.

**ACKNOWLEDGMENTS**

The Shanghai Genome Wide Associations Studies (SGWAS) would like to thank the dedicated investigators and staff members from the research teams at Vanderbilt University, the Shanghai Cancer Institute and the Shanghai Institute of Preventive Medicine and, most of all, the study participants for their contributions to this work. Genotyping assays and statistical analyses for the SGWAS were primarily supported by grants from the US National Institutes of Health (NIH; R01 CA064277, R37 CA070867, R01 CA098989, R01 CA118229, R01 CA092585 and R01 CA122756), as well as by Ingram professorship funds, Allen Foundation funds and a Vanderbilt Clinical and Translational Science Award (CTSA; UL1 RR024975) from the National Center for Research Resources (NCRR) at the NIH. NIH grants provided support for the participating studies, including the Shanghai Breast Cancer Study (R01 CA064277), the Shanghai Breast Cancer Survival Study (R01 CA118229) and the Shanghai Endometrial Cancer Study (R01 CA092585). The KARE project was supported by grants from the Korea Centers for Disease Control and Prevention (4845-301, 4851-302 and 4851-307). The Singapore Prospective Study Program (SP2) was funded through grants from the Biomedical Research Council of Singapore (BMRC, 05/136/19/413 and 03/127/18/216) and the National Medical Research Council of Singapore (NOMR, NMRC/117/4/2008). E.S.T. also received support from the NMRC through a clinician scientist award (NMRC/CSA/008/2009). The Singapore Malay Eye Study (SIMES) was funded by the NMRC (NMRC/0796/2003 and NMRC/StAR/0003/2008) and the BMRC (09/1/35/19/616). The CAGE Network Studies were supported by grants for the Core Research for Evolutional Science and Technology (CREST) from the Japan Science Technology Agency, the Program for Promotion of Fundamental Studies in Health Sciences, the National Institute of Biomedical Innovation Organization (NIBIO) and the National Center for Global Health and Medicine (NCGM). L.Q. is supported by a grant from the NIH (HL071981), an American Heart Association Scientist Development Award and the Boston Obesity Nutrition Research Center (DK46200). The Genetic Epidemiology Network of Salt Sensitivity (GenSalt) is supported by research grants from the National Heart, Lung, and Blood Institute at the NIH (HL072307, HL087263 and HL090682). SINDi was funded by grants from the BMRC (09/135/19/616 and 08/135/19/550) and the NMRC (NMRC/StAR/0003/2008). SCORM was funded by the NMRC (0975/2005), the BMRC (06/121/19/466) and the Centre for Molecular Epidemiology at the National University of Singapore. The SH was supported by the Chinese National Key Program for Basic Research (973:2004CB518603) and the Chinese National High Tech Program (863:2009AA022703). The MEC was supported by grants from the National Cancer Institute (NCI; CA063464, CA054281 and CA132839) and from the NIH Genes, Environment and Health Initiative (GEI; HG004726). Assistance with genotype cleaning for the MEC Japanese prostate cancer study was provided by the GeGene Environment Association Studies (GENEVA) Coordinating Center (HG004446). Assistance with data cleaning was provided by the National Center for BioTechnology Information. Funding support for genotyping, which was performed at the Broad Institute of MIT and Harvard University, was provided by the GEI (HG004424).
AUTHOR CONTRIBUTIONS

T.A., Y.-S.C., Y.-T.G., D.-G., B.-G.H., J.H., F.B.H., N. Kamatani, N. Kato, L.-L.-M., J.-Y.L., W.L., Z.M., Y.N., D.P.-K.N., L.Q., S.-M.S., X.-O.S., E.-S.T., F.-J.T., T. Tanaka, E.J.T., Y.-T.W., J.-Y.W., B.-X.J., J.X., W.Z. and D.Z. supervised the experiments. Y.-S.C., D.G., J.H., Y.H., N. Kato, J. Liang, Z.M., Y.N., L.Q., M.S., X.-O.S., H.S., E.S.T., T. Tanaka, T.-Y.W., Z.Y. and D.Z. conceived and designed the experiments. J.H., Y.H., M.K., J. Liang, M.S., J.S., M.Y. and Y.Z. performed the experiments. L.-C.C., C.-H.C., G.K.C., R.D., M.-J.G., H.Y., M.L., C.-J., L. YO, L.Q., M.-H.S., Y.T., A.T., T. Tsunoda and W.W. performed the statistical analyses. The GIANT Consortium: C.-H.C., R.D., R.O., D.C.H., H.S., M.H., Y.N., L.J. Long, T.M.Y., R.O.H., L.Q., X.S., M.S. and Y.T. analyzed the data. T.A., Q.C., Y.-T.G., C.A.H., B.E.H., N.I., N. Kato, Y.K., L.L.-M., J. Liang, J.-J.L., W.L., D.P.-K.N., L.Q., S.-M.S., M.S., X.-O.S., H.S., E.S.T., F.-J.T., T.-Y.W., J.-Y.W., B.-X.Y., K.Y., M.Y., C.S.J.F. and W.Z. contributed reagents, materials and/or analysis tools. R.J.D., Y.O., X.-O.S., E.S.T., T. Tanaka, W.W. and W.Z. wrote the manuscript. S.M. reviewed the manuscript for important intellectual content. All authors reviewed and approved the final version of the manuscript.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Published online at http://www.nature.com/naturegenetics/

Reprints and permissions information is available online at http://www.nature.com/reprints/index.html.

1. Frayling, T.M. et al. A common variant in the FTO gene is associated with body mass index and predisposes to childhood and adult obesity. Science 316, 889–894 (2007).
2. Scuteri, A. et al. Genome-wide association scan shows genetic variants in the FTO gene are associated with obesity-related traits. PLoS Genet. 3, e115 (2007).
3. Loos, R.J.F. et al. Common variants near MC4R are associated with fat mass, weight change, and weight gain. Nat. Genet. 39, 770–775 (2007).
4. Thorleifsson, G. et al. Common variants in the MC4R gene are associated with obesity-related traits. Nat. Genet. 40, 1–7 (2008).
5. Willer, C.J. et al. A genome-wide association study of body mass index in European populations. Nat. Genet. 44, 511–517 (2012).
6. Meyre, D. et al. Genome-wide association study for early-onset and morbid adult obesity identifies three new risk loci in European populations. Nat. Genet. 41, 157–159 (2009).
7. Scherag, A. et al. Two new loci for body-weight regulation identified in a joint analysis of genome-wide association studies for early-onset extreme obesity in French and German body groups. PLoS Genet. 6, e1000916 (2010).
8. Spielhofer, 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).
9. Chambers, J.C. et al. Common genetic variation near MC4R is associated with waist circumference and insulin resistance. Nat. Genet. 40, 716–718 (2008).
10. McCarthy, M.I. Genomics, type 2 diabetes, and obesity. N. Engl. J. Med. 363, 2339–2350 (2010).
11. Cho, I.S. et al. A large-scale genome-wide association study of Asian populations uncovers genetic factors influencing eight quantitative traits. Nat. Genet. 41, 527–534 (2009).
12. Deurenberg, P., Deurenberg-Yap, M. & Guricci, S. Asians are different from Caucasians and from each other in their body mass index/body fat percentage relationships. Obes. Rev. 3, 141–146 (2002).
13. de Bakker, P.I. et al. Practical aspects of imputation-driven meta-analysis of genome-wide association studies. Hum. Mol. Genet. 17, R122–R128 (2008).
14. W.H.O. Expert Consultation. Appropriate body-mass index in populations and its implications for policy and intervention strategies. Lancet 363, 157–163 (2004).
15. Voight, B.F. et al. Twelve type 2 diabetes susceptibility loci identified through large-scale association analysis. Nat. Genet. 42, 579–589 (2010).
16. Zeggini, E. et al. Replication of genome-wide association signals in UK samples reveals risk loci for type 2 diabetes. Science 316, 1336–1341 (2007).
17. Steinhodttdorf, V. et al. A variant in CDKAL1 influences insuline response and risk of type 2 diabetes. Nat. Genet. 39, 770–775 (2007).
18. Winkler, C. et al. BMI at age 8 years is influenced by the type 2 diabetes gene HHEX in non-British European populations. Diabetes 59, 1063–1067 (2010).
19. Veyrieras, J.B. et al. High-resolution mapping of expression-QTLs yields insight into human gene regulation. PLoS Genet. 4, e1000214 (2008).
20. Okada, Y. et al. Common variants at CDKAL1 and KLF19 loci are associated with body mass index in Japanese populations. Nat. Genet. published online, doi:10.1038/nbg.1086 (12 February 2012).
21. Benzinou, M. et al. Common nonsynonymous variants in PCSK1 confer risk of obesity. Nat. Genet. 40, 943–948 (2008).
22. Willer, C.J. et al. TRANSACT, an integrated system for gene expression regulation. Nucleic Acids Res. 38, 316–319 (2010).
23. Gamanon, E.R. et al. SCAN: SNP and copy number annotation. Bioinformatics 26, 259–262 (2010).
24. McDonnell, J.P. et al. Increased plasma concentration of macrophage migration inhibitory factor (MIF) and MIF mRNA in mononuclear cells in the obese and the suppressive action of mifmetin. J. Clin. Endocrinol. Metab. 89, 5043–5047 (2004).
25. Yang, T.P. et al. Genevar: a database and Java application for the analysis of genome-wide association studies. Nat. Genet. 44, 579–589 (2010).
ONLINE METHODS

Study design. This study had three stages. Stage 1 was a meta-analysis of study-specific results on the association between SNPs and BMI from eight GWAS that participated in the consortium and included a total of 27,715 individuals of east Asian ancestry. Each participating study was approved by the local institutional review board, and informed consent was obtained from participants. Promising SNPs selected from the stage 1 meta-analysis were further examined by in silico (stage 2) and de novo (stage 3) replication analyses. Basic information for all participating studies is summarized in Supplementary Figure 1, Supplementary Tables 1 and 2 and the Supplementary Note.

Stage 1 samples and genotyping. The sample sizes of the eight GWAS in stage 1 varied between 821 and 8,838 participants, with a total of 27,715 individuals. For genotyping, two studies used Affymetrix arrays, and six studies used an Illumina platform (detailed information is provided in the Supplementary Note). To allow for the combination of the data derived from different genotyping platforms and to improve coverage of the genome, genotype imputation was performed by each participating study using either MACH or IMPUTE.

Stage 1 statistical analysis. A uniform statistical analysis protocol was followed by each participating study. To calculate BMI, each study collected weight and height measurements. To improve the normality of the BMI distribution and alleviate the impact of outliers, the rank-based inverse normal transformation (INT) was applied separately to BMI values for each gender in each study. INT involves ranking all BMI values, transforming these ranks into quantiles and converting the resulting quantiles into normal deviates. The association between SNPs and the inverse normal transformed BMI values was analyzed with a linear regression model. The association between SNPs and obesity as a dichotomous outcome, which defined obesity as BMI ≥ 27.5 (ref. 14), was analyzed with a logistic regression model assuming an underlying additive genetic mode and adjusting for age (continuous), age squared, gender (if applicable) and ancestry (if applicable). Stratified analyses by gender and disease status (for cancer or type 2 diabetes) were also performed by each study.

Next, we carried out meta-analyses using two methods in parallel to cross-check the results: one approach combined effects weighted by the inverse variance and the second combined P values weighted by the square root of the sample size for each study. Both meta-analysis procedures were implemented in the METAL software package (see URLs). The final P values obtained from these two methods were highly congruent (Pearson correlation r = 0.98). P values derived from the effect-size–based method are reported here, as this method is preferred, in general, to the P value–based method and also provides combined regression coefficients and their standard errors26. The meta-analyses were carried out with all data combined and were also stratified by gender and disease status. The presence of heterogeneity across studies and between genders was tested with Cochran’s Q statistic27.

To correct each study for residual population stratification or cryptic relatedness, the meta-analyses were performed with genomic control correction20 by adjusting for the study-specific inflation factor (λ), which ranged from 1.000 to 1.075 in stage 1 (Supplementary Table 3). After study-specific genomic control adjustment, the estimated inflation factor for the stage 1 meta-analysis statistic was 1.056, which was further adjusted when combining stage 1 results with stage 2 replication data.

On the basis of the stage 1 meta-analysis on the association between SNPs and BMI in all participants, we selected for stage 2 replication a total of 848 SNPs, which included 798 SNPs with P < 1.0 × 10−5 and 50 SNPs located in previously reported obesity-related loci that had P > 1.0 × 10−7. The cutoff of P < 1.0 × 10−5 was chosen so that the overall P value would reach genome-wide significance (P < 5.0 × 10−8) given the sample sizes of stages 1 and 2.

Stage 2 in silico replication. The 848 SNPs selected for replication were investigated in an independent set of 37,691 individuals of east Asian ancestry from seven additional GWAS. The sample sizes of the seven additional studies varied between 901 and 27,284 subjects. The RIKEN study was the main source of the replication data and included 27,284 individuals. One study (SCORM) included only children (9 year olds). All studies used the Illumina platform except for the GenSalt study, which used an Affymetrix array. Genotype imputation was also performed by each study using either MACH or IMPUTE, as for studies included in stage 1.

Each study individually conducted a similar analysis of the SNPs selected from stage 1, using the same protocol as in stage 1. The stage 2 data were combined using meta-analysis methods with study-specific genomic control adjustment in a manner similar to that performed in stage 1. Finally, we used meta-analysis to combine all data from stages 1 and 2, with further adjustment for the estimated inflation factor for the stage 1 meta-analysis statistic.

Stage 3 de novo replication. Seven SNPs that were associated with BMI according to the analysis of combined stage 1 and 2 data, including four SNPs at four newly identified loci and three loci that overlapped with loci that were reported by the GIANT consortium during the course of our study, were further validated in our stage 3 de novo replication studies. These analyses were conducted with data from three study sites (Supplementary Table 1), and the genotyping for the seven SNPs was carried out for a total of 17,642 east Asians. The results from stages 1, 2 and 3 were combined and analyzed using meta-analysis methods.

Quality control procedures. The following quality control procedures were recommended for each participating study. SNPs were excluded, either in the primary analysis conducted by each participating study or at the meta-analysis stage (Supplementary Table 3), if they (i) had a call rate of <90%, (ii) deviated from Hardy-Weinberg equilibrium with P < 1.0 × 10−5, (iii) had a MAF of <1%, (iv) had low imputation quality (for imputed SNPs; r-hat < 0.3 for MACH or proper-info < 0.5 for IMPUTE) or (v) were potentially contaminated. Samples from individuals were removed if they had a call rate of <90%, if they showed first-degree cryptic relationships in an identity–by-descent (IBD) analysis or if they were potentially contaminated. The specific quality control procedures adopted by each study are summarized in Supplementary Table 3.

Conditional analysis. To investigate the independent association of SNPs in the same locus, conditional analyses were conducted by including both SNPs at the same locus in the same regression model for mutual adjustment. The normal transformation of BMI values and the adjustment of covariates were applied in the same manner as in the stage 1 analysis. These conditional analyses were conducted among 57,931 (88.6%) subjects from 11 of the 15 studies in stages 1 and 2.

Estimation of the explained variance. The variation in BMI explained by an individual SNP was estimated by 2βj(1 − f) (ref. 29), where f is the frequency of the variant and β is its additive effect estimated from the stage 2 studies. We subsequently estimated the overall fraction of variance that can be explained by all significantly associated SNPs found in the current meta-analyses by calculating the genetic score with

\[
\text{Score}_i = \sum_{j=1}^{m} \beta_j A_j
\]

where m is the number of SNPs, βj is the effect of an allele at locus j, estimated from the stage 2 data, and Ai is the number of reference alleles of individual i at locus j. The measure of variance explained (adjusted R2) was estimated from a linear regression model incorporating the score as the predictor and the covariate-adjusted inverse normal transformed BMI residuals as the outcome. We reported the average explained variance weighted by the sample size of each study. These analyses to estimate the explained variance were conducted among 57,931 (88.6%) subjects from 11 of the 15 studies in stages 1 and 2.

Analyses of coding SNPs, eQTLs and copy-number variant (CNVs). Variants with potential functional impact were evaluated using the SNP Function Prediction tool, which is part of the SNPinfo Web Server30. This tool identifies SNPs with potential functional consequences, including those resulting in coding changes. The severity of a coding change was evaluated with PolyPhen, although SNPs reported to result in coding changes of consequence were independently verified using the PolyPhen website31 (see URLs).

eQTLs were evaluated using the SCAN database23 and the GeneVar program25, with RNA sequencing and genotyping experiments conducted as described previously32.

To test for CNVs in the proximity of the specific variants of interest, we used the UCSC Genome Browser33. Although direct access to individual-level SNP
genotype information was unavailable to directly test for the presence of CNVs, we used the phased genotypes from HapMap 2 release 22 data\textsuperscript{24}, which show estimated LD values for the CHB and JPT HapMap populations (hapmapLdPhChbJpt table), in combination with the UCSC Genome Browser track from the Database of Genomic Variants (dgv table)\textsuperscript{35} to identify the presence of known CNVs that intersected regions in strong LD with GWAS variants.

26. Zeggini, E. & Ioannidis, J.P. Meta-analysis in genome-wide association studies. \textit{Pharmacogenomics} \textbf{10}, 191–201 (2009).
27. Cochran, W.G. The combination of estimates from different experiments. \textit{Biometrics} \textbf{10}, 101–129 (1954).
28. Devlin, B. & Roeder, K. Genomic control for association studies. \textit{Biometrics} \textbf{55}, 997–1004 (1999).
29. Park, J.H. \textit{et al.} Estimation of effect size distribution from genome-wide association studies and implications for future discoveries. \textit{Nat. Genet.} \textbf{42}, 570–575 (2010).
30. Xu, Z. \& Taylor, J.A. SNPinf: integrating GWAS and candidate gene information into functional SNP selection for genetic association studies. \textit{Nucleic Acids Res.} \textbf{37}, W600–W605 (2009).
31. Ramensky, V., Bork, P. \& Sunyaev, S. Human non-synonymous SNPs: server and survey. \textit{Nucleic Acids Res.} \textbf{30}, 3894–3900 (2002).
32. Montgomery, S.B. \textit{et al.} Transcriptome genetics using second generation sequencing in a Caucasian population. \textit{Nature} \textbf{464}, 773–777 (2010).
33. Fujita, P.A. \textit{et al.} The UCSC Genome Browser database: update 2011. \textit{Nucleic Acids Res.} \textbf{39}, D876–D882 (2011).
34. Frazer, K.A. \textit{et al.} A second generation human haplotype map of over 3.1 million SNPs. \textit{Nature} \textbf{449}, 851–861 (2007).
35. Feuk, L., Carson, A.R. \& Scherer, S.W. Structural variation in the human genome. \textit{Nat. Rev. Genet.} \textbf{7}, 85–97 (2006).