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

秋山, 雅人

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

Supplementary Information

Supplementary Figures:

Supplementary Fig. 1 Overview of the study (p. 2)

Supplementary Fig. 2 QQ plot, LD score plot and Manhattan plot of GWAS (p. 3)

Supplementary Fig. 3 Results of sex-stratified GWAS (p. 5)

Supplementary Fig. 4 Comparison of variants between male and female (p. 7)

Supplementary Fig. 5 Regional association plots of secondary signals (p. 8)

Supplementary Fig. 6 Different patterns of signals between Japanese and Europeans (p. 10)

Supplementary Fig. 7 Manhattan plots of trans-ethnic meta-analysis (p. 12)

Supplementary Fig. 8 Comparison of variants between European and Asian (p. 13)

Supplementary Fig. 9 Regional association plots for sex-stratified trans-ethnic meta-analysis (p. 14)

Supplementary Fig. 10 Phenotypic variance explained by each chromosome (p. 16)

Supplementary Fig. 11 Phenotypic variance explained by subsets of variants based on *P* value (p. 17)

Supplementary Fig. 12 Comparison of the variants between T2D and others (p. 18)

Supplementary Fig. 13 Forest plots of the effect sizes for BMI in T2D and in non-diabetics (p. 19)

Supplementary Fig. 14 Comparison of T2D associated variants between GWAS and replication sets (p. 20)

Supplementary Note:

1. Descriptions of the participating cohorts (p. 21)

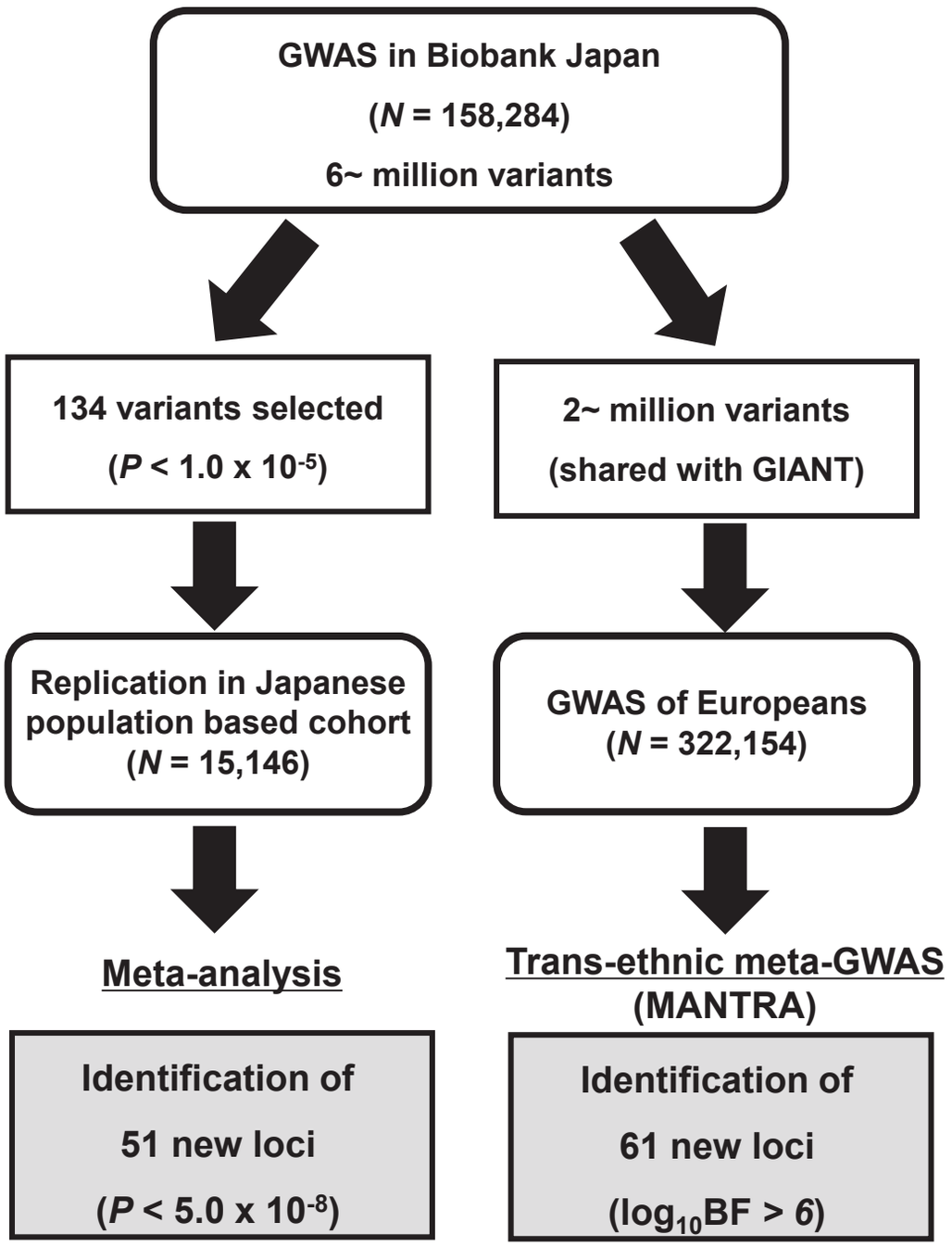
2. Characteristics of samples analyzed in GWAS and Japanese population-based cohorts (p. 23)

3. Evaluation of enhancer overlaps in each cell-type (p. 24)

4. GWASs used for cross trait LD score regression analysis (p. 26)

Reference (p. 29)

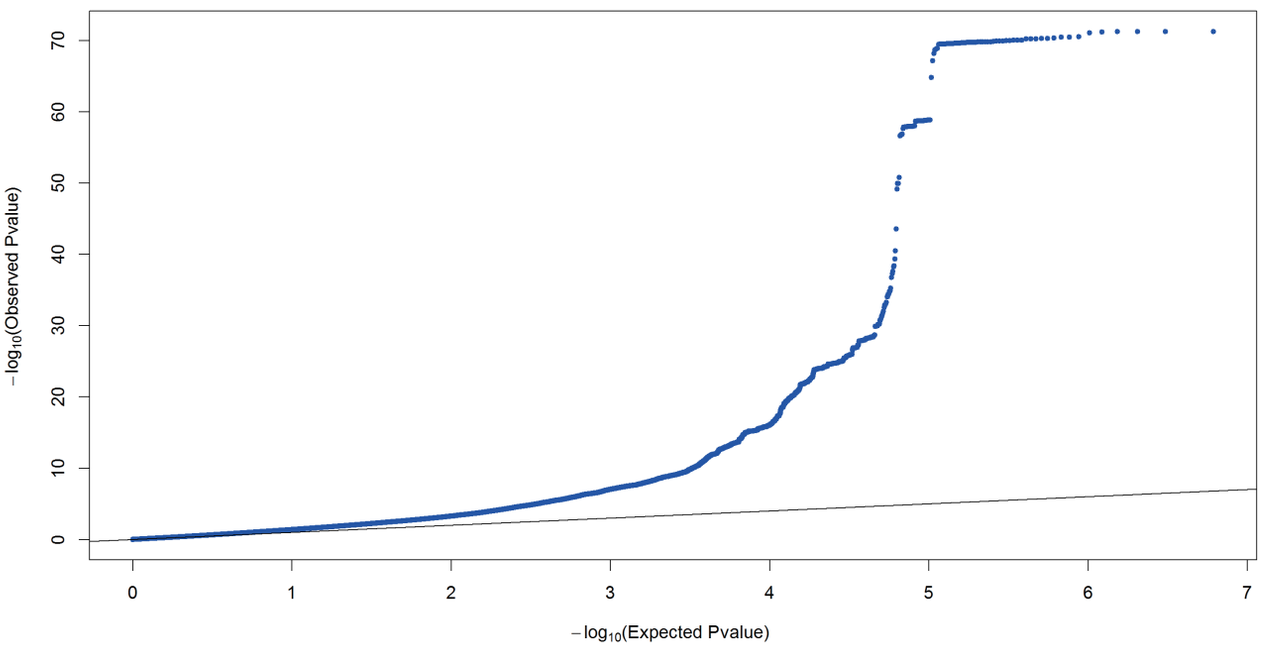
Nature Genetics: doi:10.1038/ng.3951



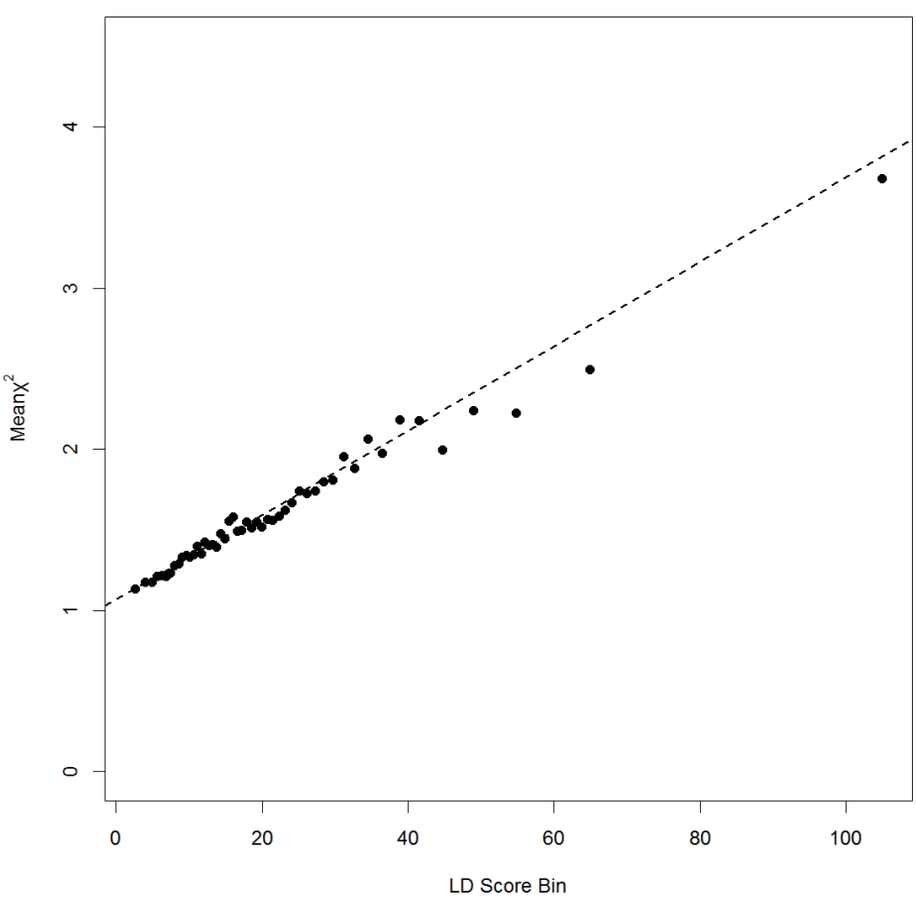
Supplementary Fig. 1 Overview of the study

Nature Genetics: doi:10.1038/ng.3951

a) QQ plot

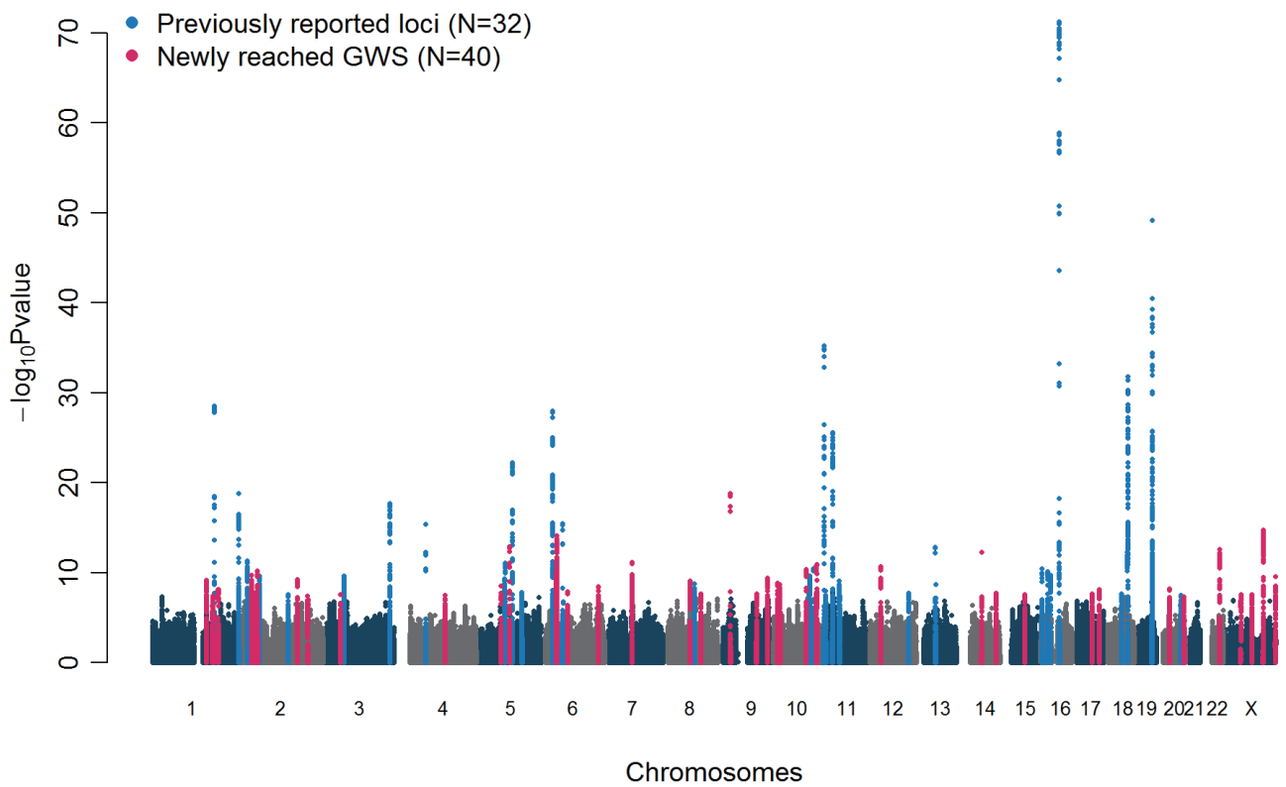


b) LD score plot



Nature Genetics: doi:10.1038/ng.3951

c) Manhattan plot



Supplementary Fig. 2

QQ plot, LD score plot and Manhattan plot of GWAS

a) QQ plot of GWAS (lambda GC = 1.44).

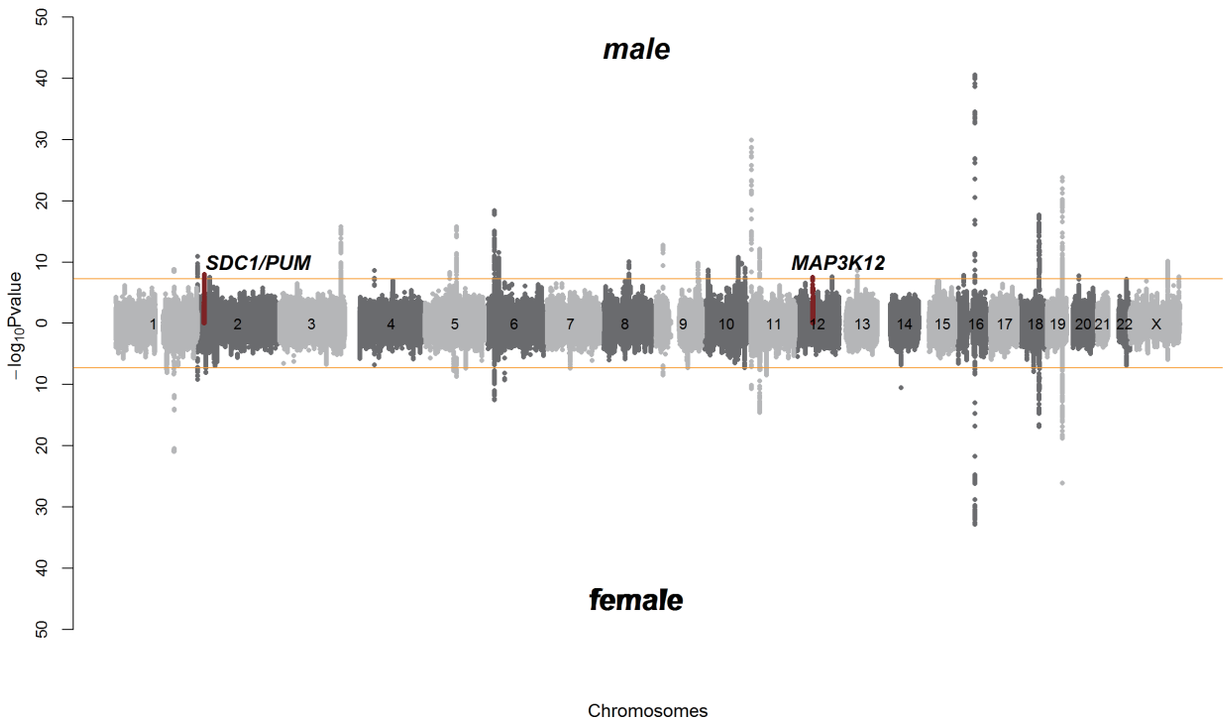
b) We generated LD score plot of GWAS. We plotted mean χ^2 statistics for each LD score quantile. The dashed line denotes the LD score regression line (intercept = 1.07, mean χ^2 = 1.63, and the resulted h^2 = 0.17).

c) Manhattan plot of GWAS. We plotted loci newly reached genome-wide significance (lead variants ± 1Mb) in pink. Previously reported loci were colored blue.

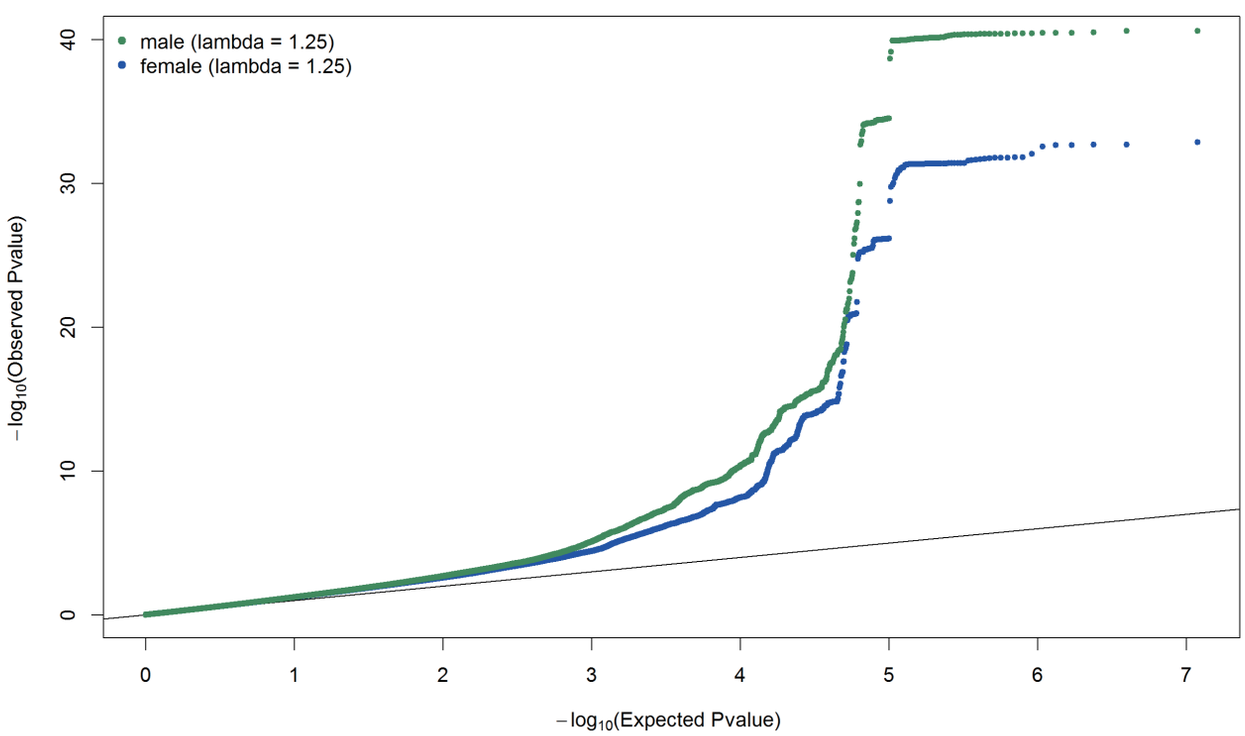
GWS; genome-wide significance.

Nature Genetics: doi:10.1038/ng.3951

a) Manhattan plot of sex-stratified GWAS

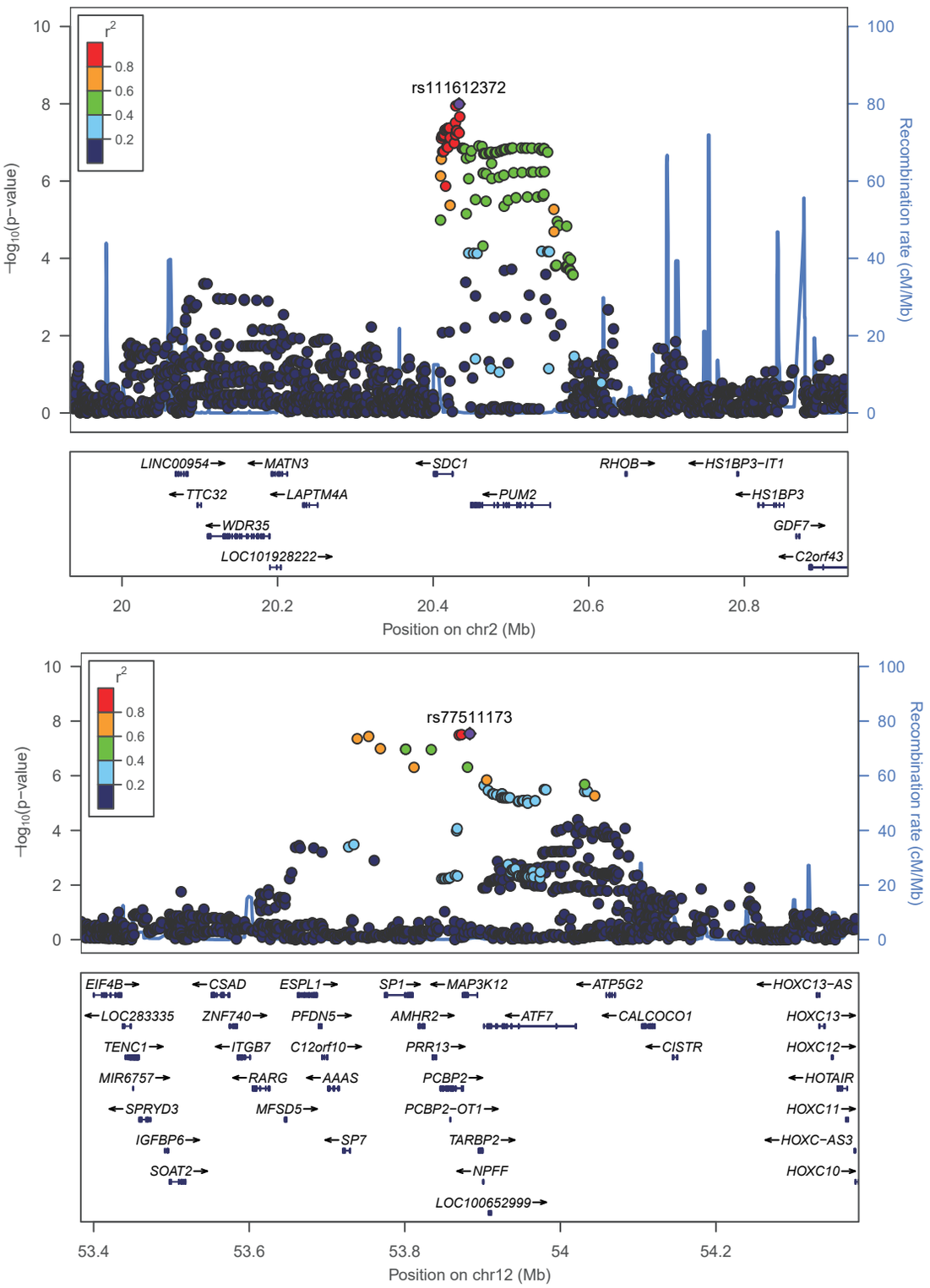


b) QQ plot for each sex



Nature Genetics: doi:10.1038/ng.3951

c) Regional association plot

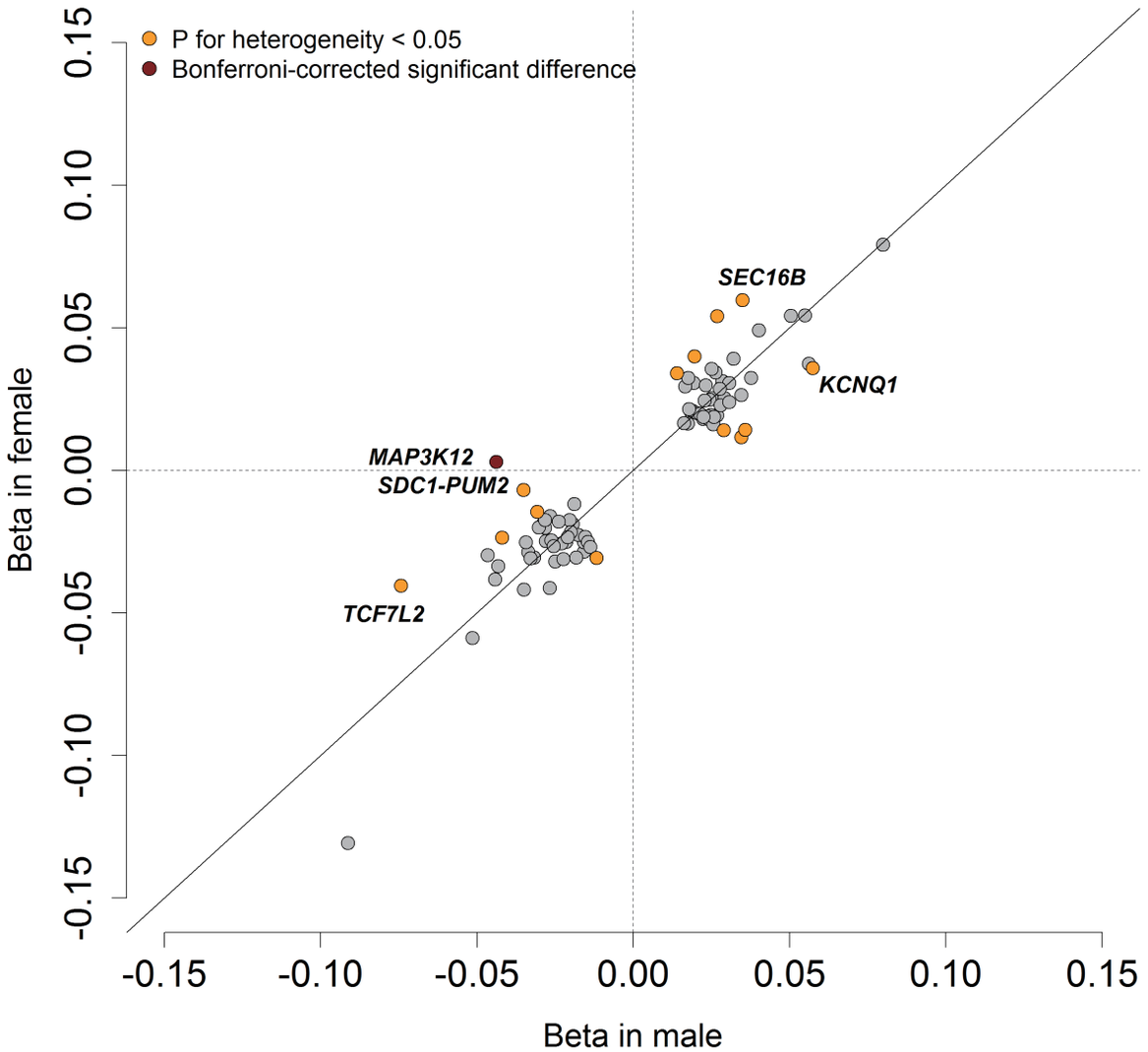


Supplementary Fig. 3

Results of sex-stratified GWAS

a) Manhattan plot of the sex-stratified GWAS. Associations ($-\log_{10}P$) were plotted positive y- and negative y-axes for male and female, respectively. b) QQ plots of sex-stratified GWAS are colored green for male and blue for female. c) Regional association plots of significant loci in men were shown. Colors of plots indicate linkage disequilibrium measure r^2 with lead SNPs.

Nature Genetics: doi:10.1038/ng.3951



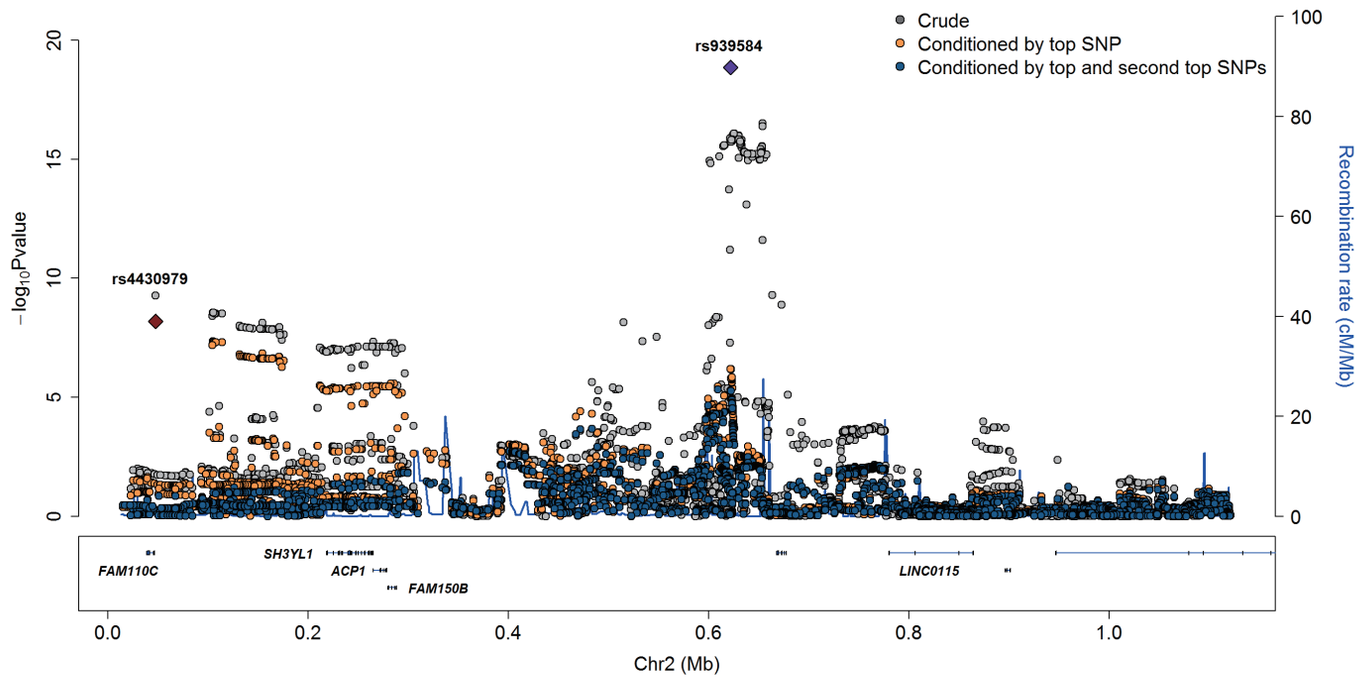
Supplementary Fig. 4

Comparison of variants between male and female

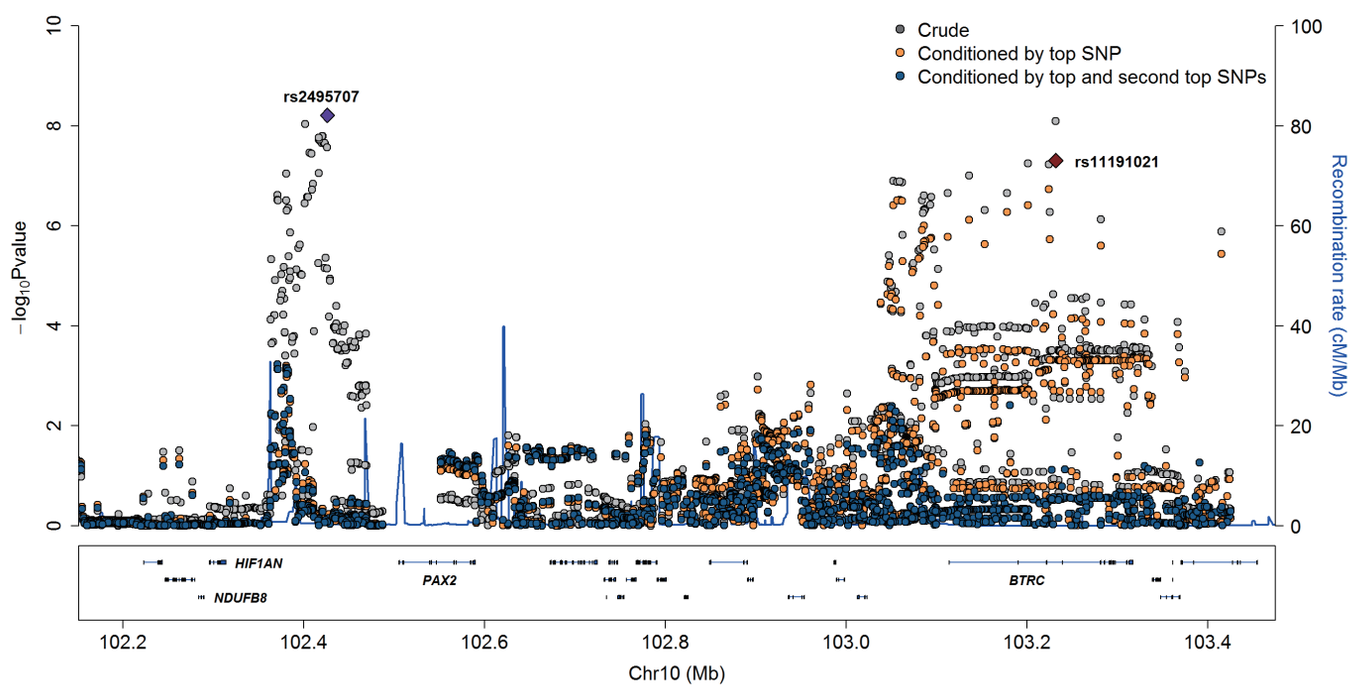
Beta coefficients of significantly associated variants after stratified by sex. The variants which showed nominal difference between sexes were plotted in yellow. Variants showed Bonferroni-corrected level of significant difference (MAP3K12; α= 0.05/85) were plotted in red. Solid line denotes y = x.

Nature Genetics: doi:10.1038/ng.3951

a)

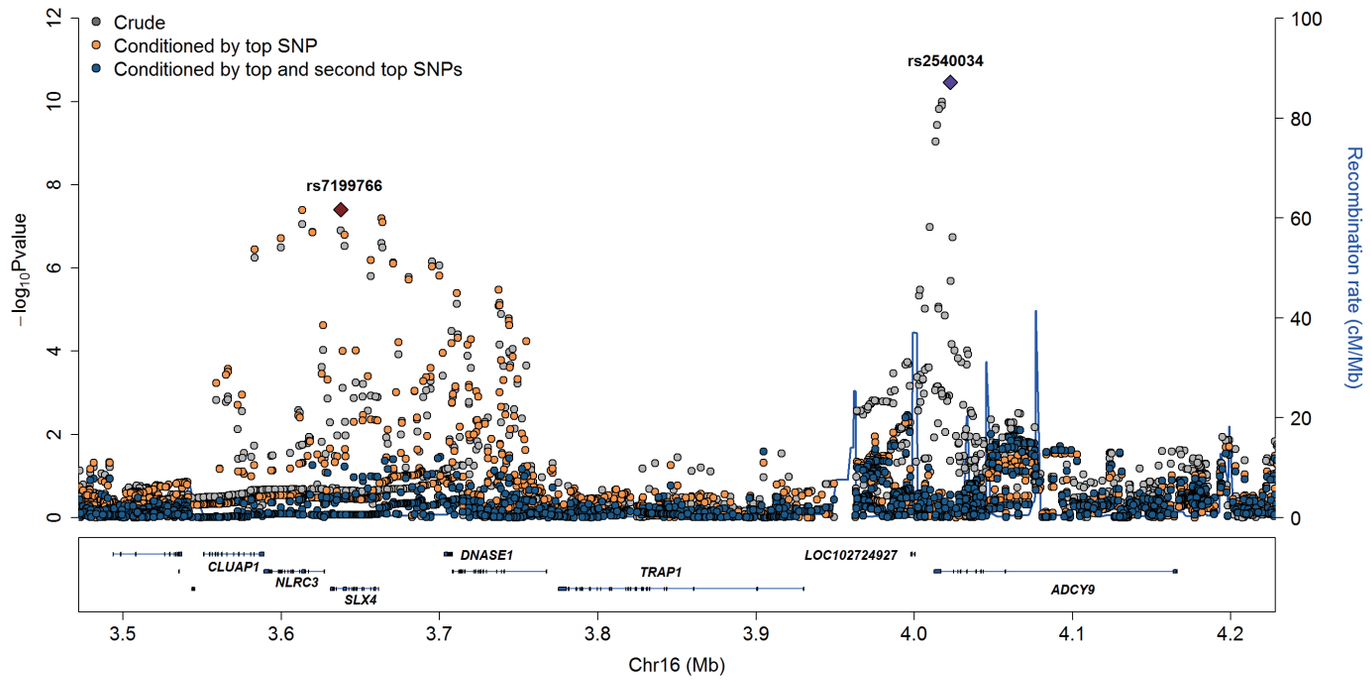


b)

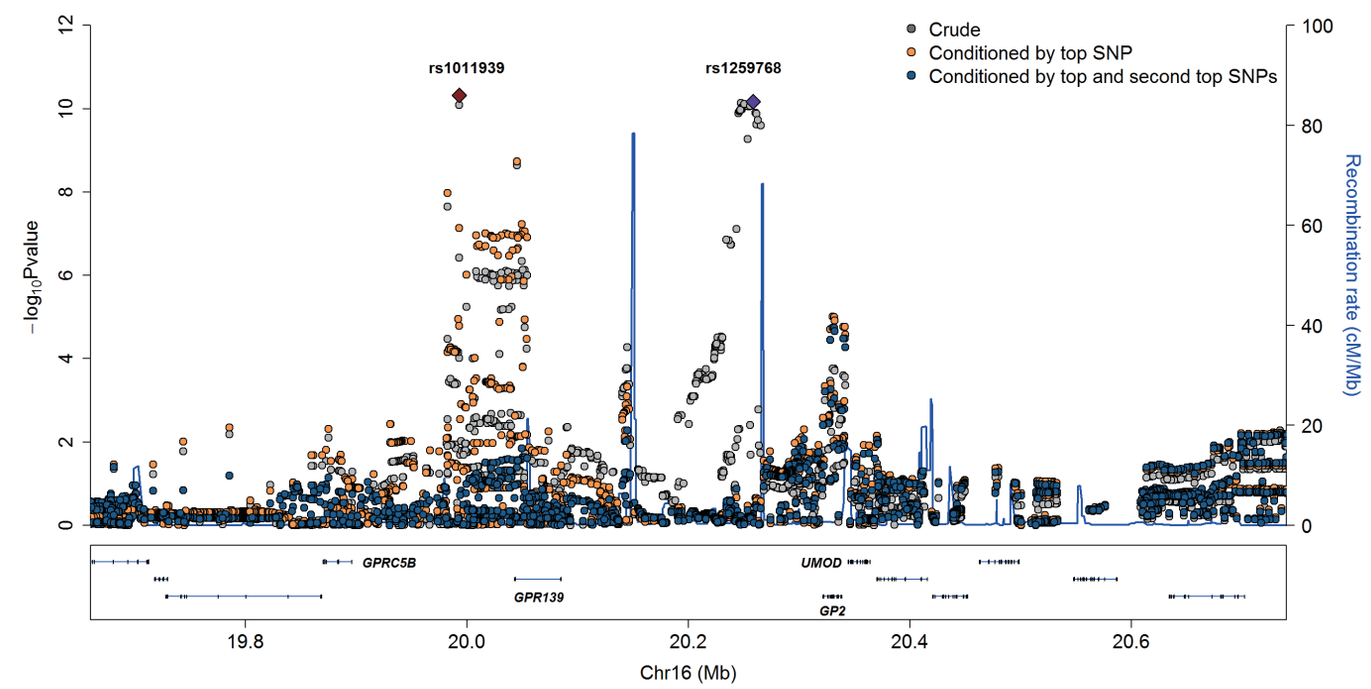


Nature Genetics: doi:10.1038/ng.3951

c)



d)

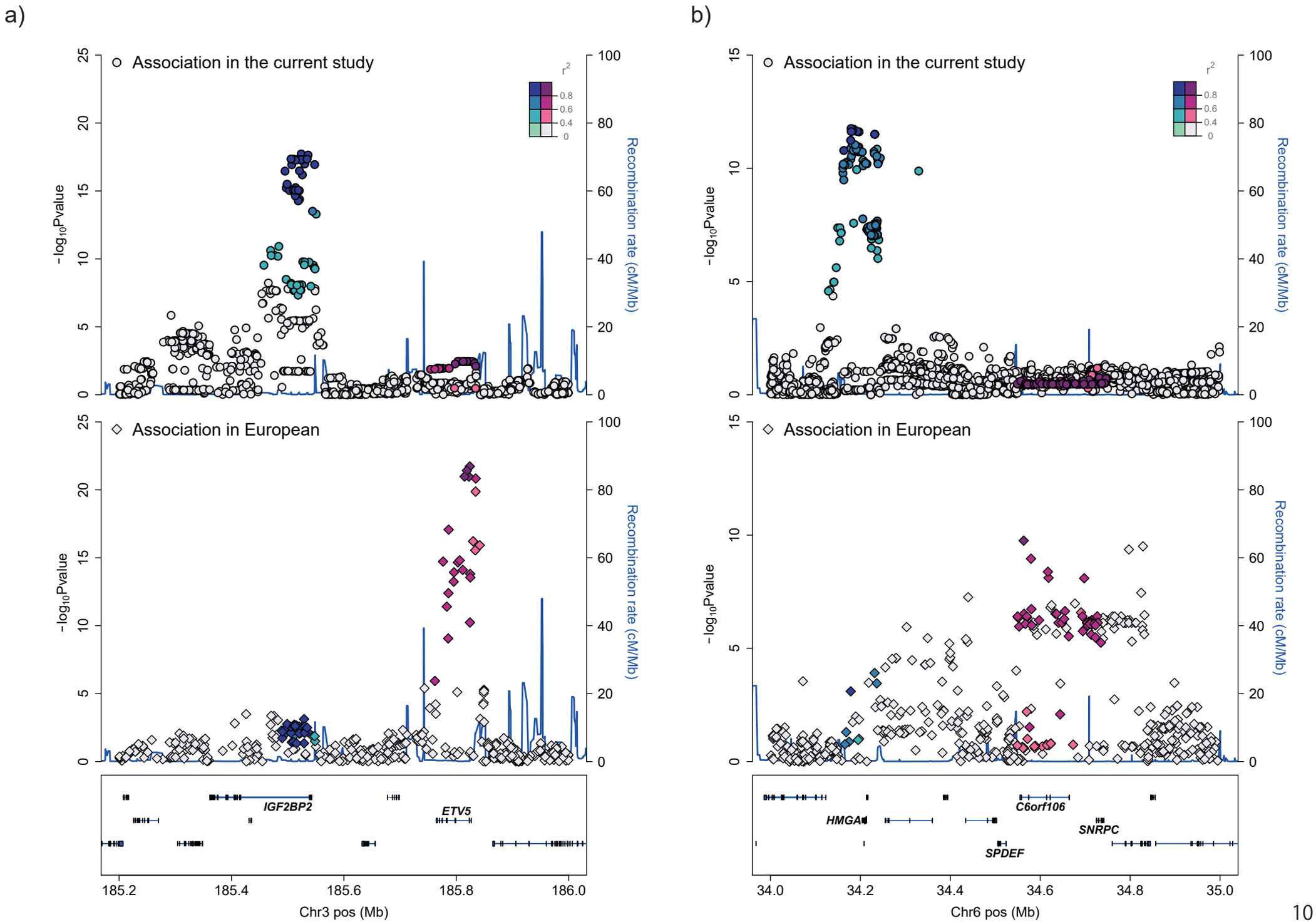


Supplementary Fig. 5

Regional association plots of secondary signals

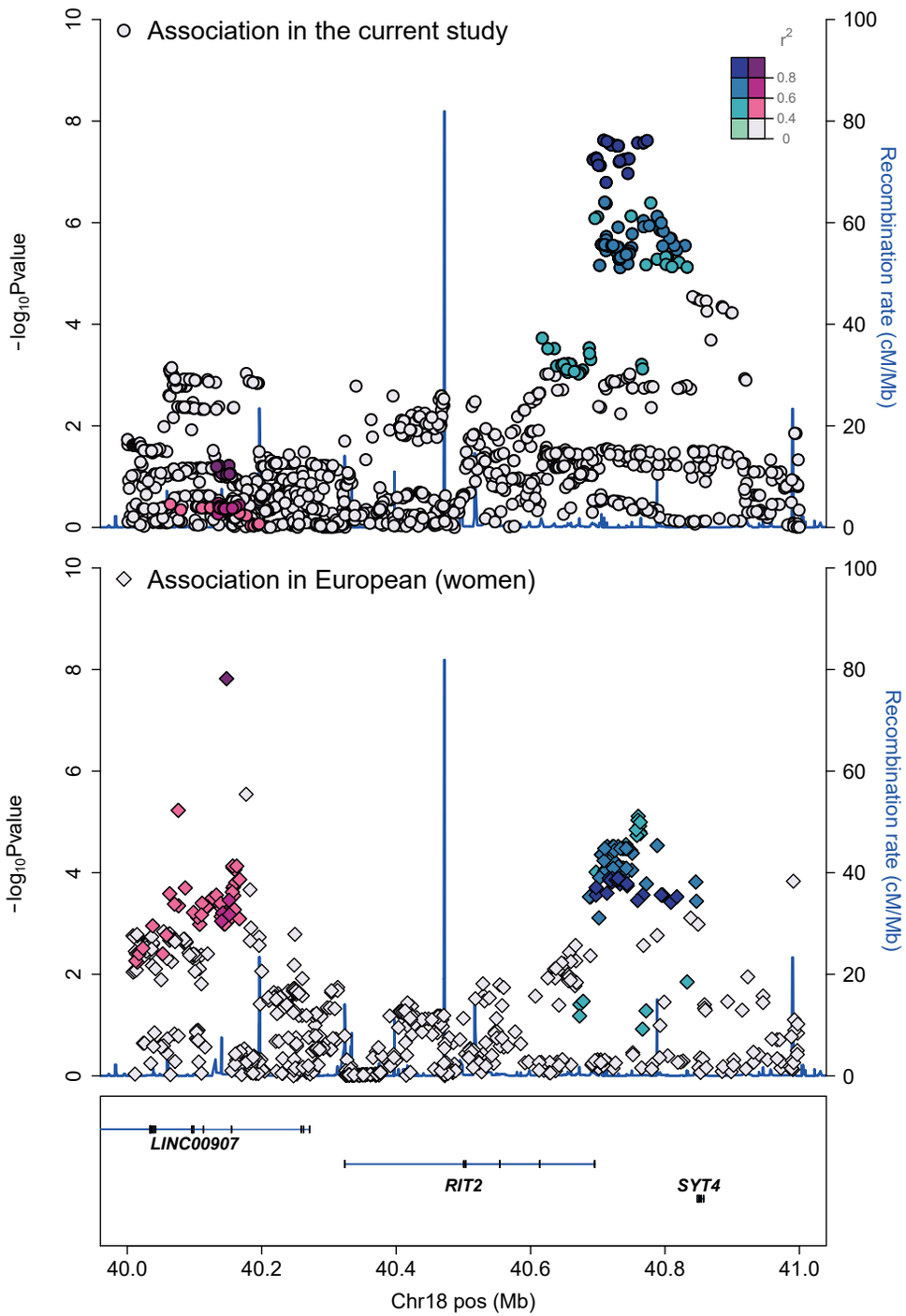
a) – d) Regional association plots of conditional analysis. We plotted crude P value (colored gray), P values of conditioned by top SNPs of each region (colored yellow) and conditioned by top and secondary associated SNP (colored blue). Red diamonds indicate top SNPs in each region, and purple diamonds denote second hits after conditioned by top hits.

Nature Genetics: doi:10.1038/ng.3951



Nature Genetics: doi:10.1038/ng.3951

c)



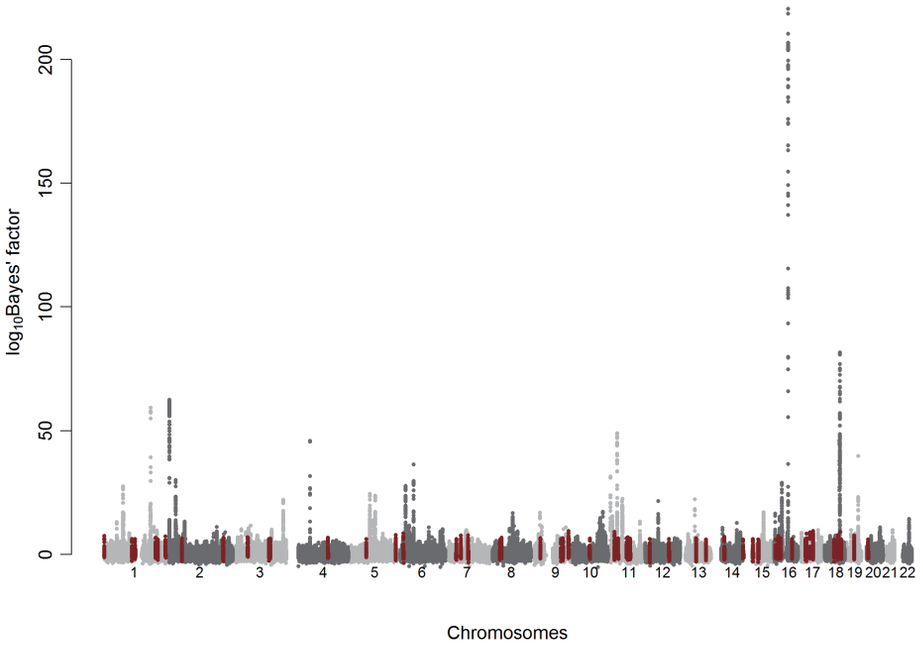
Supplementary Fig. 6

Distinct signals between Japanese and Europeans

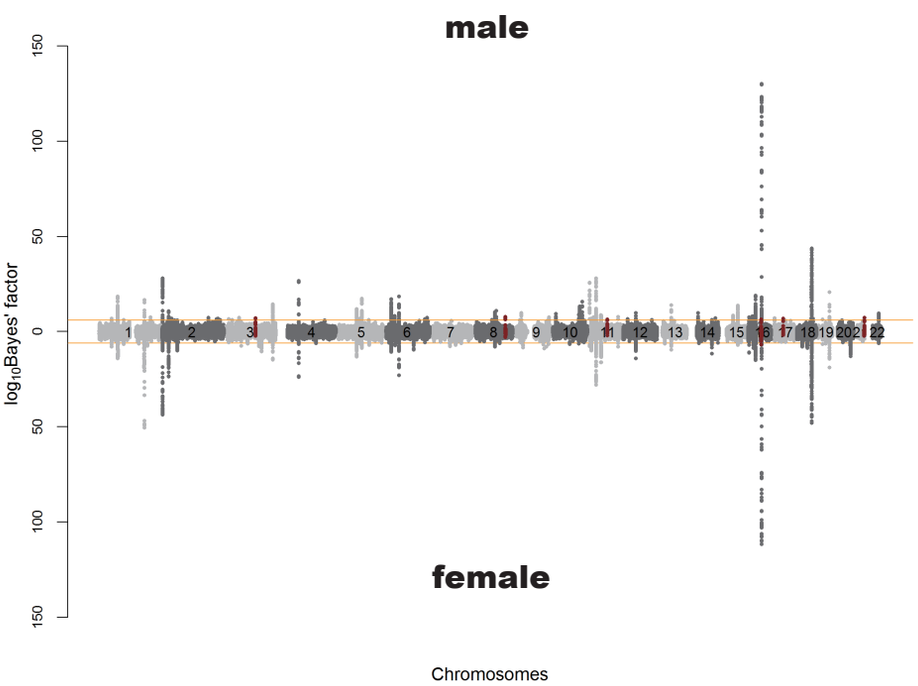
a) – c) Regional association plots which showed different patterns of association signals between current study and European meta-GWAS conducted by GIANT consortium. We downloaded GWAS summary statistics of European GWAS from their download site. Chromosomal positions of SNPs were converted from hg18 to hg19 using lift over tool implemented in UCSC genome browser. Association results in GIANT GWAS were plotted as diamonds. Association results in the current GWAS were plotted as circles. Each plot were color coded according to the pairwise LD (r^2) with the lead variants of each study. In RIT2 region (c), we plotted the association results of female which was reported as significant region in the original report.

Nature Genetics: doi:10.1038/ng.3951

a) Manhattan plot of meta-GWAS without stratification by sex



b) Manhattan plot of meta-GWASs with sex-stratification



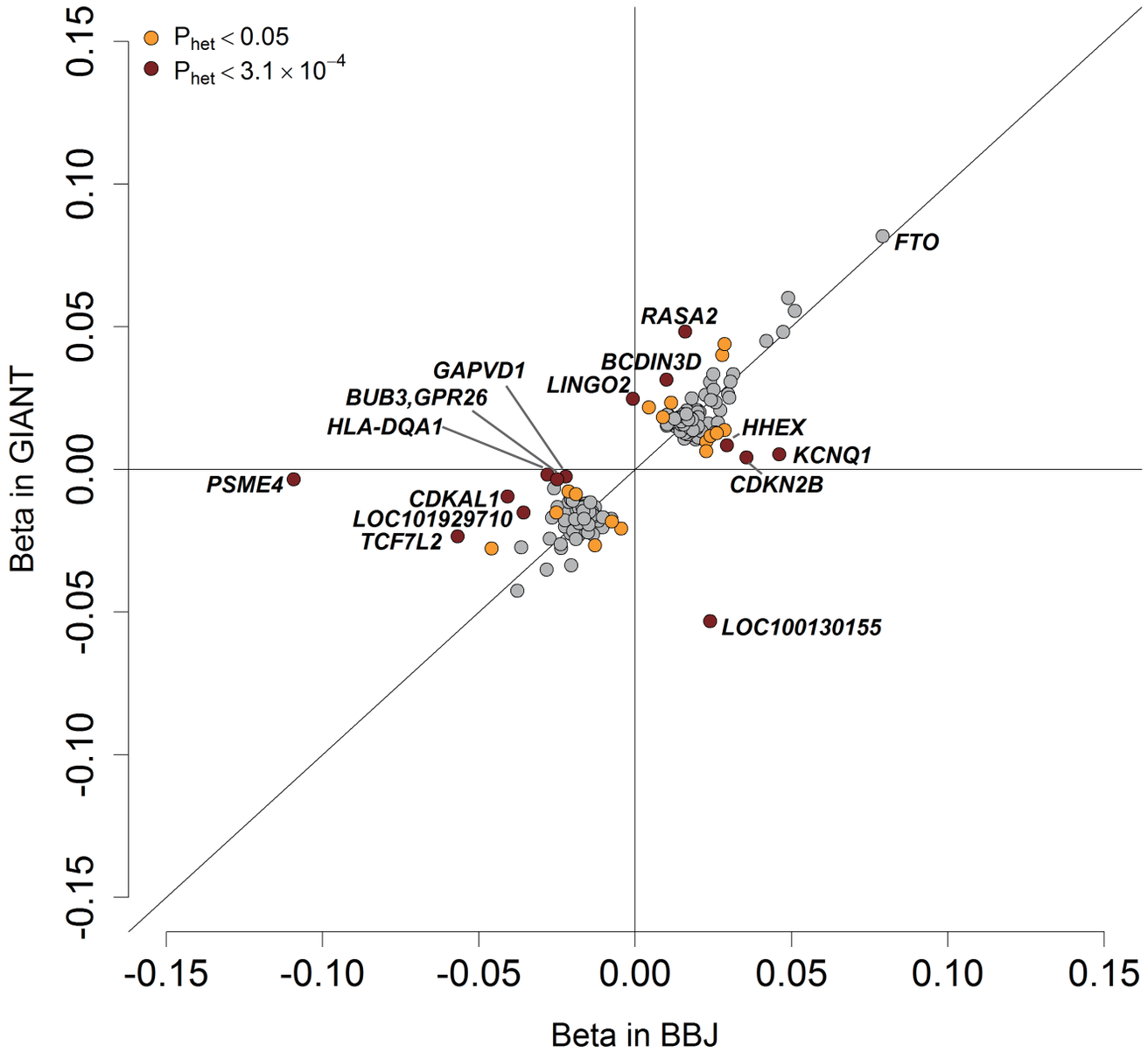
Supplementary Fig. 7

Manhattan plots of trans-ethnic meta-analysis

We plotted the result of trans-ethnic meta-analysis of the GWASs using current Japanese and publicly available results of Europeans.

Manhattan plots of the GWAS wich was not stratified by sex (a) and sex-stratified were shown. Y axis denotes $\log_{10}$Bayes' s factor. Newly identified loci were colored red.

Nature Genetics: doi:10.1038/ng.3951



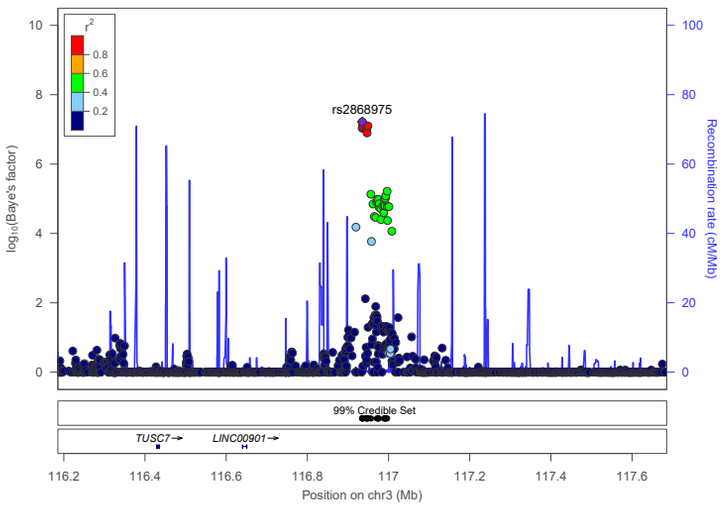
Supplementary Fig. 8

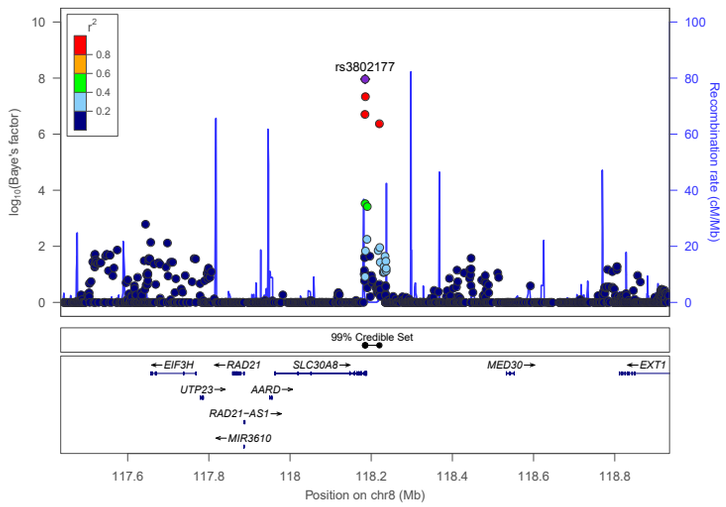
Comparison of variants between European and Asian

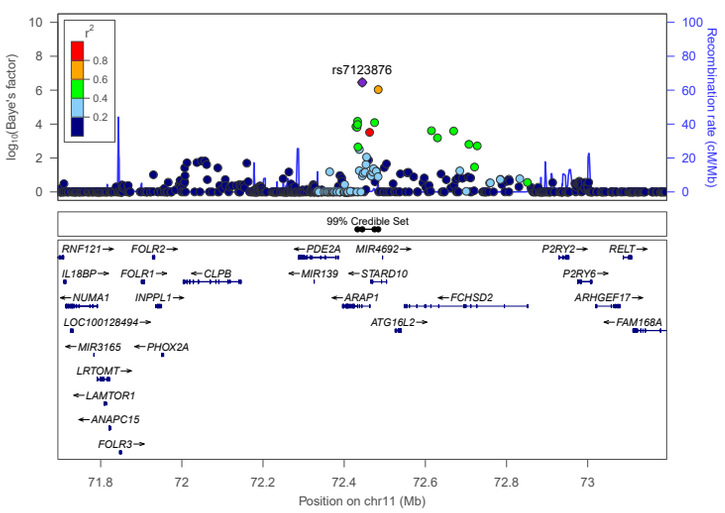
We evaluated differences in effects of 163 loci which reached significant level in the trans-ethnic meta-analysis between the current GWAS and European GWAS conducted by GIANT consortium (Locke, A. E. et al. *Nature* **518**, 197–206, 2015). Each plot were color coded according to *P* for heterogeneity between studies. Solid line denotes y = x.

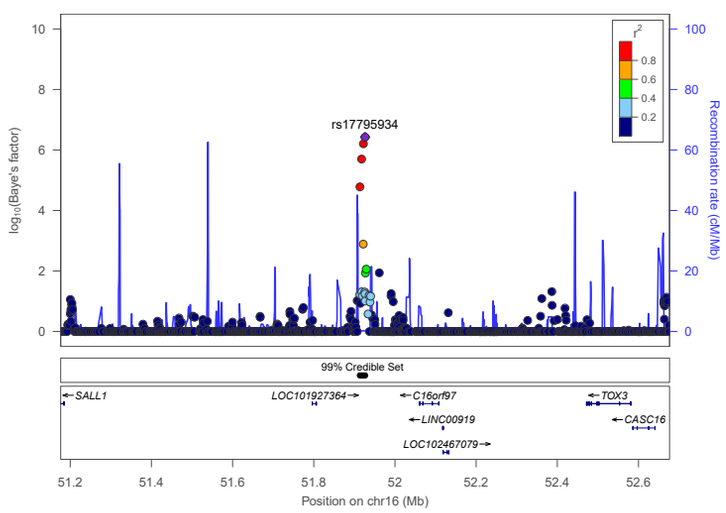
Nature Genetics: doi:10.1038/ng.3951

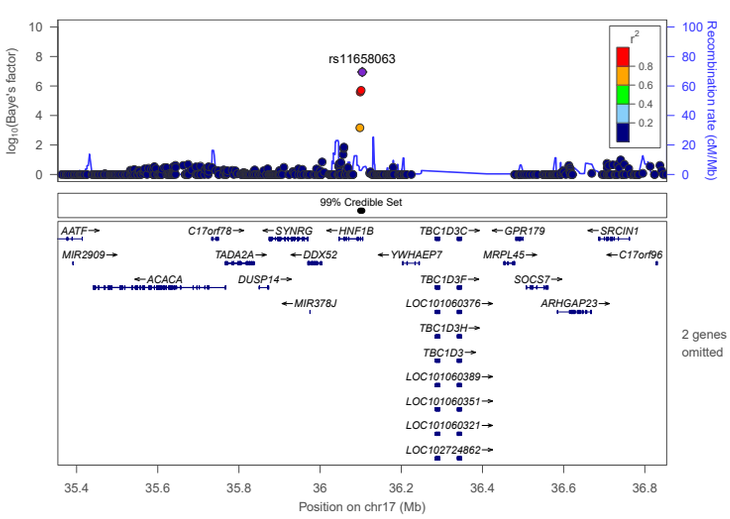
a) Regional association plots for trans-ethinc meta-analysis in male.

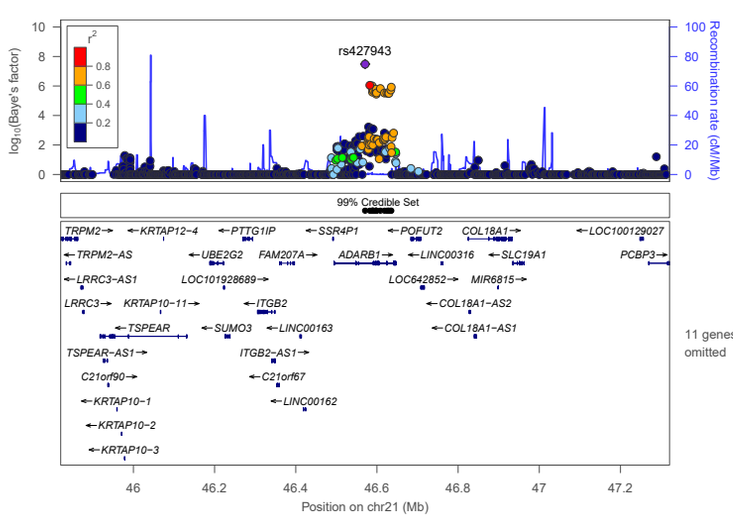






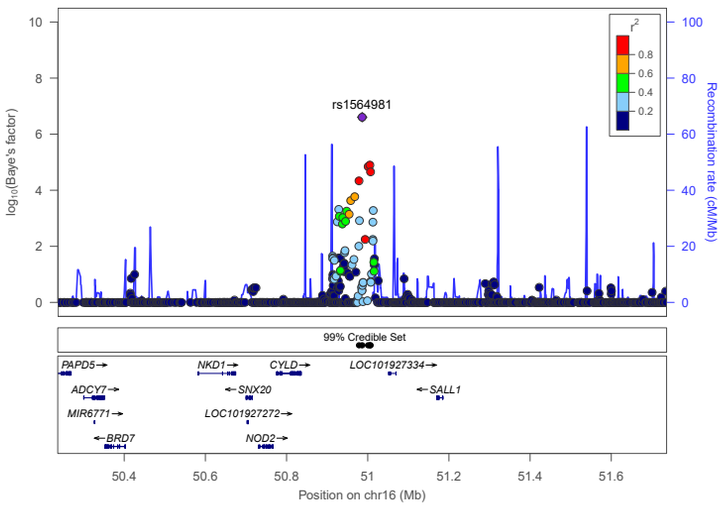




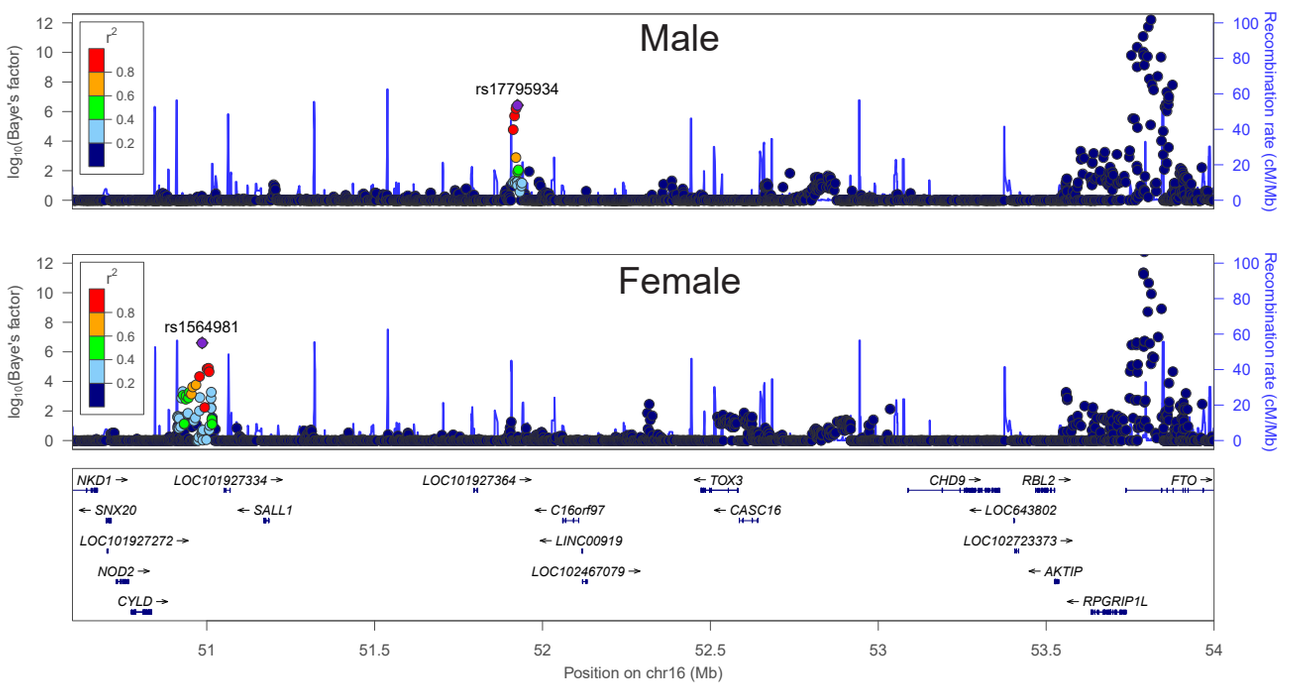


Nature Genetics: doi:10.1038/ng.3951

b) Regional association plots for trans-ethinc meta-analysis in female.



c) Regional association plots for 16q12 region in sex-stratified analysis.

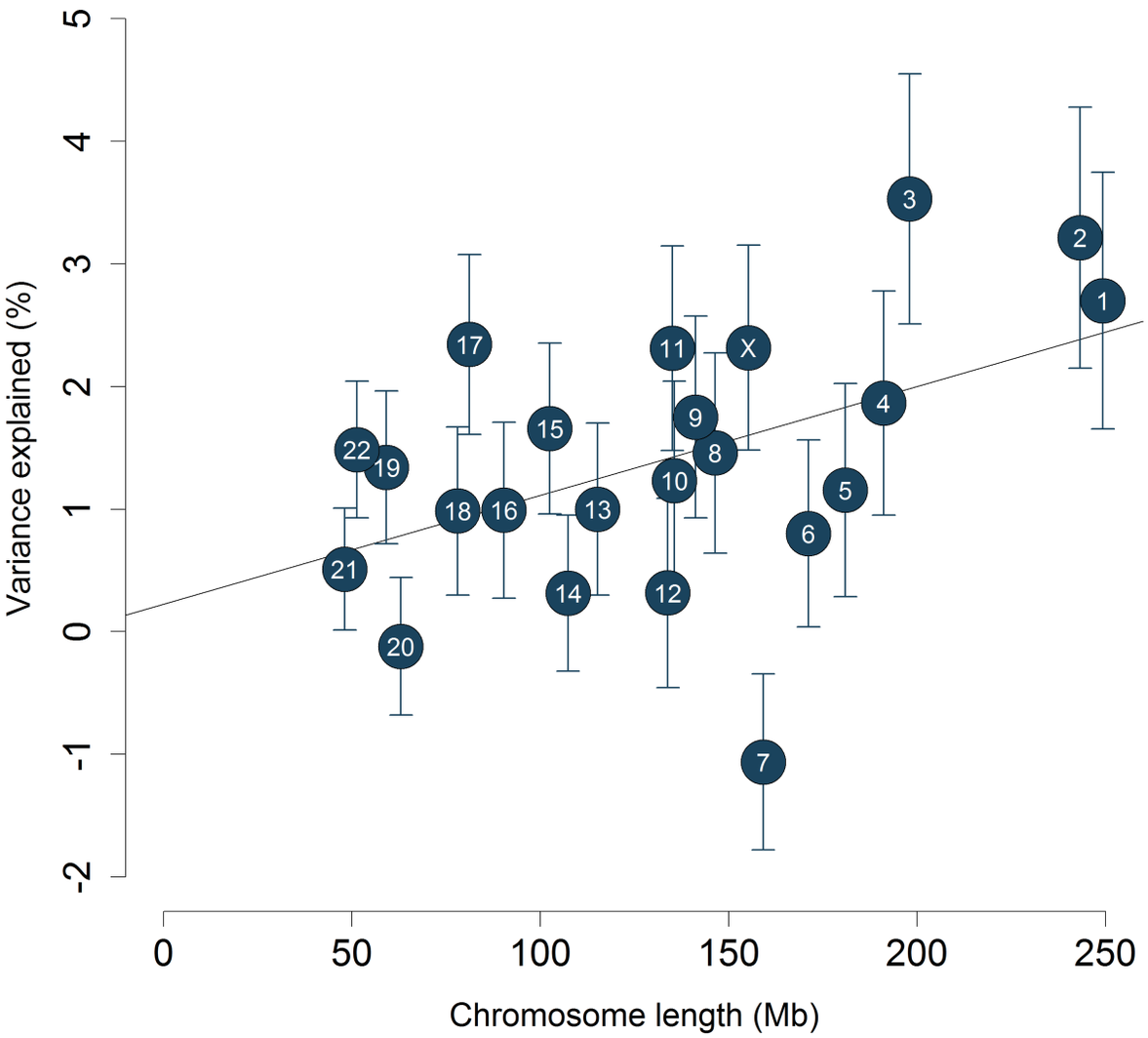


Supplementary Fig. 9

Regional association plots for sex-stratified trans-ethnic meta-analysis.

We plotted the associations of the significantly associated loci which were newly found by the meta-analysis of sex-stratified GWAS in male (a) and female (b), respectively. We additionally plotted chromosome 16q12 region to clarify positional relationship among the distinct signals (c).

Nature Genetics: doi:10.1038/ng.3951

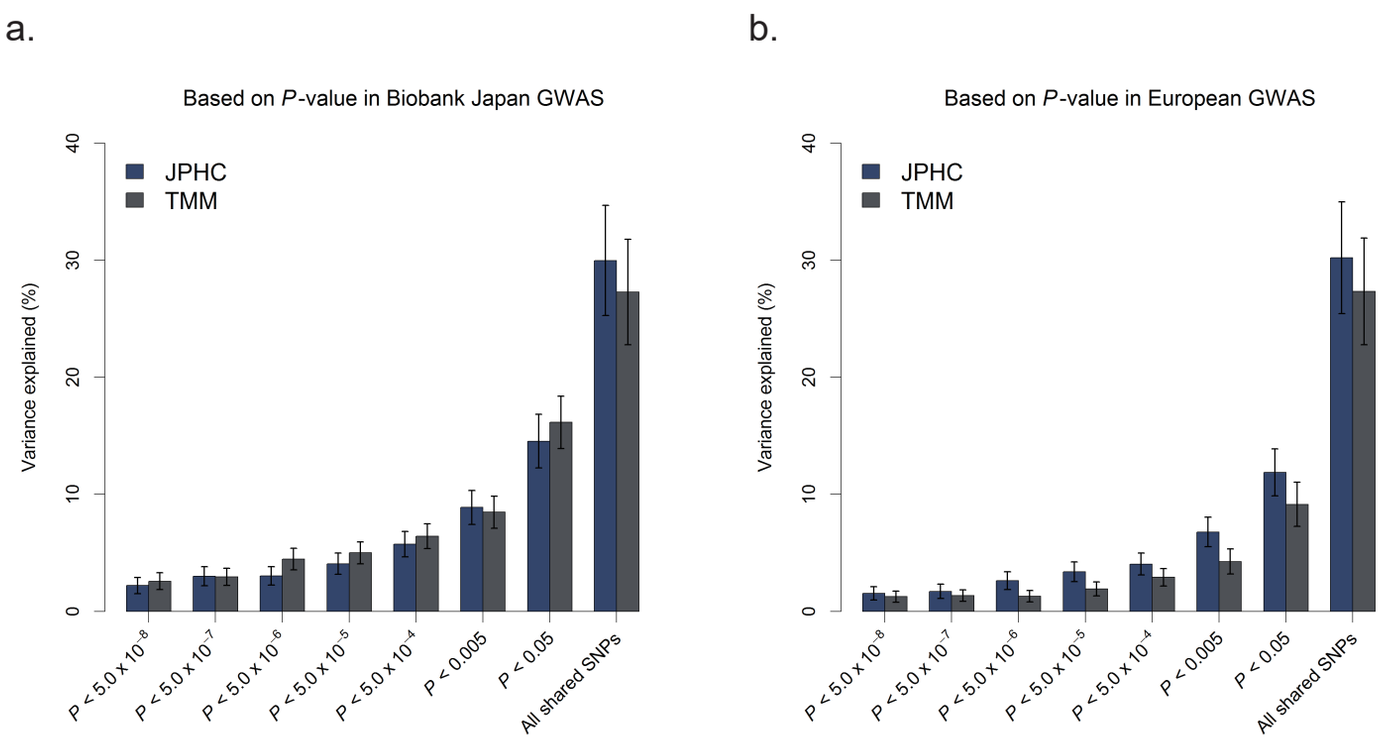


Supplementary Fig. 10

Phenotypic variance explained by each chromosome

Plot shows explained variance explained by each chromosome estimated from GREML analysis implemented by GCTA software. X axis is the chromosome length of each chromosome in hg19.

Nature Genetics: doi:10.1038/ng.3951

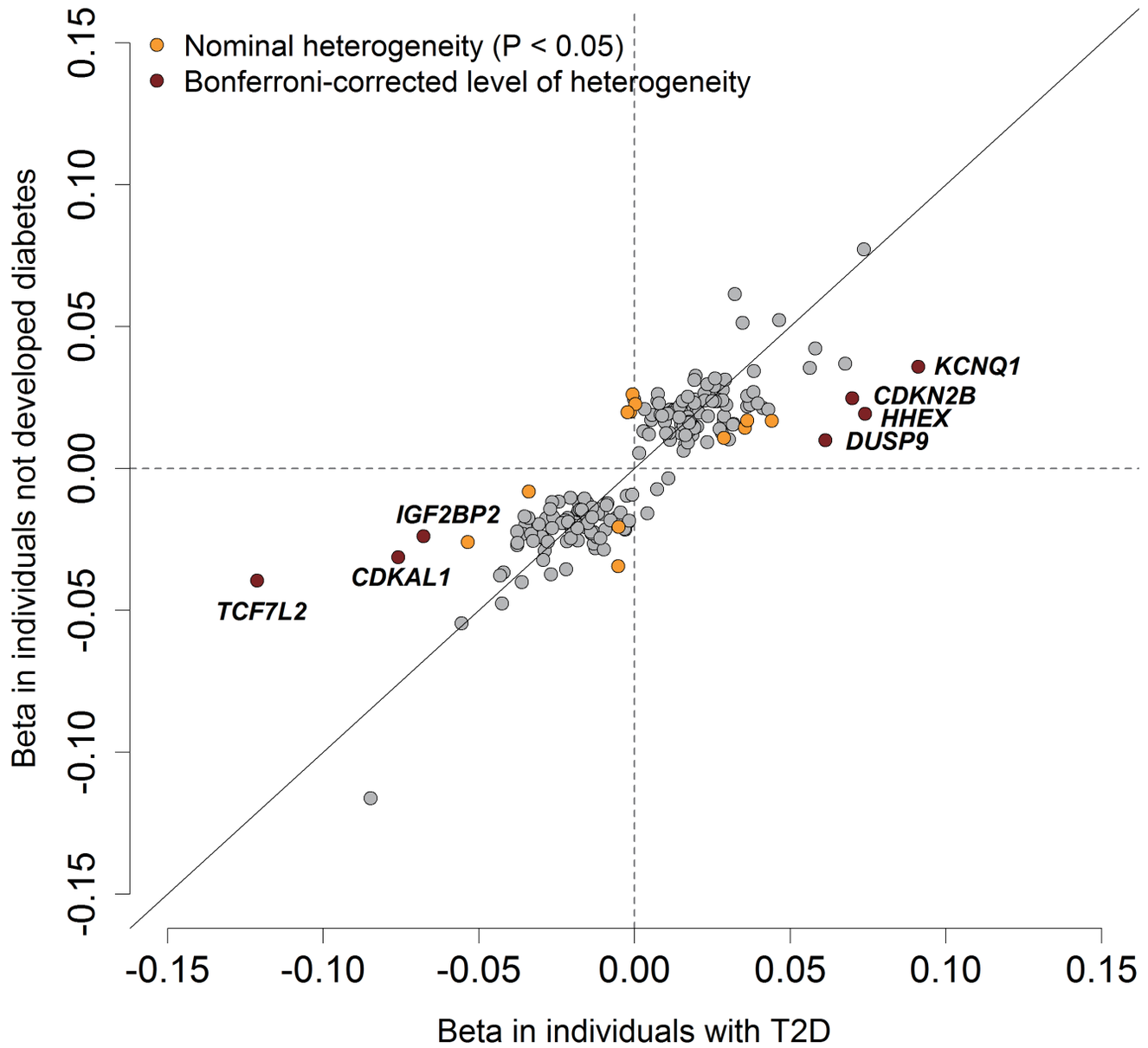


Supplementary Fig. 11

Phenotypic variance explained by subsets of variants based on *P* value

We estimated variance explained by subset of variants below *P* value thresholds for two Japanese population cohort (JPHC and TMM). a) Estimates from *P* value based on current GWAS, b) Estimates based on GIANT GWAS. Error bars indicate standard errors.

Nature Genetics: doi:10.1038/ng.3951

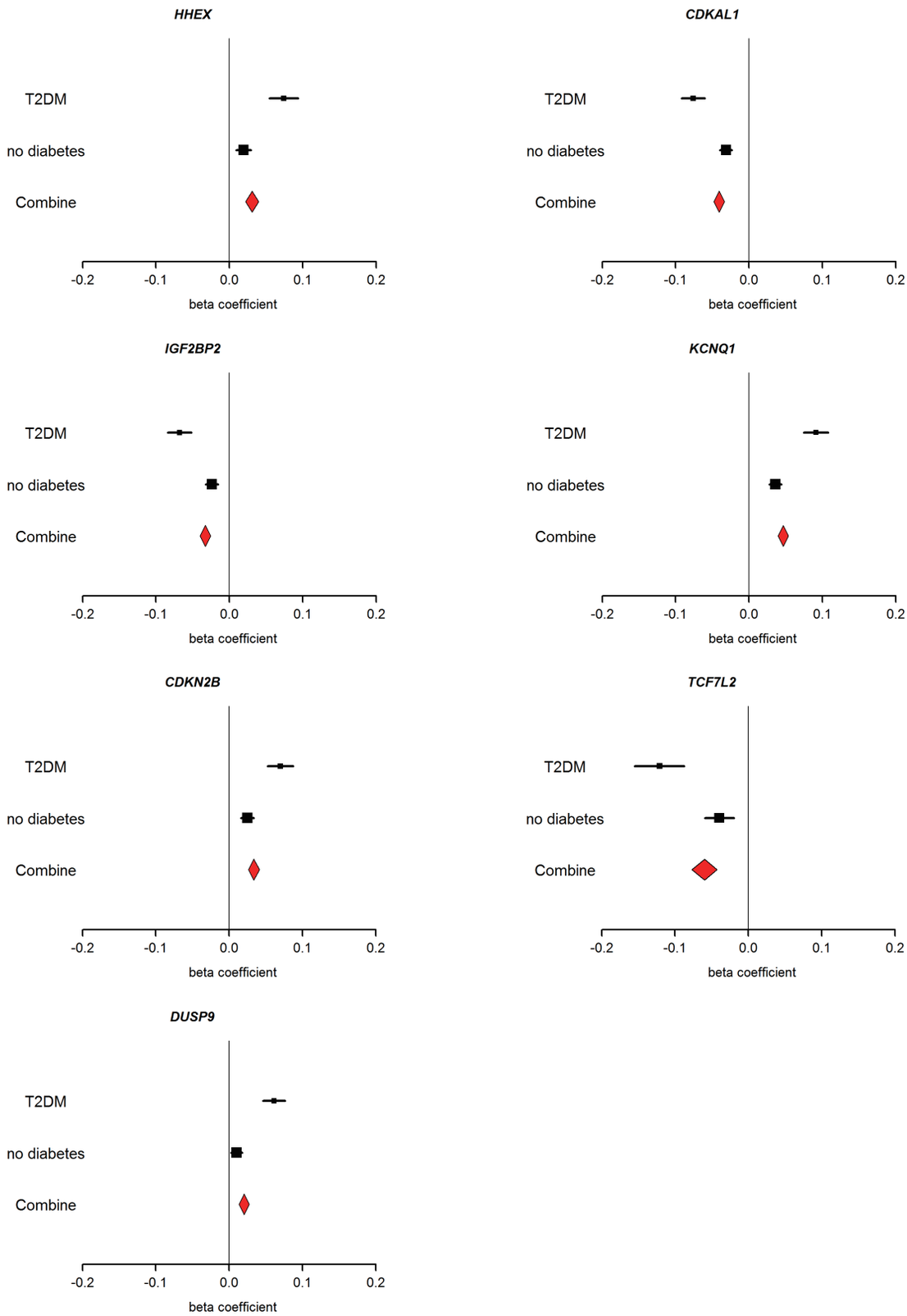


Supplementary Fig. 12

Comparison of the variants between T2D and others

Comparison of beta coefficients for BMI between individuals with type 2 diabetes and individuals who were not diagnosed as T2D. Loci which showed nominally significant difference were plotted in yellow, Bonferroni-corrected level of differences were plotted in red.

Nature Genetics: doi:10.1038/ng.3951

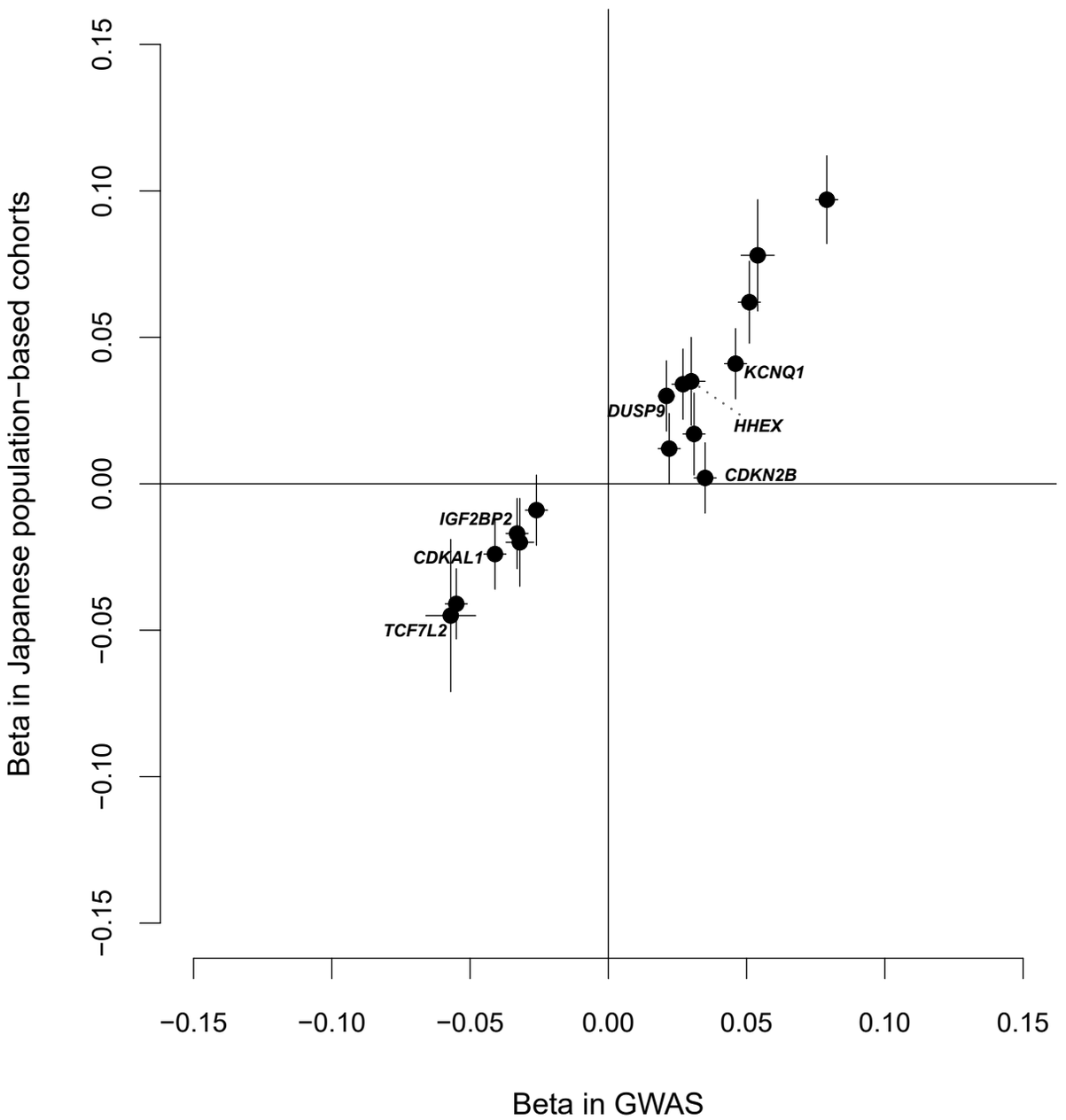


Supplementary Fig. 13

Forest plots of the effect sizes for BMI in T2D and in non-diabetics

Forest plot of the effect sizes for BMI at seven loci which showed significant differences between individuals with type 2 diabetes (T2D) and individuals who were not developed T2D are shown.

Nature Genetics: doi:10.1038/ng.3951



Supplementary Fig. 14

Comparison of T2D associated variants between GWAS and replication sets

We evaluated differences in effect sizes for BMI of 16 variants which reached genome-wide significant level of association for type 2 diabetes (T2D) between GWAS and Japanese population based cohorts (replication sets).

Nature Genetics: doi:10.1038/ng.3951

Supplementary Note

1. Descriptions of the participating cohorts

The Biobank Japan (BBJ)

The BBJ Project (http://biobankjp.org) started at the Institute of Medical Science, the University of Tokyo in 2003. To date, the BBJ Project has collected around 200,000 individuals with disease cases consisting of 47 various diseases. These subjects were recruited from 12 medical institutes in Japan including, Osaka Medical Center for Cancer and Cardiovascular Diseases, the Cancer Institute Hospital of Japanese Foundation for Cancer Research, Juntendo University, Tokyo Metropolitan Geriatric Hospital, Nippon Medical School, Nihon University School of Medicine, Iwate Medical University, Tokushukai Hospitals, Shiga University of Medical Science, Fukujuji Hospital, National Hospital Organization Osaka National Hospital, and Iizuka Hospital.

Japan Public Health Center-based Prospective Study (JPHC)

The JPHC samples were derived from a cohort of 33,736 residents in 9 public health center (PHC) areas who not only returned a self-administered questionnaire but also donated 10 mL of venous blood at the baseline survey. For the first step of sample selection, we stratified the cohort by sex, 5-year age categories, and 9 PHC areas, and then conducted a random sampling, in which a similar proportion of subjects was selected from each stratum. Consequently, we determined 9,296 subjects for the present GWAS. Before using the JPHC samples for genetic research, we obtained an approval from the institutional review board of the National Cancer Center (approval number: 2011-044), Tokyo, Japan, and provided all eligible subjects with the opportunity to refuse participation in the research.

Nature Genetics: doi:10.1038/ng.3951

The Tohoku Medical Megabank (TMM) Project

The TMM Project is a ten-year reconstruction project from the Great East Japan Earthquake, 2011, conducted by Tohoku University (http://www.megabank.tohoku.ac.jp/english/) and Iwate Medical University (http://iwate-megabank.org/en/)[1]. The TMM Project conducts two prospective cohort studies in Miyagi and Iwate Prefectures, Japan; the TMM Community-Based Cohort Study (TMM CommCohort Study) and the TMM Birth and Three-Generation Cohort Study (TMM BirThree Cohort Study). The TMM CommCohort Study is a population-based adult cohort study and has recruited approximately 84,000 participants aged 20 years or over during 2013–2016. The TMM BirThree Cohort Study has recruited around 60,000 participants including fetuses and their parents, siblings, grandparents, and extended family members as of July 2016, and will recruit 70,000 or more participants by March 2017. All participants in the TMM Project consent to genetic studies. Biospecimens (blood and urine) and medical data (questionnaires, blood and urine tests, and physiological measurements) have been collected at baseline examination. These samples and information are stored in the integrated biobank of the TMM Project.

DNA samples of the participants of the TMM CommCohort Study recruited in 2013 have been analyzed by using the Illumina OmniExpressExome array (N=10,000). Information about age and sex has been collected by using self-administered questionnaires and by reviewing municipal basic resident register. Of the 10,000 persons with genotype data, height and weight were measured for 9,202 persons in a standard manner. For persons without the measurement of body height and weight (N=798), these variables were obtained from self-reported these variables were obtained from self-reported questionnaires when available (N=703). Remaining 95 persons who had neither measured nor self-reported values are excluded from the analyses of the present study. questionnaires when available (N=703). Remaining 95 persons who had neither measured nor self-reported values are excluded from the analyses of the present study.

Nature Genetics: doi:10.1038/ng.3951

2. Characteristics of samples analyzed in GWAS and population-based cohorts

Characteristics of participants

Study	Study	N (male / female)	Age [year; mean ± SD]	BMI [kg/m^2; mean ± SD]
GWAS	BBJ	158,284 (85,894 / 72,390)	62.6 ± 14.0	23.3 ± 3.7
Cohorts	JPHC	7,379 (2,475 / 4,904)	53.6 ± 7.9	23.6 ± 3.0
	TMM	7,767 (2,623 / 5,144)	61.1 ± 11.1	23.5 ± 3.5
	Total	173,430 (90,992 / 82,438)	-	-

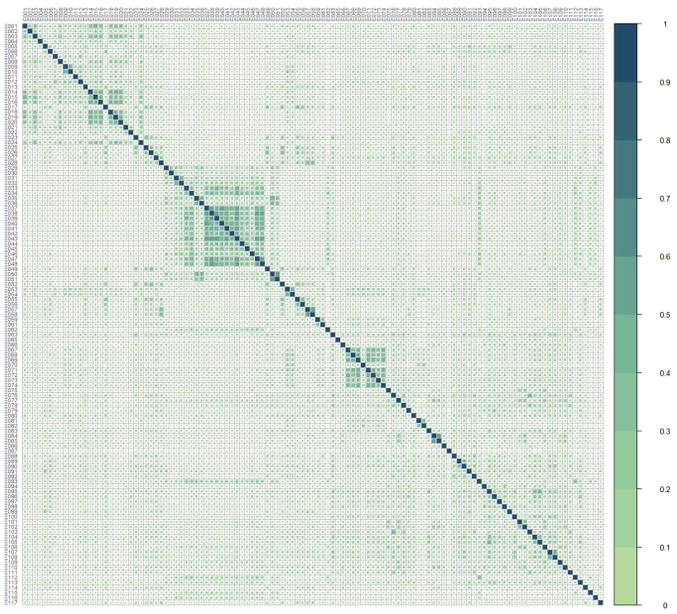
Baseline characteristics of analyzed samples were summarized above.

BMI: body-mass index, SD: standard deviation.

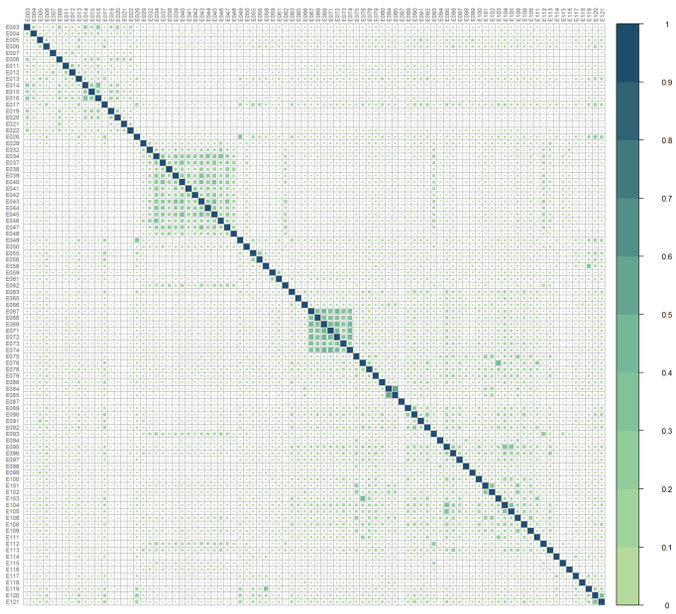
Nature Genetics: doi:10.1038/ng.3951

3. Evaluation of enhancer overlaps in each cell-type

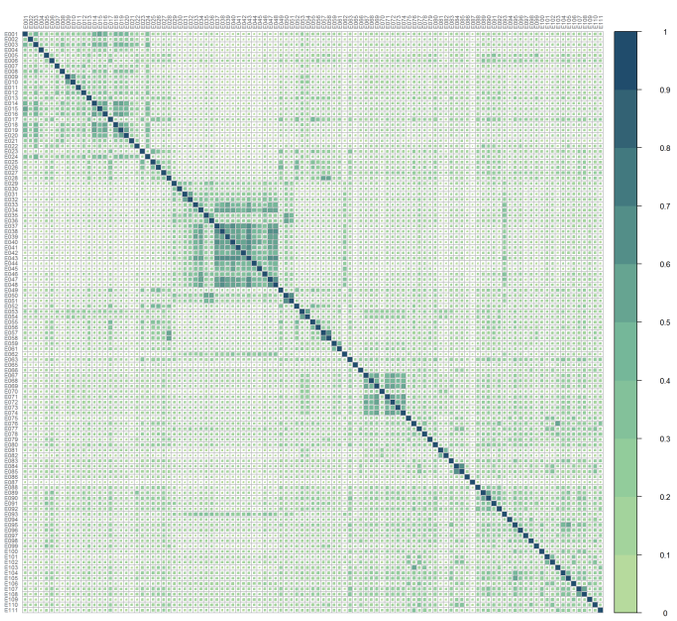
a



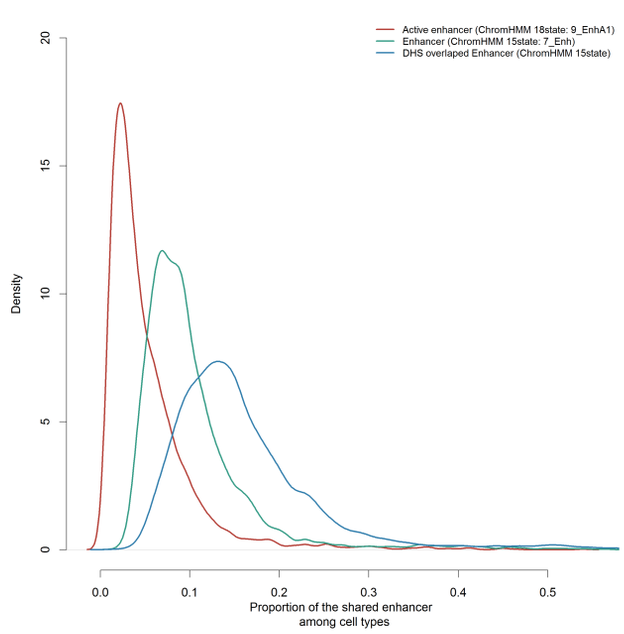
b



c



d



Nature Genetics: doi:10.1038/ng.3951

To evaluate the overlaps of enhancer in publicly available annotations, we downloaded two chromatin state models (Core-15 state model (a), and Expanded 18-state model (b)) and DNaseI-accessible enhancer (c) information from Roadmap epigenomics project. We annotated the variants included in the trans-ethnic meta-analysis (N = 2 mililon~) by each cell-type. From the chromatin state information, we selected enhancers (state 7) in 15 state model, and Enhncer A1 (state 9) in 18 state model for the annotation, respectively.

We calculated the overlaps of the enhancer between the enhancer of cell-type A (EnhA) and cell-type B (EnhB) as follows: Overlap = N(EnhA ∩ EnhB) / N(EnhA ∪ EnhB).

The figures (a - c) shows the overlaps amaong cell-types. We also plotted the density of the overlaps in each pair of cell-types (d). Cell-types were coded according to the epigenome ID of the Roadmap project. As a result, the mean overlaps were 10.3 % for 15-state model enhancer, 5.3 % for active enhancer of 18-state model, and 16.0 % in DNaseI-accessible enhancer , respectively.

Nature Genetics: doi:10.1038/ng.3951

4. GWASs used for Cross-trait LD score regression analysis

a. Case-control study

1) Summary of the published GWAS used in cross-trait LD score regression analysis (n = 8)

Traits	Number of samples		Imputation reference panel	Rsq cut off	PMID
	case	control			
Adolescent idiopathic scoliosis[2]	2,109	11,140	1KGP EAS	0.9	26211971
Age-related macular degeneration[3,4]	827	3,323	HapMap2	0.7	23455636
Atopic Dermatitis[5]	1,472	7,966	1KGP EAS	0.7	26482879
Crohn's disease[6]	372	3,321	1KGP EAS	0.9	26511940
Inflammatory bowel disease[6]	743	3,321	1KGP EAS	0.9	26511940
OPLL[7]	1,130	7,135	1KGP EAS	0.9	25064007
Rheumatoid arthritis (Asian samples only)[8]	4,873	17,642	1KGP EAS	0.5	24390342
Ulcerative colitis[6]	371	3,321	1KGP EAS	0.9	26511940

OPLL: Ossification of posterior longitudinal ligament of the spine.

Citation of these studies were shown in below reference section.

Nature Genetics: doi:10.1038/ng.3951

2) Summary of the unpublished studies used in cross-trait LD score regression analysis (n = 25)

Diseases	Number of samples		Rsq cutoff	Covariates	Case recruitment	Controls
	case	ctrl				
Arterial fibrillation	8,180	28,612	0.9	top 2 PCs	BBJ	Cohort
Asthma	7,570	28,870	0.7	age, sex, top 10 PCs	BBJ	Cohort
Basedow disease	1,961	7,968	0.7	none	BBJ	case-mix
Bipolar disorder	2,964	61,887	0.9	top 2 PCs	Collaborators	case-mix
Breast cancer	5,272	16,496	0.7	top 2 PCs	BBJ	Cohort (female only)
Cerebral Aneurysm	2,597	28,870	0.7	age, sex, top 10 PCs	BBJ	Cohort
Chronic hepatitis B	1,092	28,870	0.7	top 10 PCs	BBJ	Cohort
Chronic hepatitis C	4,988	28,870	0.7	top 10 PCs	BBJ	Cohort
Hepatocellular carcinoma in individuals with HCV	1,001	1,968	0.7	top 2 PCs	BBJ	Individuals with HCV+ but not developed HCC.
Endometrial cancer	1,931	17,492	0.7	top 2 PCs	BBJ	Cohort (female only)
Endometriosis	705	17,492	0.7	top 2 PCs	BBJ	Cohort (female only)
Esophageal cancer	1,225	27,178	0.7	top 2 PCs	BBJ	Cohort (female only)
Gastric cancer	6,171	27,178	0.7	top 2 PCs	BBJ	Cohort
Glaucoma	3,980	18,815	0.7	top 10 PCs	BBJ	Cohort
Ischemic stroke	16,256	27,294	0.7	age, sex, top 10 PCs	BBJ	Cohort
Lung cancer	3,874	27,178	0.7	top 2 PCs	BBJ	Cohort
Myocardial infarction	12,494	28,870	0.7	age, sex, top 20 PCs	BBJ	Cohort
Osteoporosis	7,099	28,870	0.7	age, sex, top 10 PCs	BBJ	Cohort
Ovarian cancer	681	17,492	0.7	top 2 PCs	BBJ	Cohort (female only)
Peripheral artery disease	3,382	28,870	0.7	age, sex, top 5 PCs	BBJ	Cohort
Prostate cancer	5,088	10,682	0.7	top 2 PCs	BBJ	Cohort (male only)
Colorectal cancer	6,692	27,178	0.7	top 2 PCs	BBJ	Cohort
Schizophrenia	1,987	9,788	0.7	Top 2 PCs	Collaborators	case-mix
Type 2 diabetes	36,832	28,870	0.7	age, sex, top 20 PCs	BBJ	Cohort
Uterine Fibroid	5,720	17,492	0.7	top 10 PCs	BBJ	Cohort (female only)

HCV: hepatitis C virus.

Nature Genetics: doi:10.1038/ng.3951

b. GWAS for hematological traits

Traits used in cross-trait LD score regression analysis (N = 8)

Traits	N	Phenotype standardization	Covariates
RBC	111,268	Z-score†	Age, Sex, Smoking (ever or never), top 10 PCs, affected disease
WBC	110,397	Z-score†	Age, Sex, Smoking (ever or never), top 10 PCs, affected disease
Platelet	110,659	Z-score†	Age, Sex, Smoking (ever or never), top 10 PCs, affected disease
Lymphocytes	63,197	Rank-based normal transformation	Age, Sex, Smoking (ever or never), top 10 PCs, affected disease
Neutrophil	63,197	Rank-based normal transformation	Age, Sex, Smoking (ever or never), top 10 PCs, affected disease
Basophil	63,197	Rank-based normal transformation	Age, Sex, Smoking (ever or never), top 10 PCs, affected disease
Eosinophil	63,197	Rank-based normal transformation	Age, Sex, Smoking (ever or never), top 10 PCs, affected disease
Monocytes	63,197	Rank-based normal transformation	Age, Sex, Smoking (ever or never), top 10 PCs, affected disease

RBC: red blood cells, WBC: white blood cells.

†: Individuals who were out of normalized value ± 4SD were excluded.

In all unpublished studies, we obtained GWAS summary statistics which imputed by minimac using EAS samples of 1KGP as reference after standard quality controls. Diagnoses of diseases were based on medical records in collaborative hospitals. Most of cases were recruited by BBJ project with seven exceptions (Bipolar disorder and schizophrenia were collected by M.I. and N.I. and the advanced Collaborative Study of Mood Disorder (COSMO) team; the summary statistics of GWAS for rheumatoid arthritis in Asian was obtained from Y.O.; age-related macular degeneration, inflammatory bowel disease including ulcerative colitis and Crohn’ s disease, adolescent idiopathic scoliosis and OPLL were collected by collaborating hospitals). As controls, genotype data obtained by population-based cohort studies were used with ten exceptions (Basedow disease, bipolar disorder, schizophrenia, age-related macular degeneration, inflammatory bowel disease including ulcerative colitis and Crohn’ s disease, adolescent idiopathic scoliosis, OPLL and hepatocellular carcinoma: case-mix controls were used).

Nature Genetics: doi:10.1038/ng.3951

Reference

1. Kuriyama, S. et al. The Tohoku Medical Megabank Project: Design and Mission. *J. Epidemiol.* **26** (9), 493-511 (2016).
2. Ogura, Y. et al. A Functional SNP in BNC2 Is Associated with adolescent idiopathic scoliosis. *Am. J. Hum. Genet.* **97**, 337–342 (2015).
3. Arakawa, S. et al. Genome-wide association study identifies two susceptibility loci for exudative age-related macular degeneration in the Japanese population. *Nat. Genet.* **43**, 1001–4 (2011).
4. Fritsche, L. G. et al. Seven new loci associated with age-related macular degeneration. *Nat. Genet.* **45**, 433–9 (2013).
5. Hirota, T. et al. Genome-wide association study identifies eight new susceptibility loci for atopic dermatitis in the Japanese population. *Nat. Genet.* **44**, 1222–6 (2012).
6. Fuyuno, Y. et al. Genetic characteristics of inflammatory bowel disease in a Japanese population. *J. Gastroenterol.* **51**(7), 672–81 (2015).
7. Nakajima, M. et al. A genome-wide association study identifies susceptibility loci for ossification of the posterior longitudinal ligament of the spine. *Nat. Genet.* **46**, 1012–6 (2014).
8. Okada, Y. et al. Genetics of rheumatoid arthritis contributes to biology and drug discovery. *Nature* **506**, 376–81 (2014).

Nature Genetics: doi:10.1038/ng.3951