Figure S1

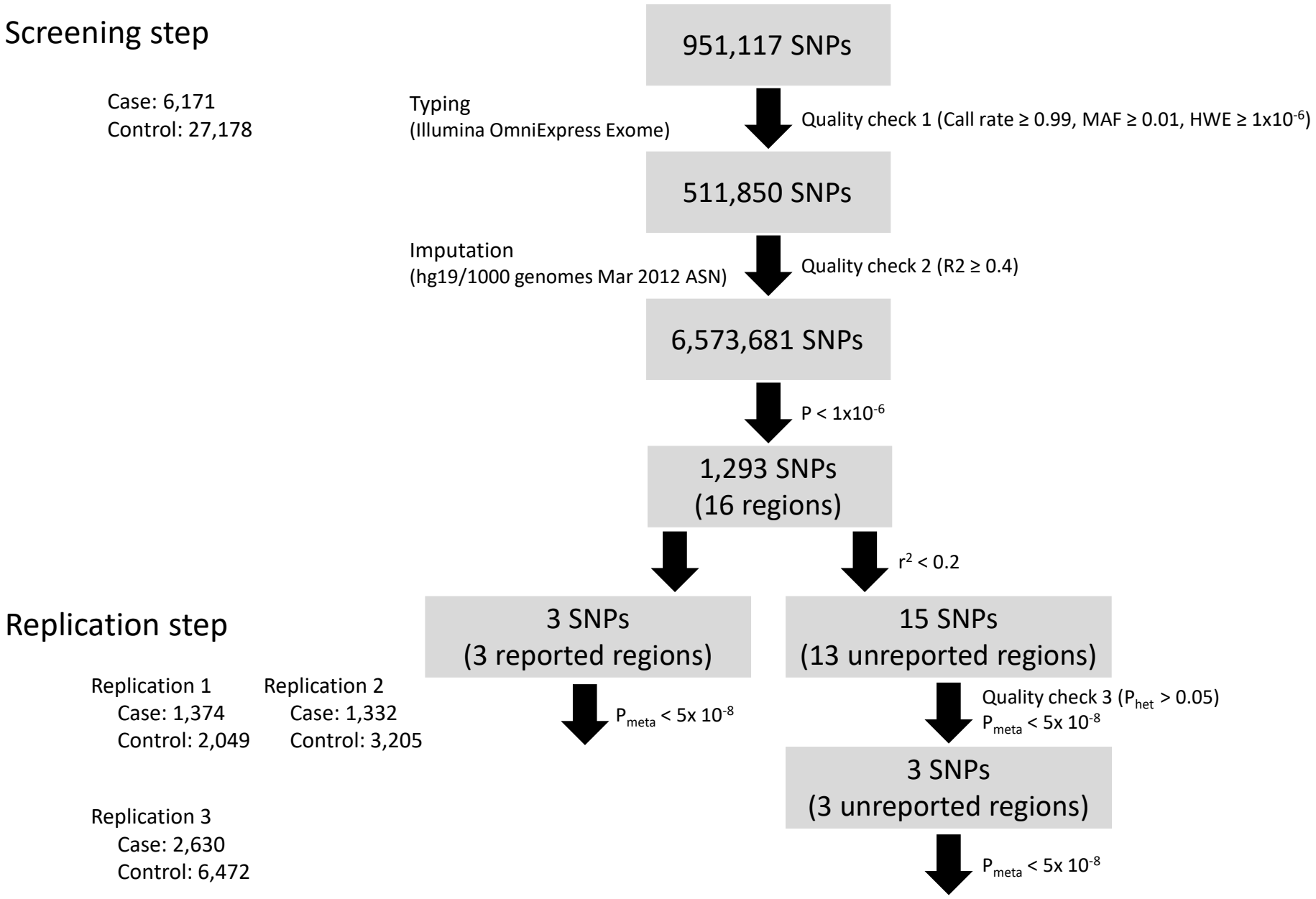
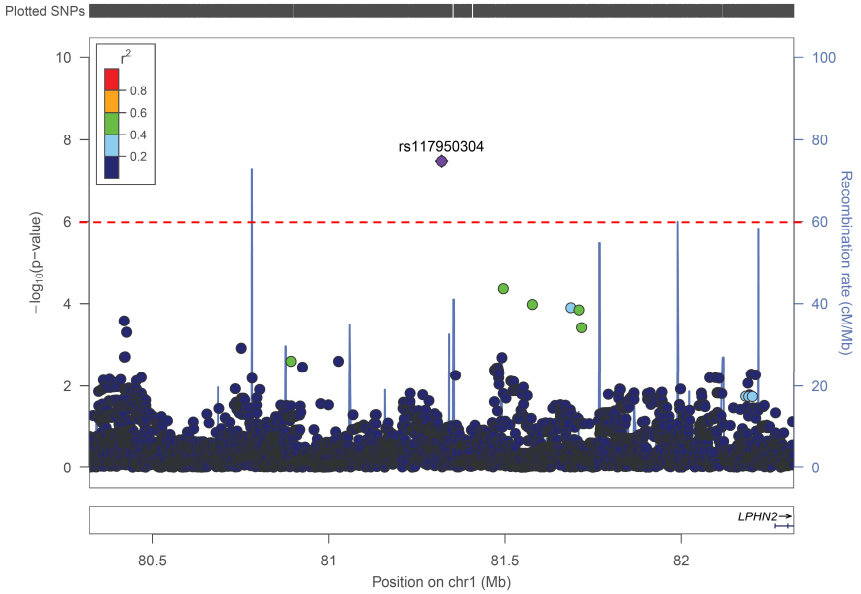


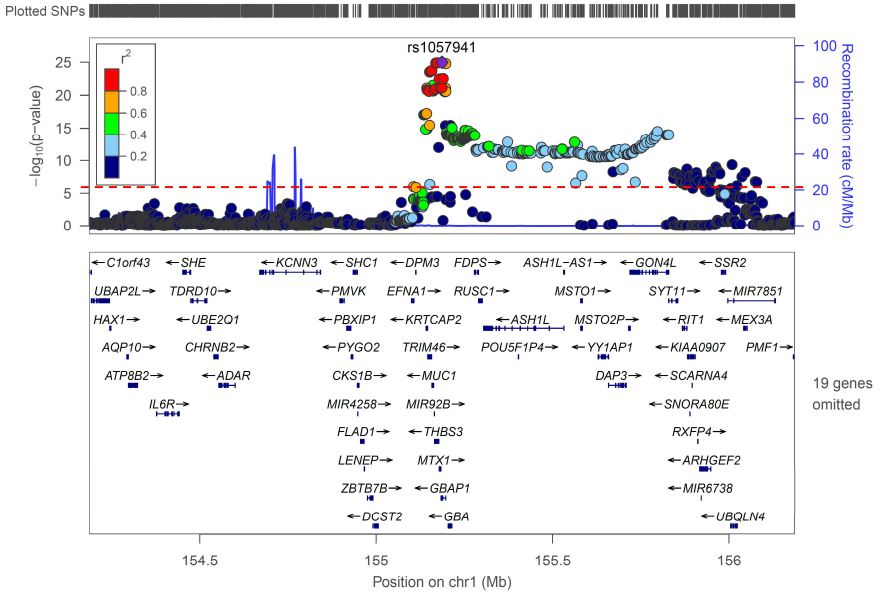
Figure S1 Scheme for study design and screening results.

Figure S2

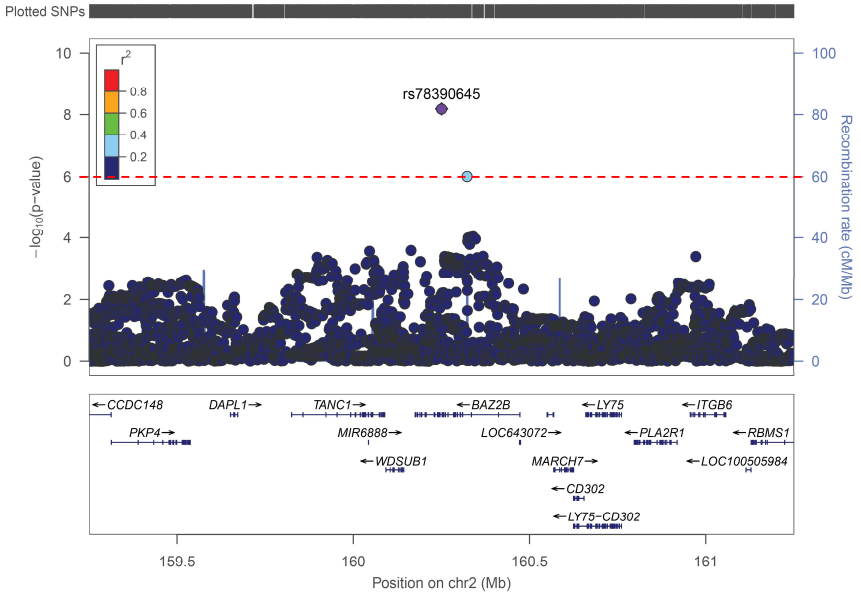
rs117950304



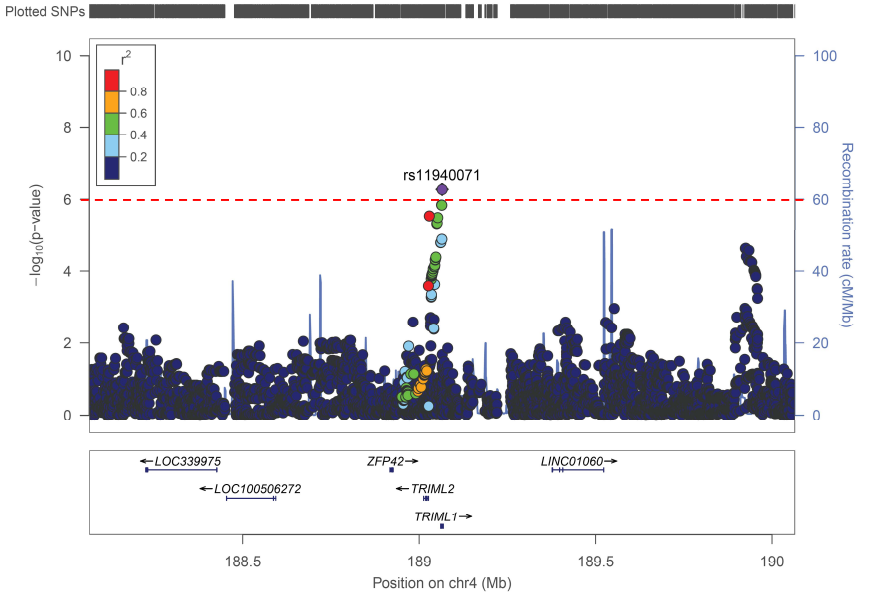
rs1057941



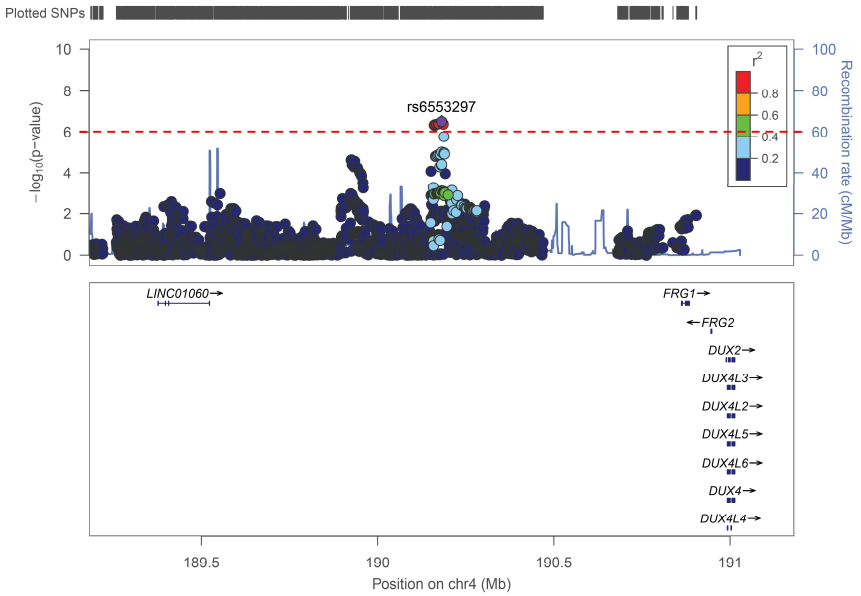
rs78390645



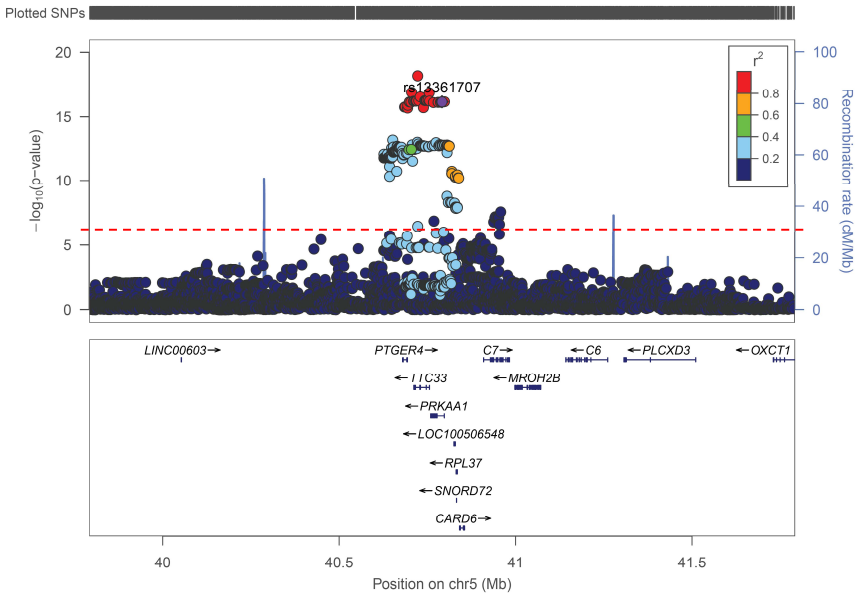
rs11940071



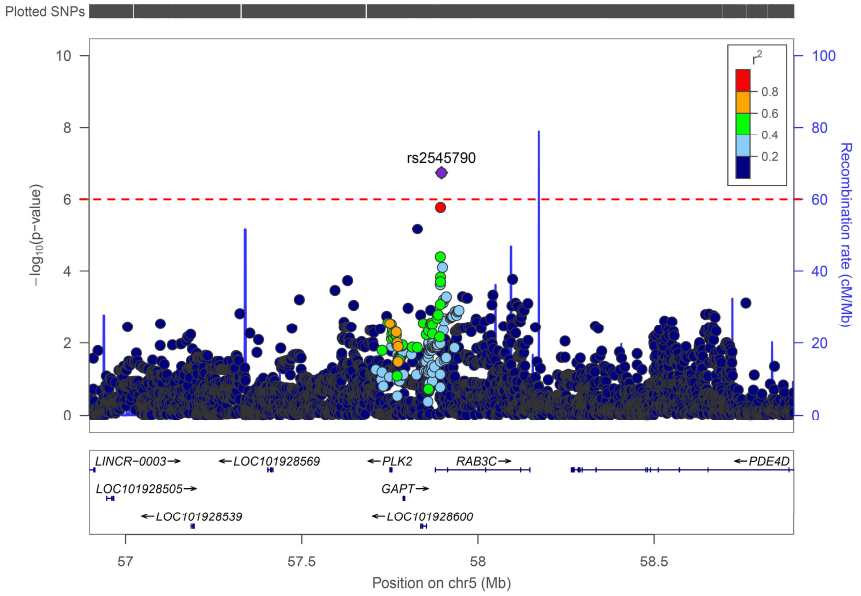
rs6553297



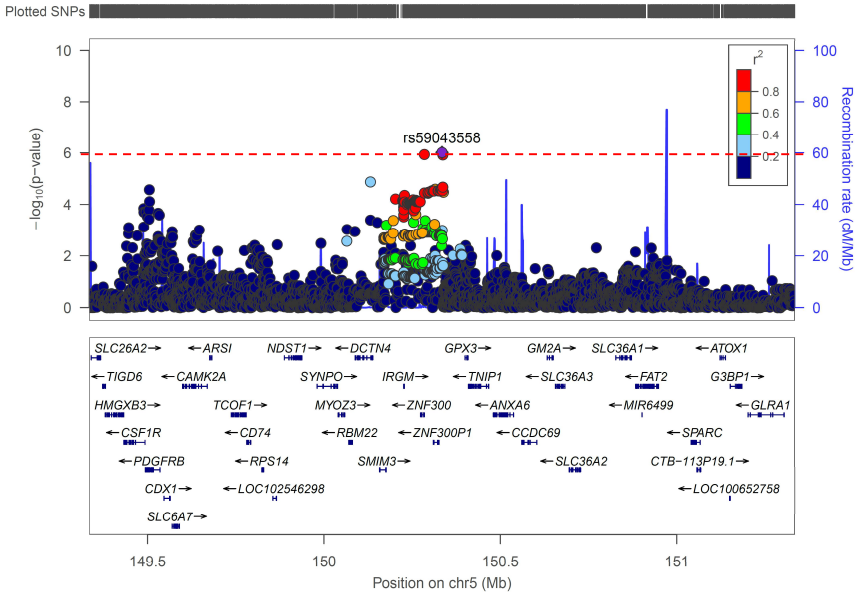
rs13361707



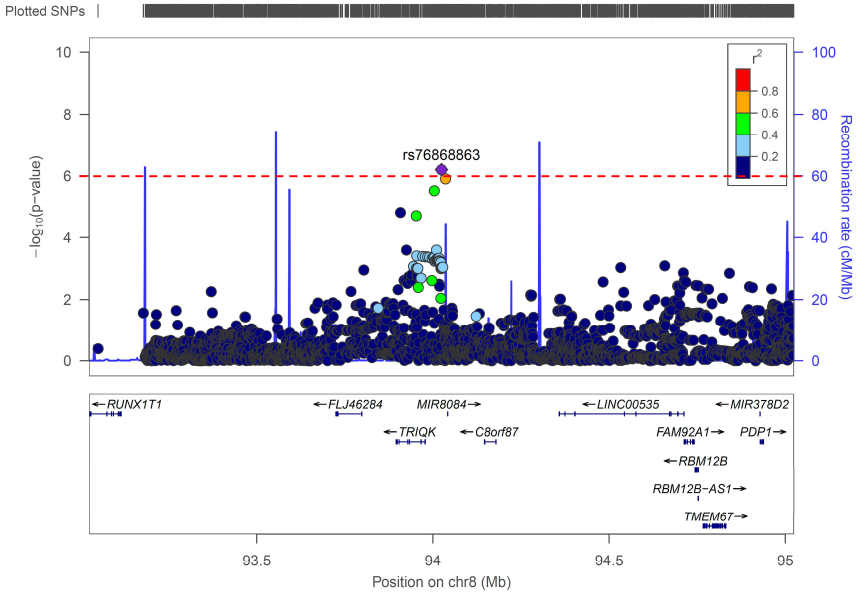
rs2545790



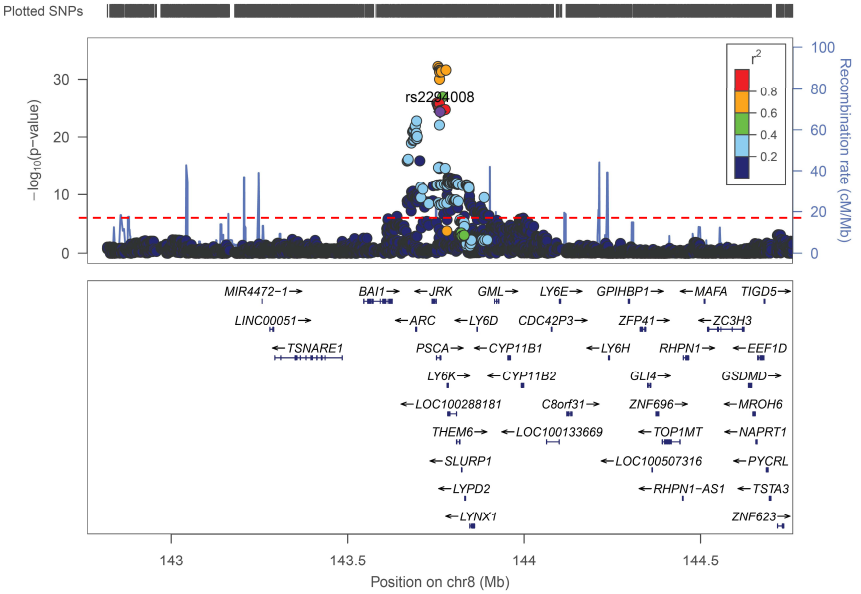
rs59043558



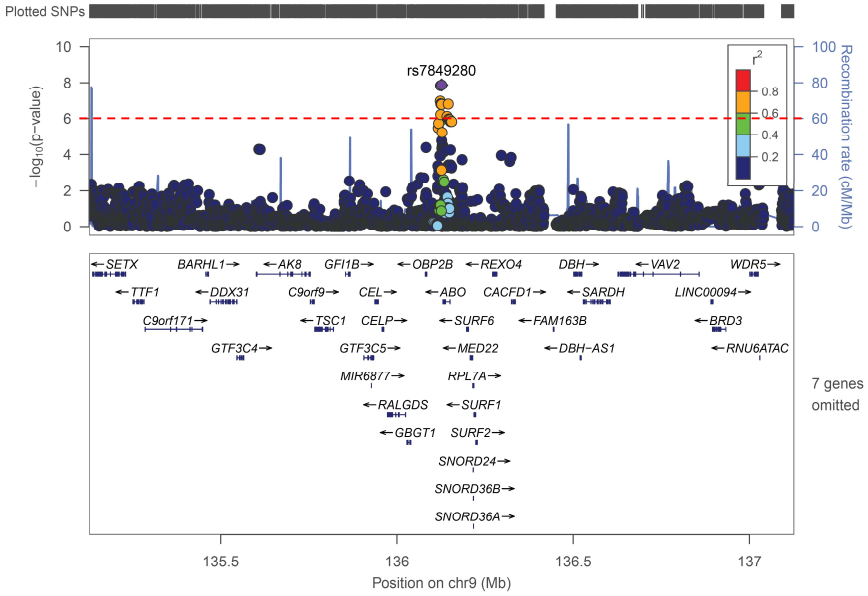
rs76868863



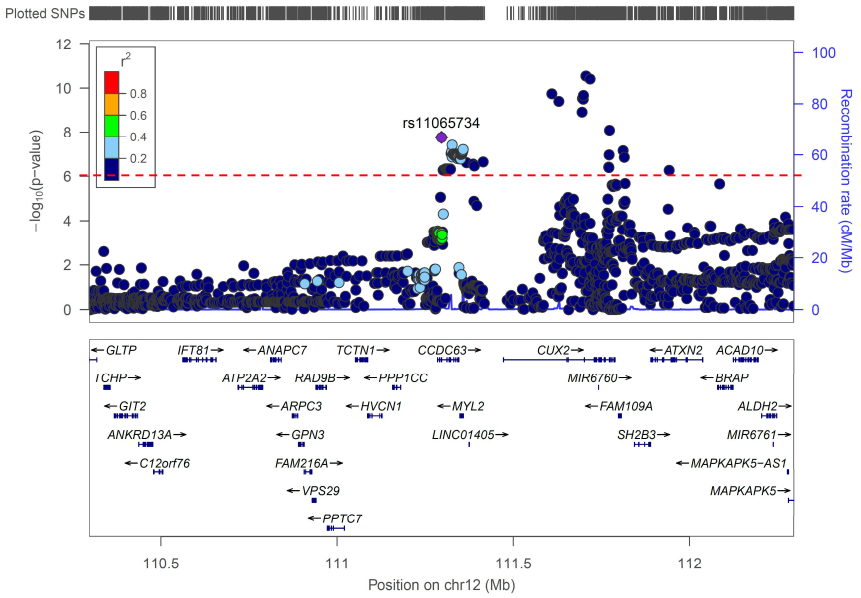
rs2294008



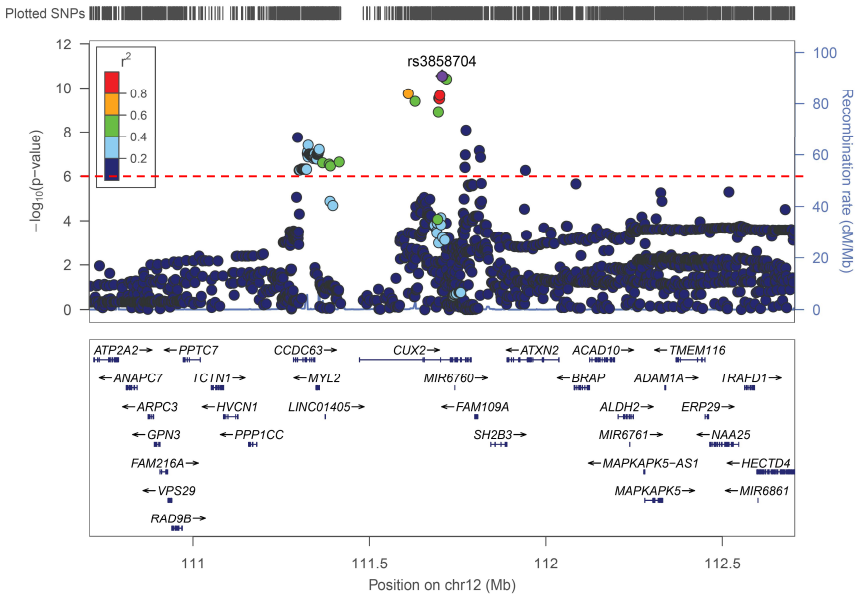
rs7849280



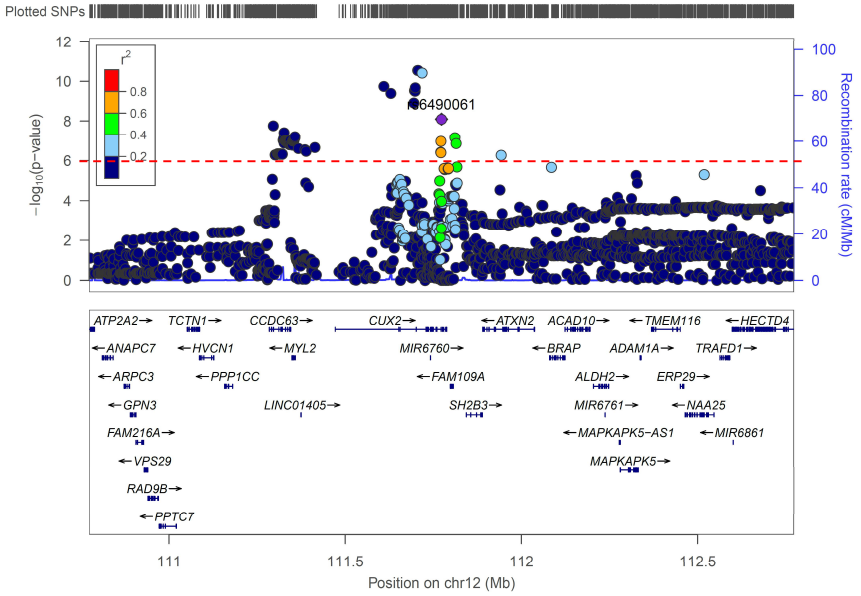
rs11065734



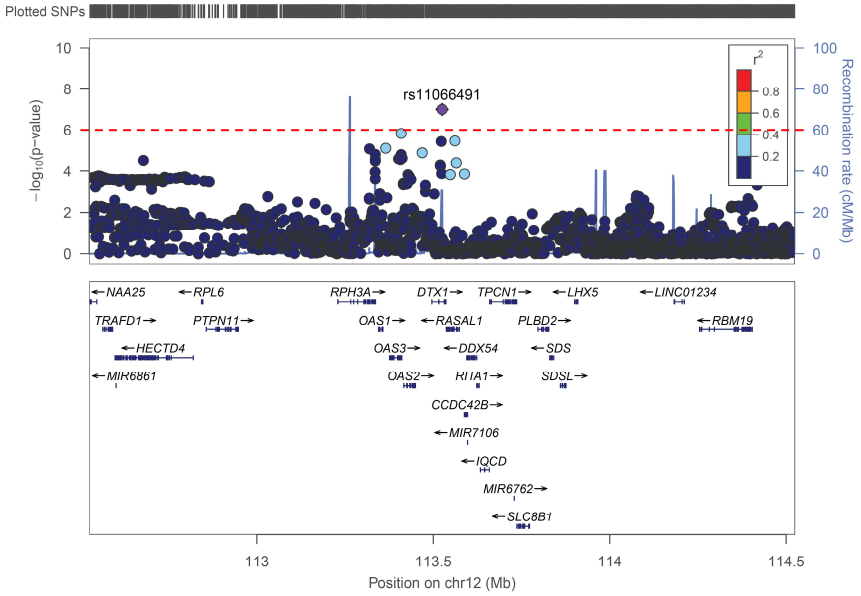
rs3858704



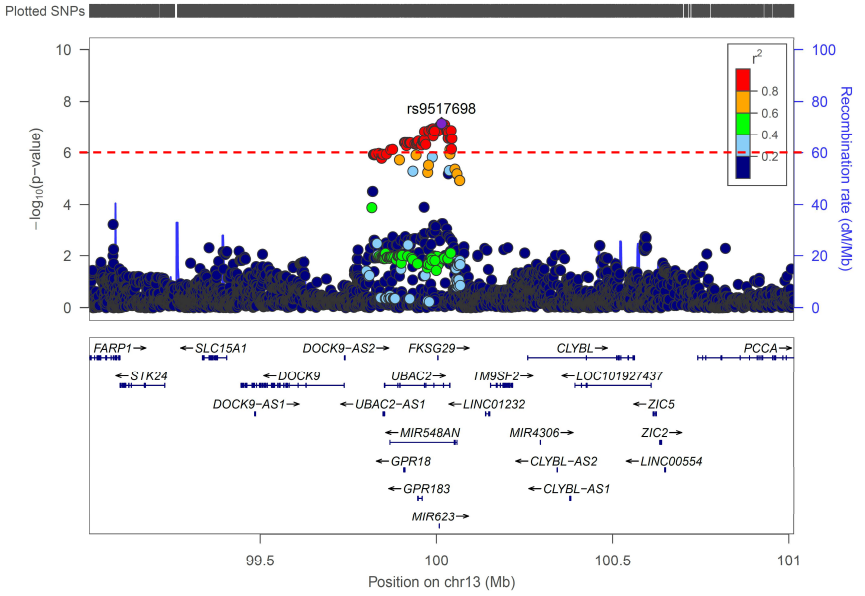
rs6490061



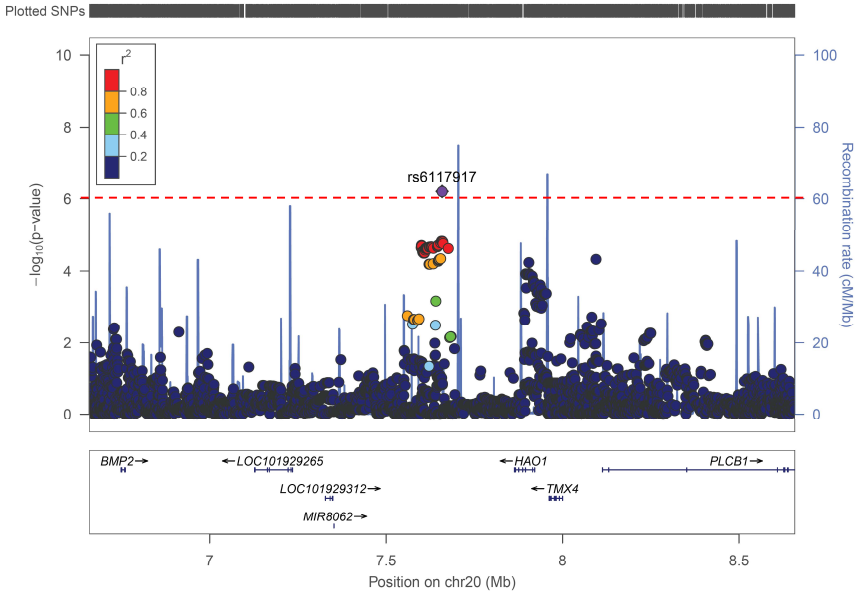
rs11066491



rs9517698



rs6117917



rs2376549

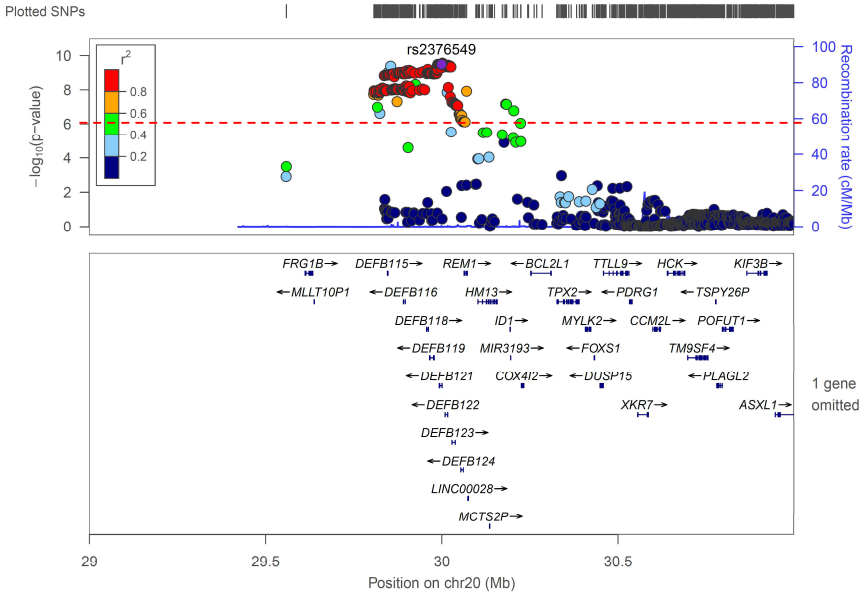
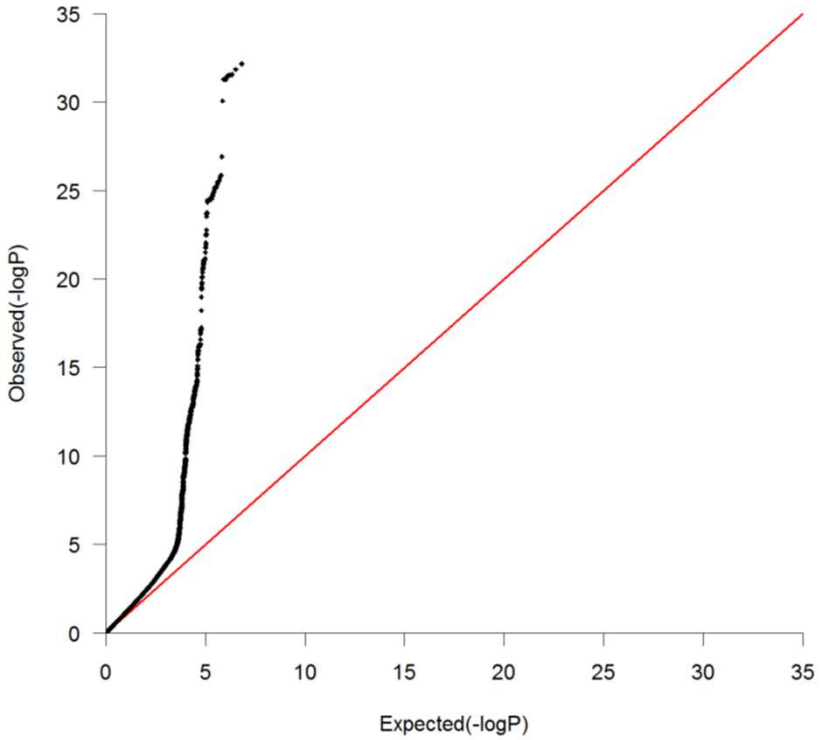


Figure S2 Regional plots of 16 candidate loci for gastric cancer. The $-\log_{10} P$ values from the screening stage in reported three loci (red) and novel 13 loci (black) are shown. Estimated recombination rates (from 1000 Genomes) are plotted in blue. The SNPs are color-coded to reflect their correlation with genotyped SNPs. Pairwise r2 values are from 1000 Genomes East Asian data (March 2012 release). Genes, position of exons and direction of transcription from UCSC genome browser (genome.ucsc.edu) are noted. Plots were generated using LocusZoom (http://csg.sph.umich.edu/locuszoom).

Figure S3



$\lambda = 1.2280$

$\lambda_{1000} = 1.0227$

Figure S3 Quantile quantile plot for imputed GWAS of gastric cancer in the Japanese population.

Figure S4

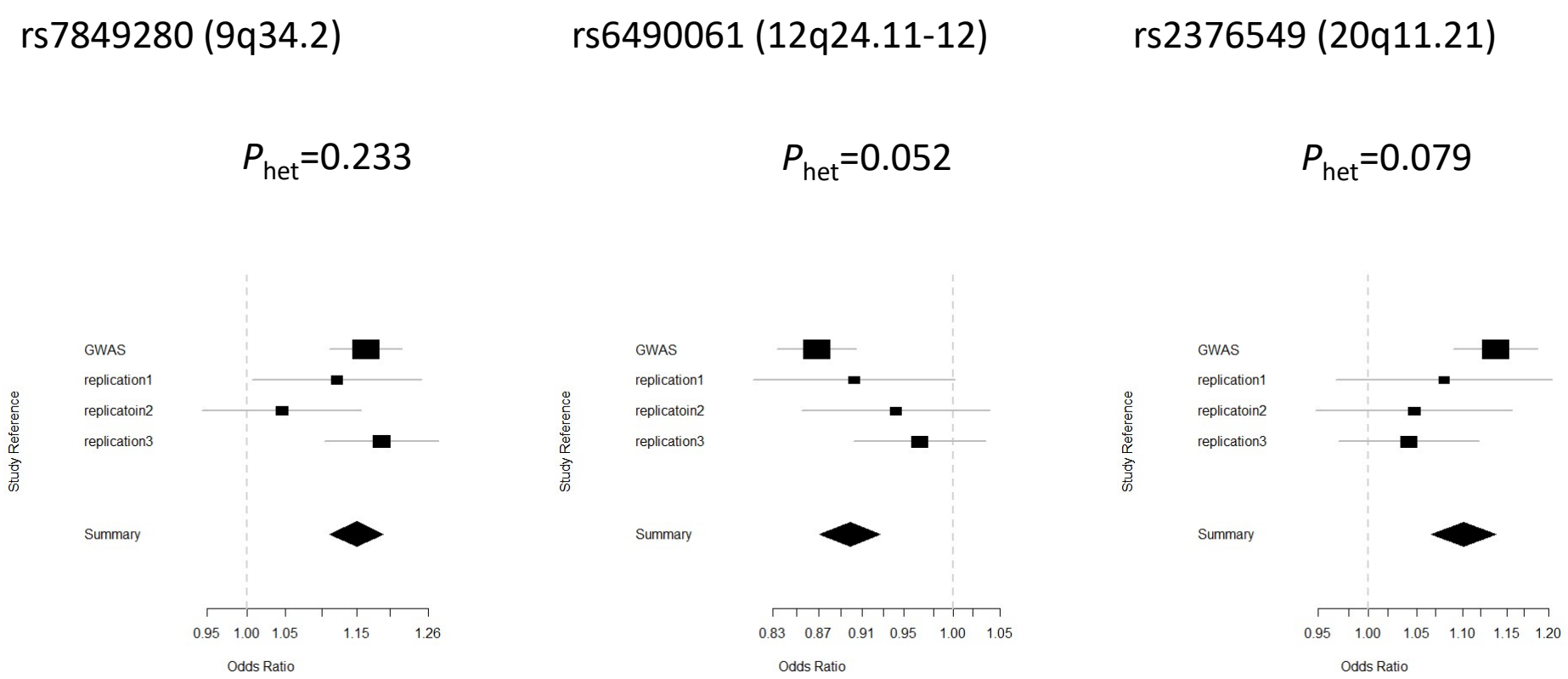
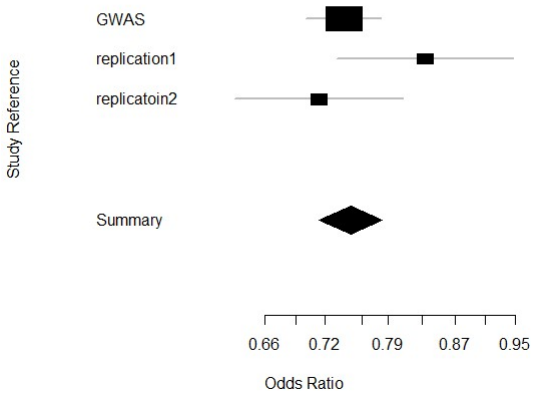


Figure S4 Forest plots for risk variants in the three novel gastric cancer risk loci. Plots show the association estimates (odds ratios) and 95% confidence intervals for the screening stage and three replication studies presented as bars. The association estimate and confidence interval for the meta-analysis combining screening and three replication results are shown as a diamond. The size of the square box is proportional to the number of cases and controls in each study site. Heterogeneity across the two stages was examined using the Cochran Q test.

Figure S5

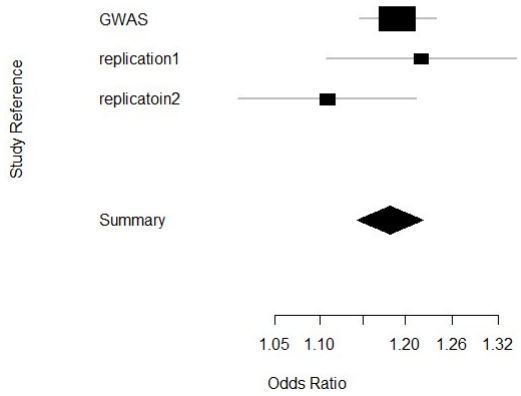
rs1057941 (1q22)

P_{het}=0.170



rs13361707 (5p13.1)

P_{het}= 0.291



rs2294008 (8q24.3)

P_{het} =0.003

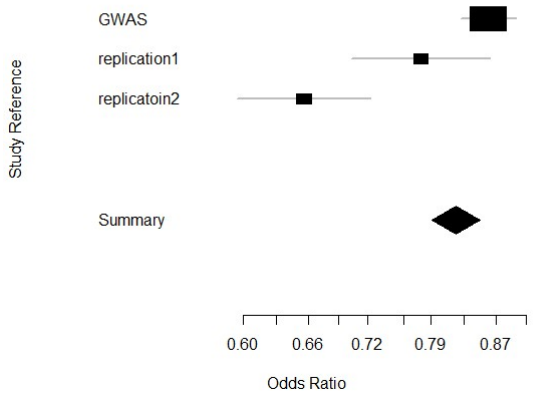
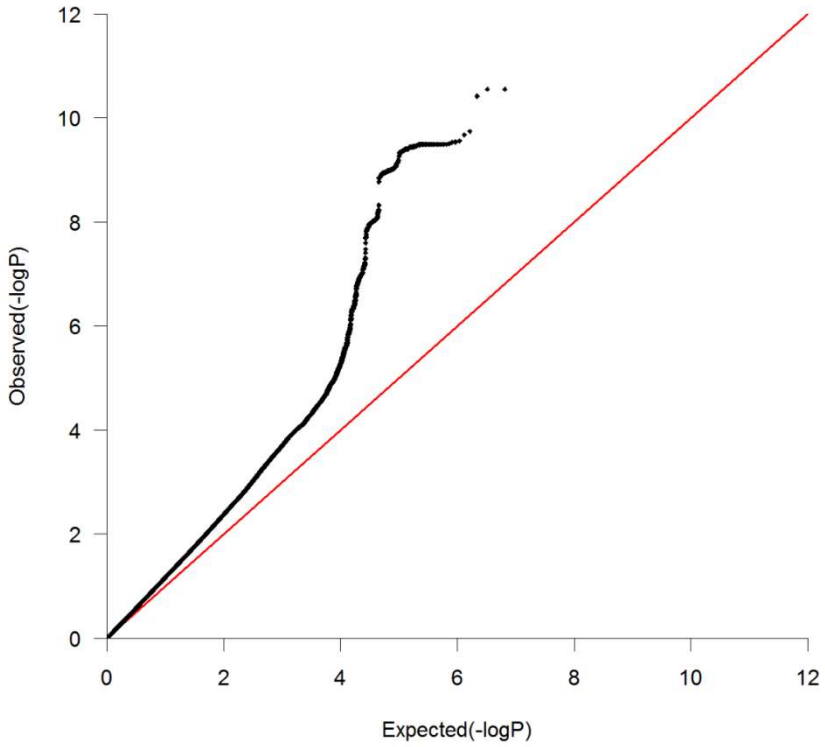


Figure S5 Forest plots for risk variants in the three previously identified gastric cancer risk loci. Plots show the association estimates (odds ratios) and 95% confidence intervals for the screening stage and two replication studies presented as bars. The association estimate and confidence interval for the meta-analysis combining screening and two replication results are shown as a diamond. The size of the square box is proportional to the number of cases and controls in each study site. Heterogeneity across the two stages was examined using the Cochran Q test.

Figure S6

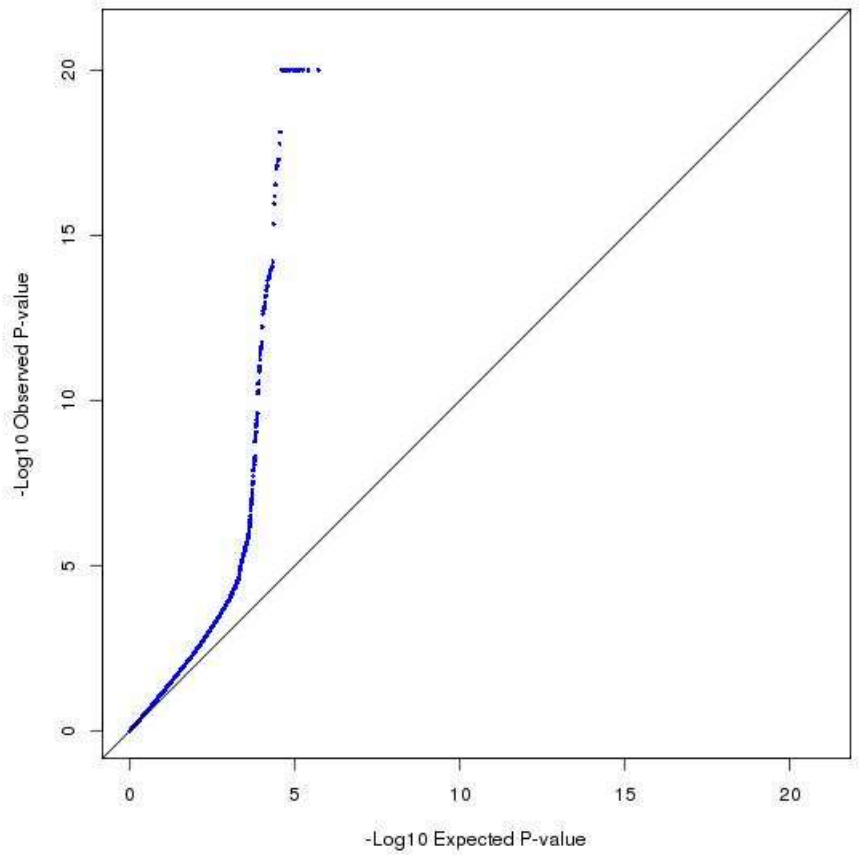


$\lambda = 1.2269$

$\lambda_{1000} = 1.0226$

Figure S6 Quantile quantile plot for imputed GWAS of gastric cancer in the Japanese population excluding three known loci; Chr1 (rs2070803±1Mb), Chr5 (rs13361707±1Mb), Chr8 (rs2294008±1Mb)).

Figure S7



λ = 1.2388

Figure S7 Quantile quantile plot for GWAS of gastric cancer in the Japanese population using 511,850 genotyped SNPs.

Figure S8

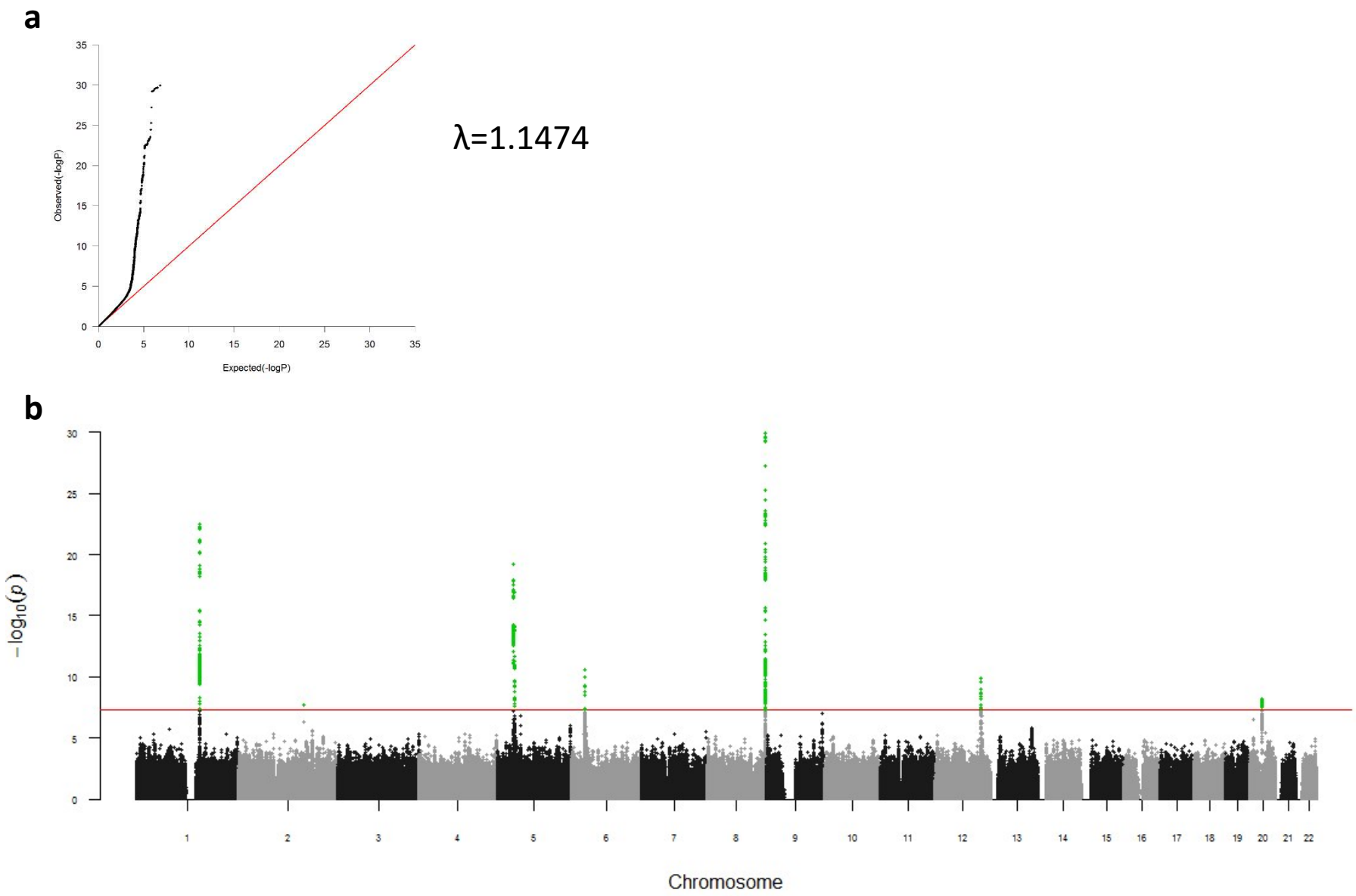
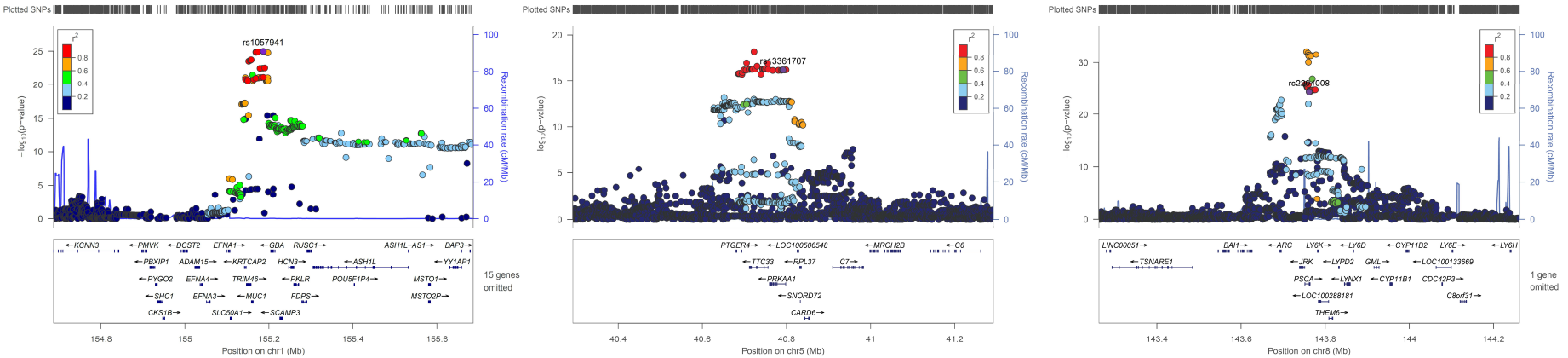


Figure S8 GWAS of gastric cancer using 10PC as covariates. (**a**) Quantile quantile plot for imputed GWAS of gastric cancer in the Japanese population. (**b**) Manhattan plot showing the genome-wide P values of association. The red horizontal line represents the genome-wide significance threshold of $P = 5.0 \times 10^{-8}$.

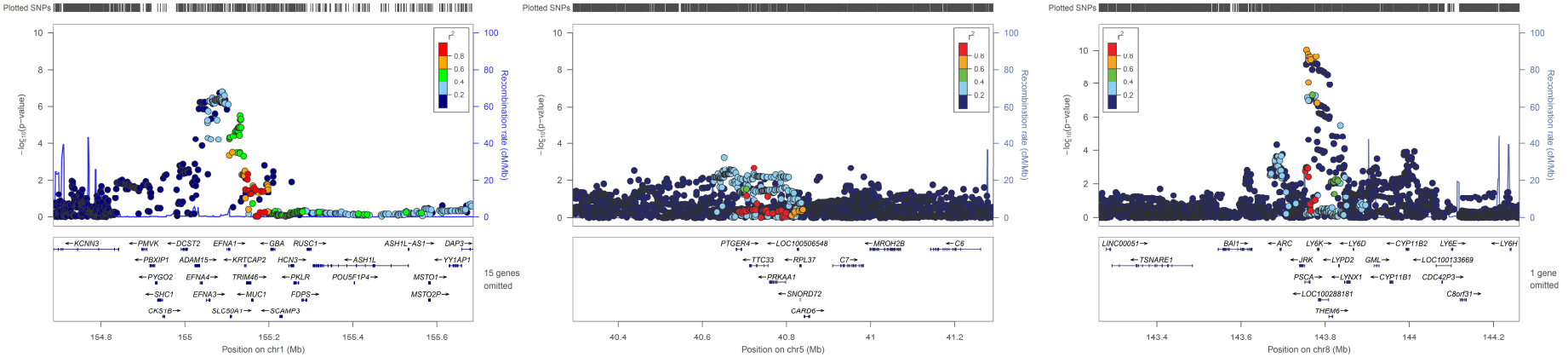
Figure S9

a

rs1057941 (1q22) | rs13361707 (5p13.1) | rs2294008 (8q24.3)



before condition | before condition | before condition



after condition | after condition | after condition

b

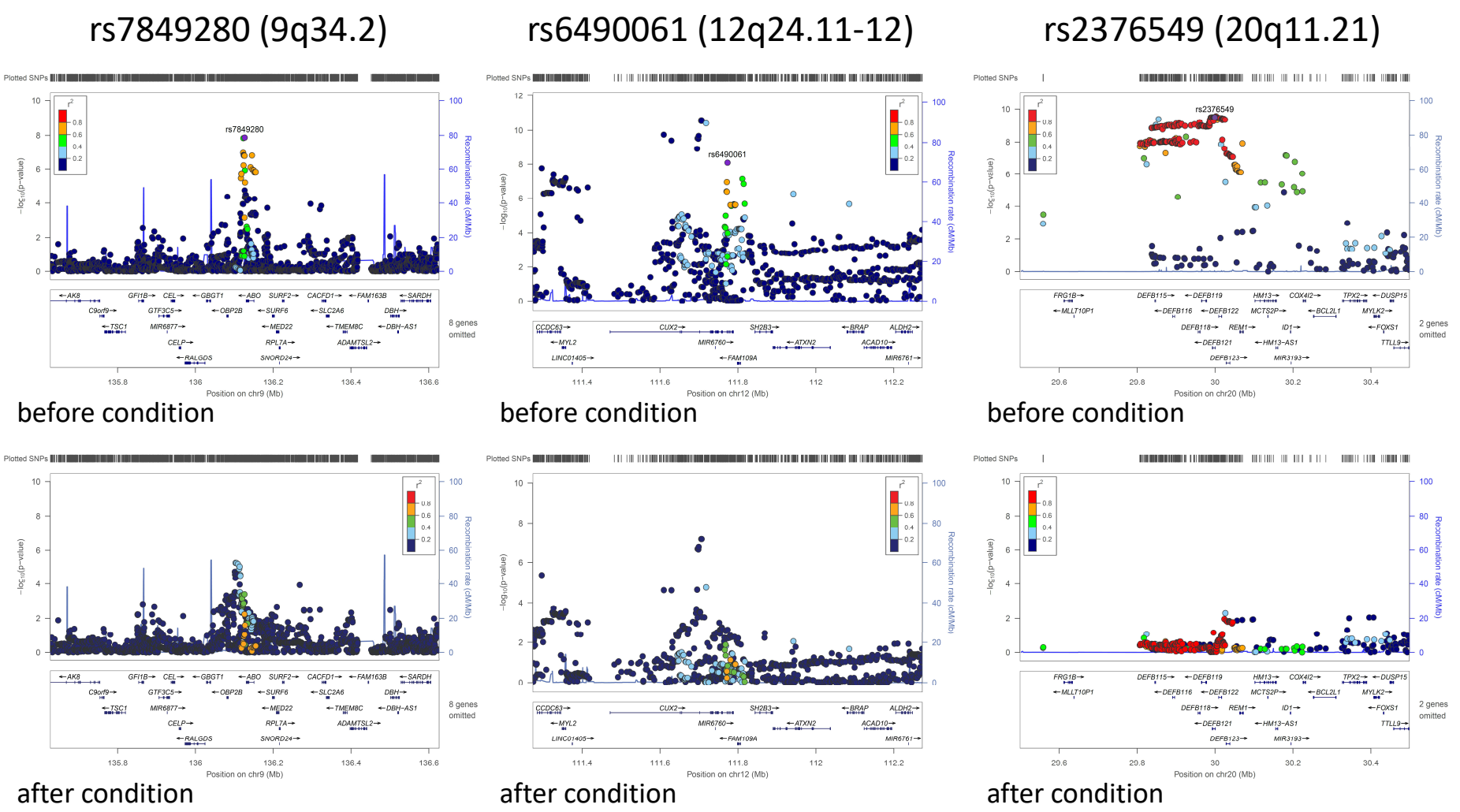


Figure S9 Regional plots of six loci for gastric cancer before and after condition. The $-\log_{10}$ *P* values from the screening stage in reported three loci (**a**) and novel three loci (**b**) are shown. Upper and lower panels are before and after condition respectively using SNPs shown in each panal. Estimated recombination rates (from 1000 Genomes) are plotted in blue. The SNPs are color-coded to reflect their correlation with genotyped SNPs. Pairwise r2 values are from 1000 Genomes East Asian data (March 2012 release). Genes, position of exons and direction of transcription from UCSC genome browser (genome.ucsc.edu) are noted. Plots were generated using LocusZoom (http://csg.sph.umich.edu/locuszoom).

Figure S10

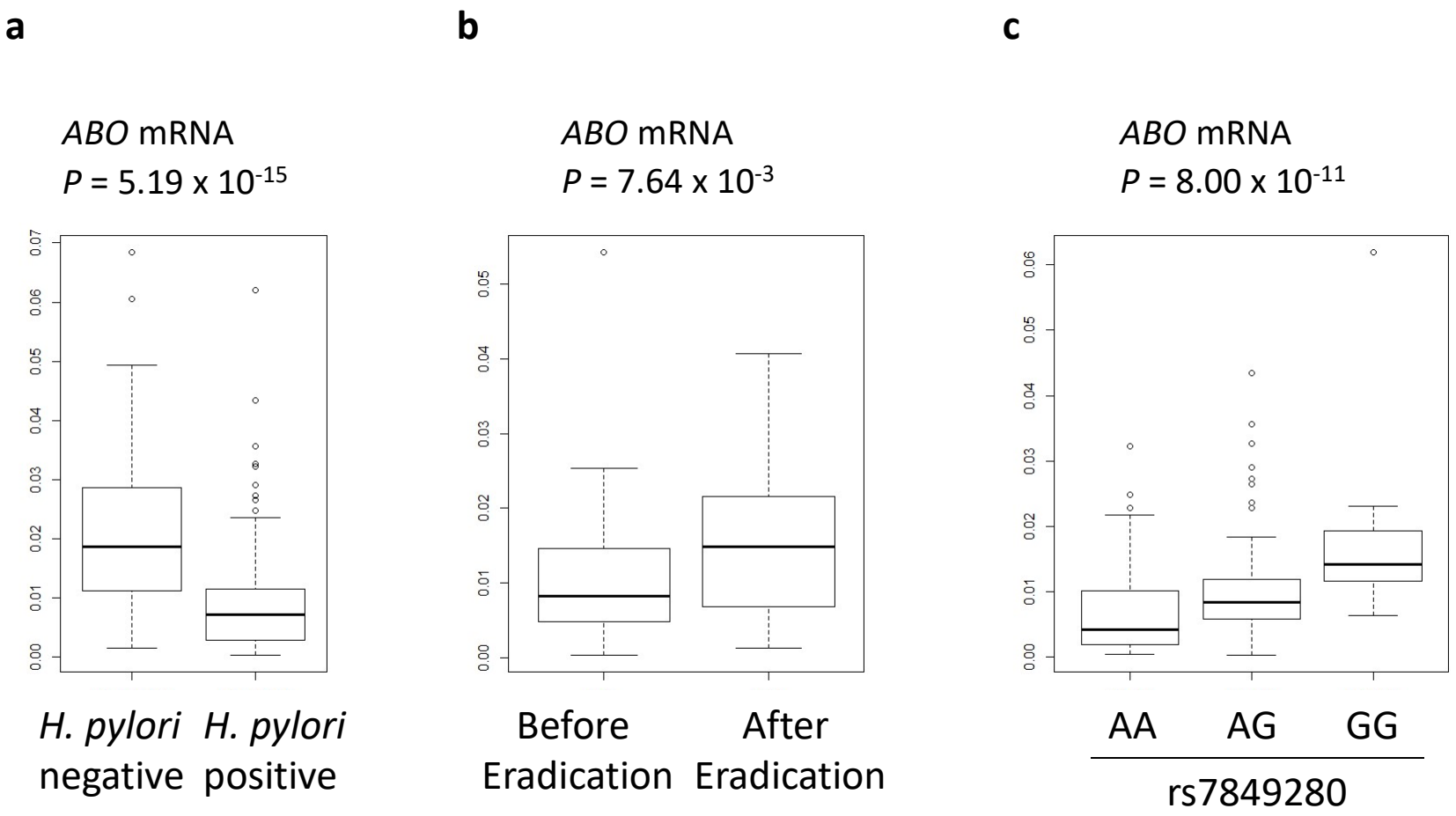


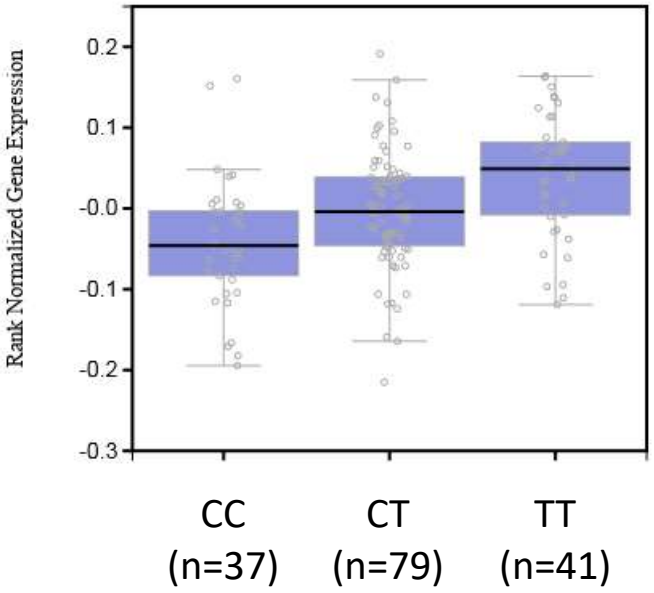
Figure S10 The regulation of *ABO* expression by *H.plylori* infection. Box-plots indicate the qRT-PCR analysis of *ABO* mRNA levels in biopsy samples from background gastric mucosa. The vertical axis indicates the expression level of *ABO* normalized against *GAPDH* expression. Box = 25th and 75th percentiles; middle line in the box = median; whiskers = min value inside 25th percentile- 1.5*inter quartile range and max value inside 75th percentile+ 1.5*inter quartile range; points = outliers. (**a**) *ABO* mRNA in *H. pylori*-negative controls (n = 28) and *H. pylori*-infected patients (n = 280). (**b**) The *ABO* expression levels in each individual patients (n = 53) before and after *H. pylori* eradication. (**c**) Association of rs7849280 with *ABO* expression in *H. pylori*-infected patients (AA; n = 160, AG; n = 103, GG; n = 17). The *P* values were calculated by t-test (**a**)(**b**) or Kruskal-Wallis test (**c**).

Figure S11

a

DEFB121 expression in testis

$P = 2.1 \times 10^{-8}$



b

DEFB119 expression in esophagus

$P = 2.1 \times 10^{-8}$

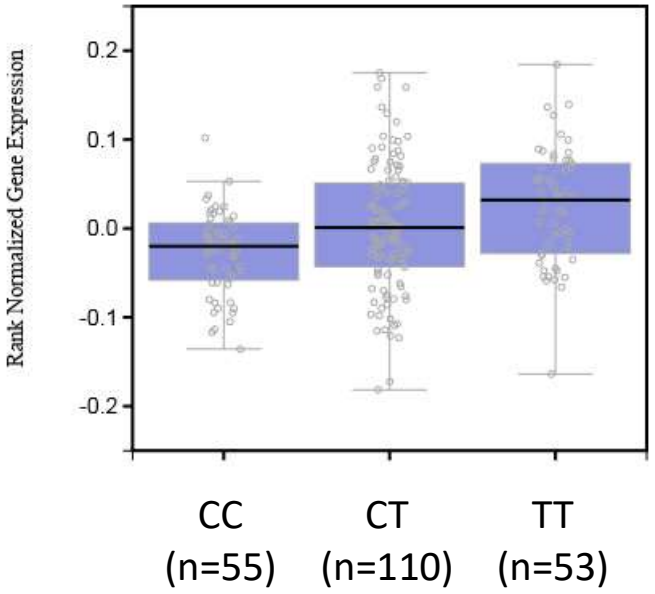


Figure S11 eQTL analysis for rs2376549. Data were derived from GTEx portal (https://www.gtexportal.org/home/). (**a**) Association of rs2376549 with *DEFB121* expression in testis. (**b**) Association of rs2376549 with *DEFB119* expression in esophagus.

Figure S12

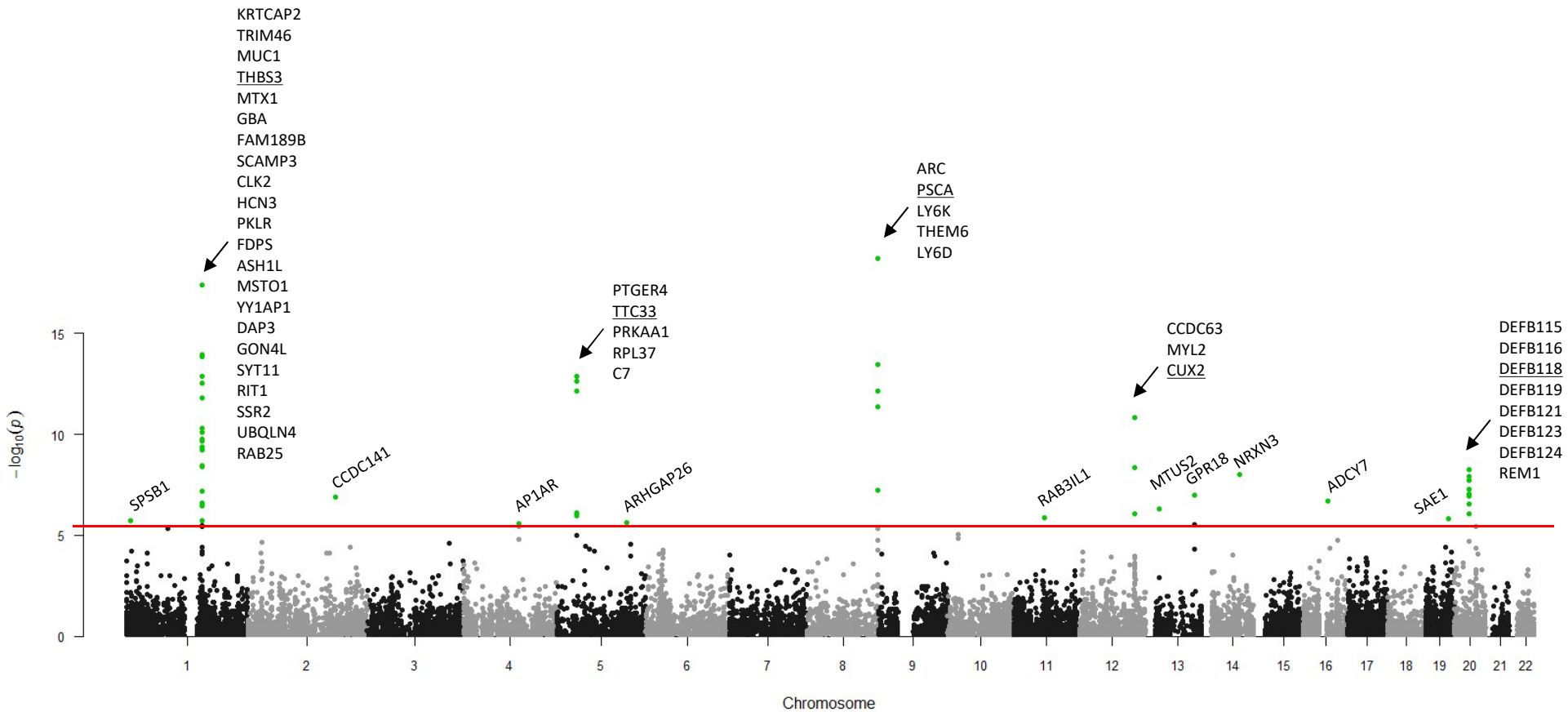


Figure S12 Manhattan plot showing the genome-wide *P* values of association. The genome-wide *P*-values of 17,599 genes in 6,171 cases and 27,178 controls from the screening phase are shown. 53 genes are identified risk genes and underlined are most significant genes in each region. The red horizontal line represents the genome-wide significance threshold of $P = 2.84 \times 10^{-6}$.

Figure S13

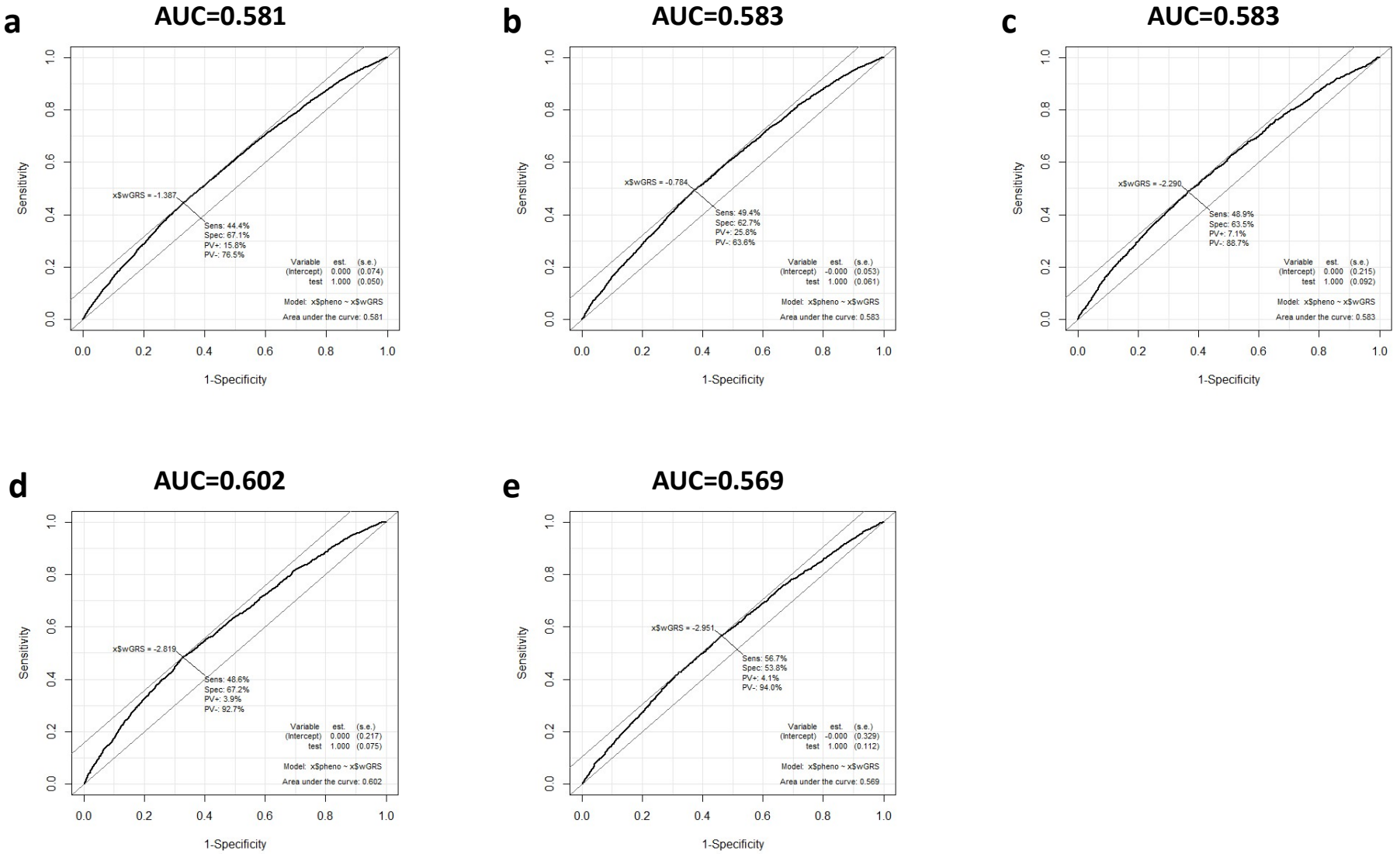


Figure S13 Risk-prediction model using 6 significant SNPs. ROC curve for wGRS using 6 genetic variations (rs1057941, rs13361707, rs2294008, rs7849280, rs6490061, and rs2376549) at the screening stage. (**a**) Total gastric cancer, (**b**) Male gastric cancer, (**c**) Female Gastric cancer, (**d**) diffuse type gastric cancer, and (**e**) intestinal gastric cancer.