

Supplemental Data

Multiethnic GWAS Reveals Polygenic

Architecture of Earlobe Attachment

John R. Shaffer, Jinxi Li, Myoung Keun Lee, Jasmien Roosenboom, Ekaterina Orlova, Kaustabh Adhikari, 23andMe Research Team, Carla Gallo, Giovanni Poletti, Lavinia Schuler-Faccini, Maria-Cátira Bortolini, Samuel Canizales-Quinteros, Francisco Rothhammer, Gabriel Bedoya, Rolando González-José, Paige E. Pfeiffer, Christopher A. Wollenschlaeger, Jacqueline T. Hecht, George L. Wehby, Lina M. Moreno, Anan Ding, Li Jin, Yajun Yang, Jenna C. Carlson, Elizabeth J. Leslie, Eleanor Feingold, Mary L. Marazita, David A. Hinds, Timothy C. Cox, Sijia Wang, Andrés Ruiz-Linares, and Seth M. Weinberg

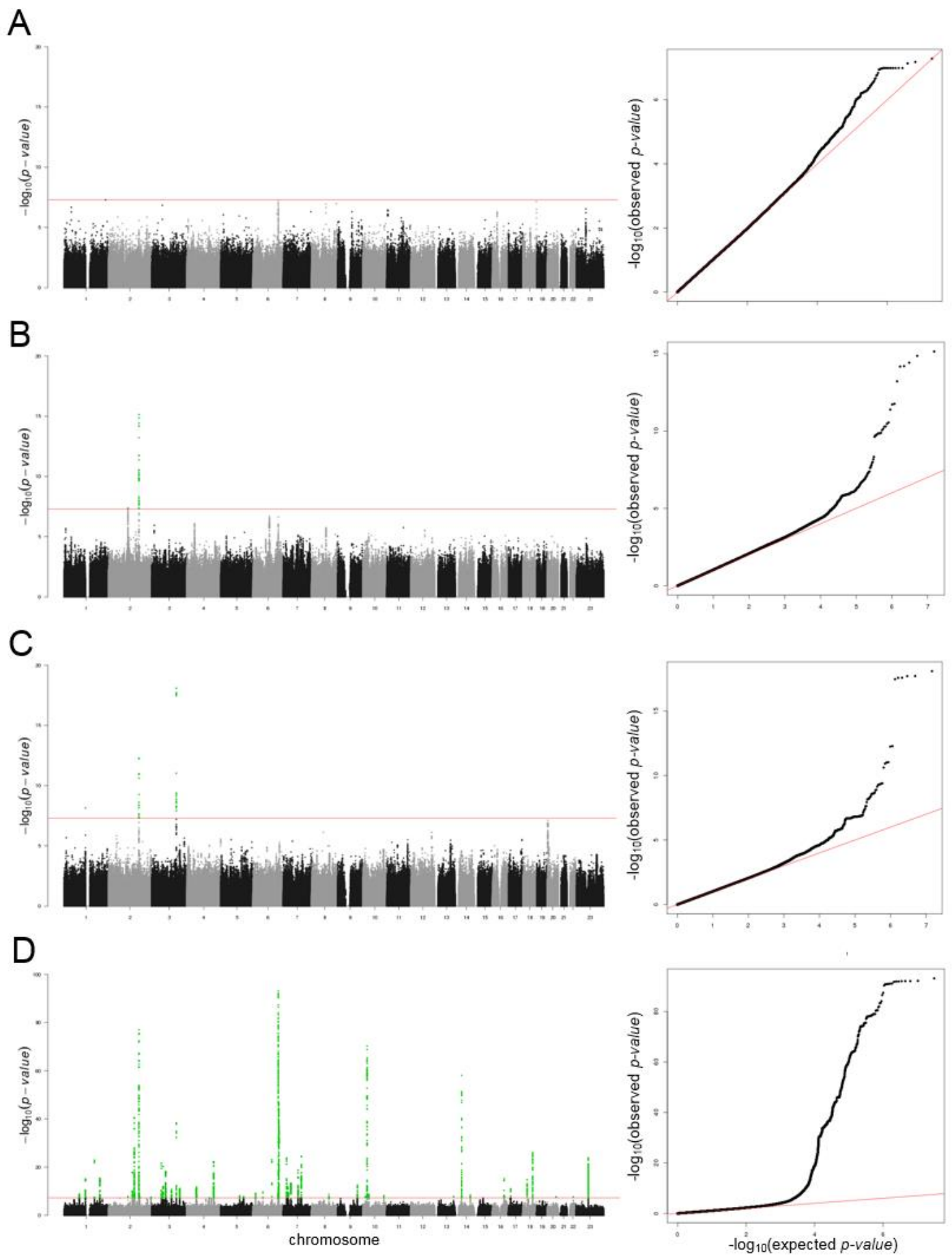
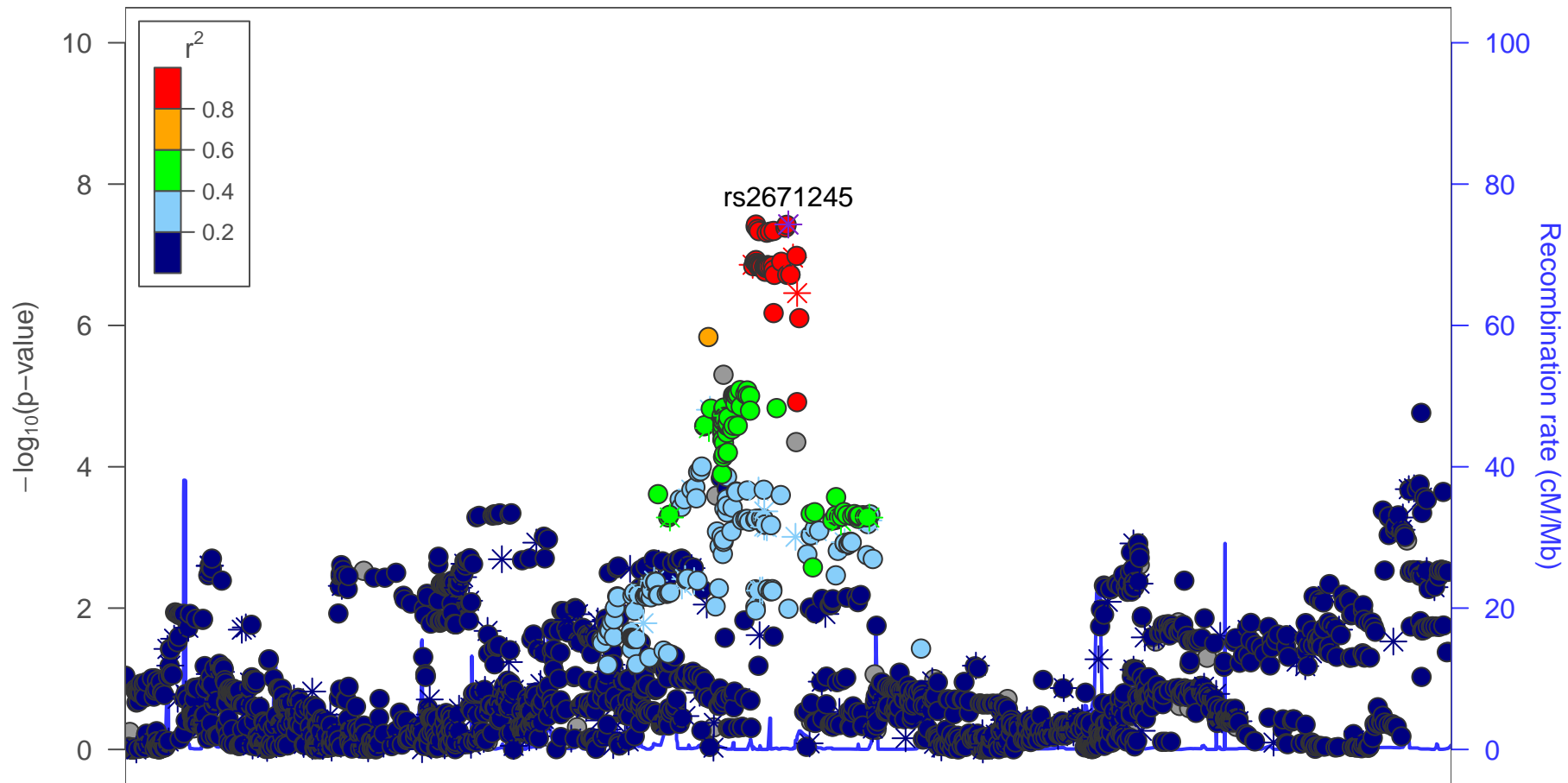


Figure S1: GWAS results shown as Manhattan (left) and quantile-quantile (right) plots for the (A) European American, (B) Latin American, (C) Chinese, and (D) 23andMe cohorts.



55.8

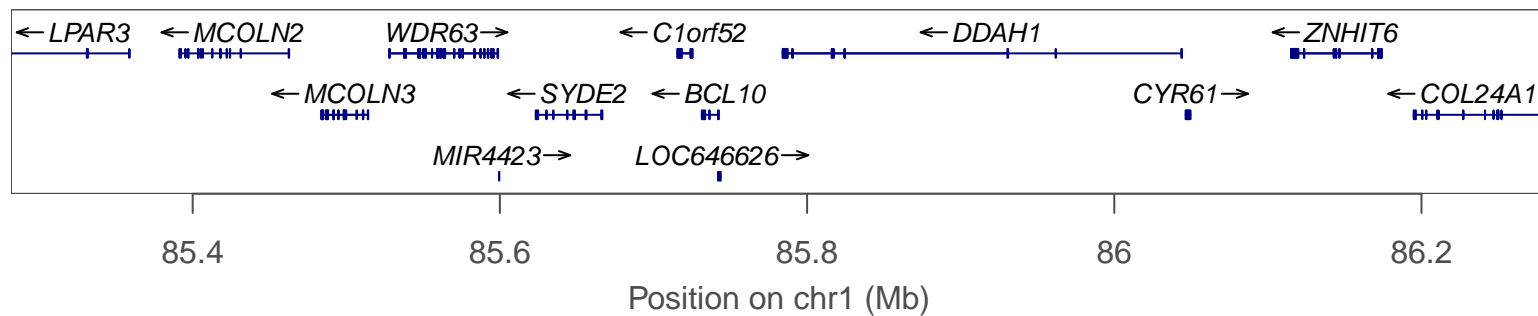
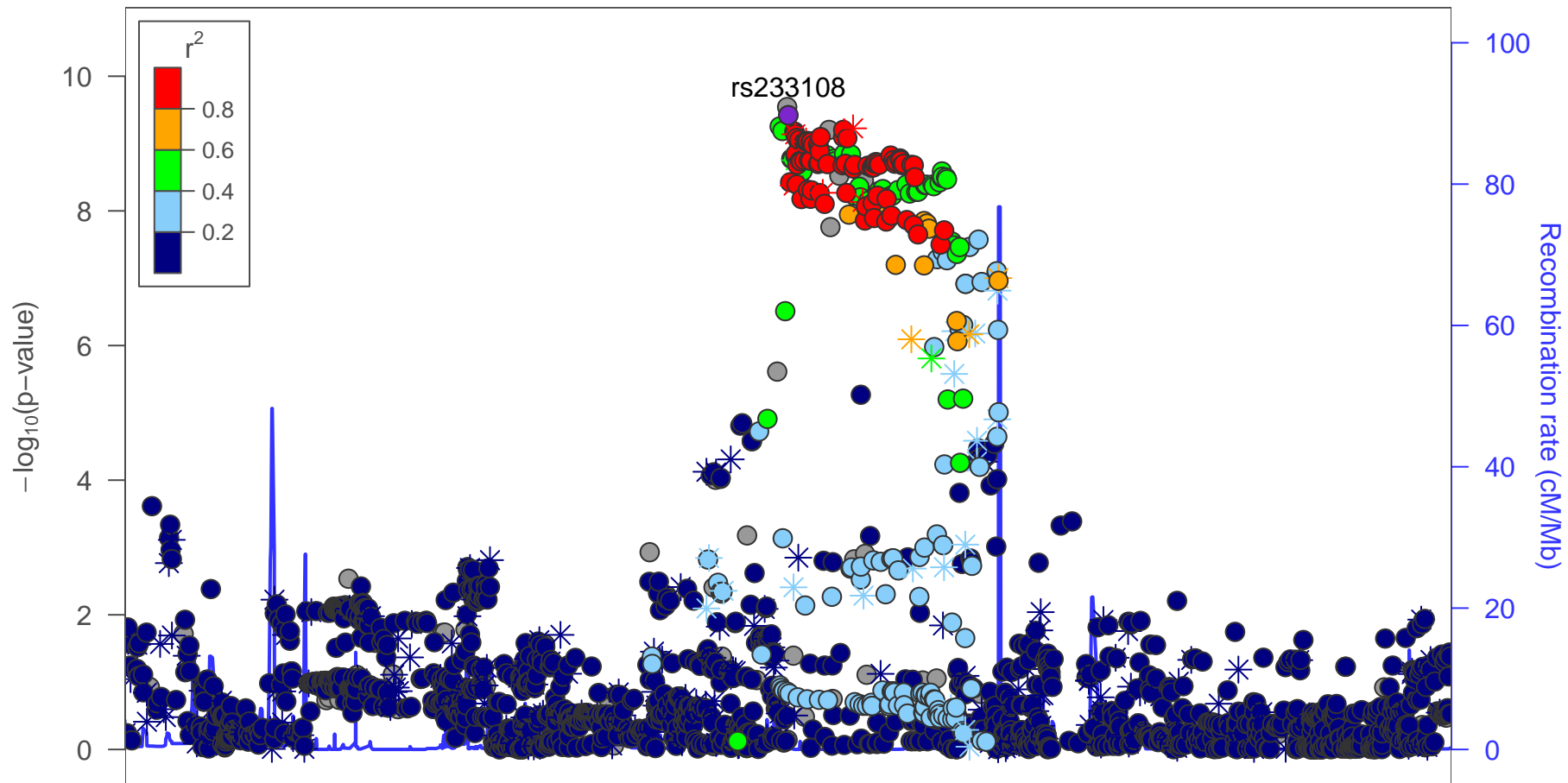
56

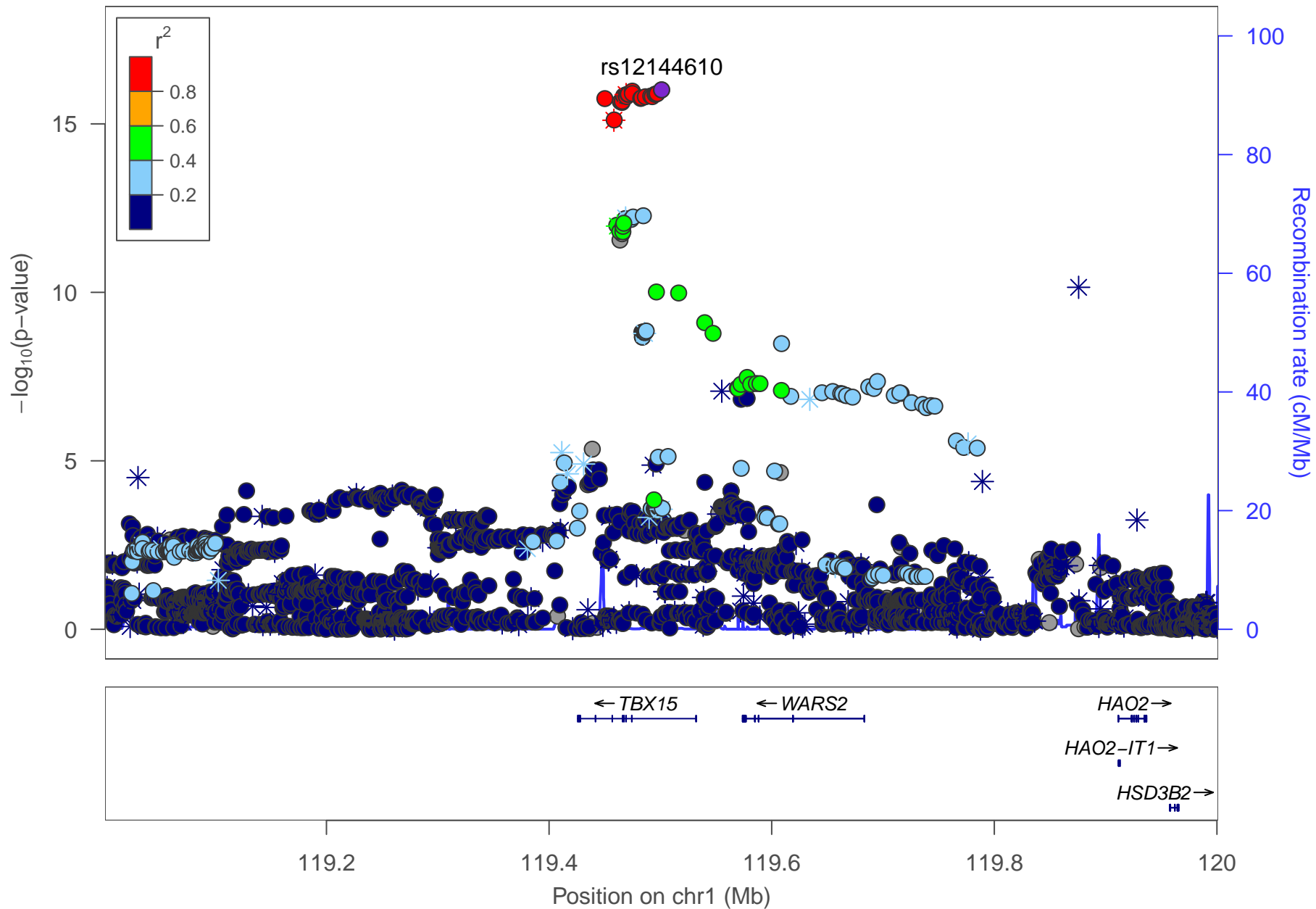
56.2

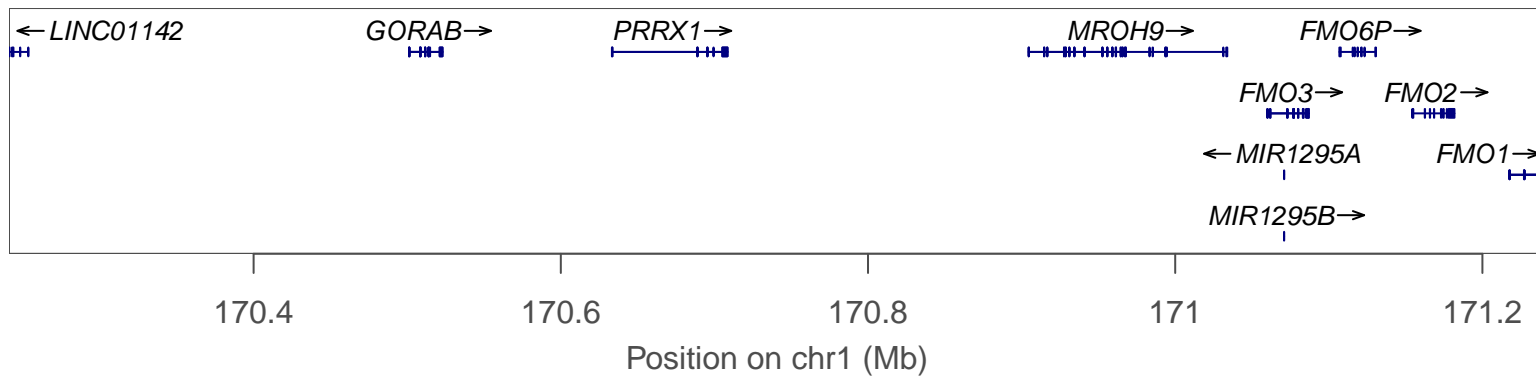
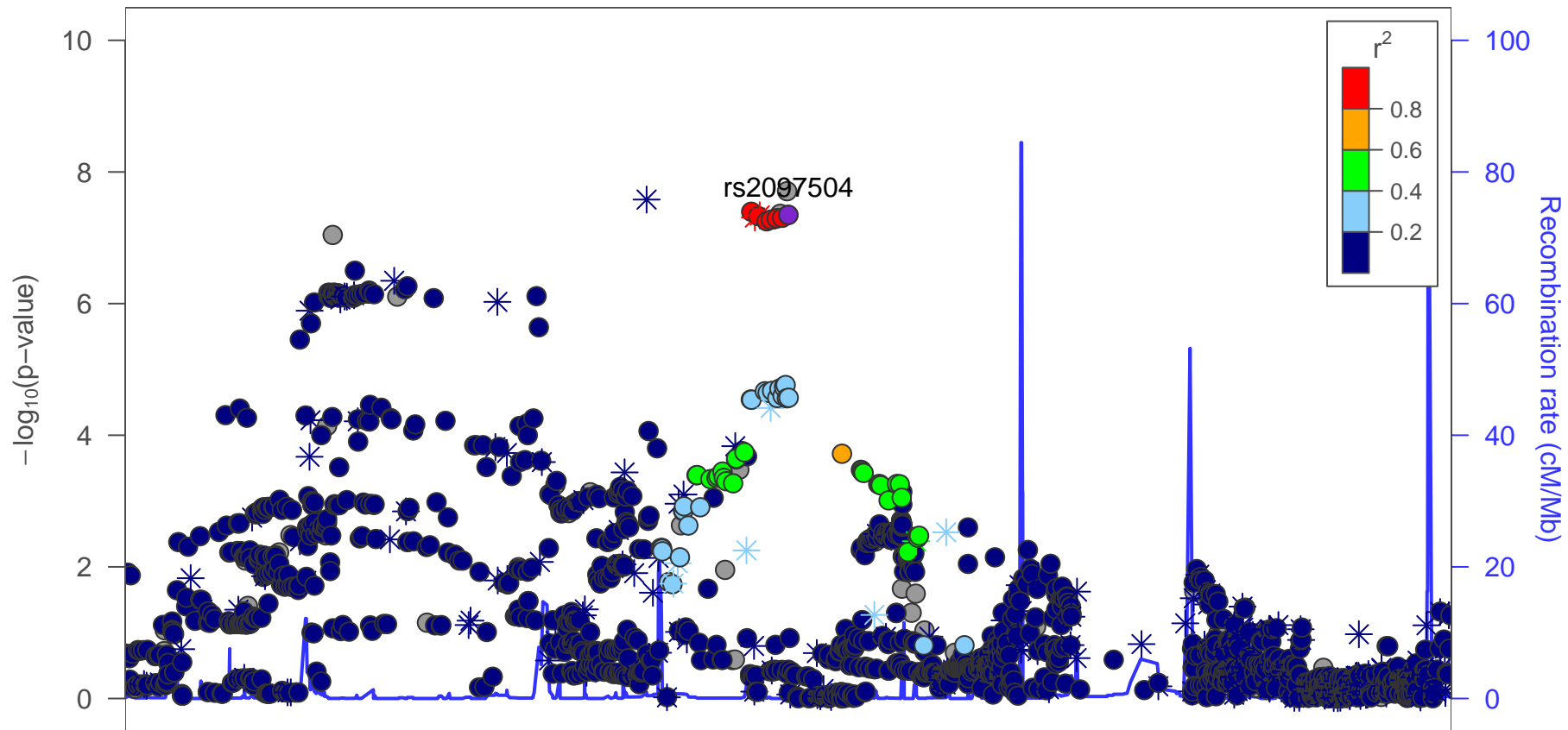
56.4

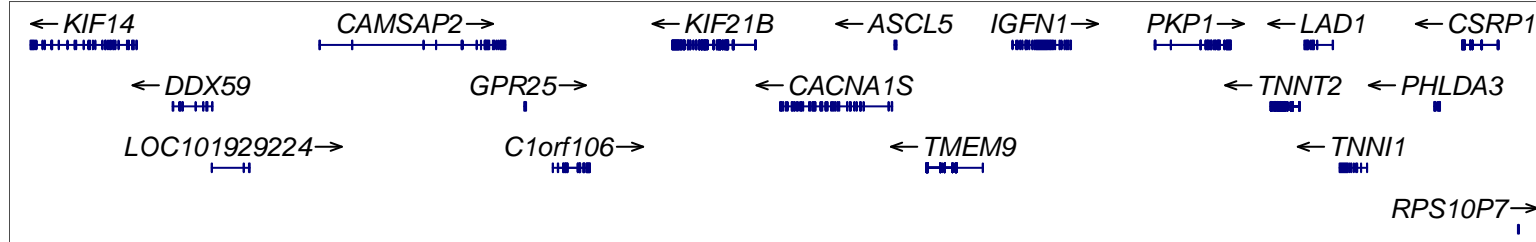
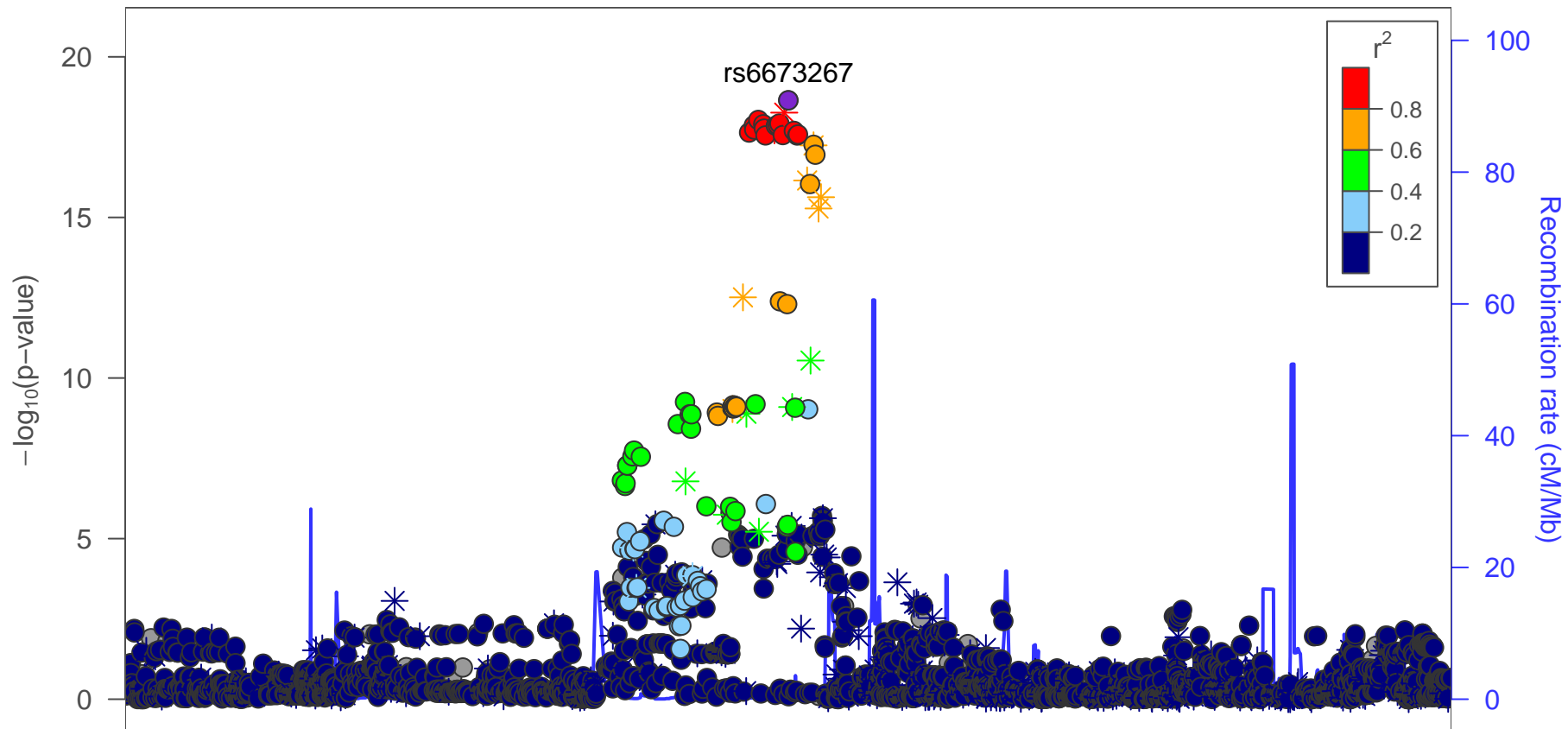
56.6

Position on chr1 (Mb)



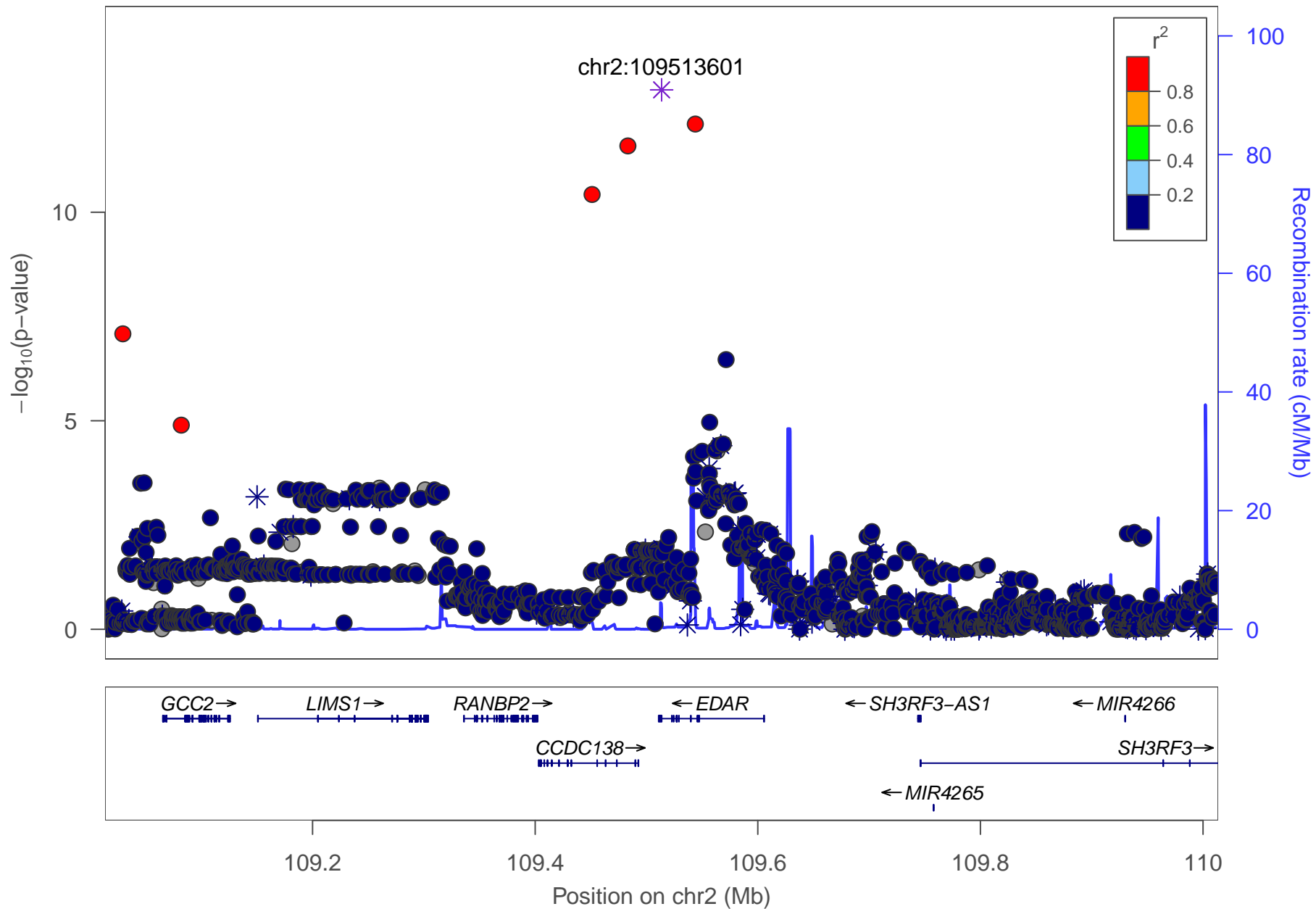


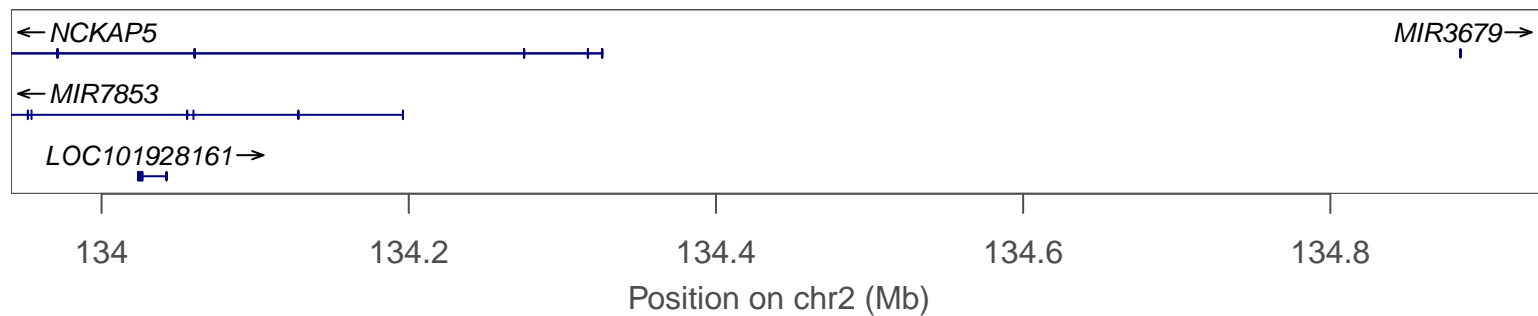
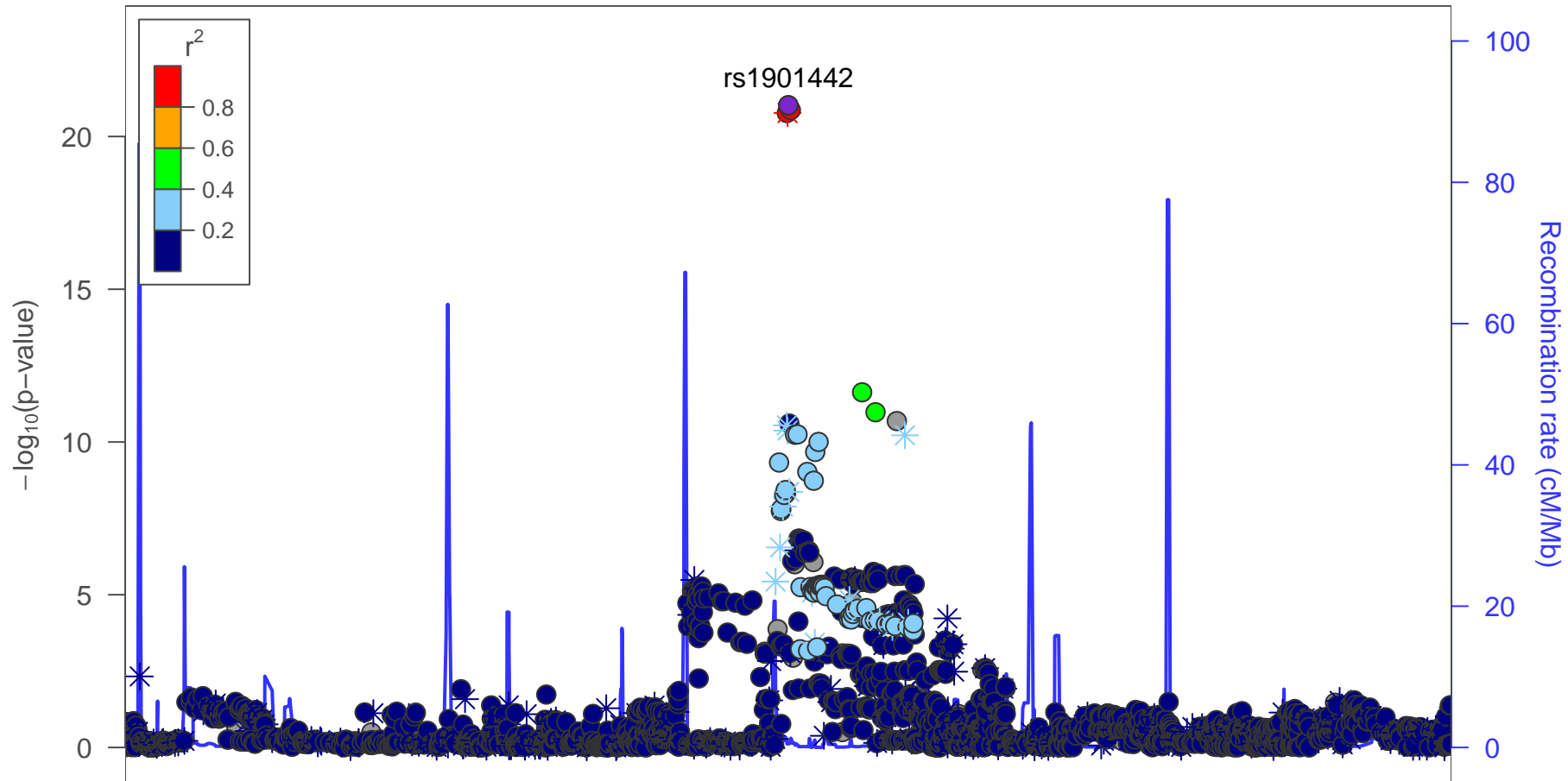


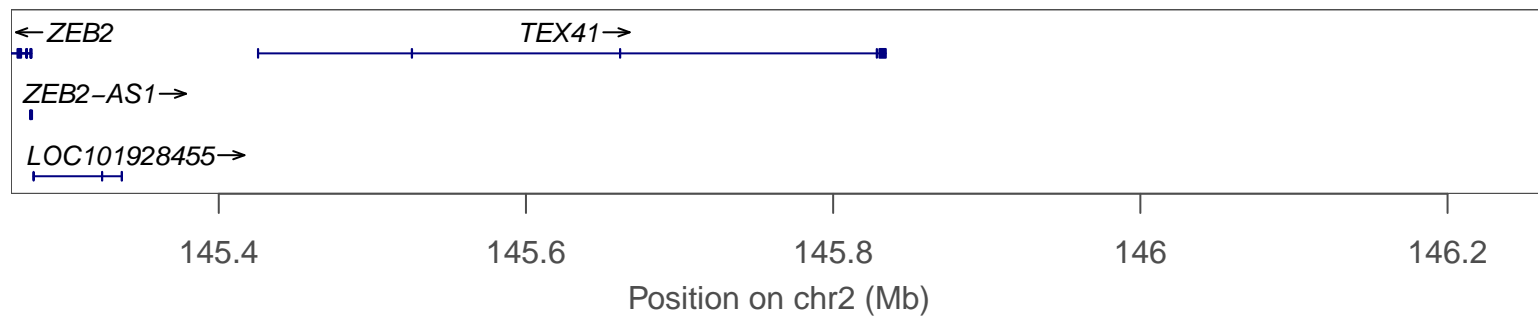
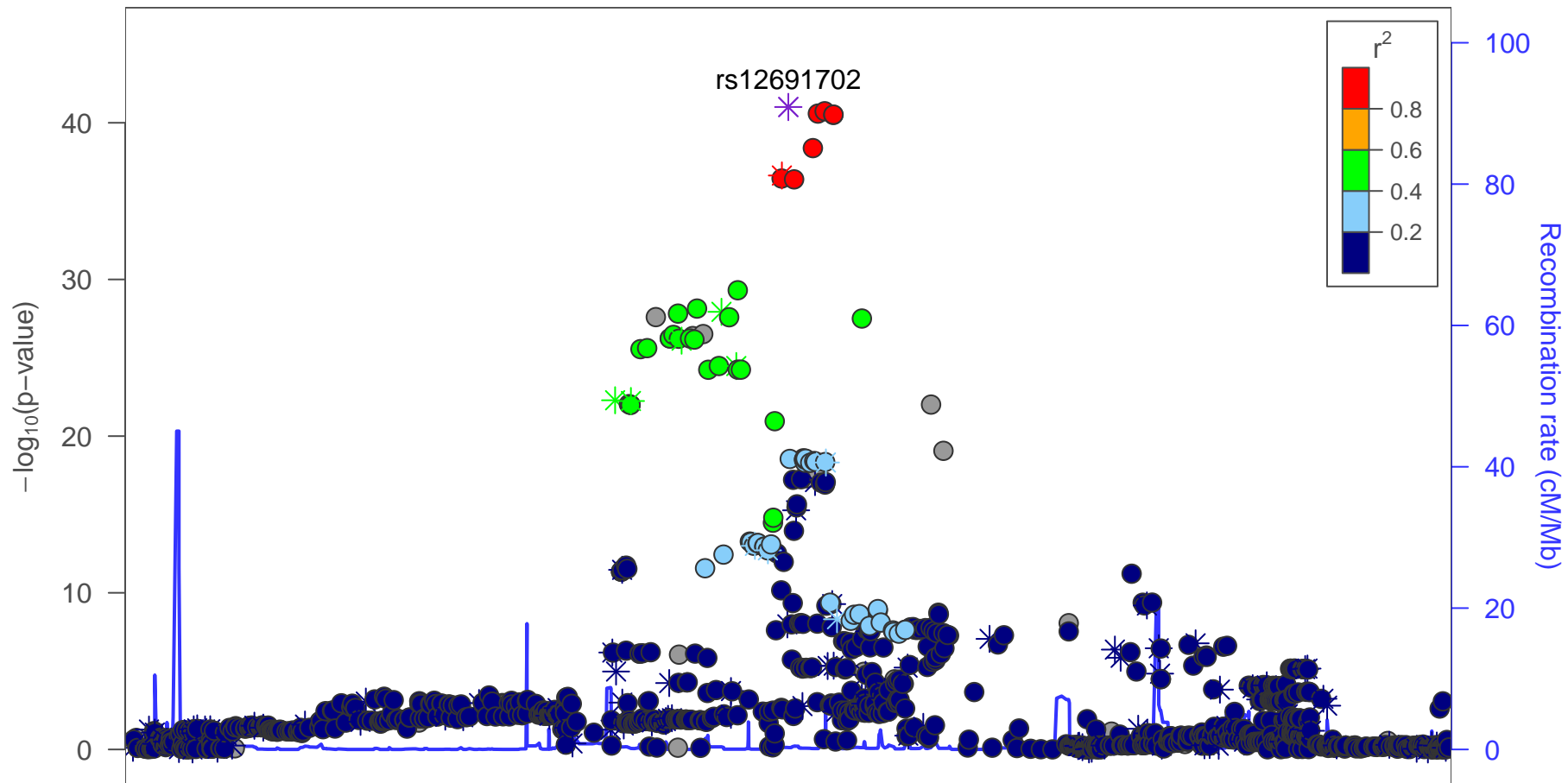


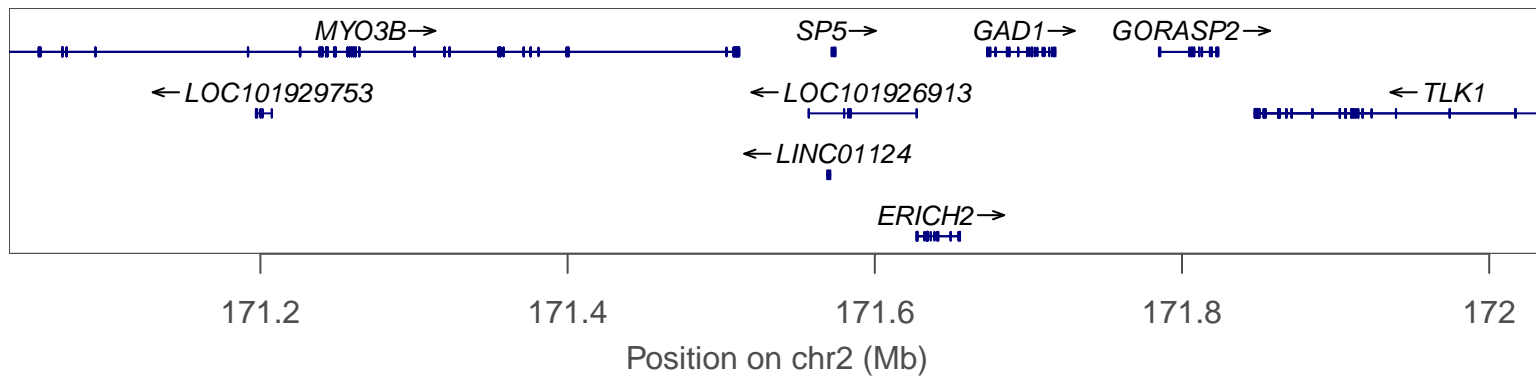
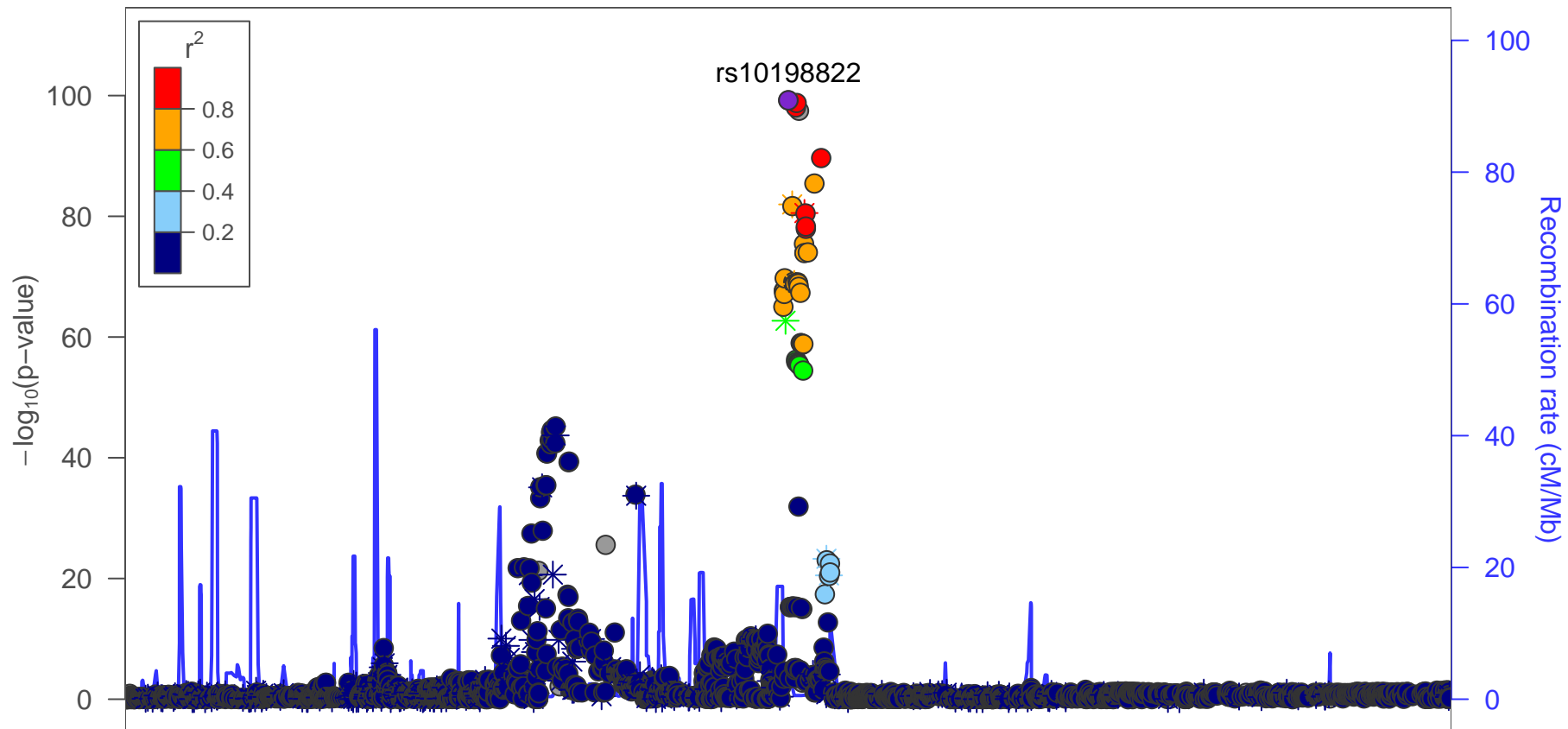
200.6 200.8 201 201.2 201.4

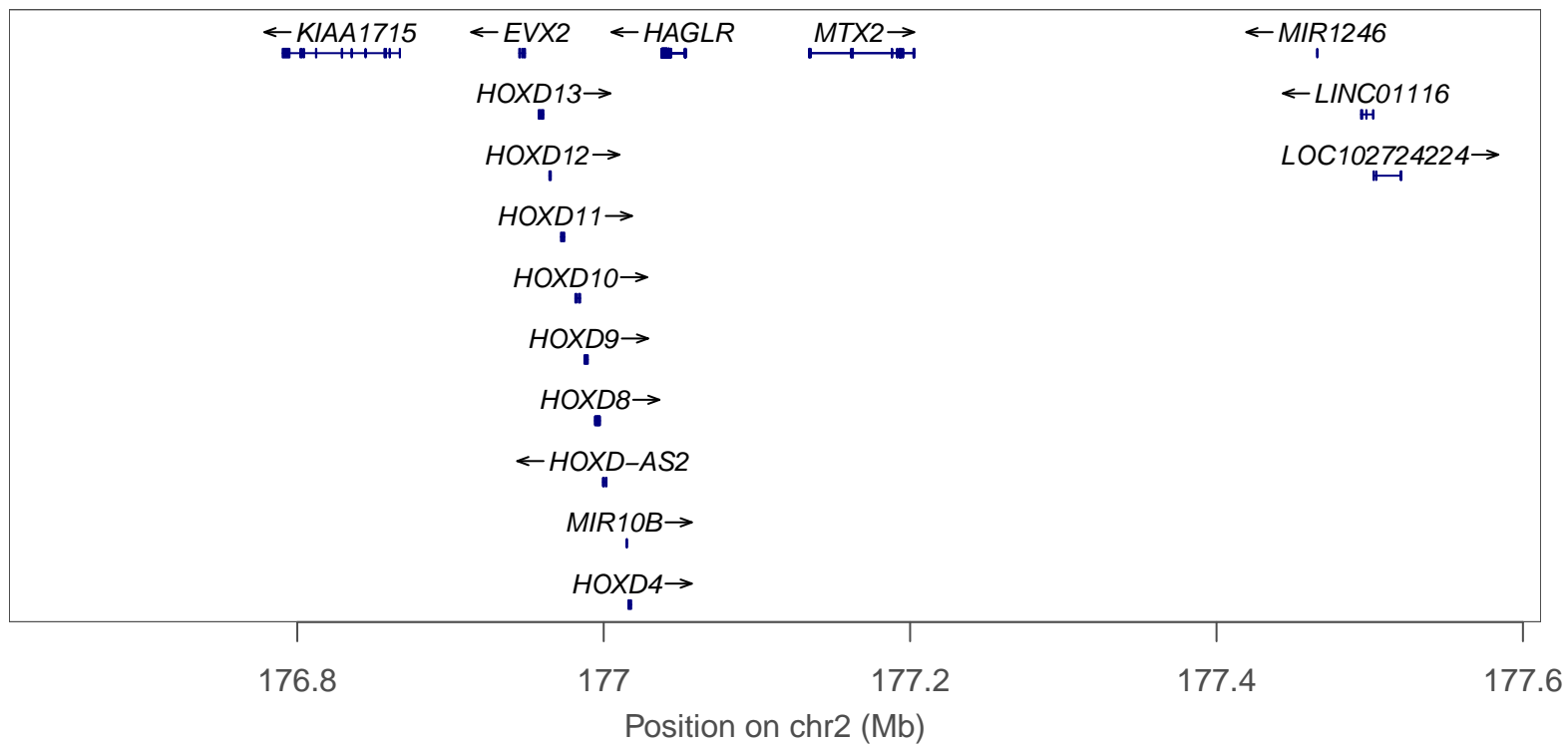
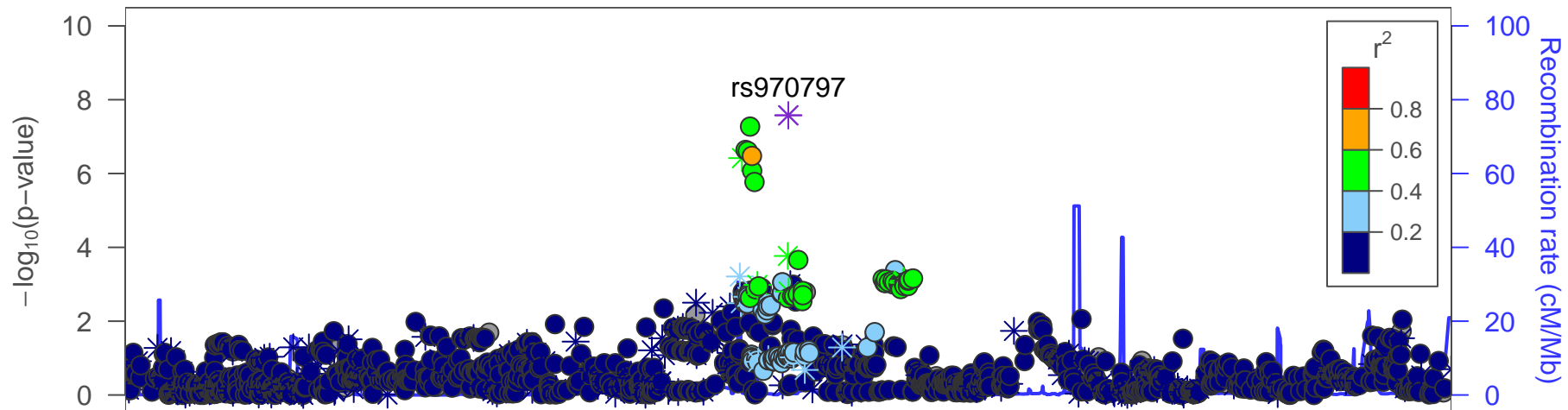
Position on chr1 (Mb)

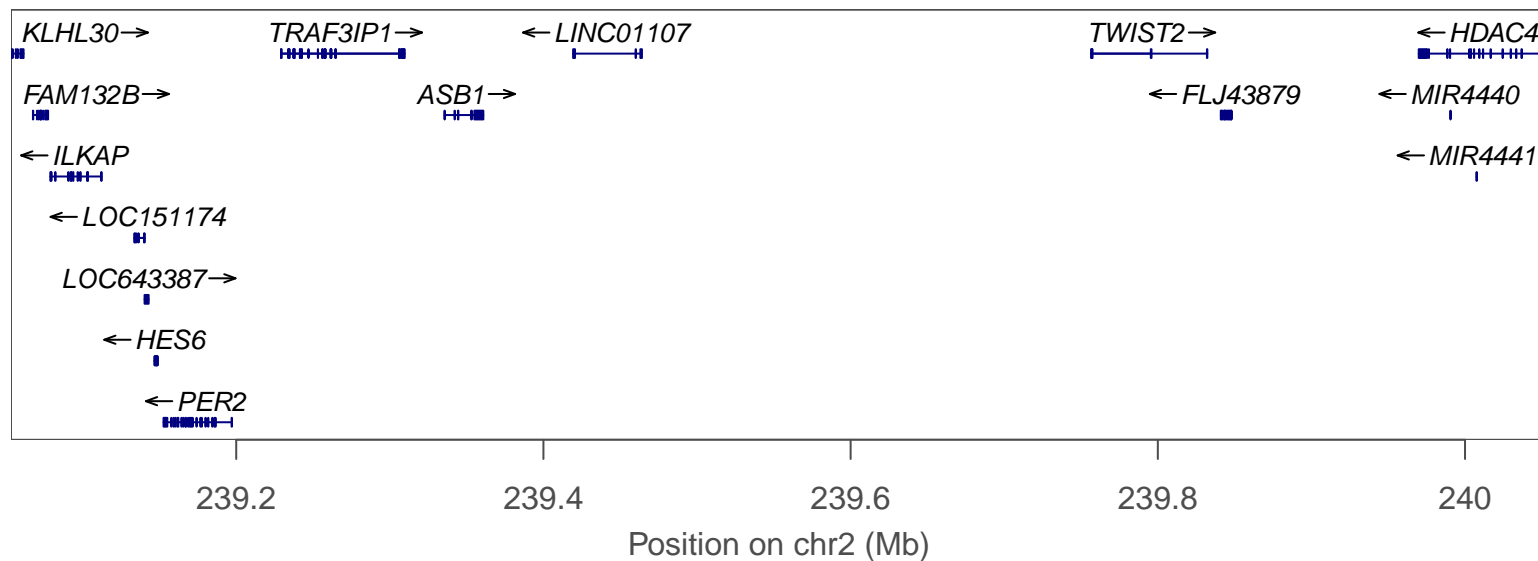
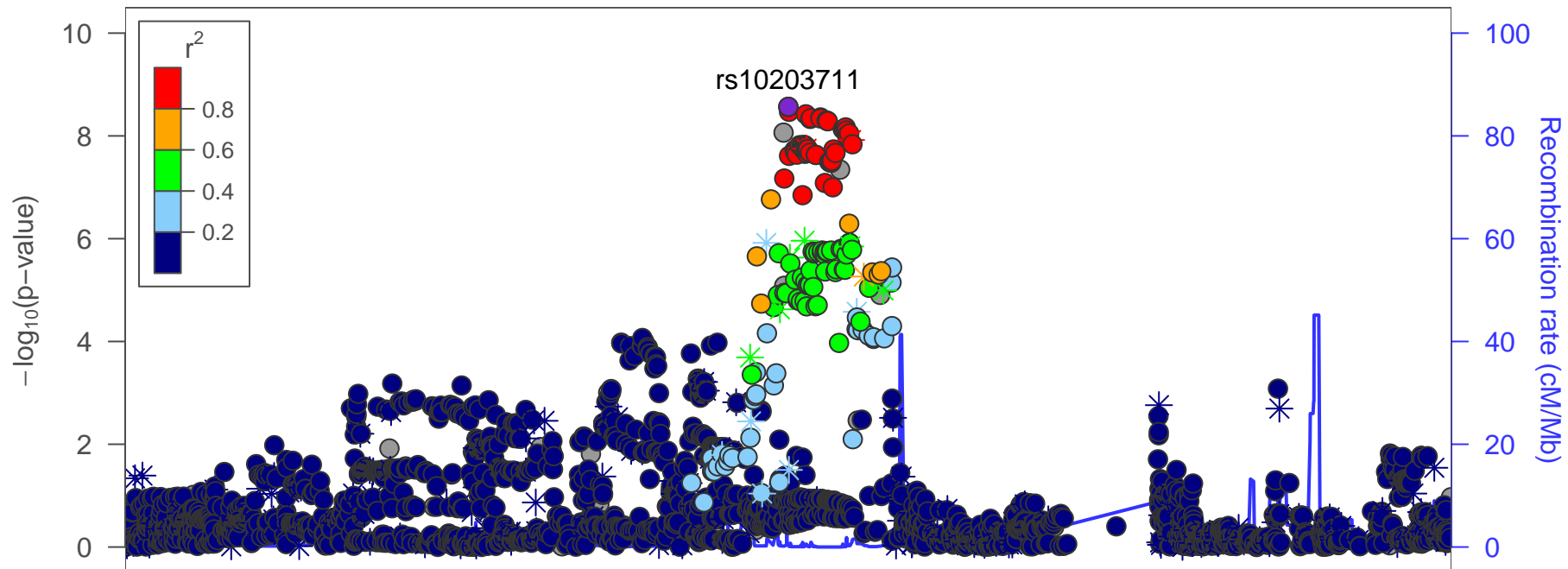


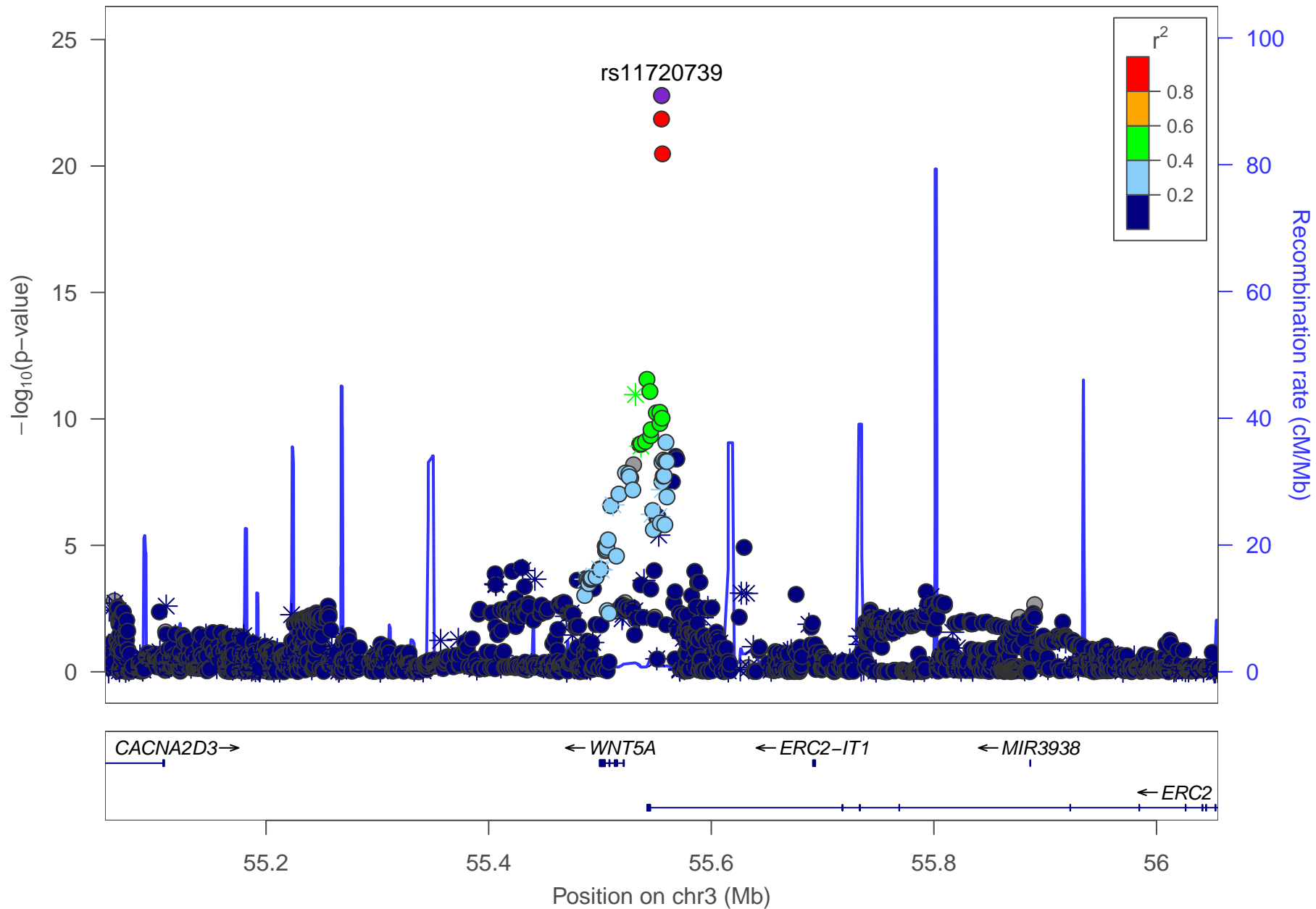


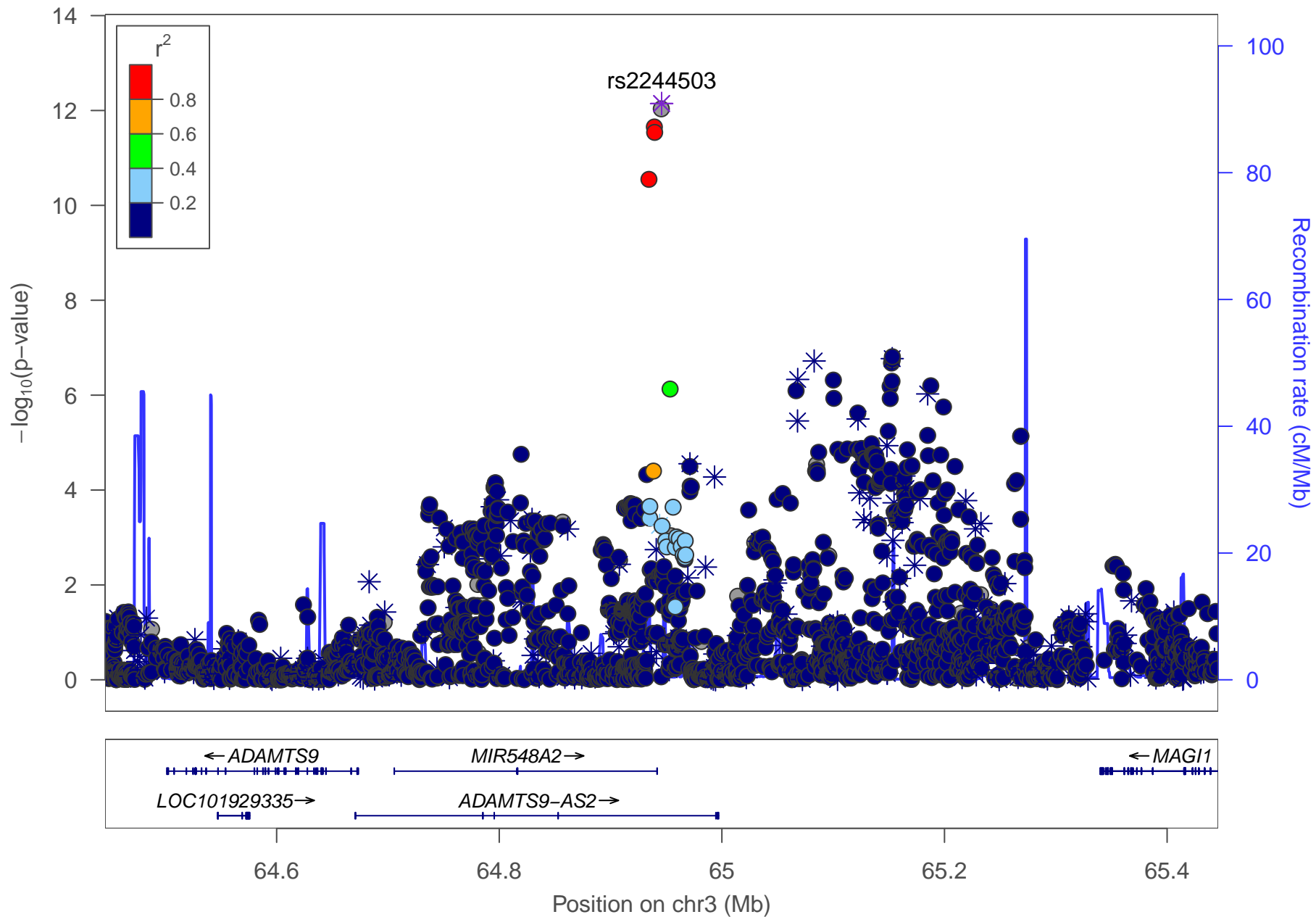


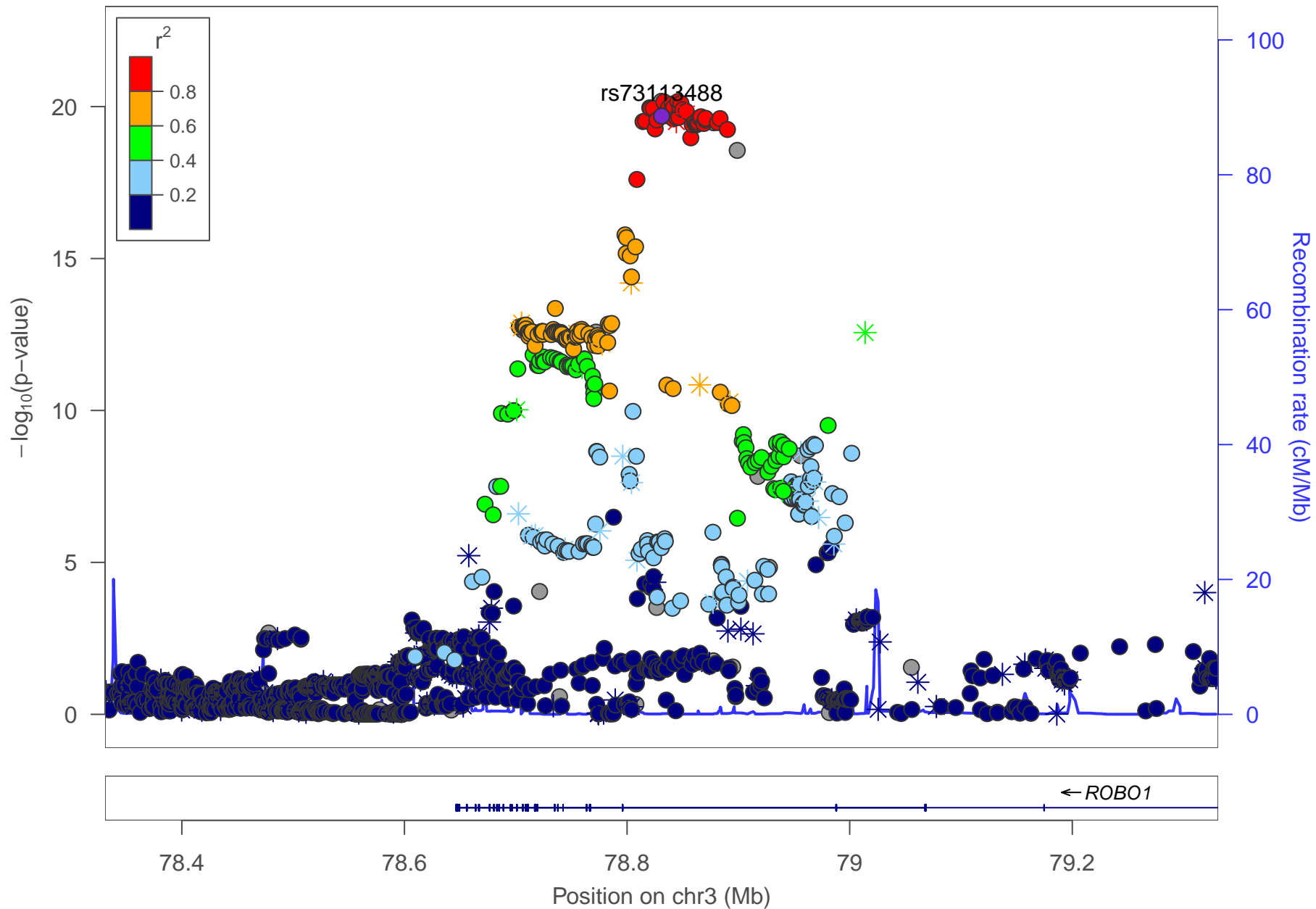


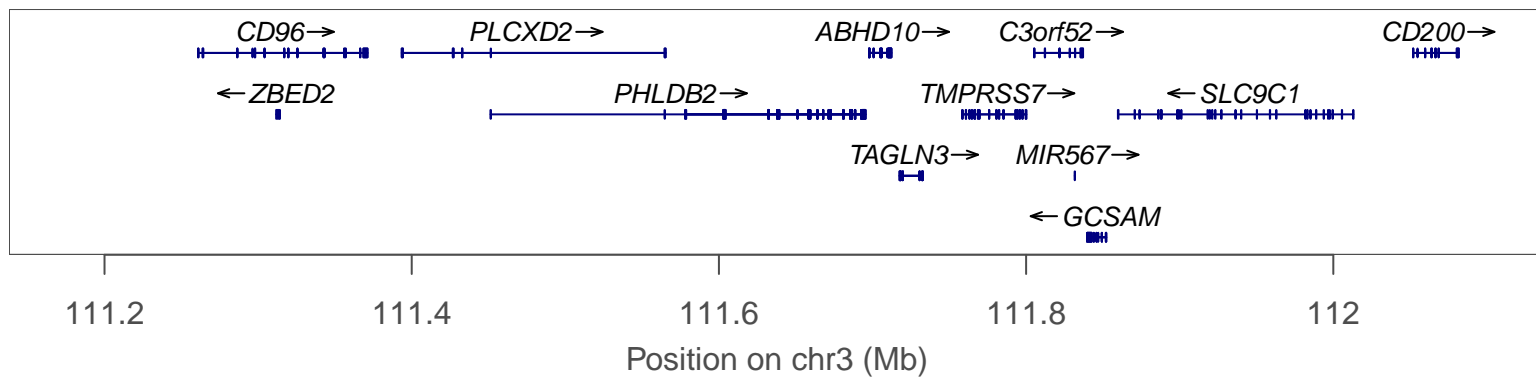
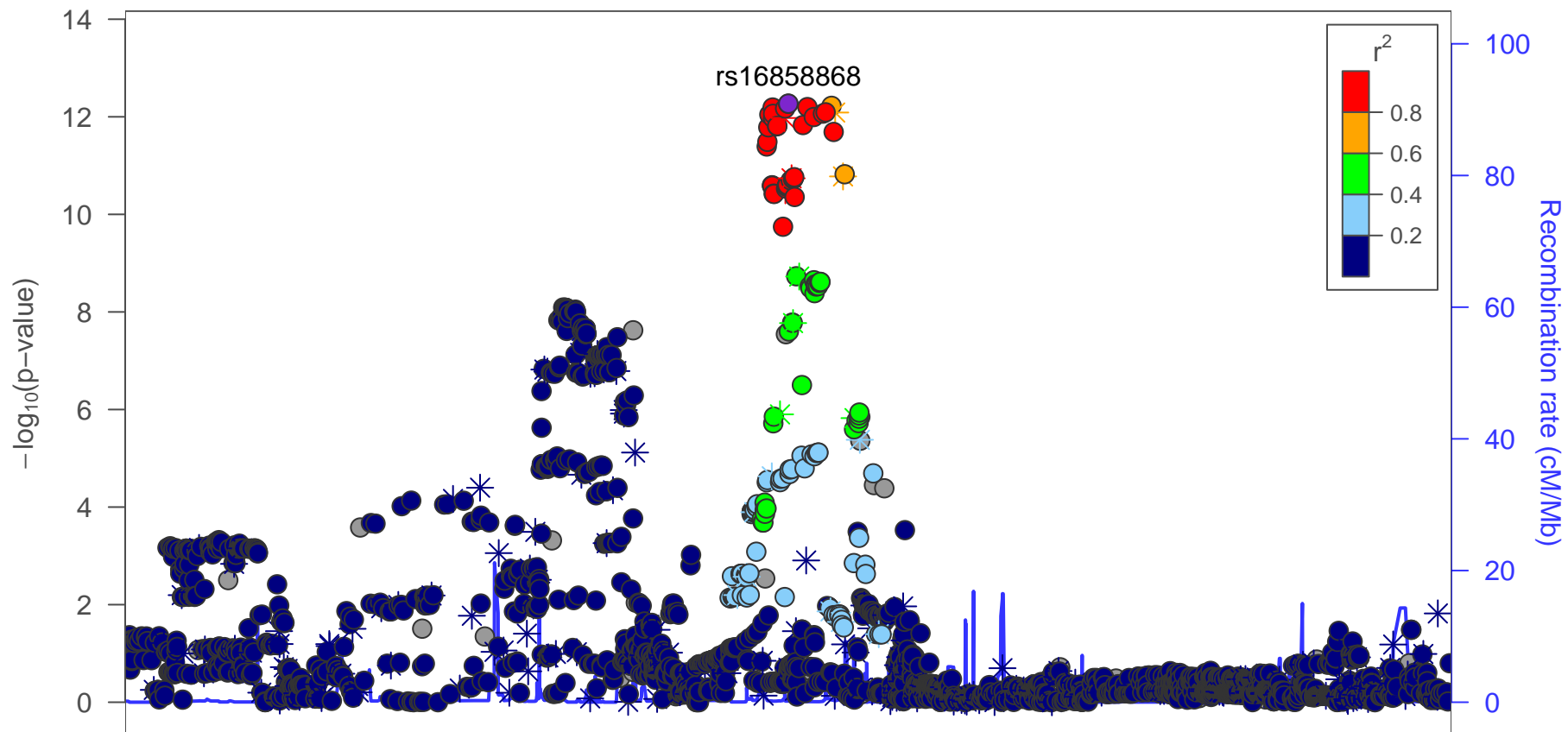


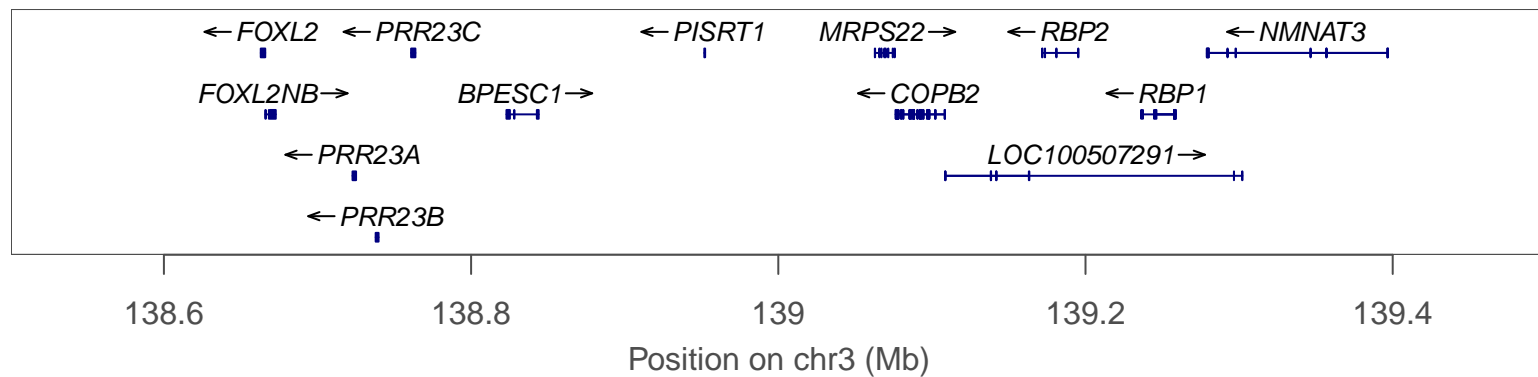
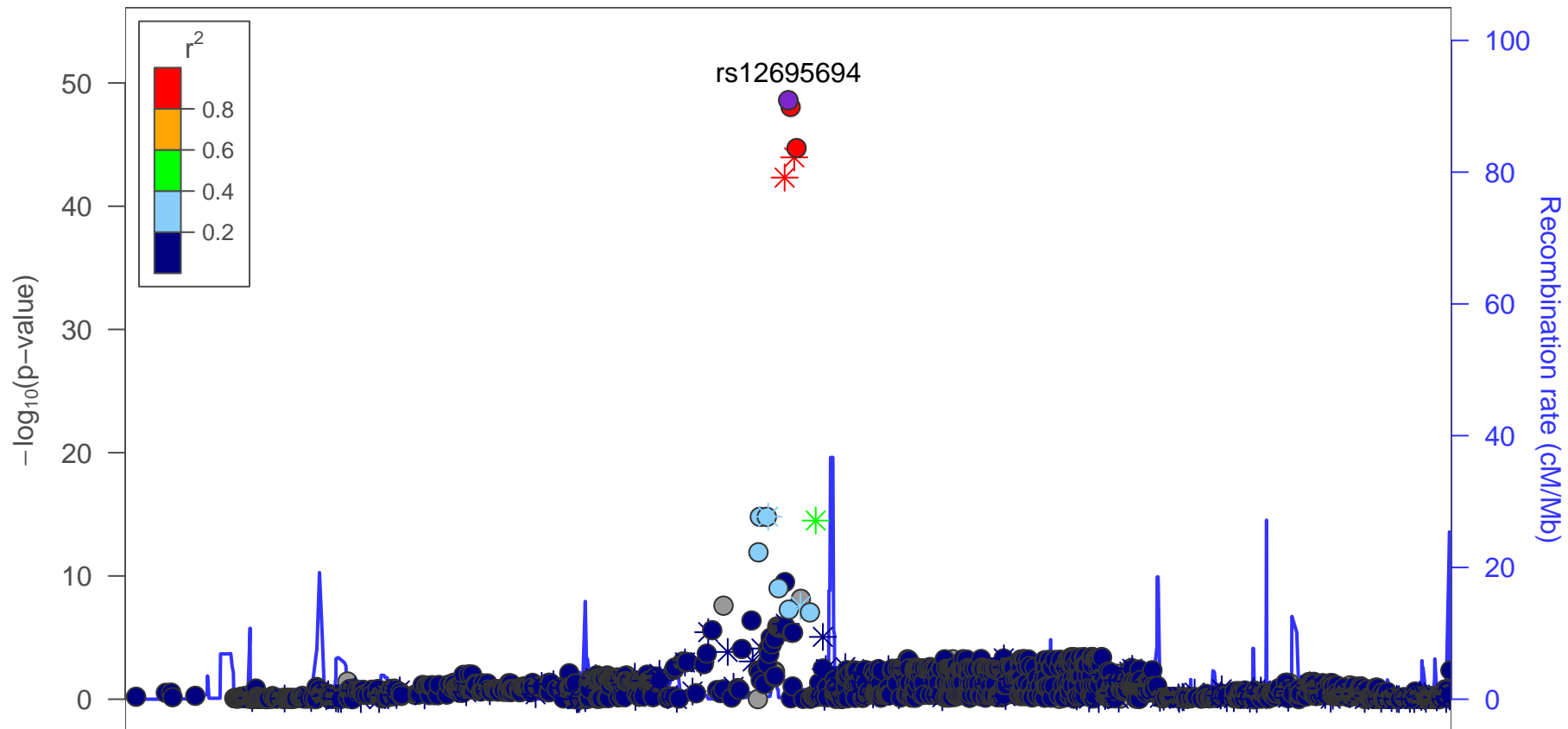


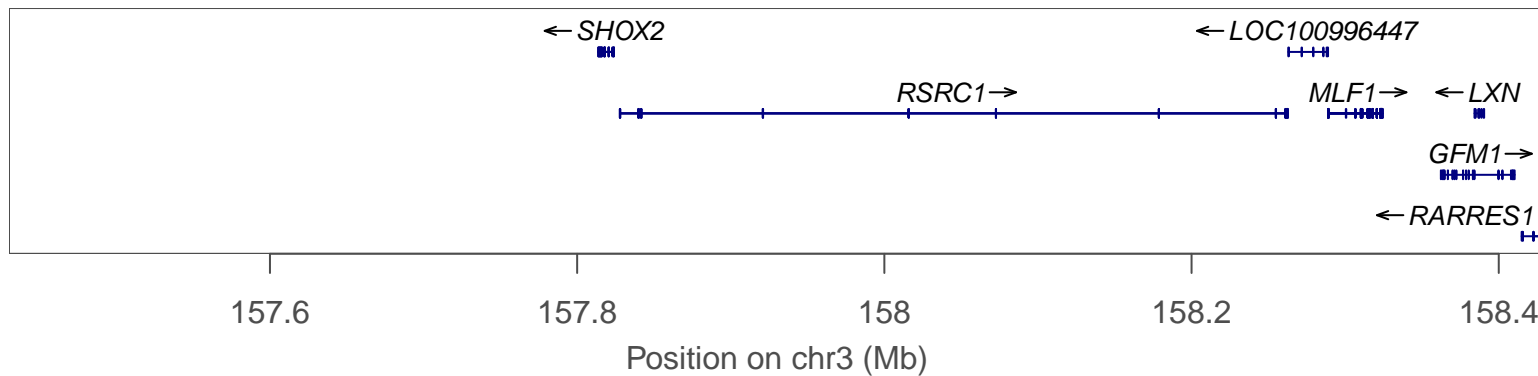
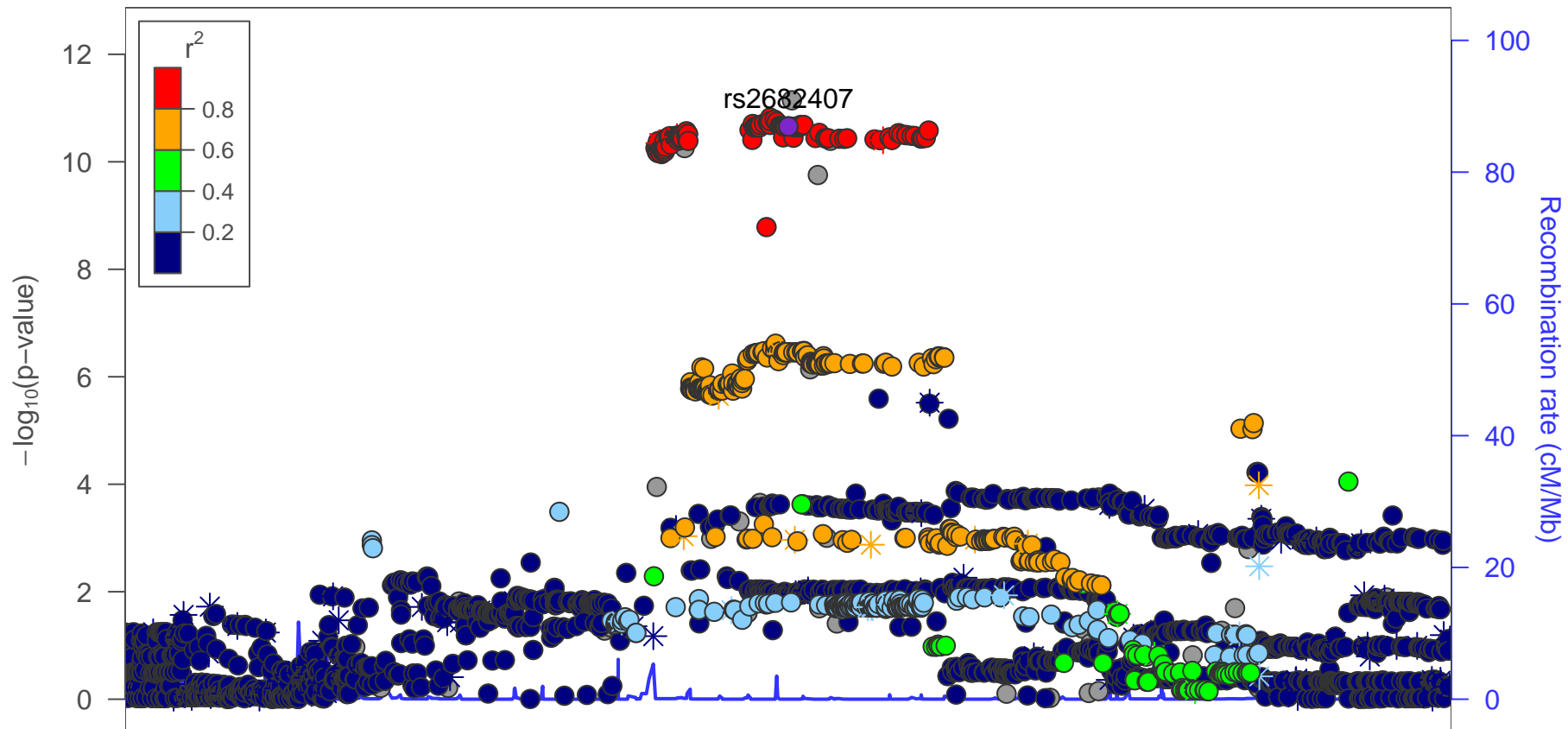


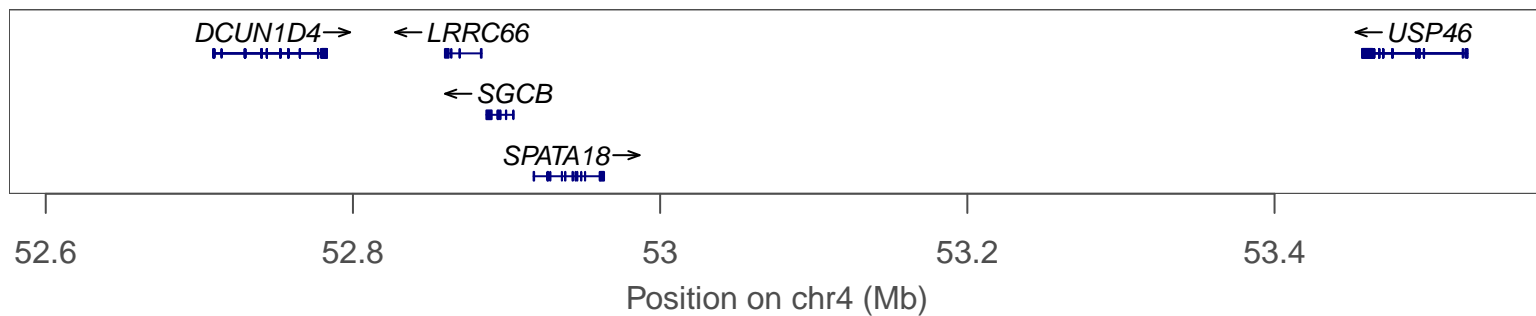
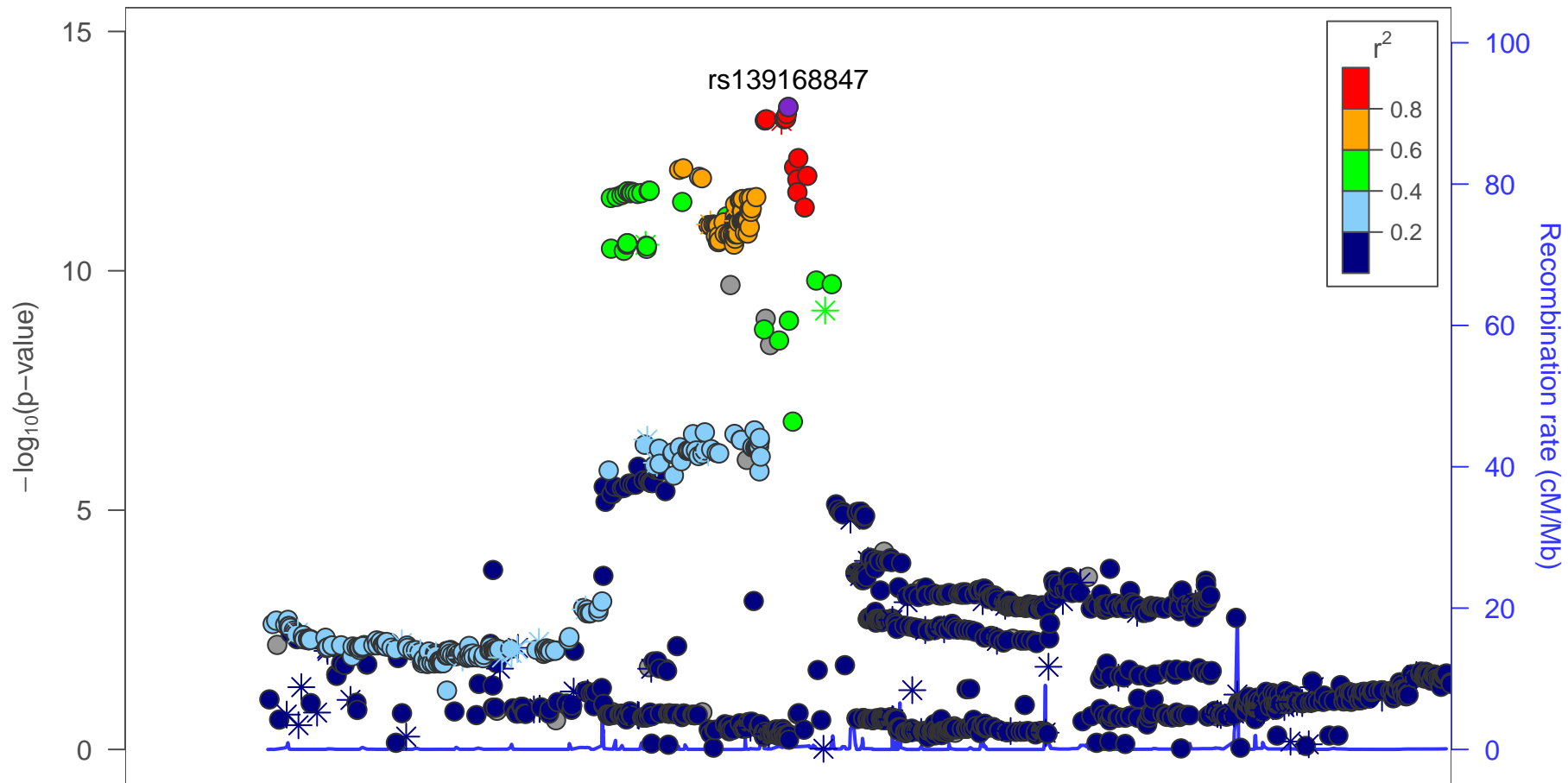


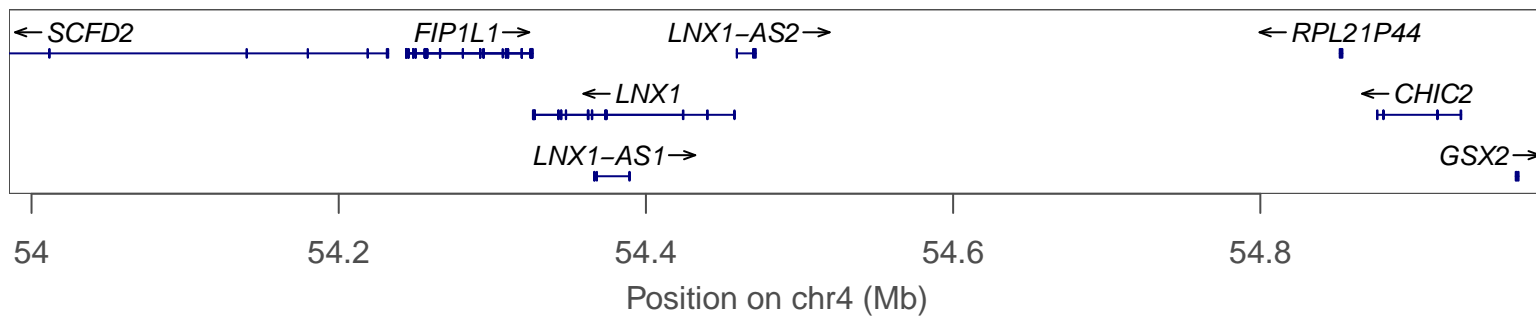
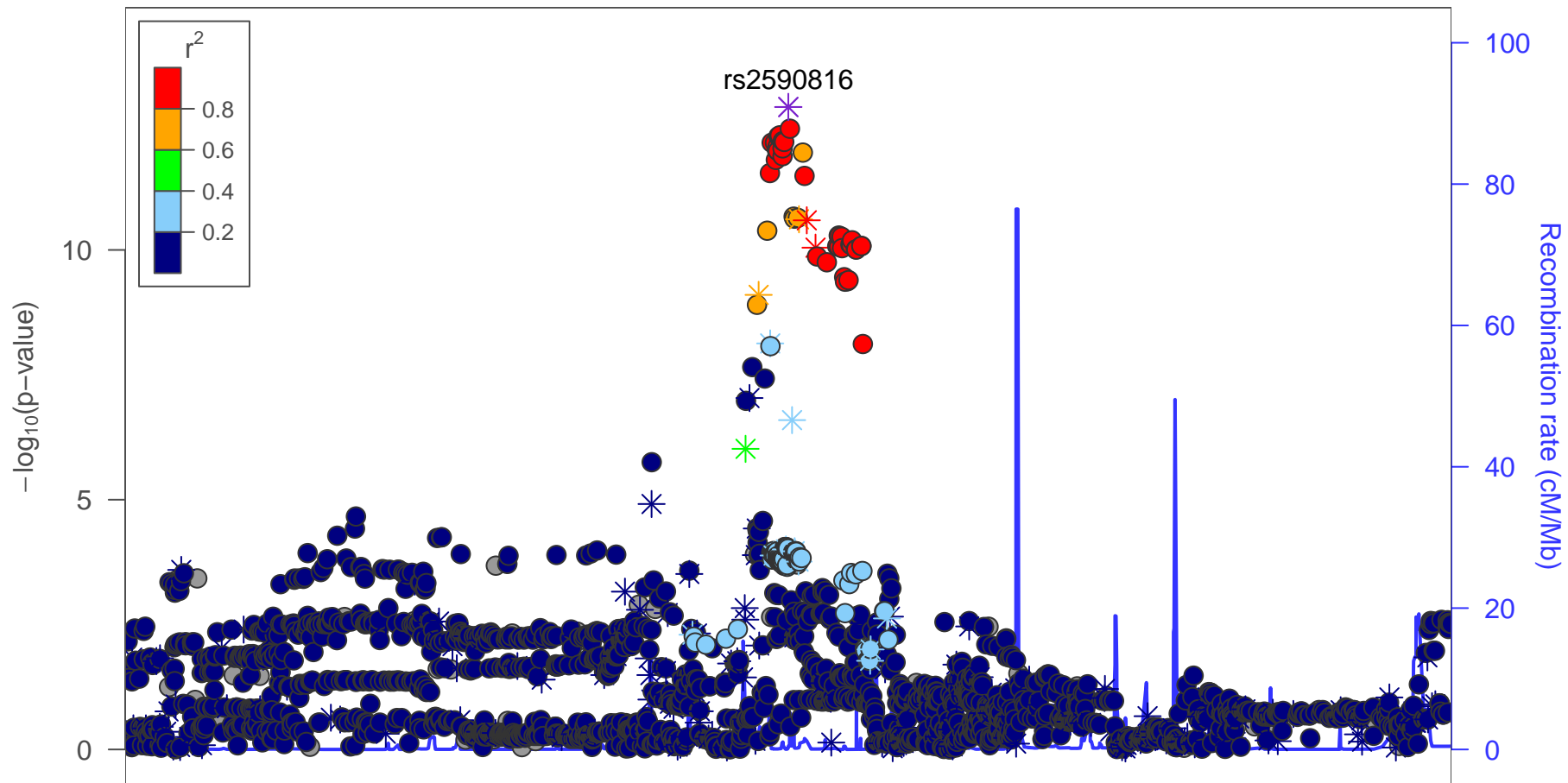


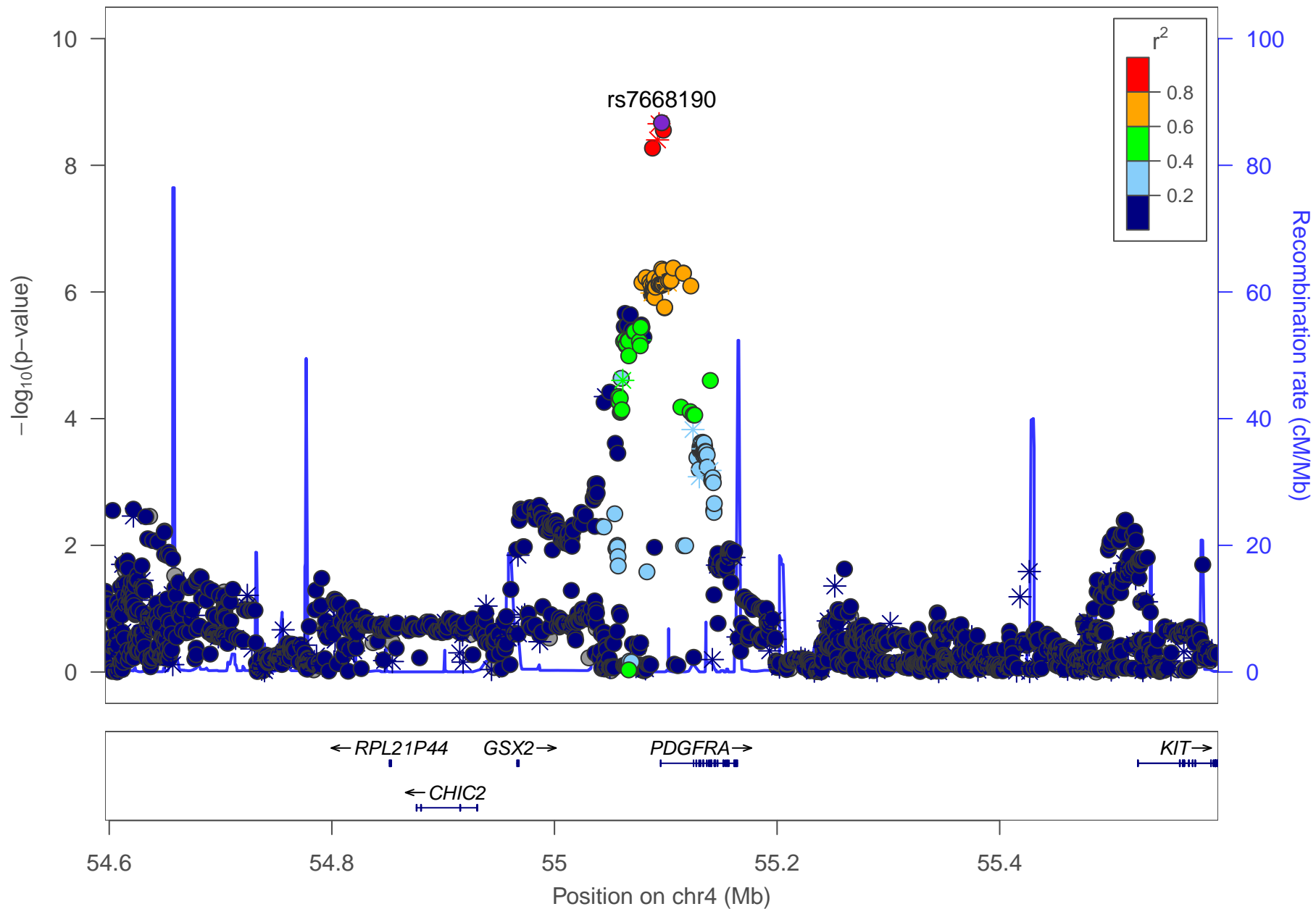


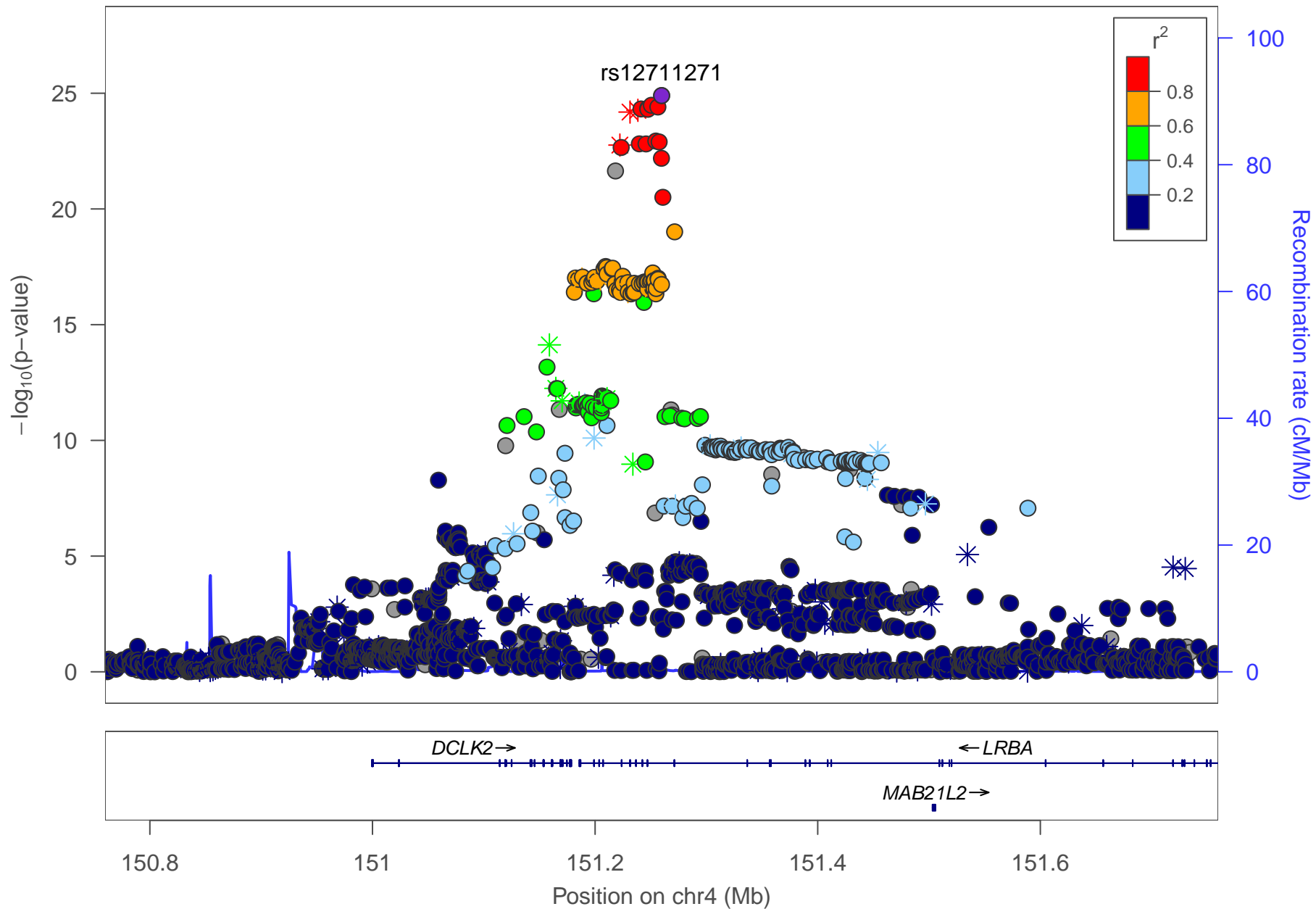


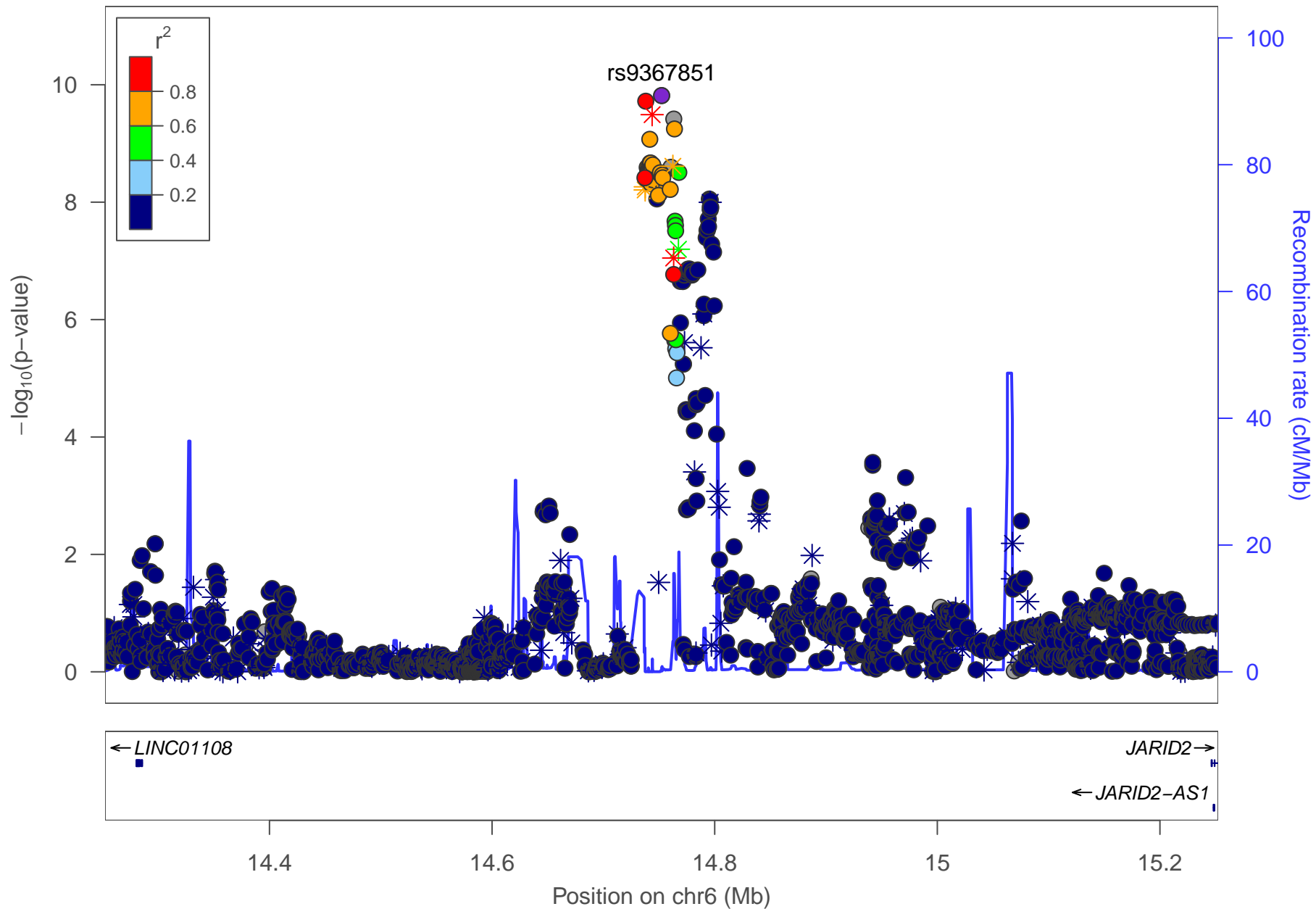


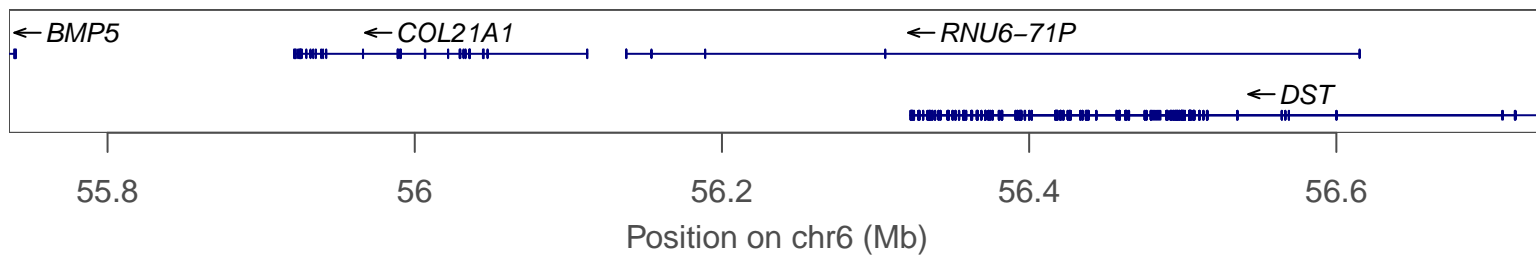
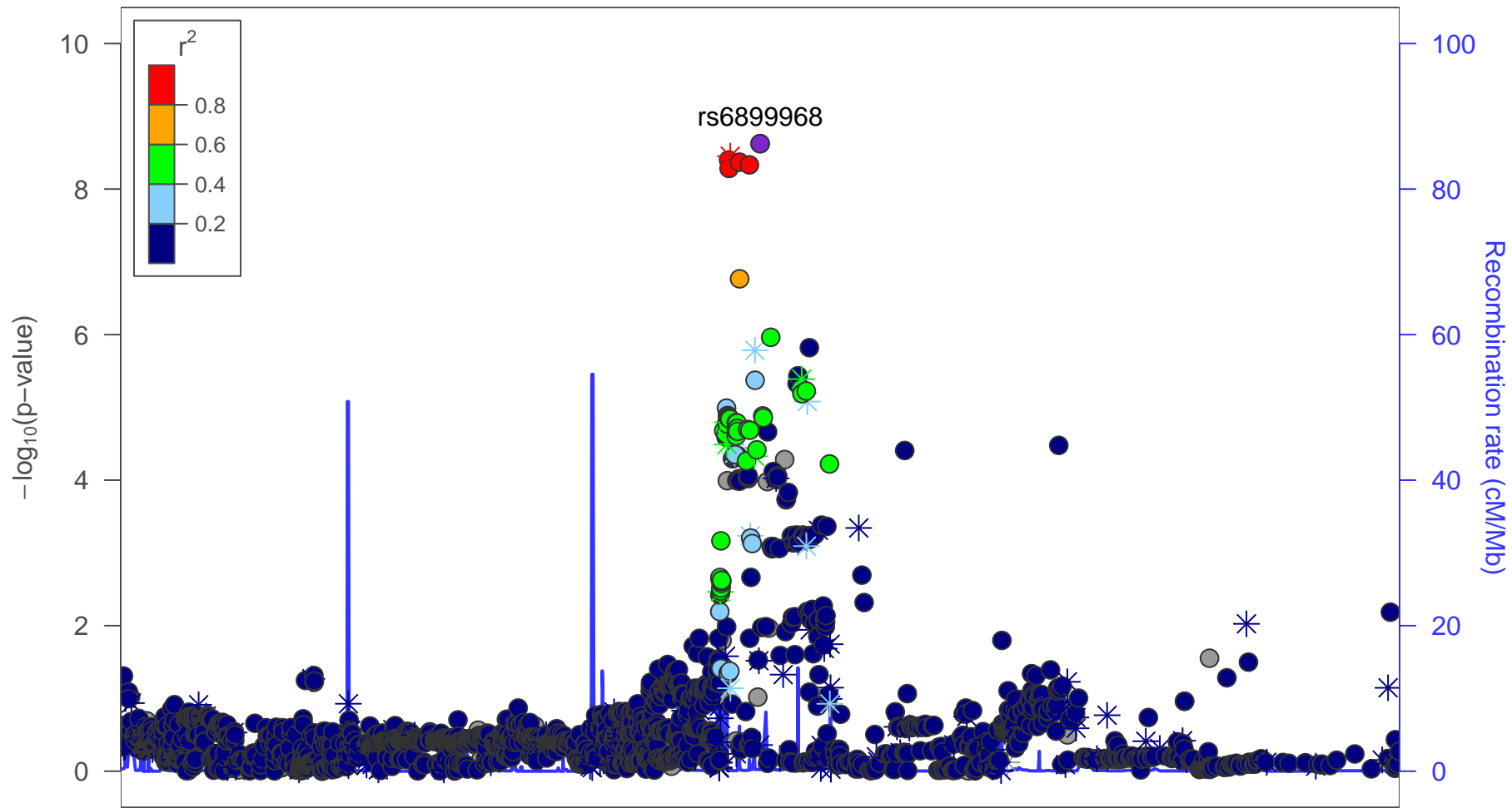


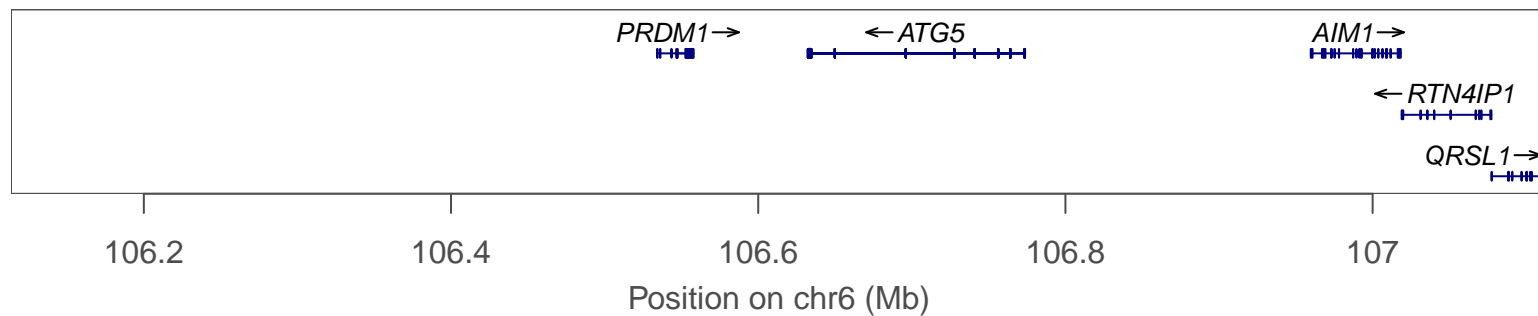
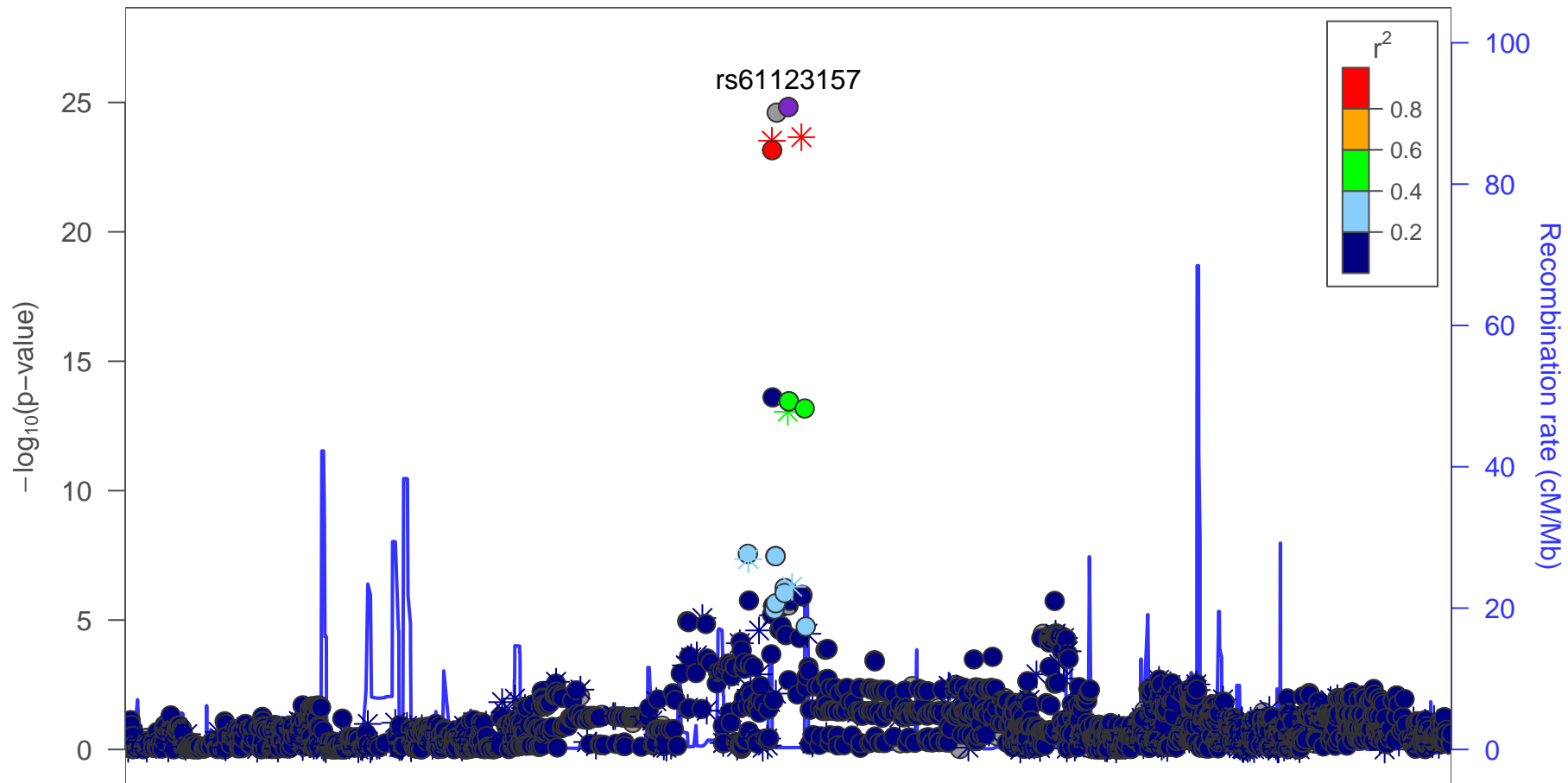


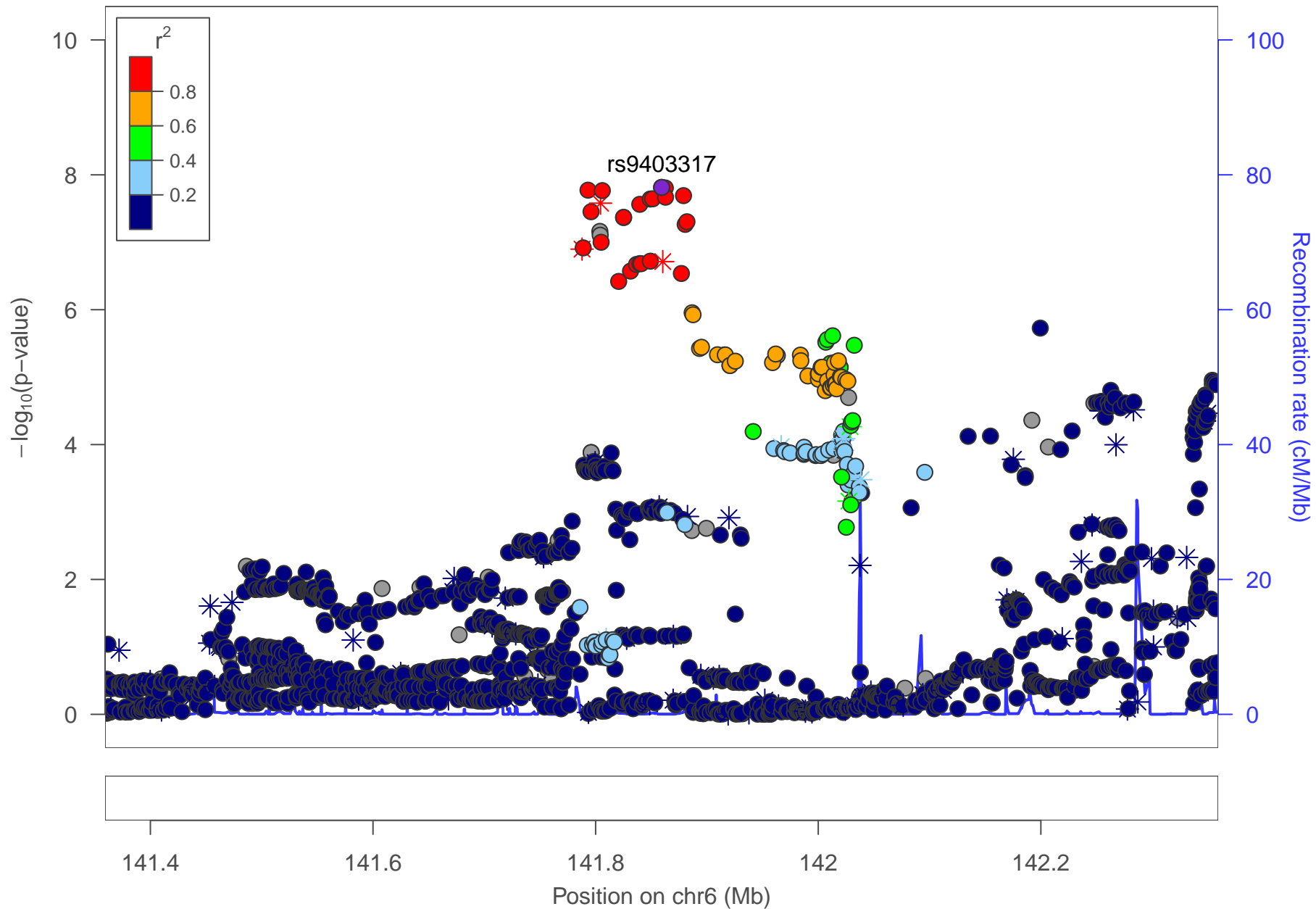


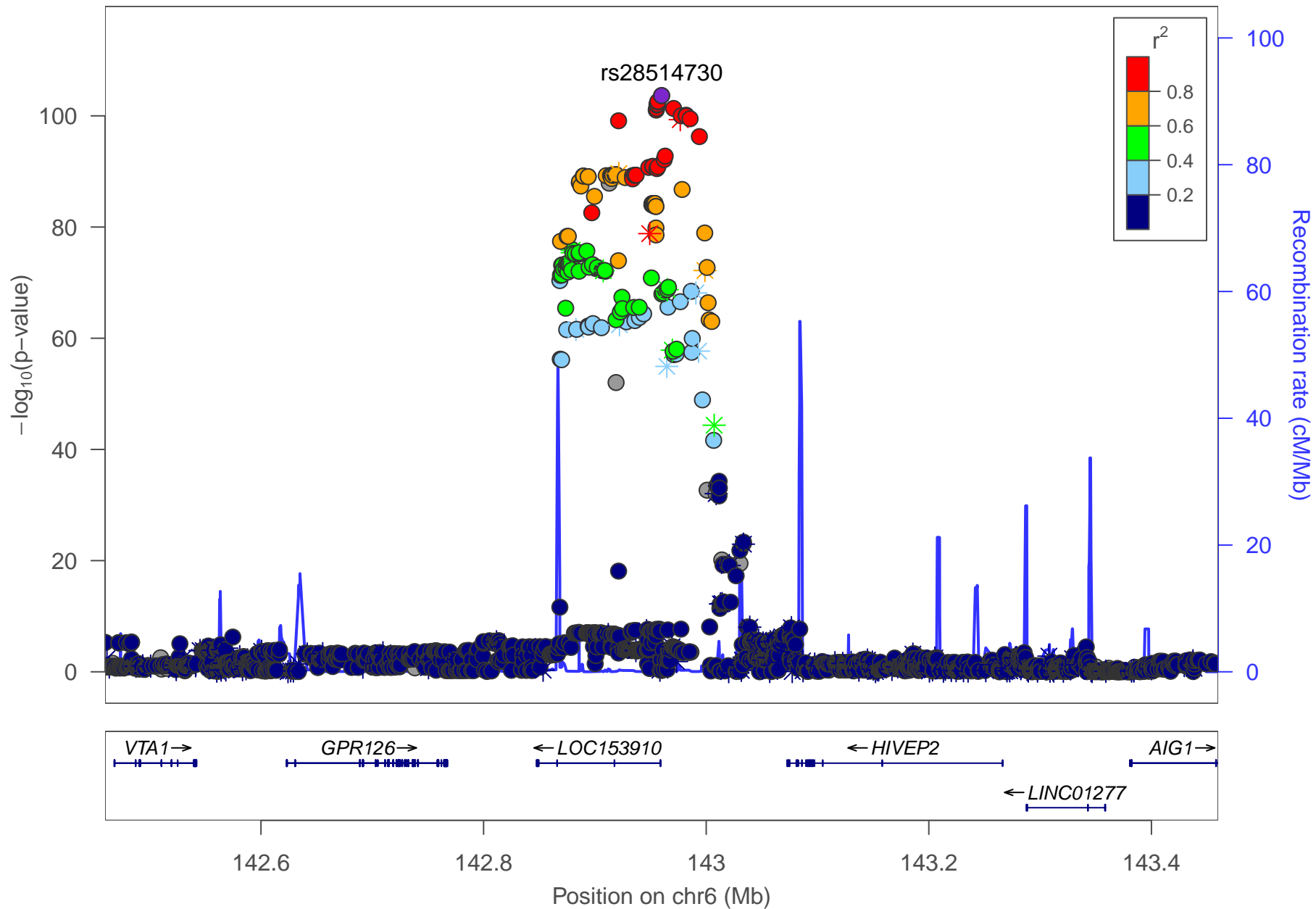


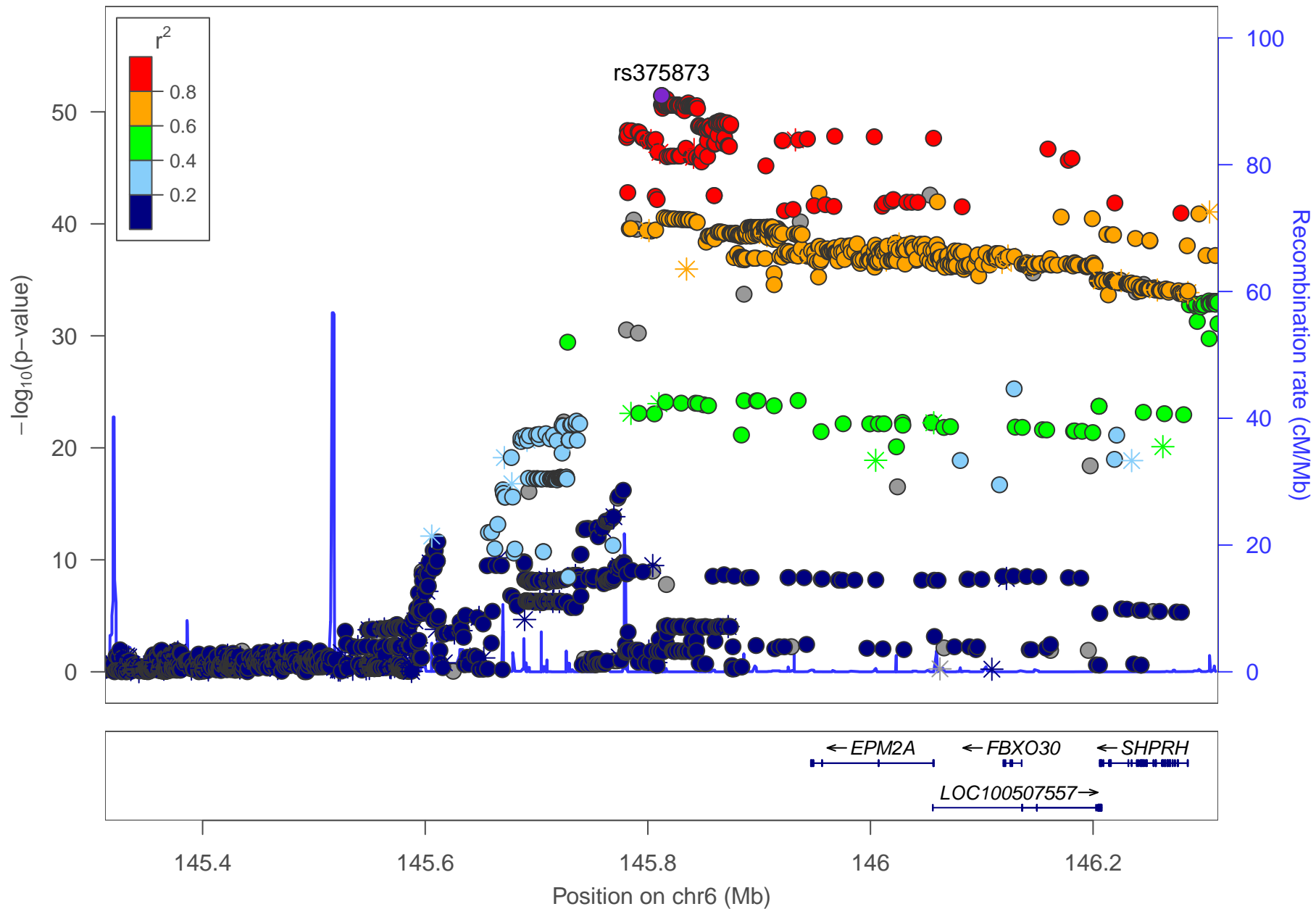


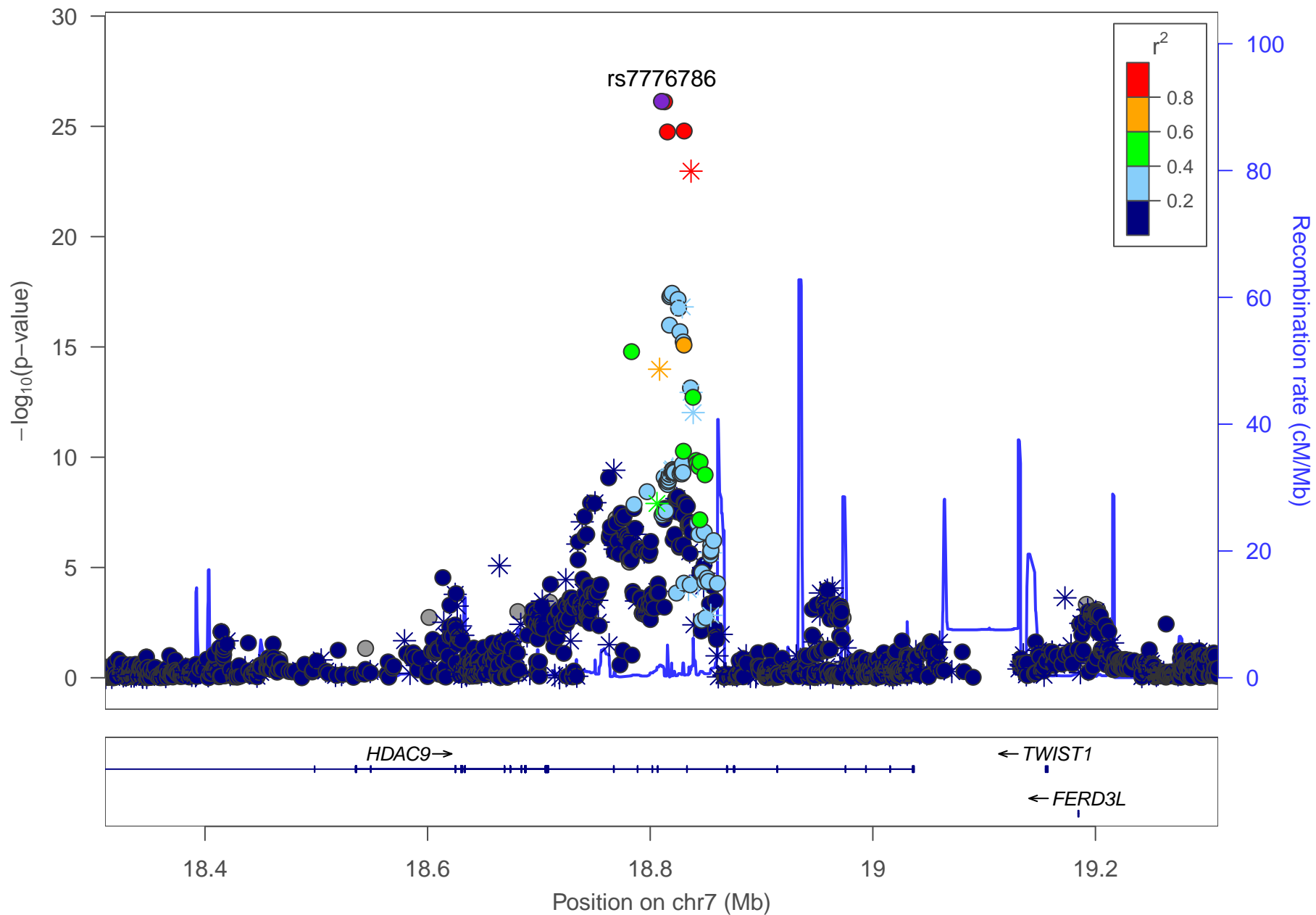


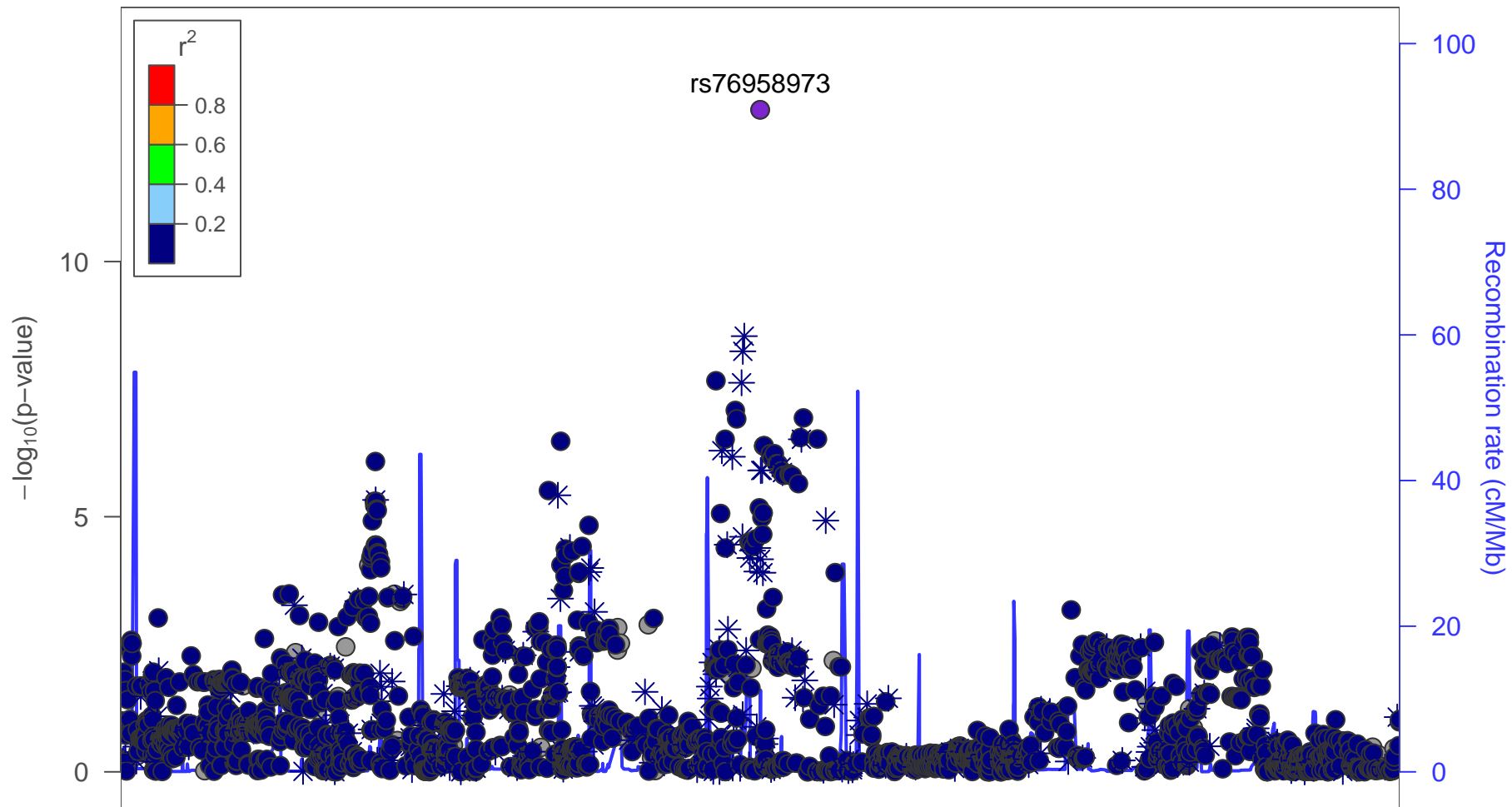












25.2

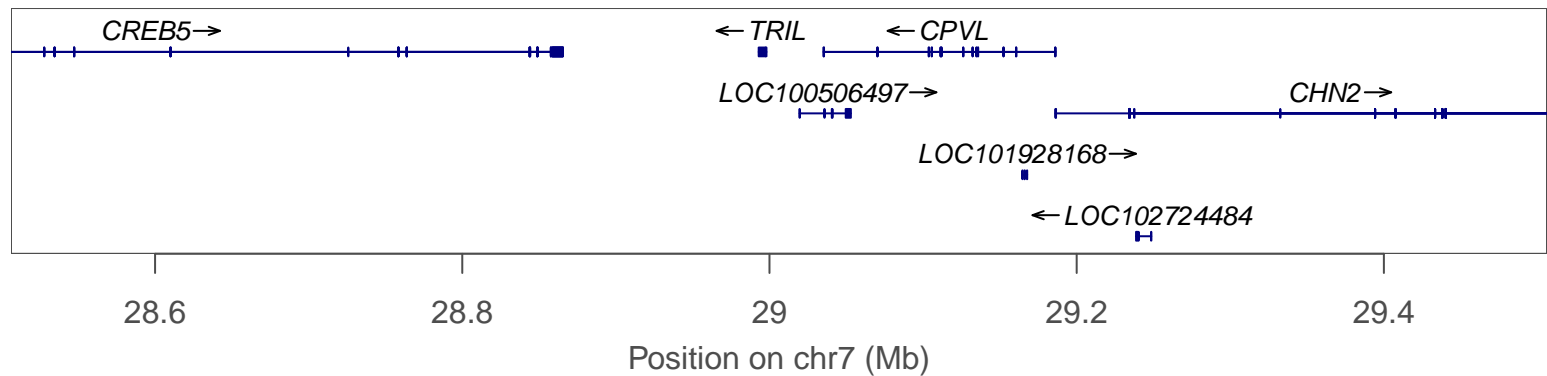
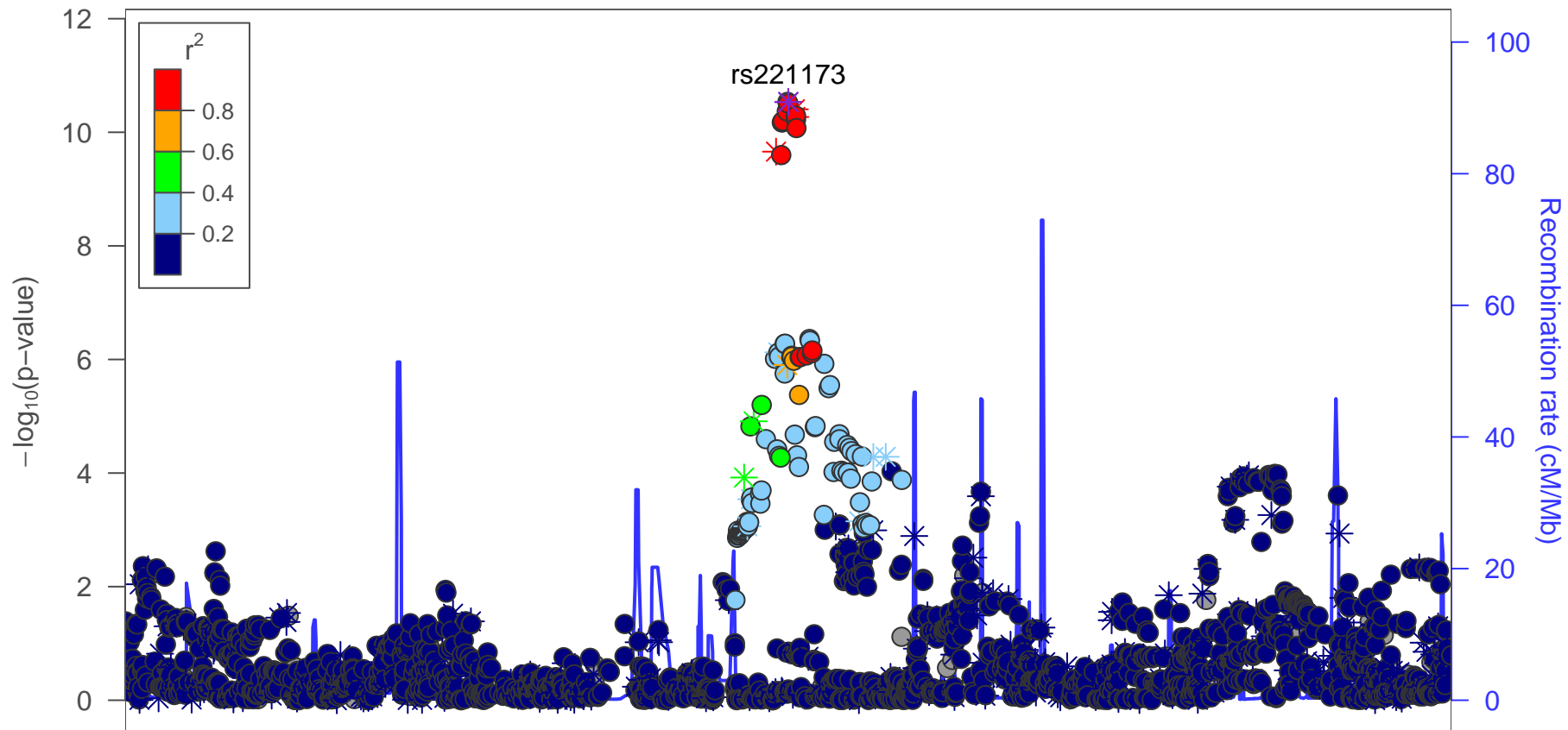
25.4

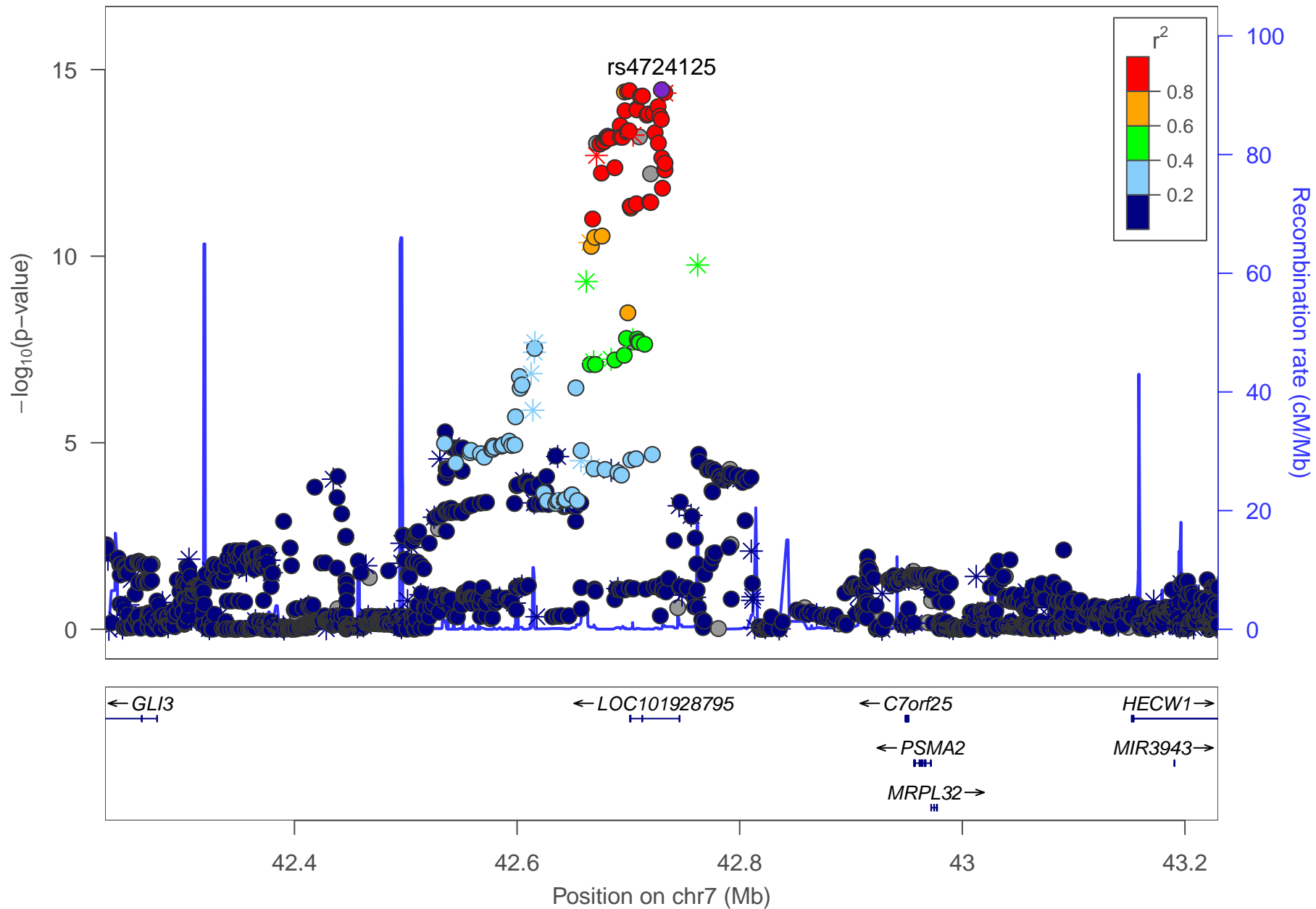
25.6

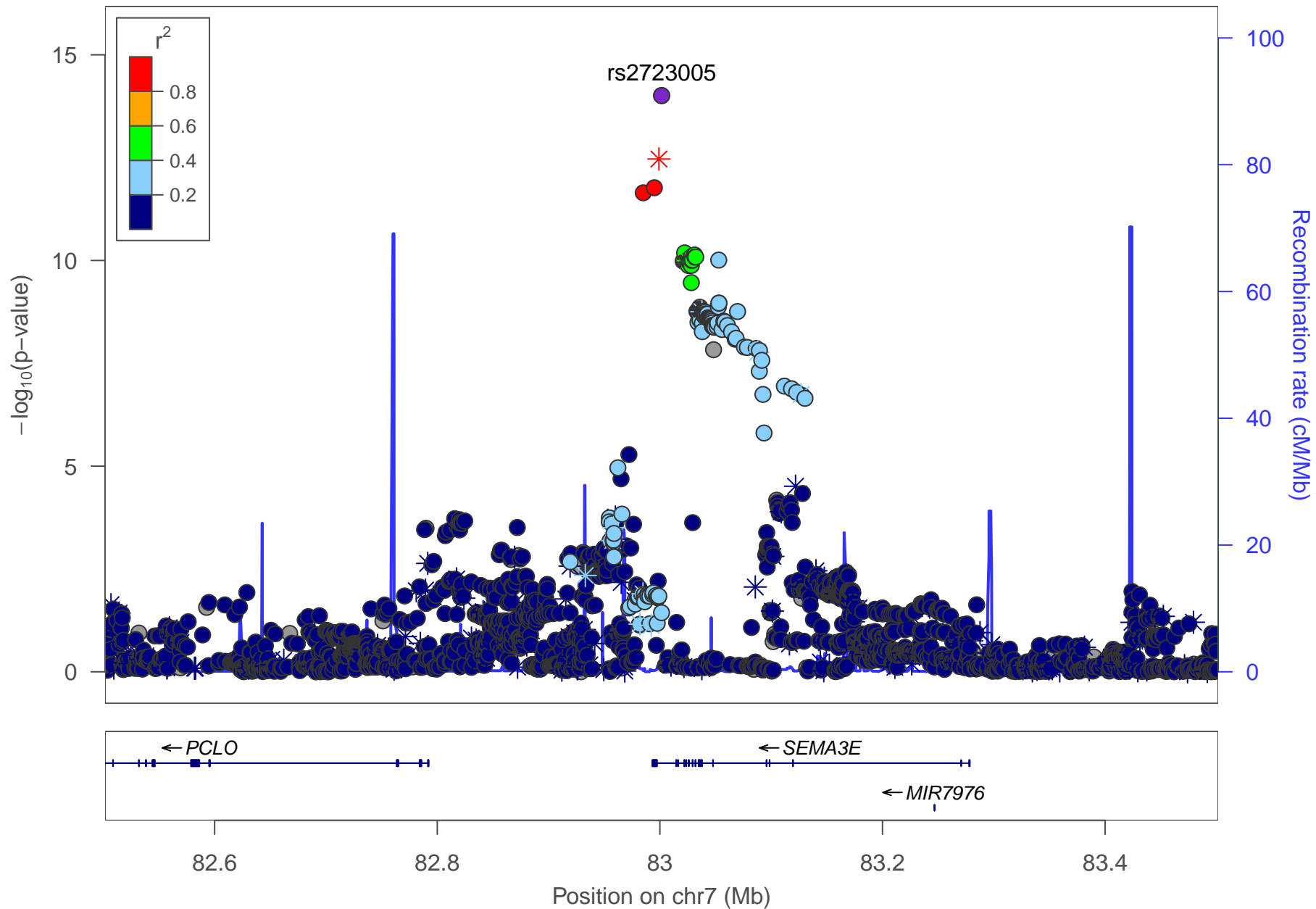
25.8

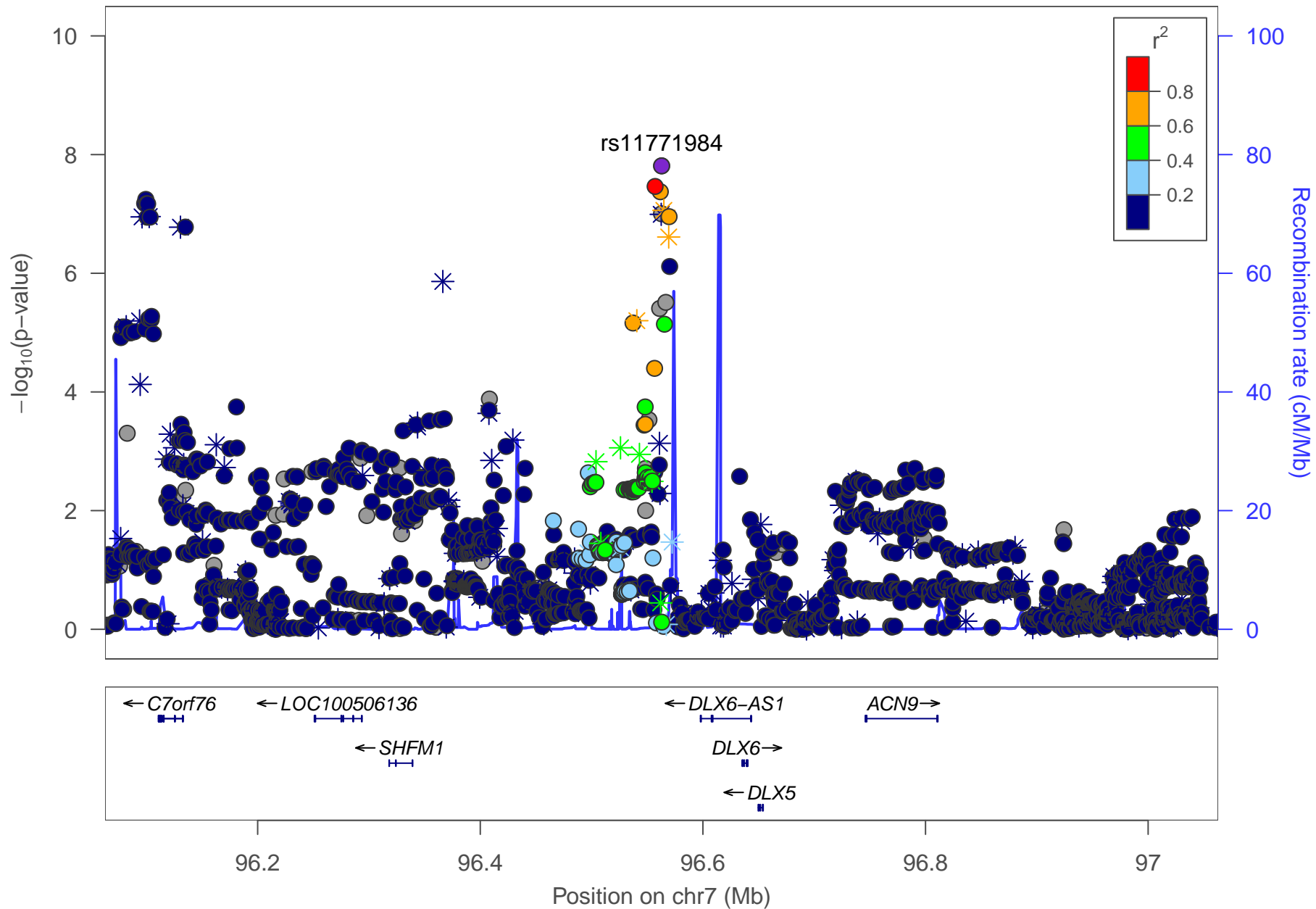
26

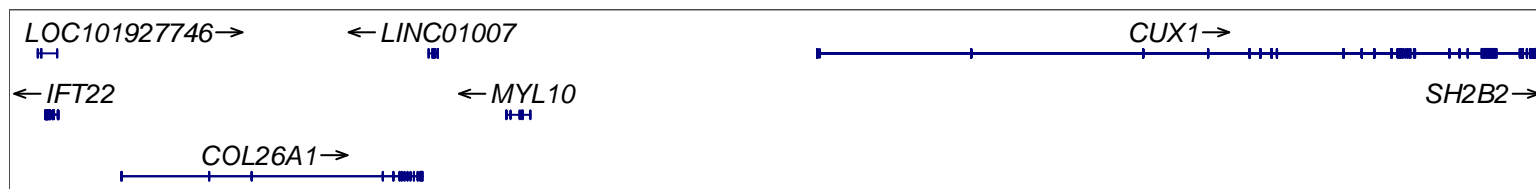
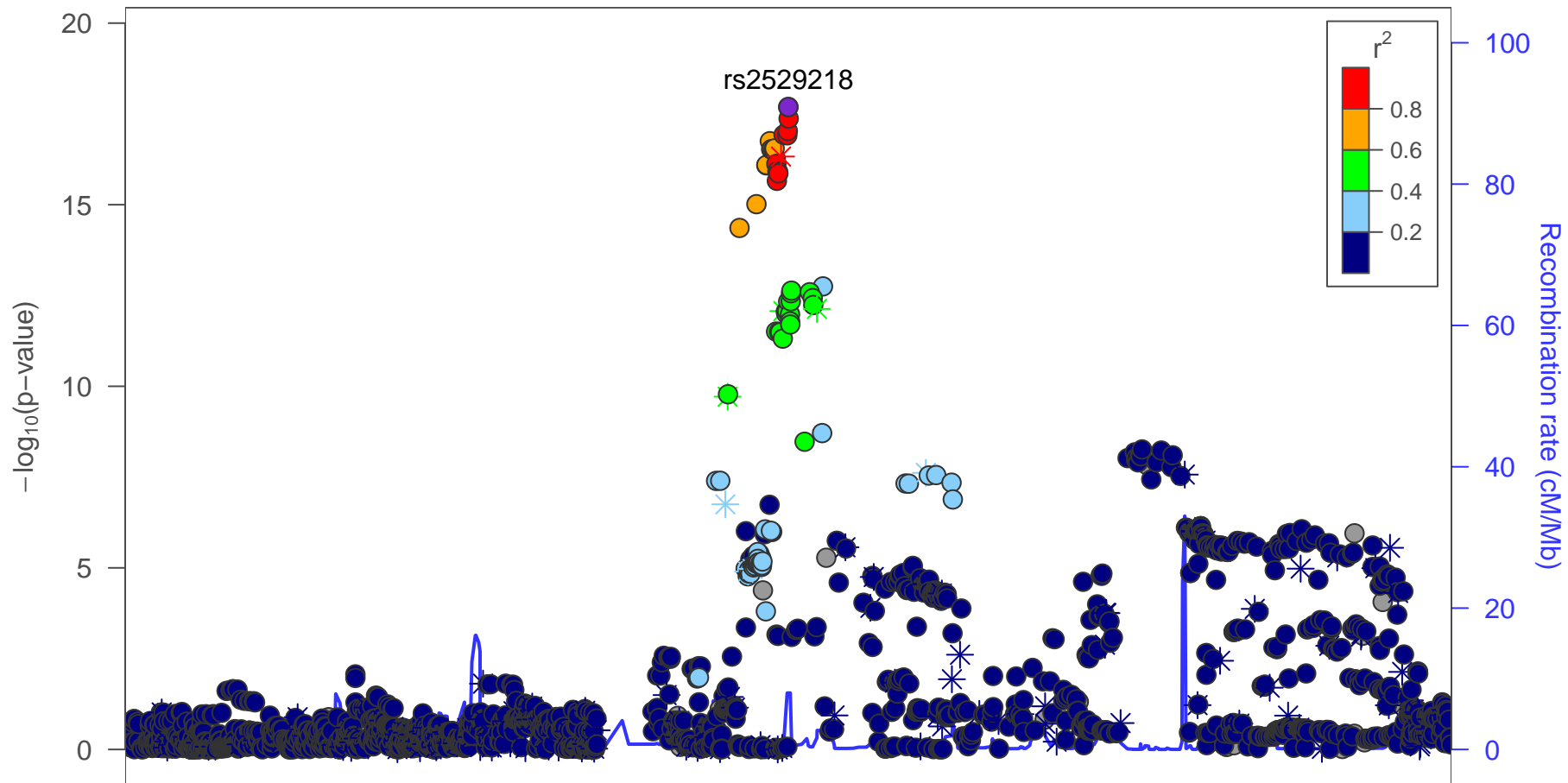
Position on chr7 (Mb)





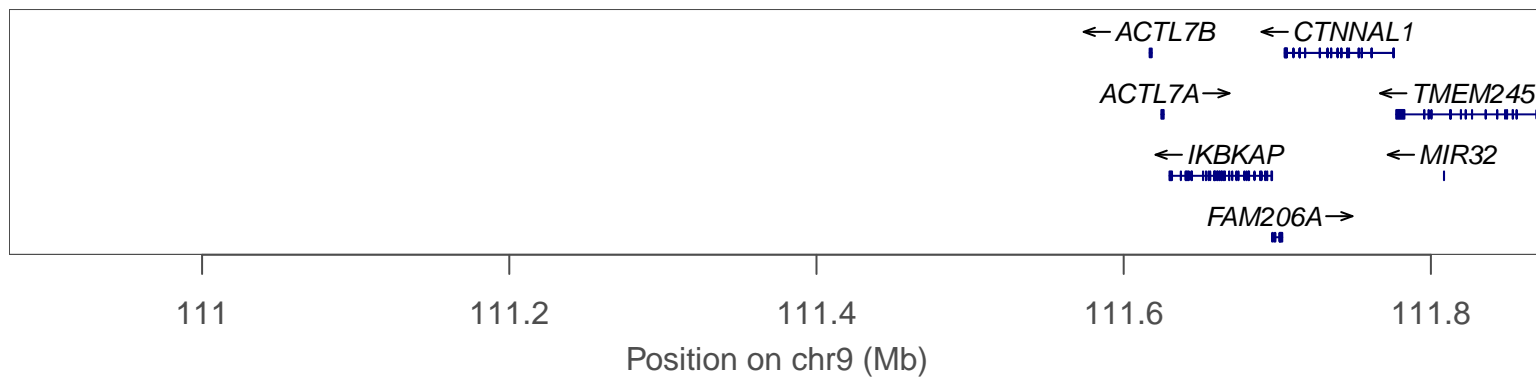
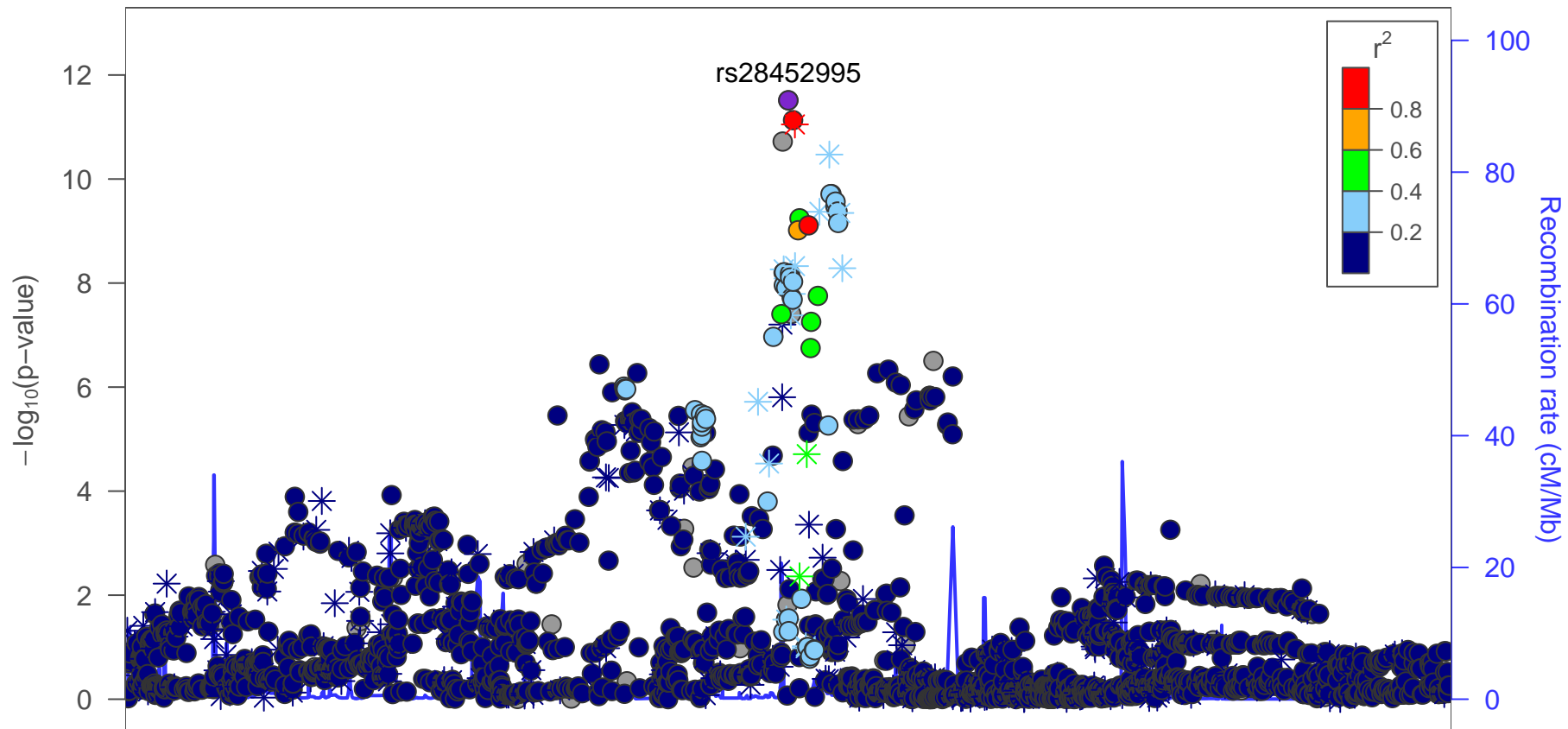


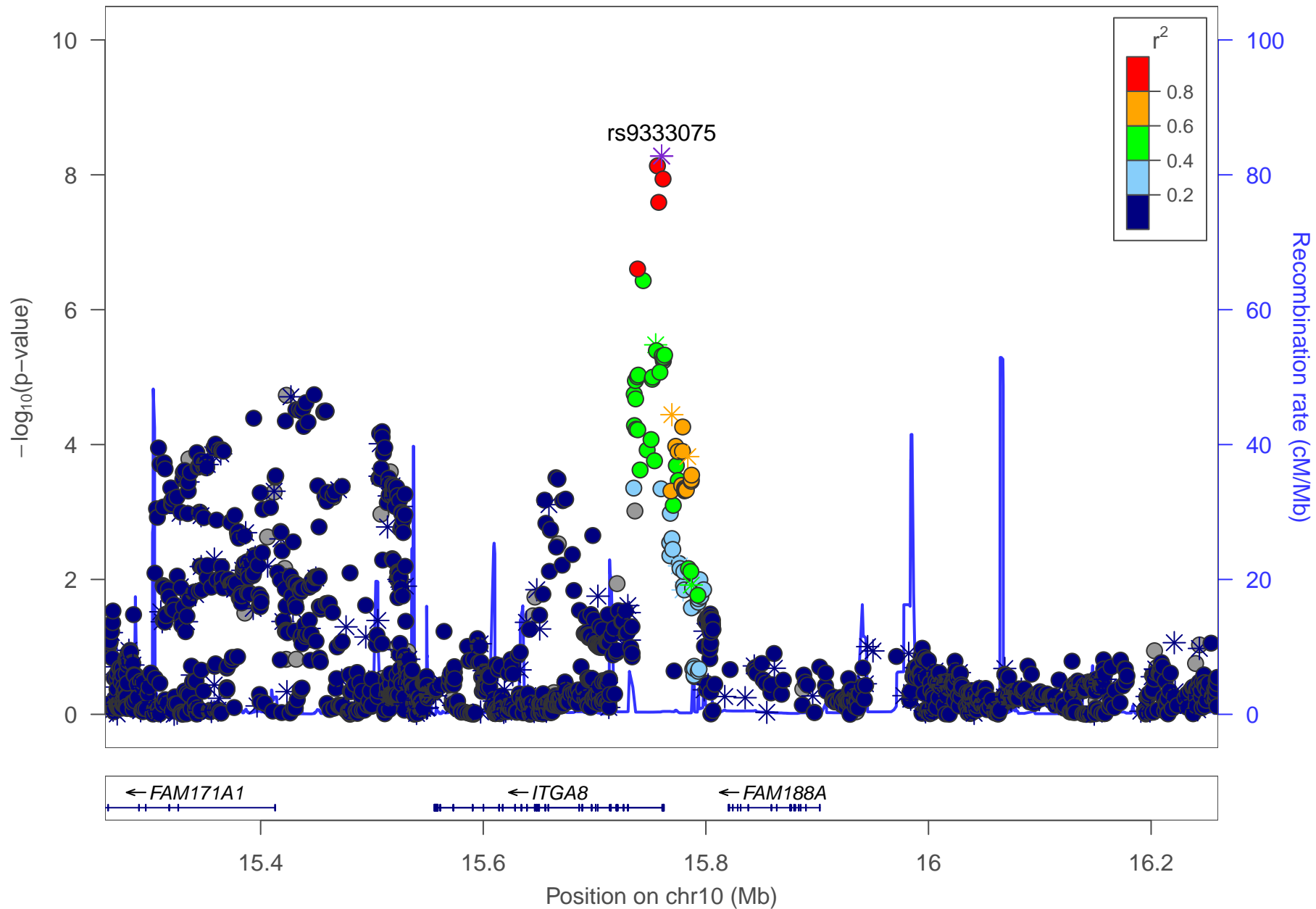


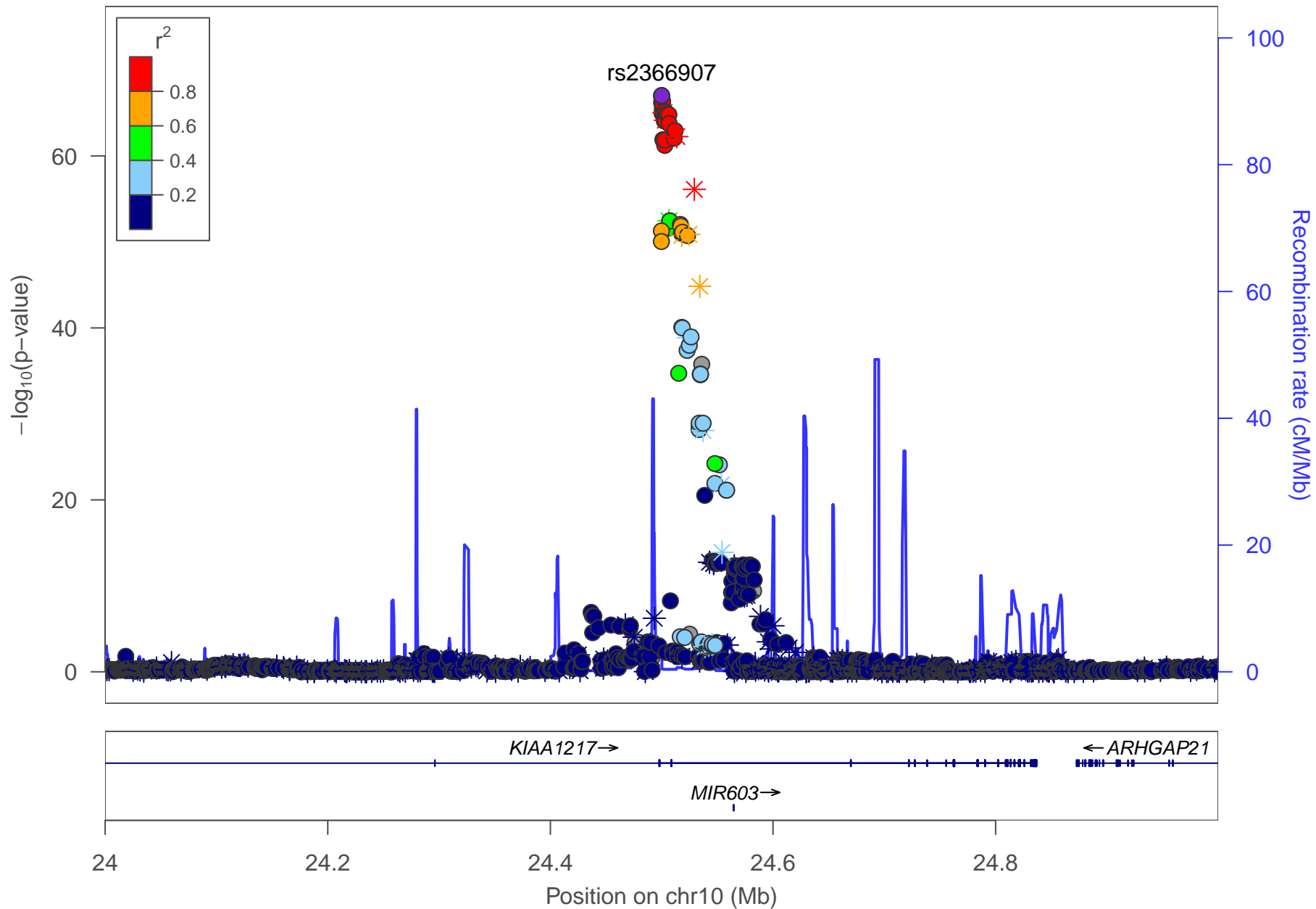


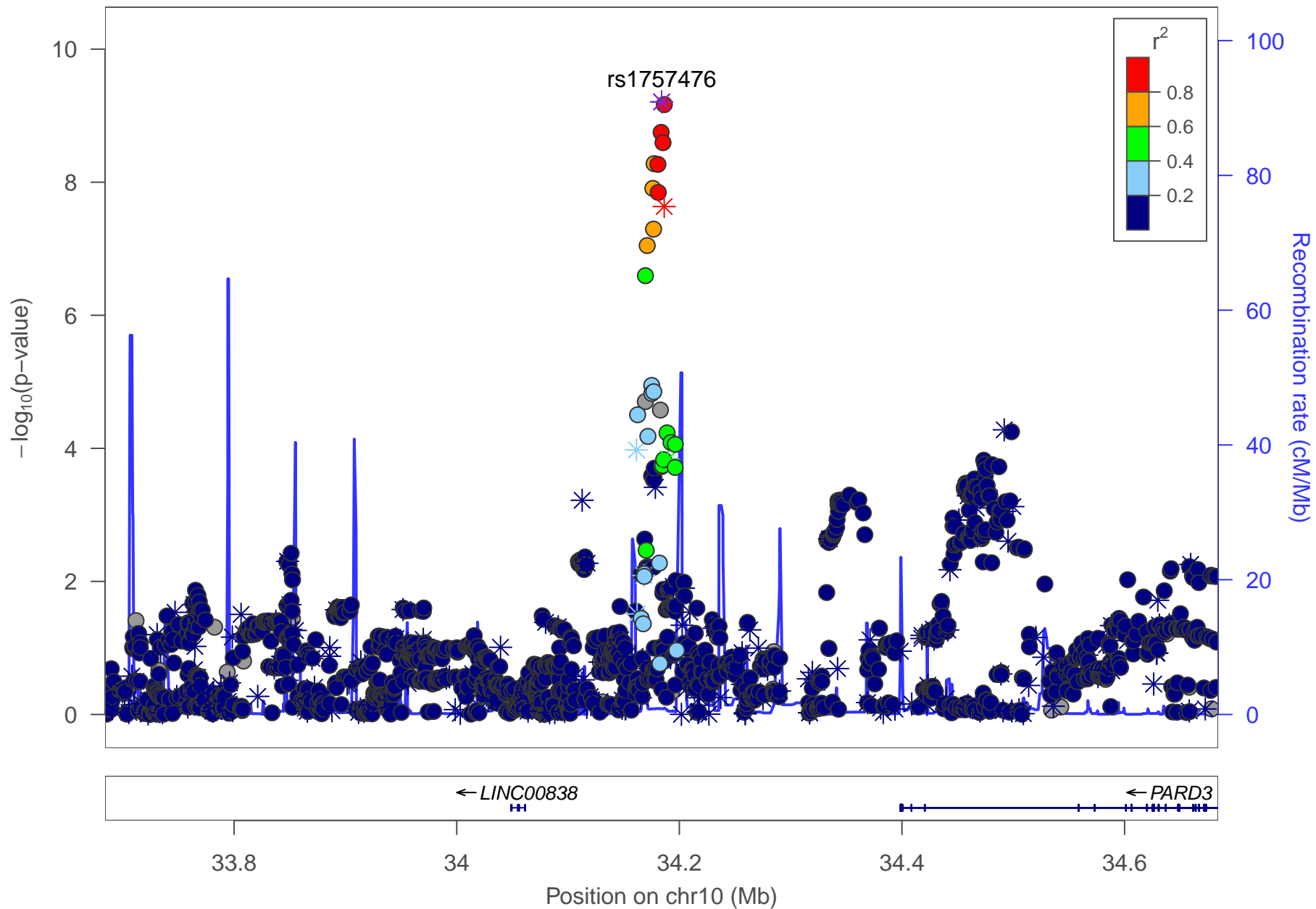
101 101.2 101.4 101.6 101.8

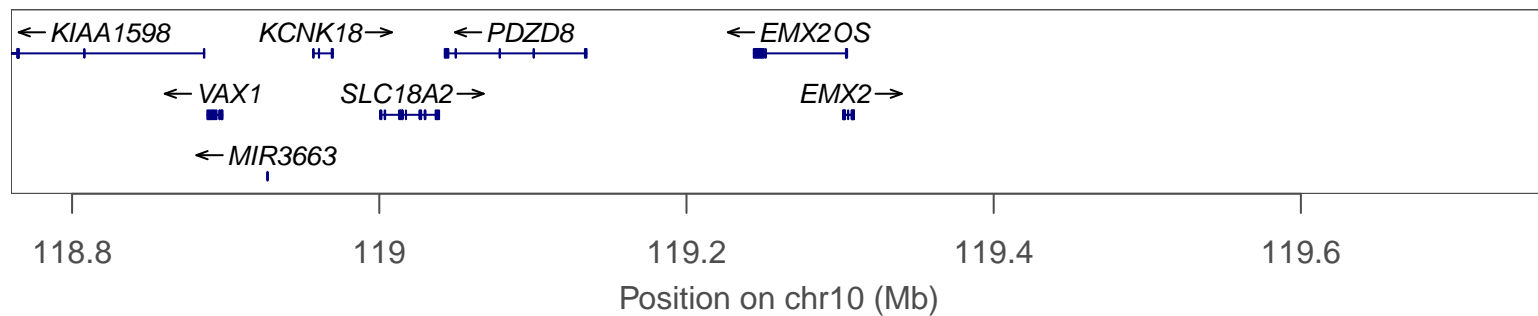
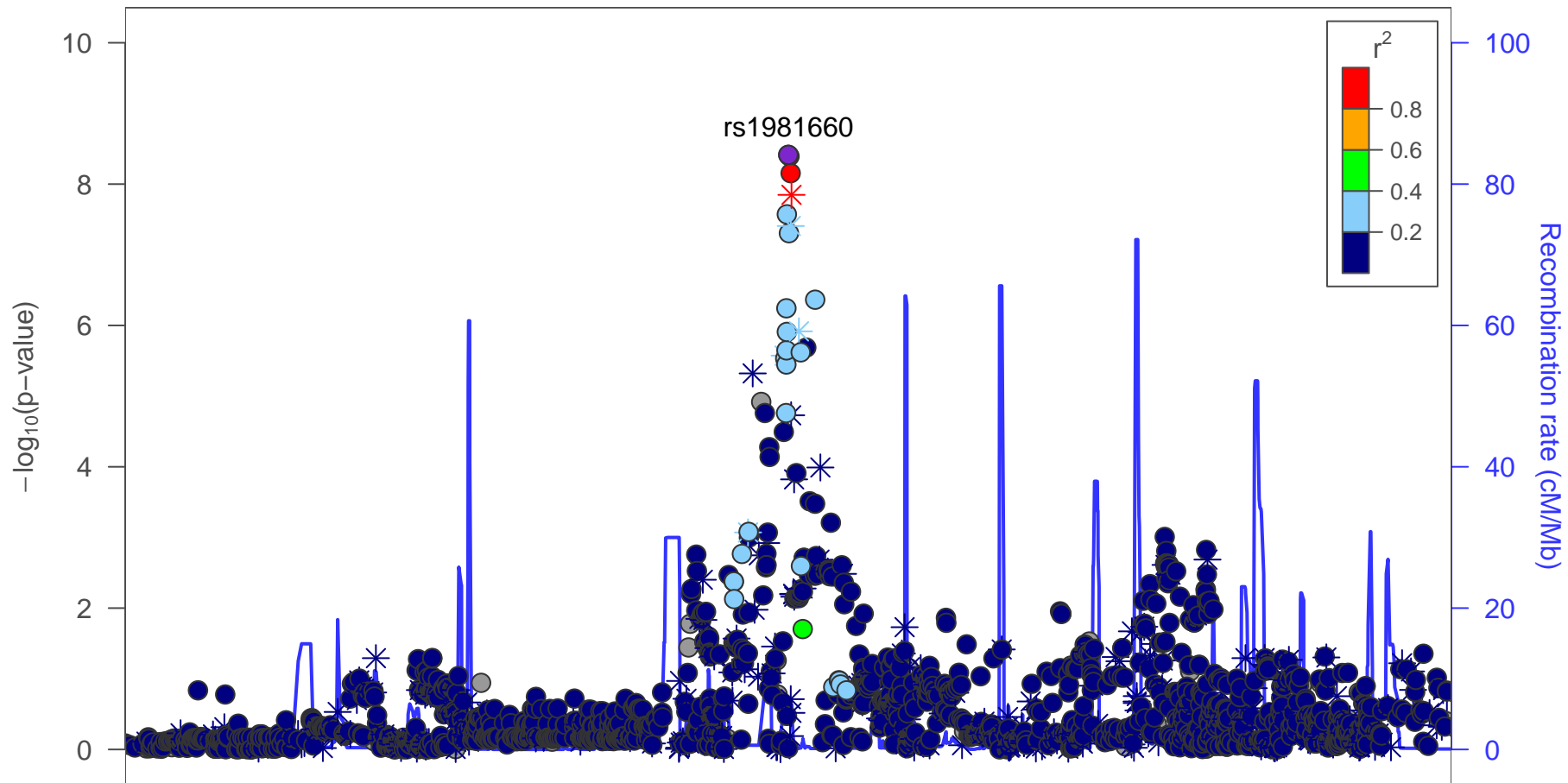
Position on chr7 (Mb)

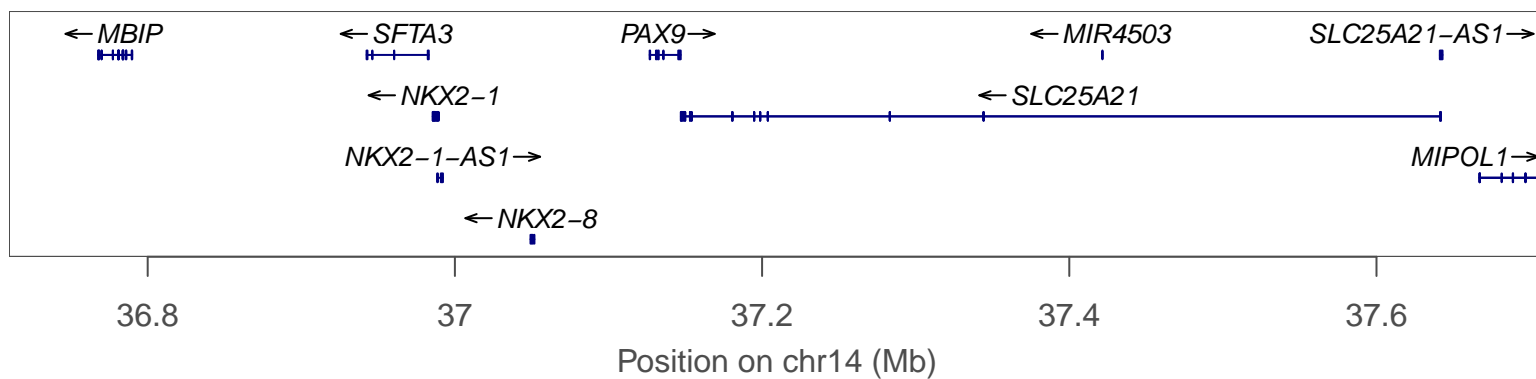
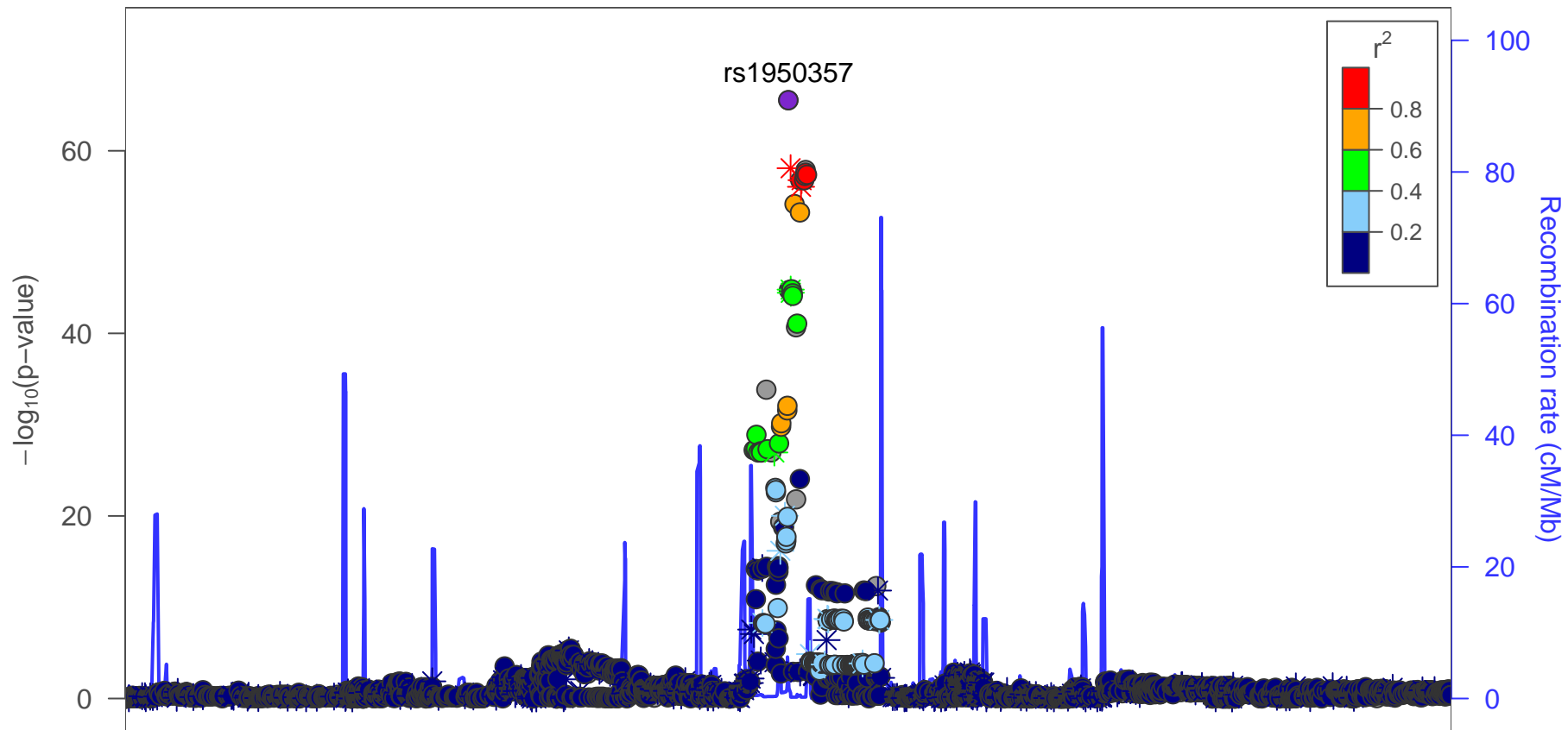


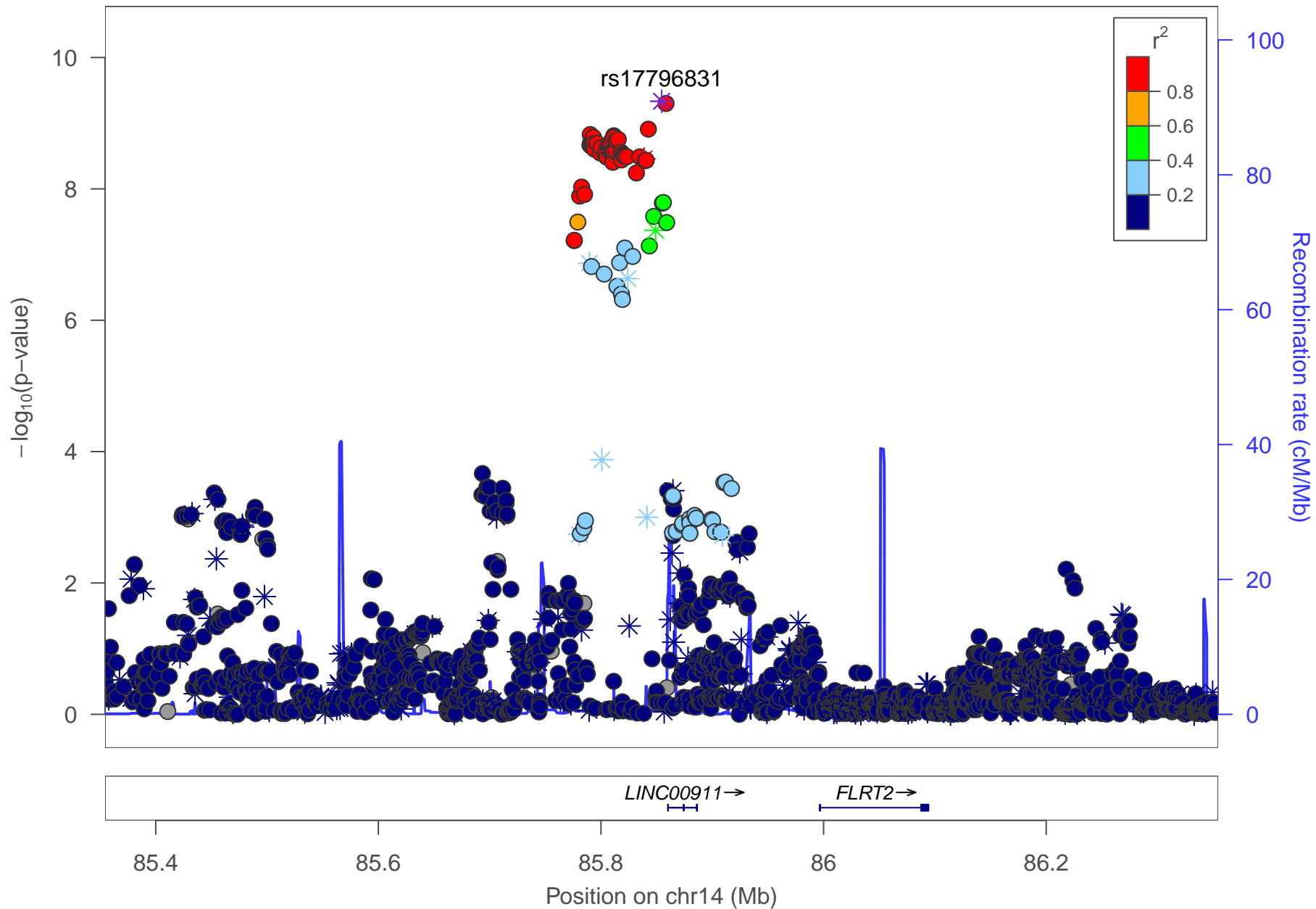


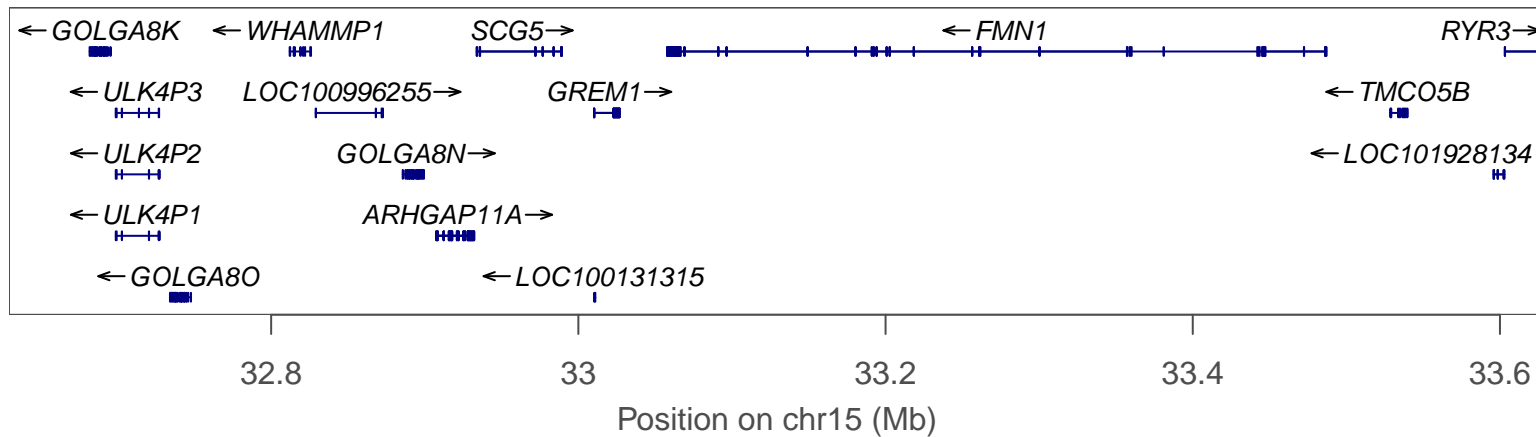
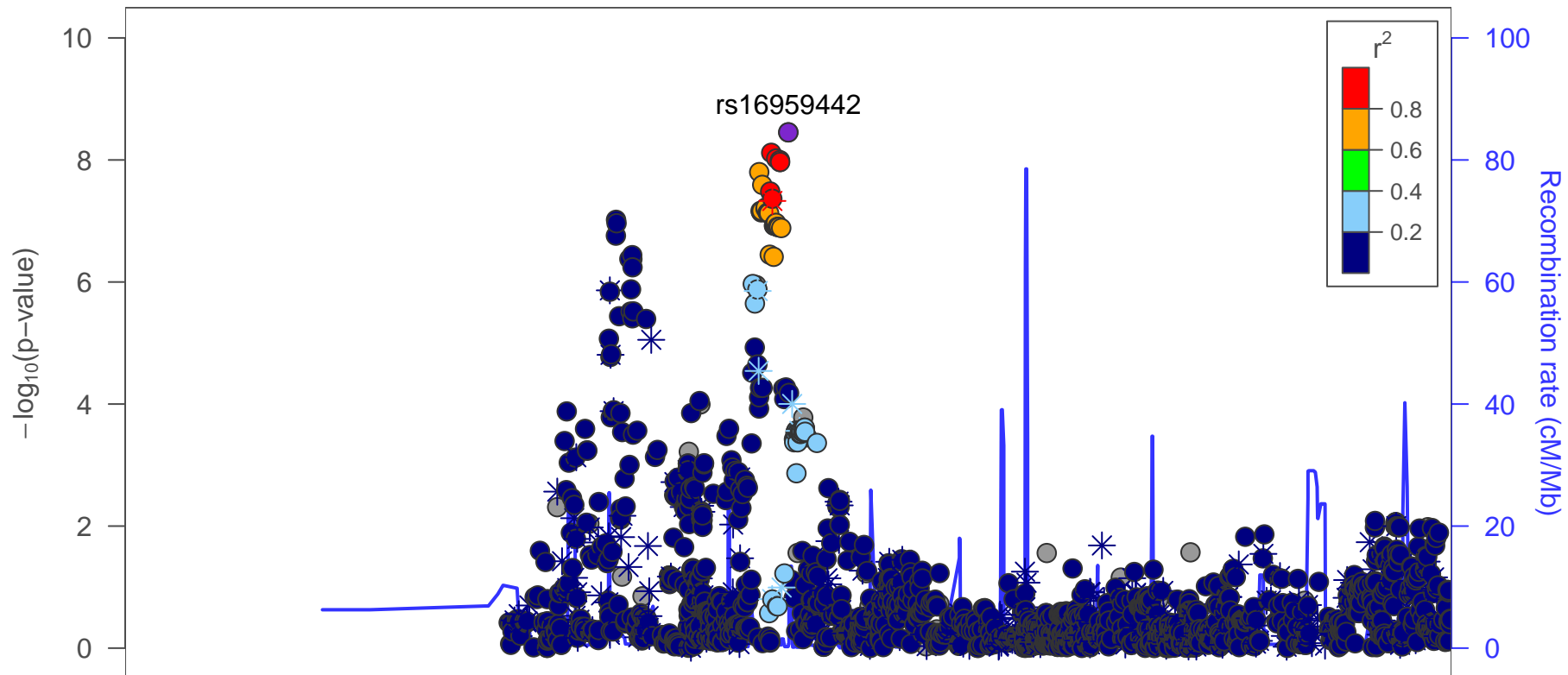


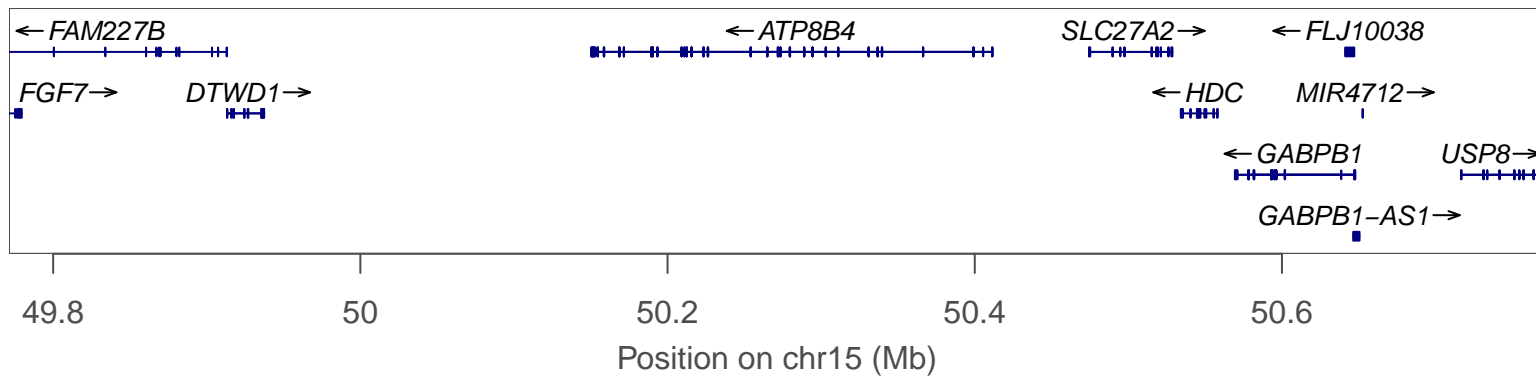
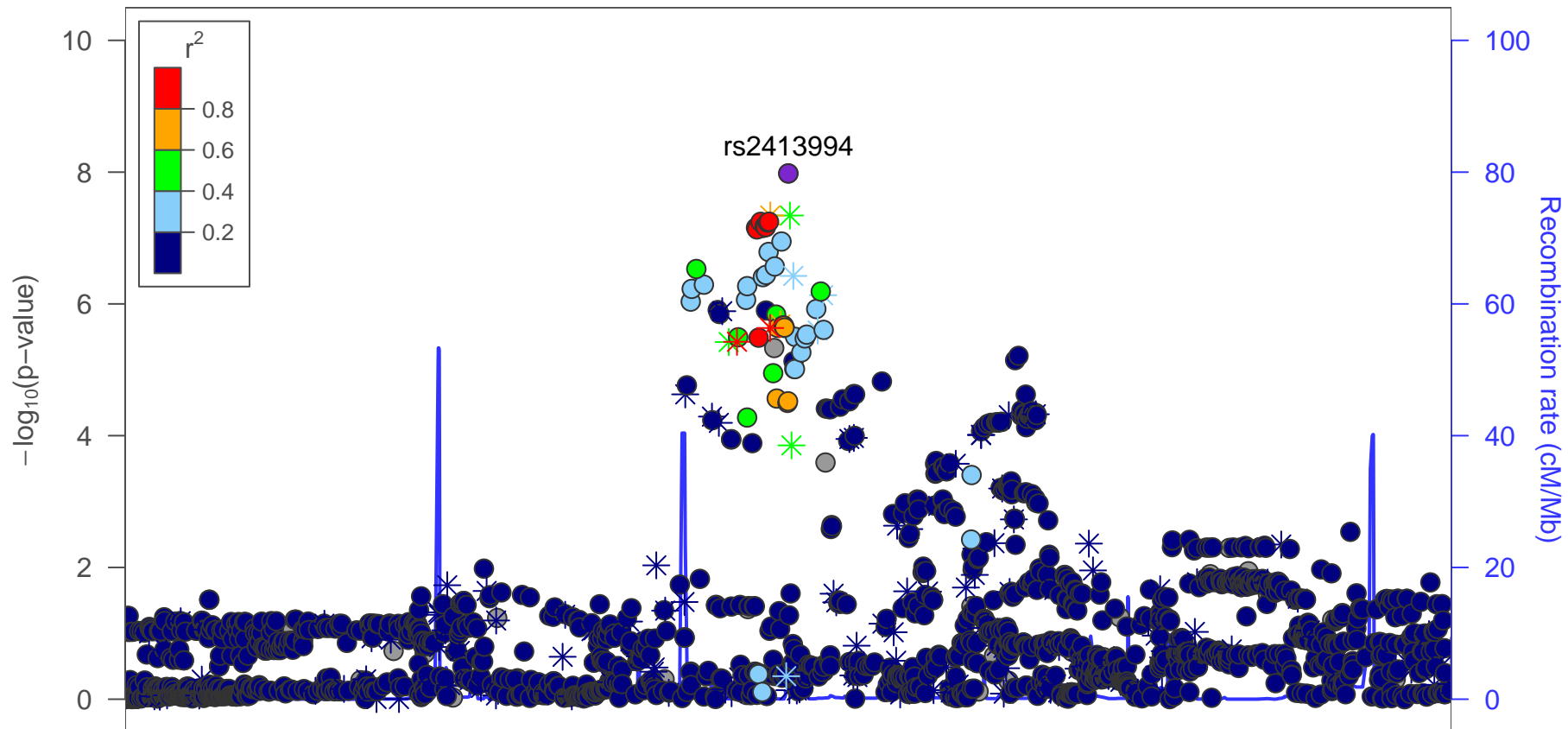


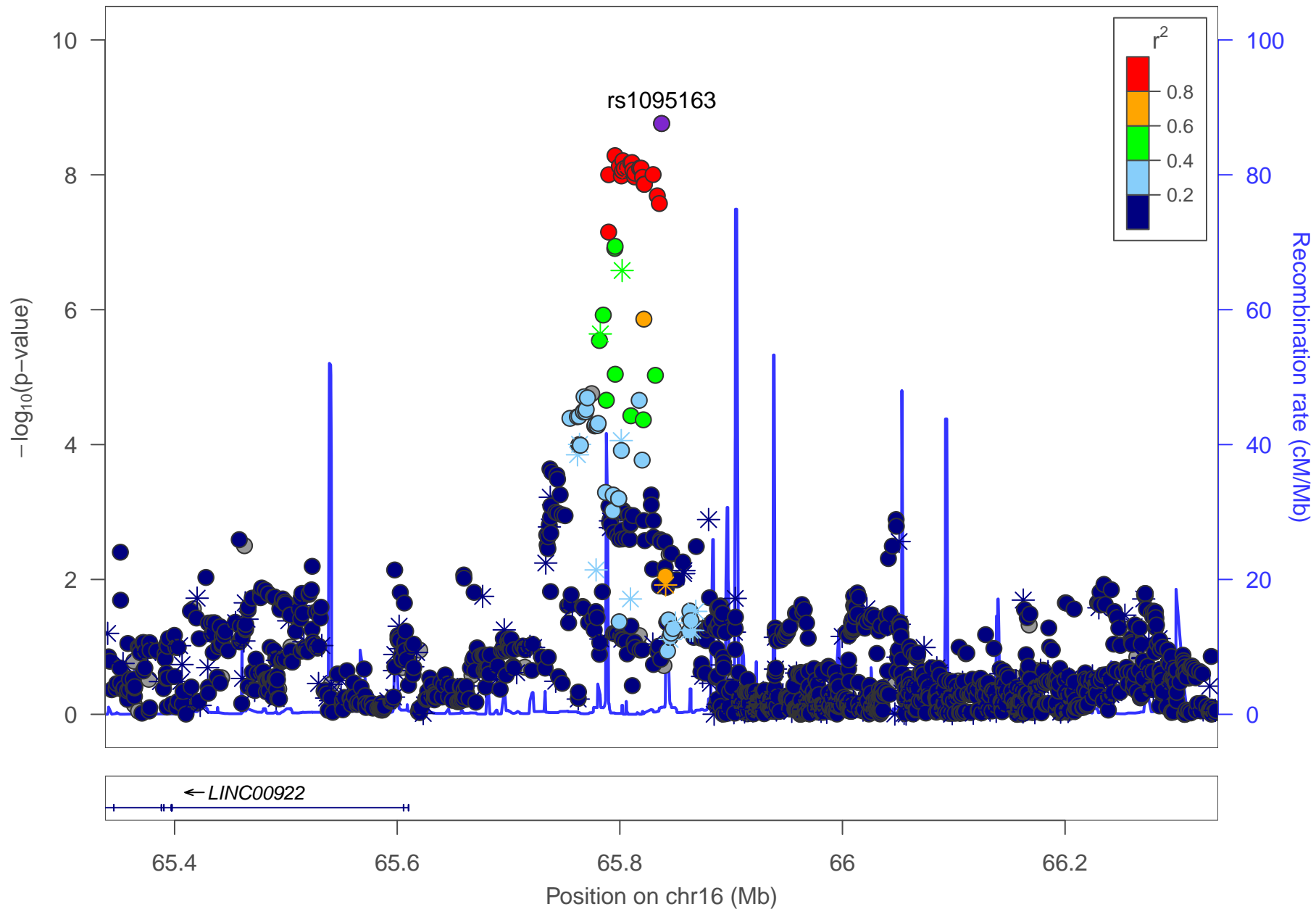


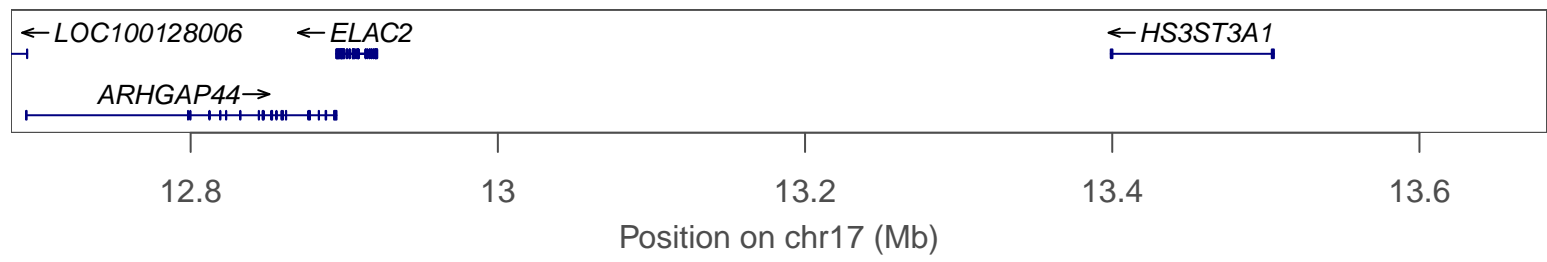
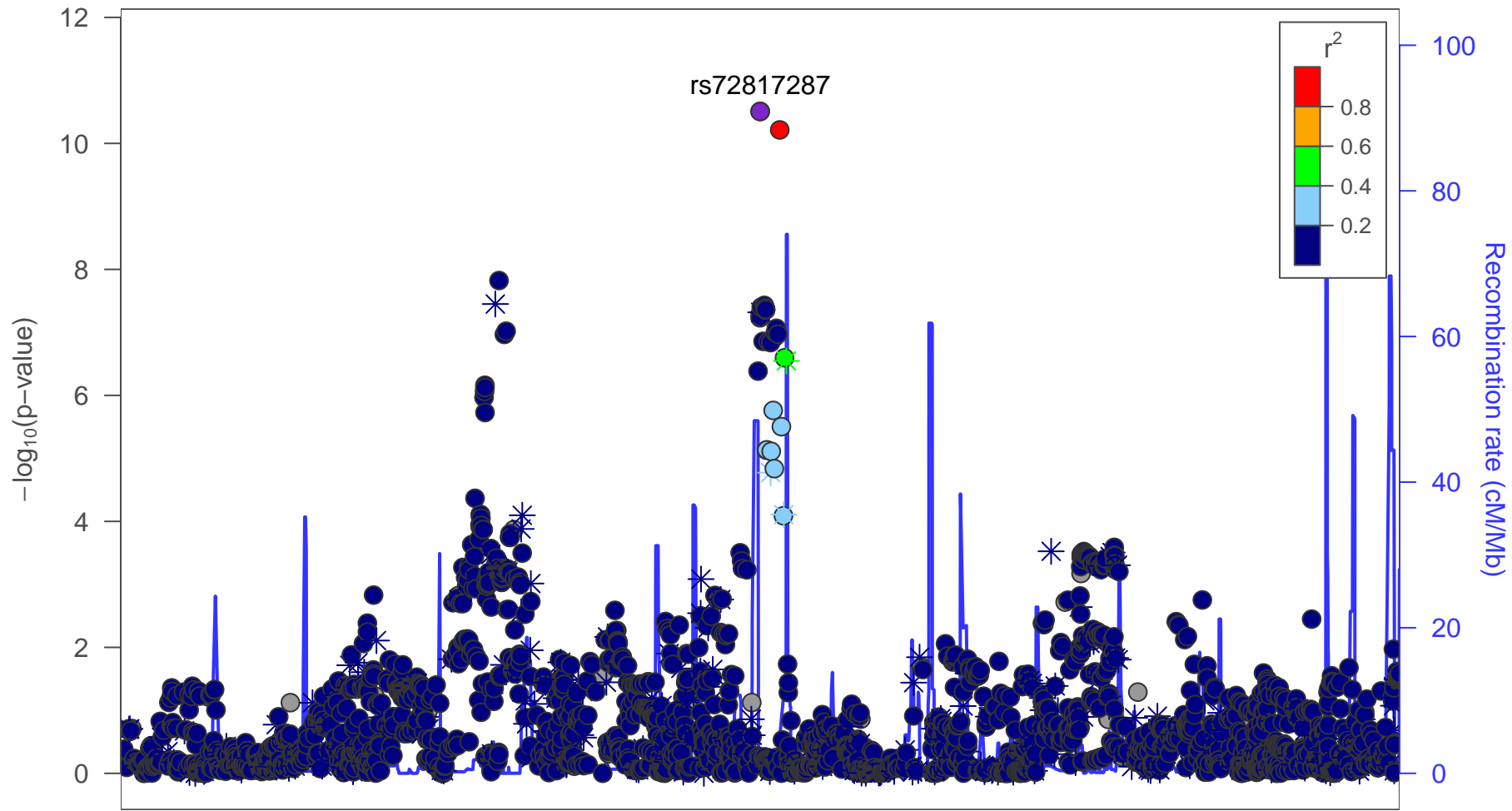


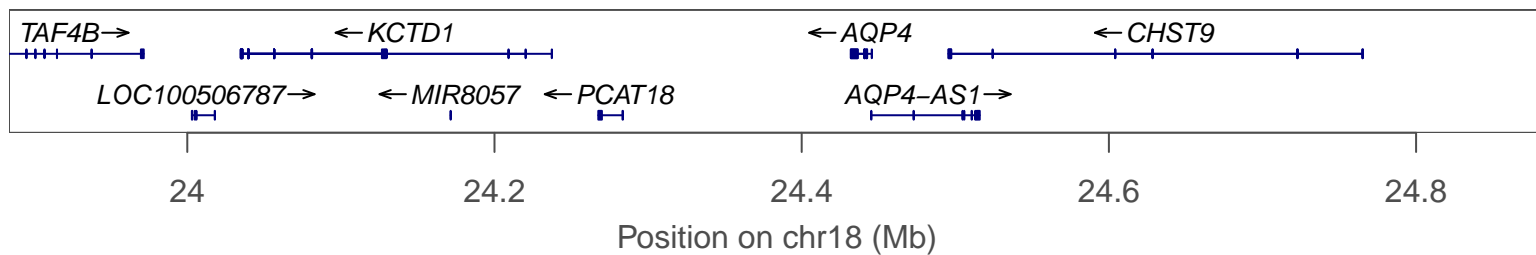
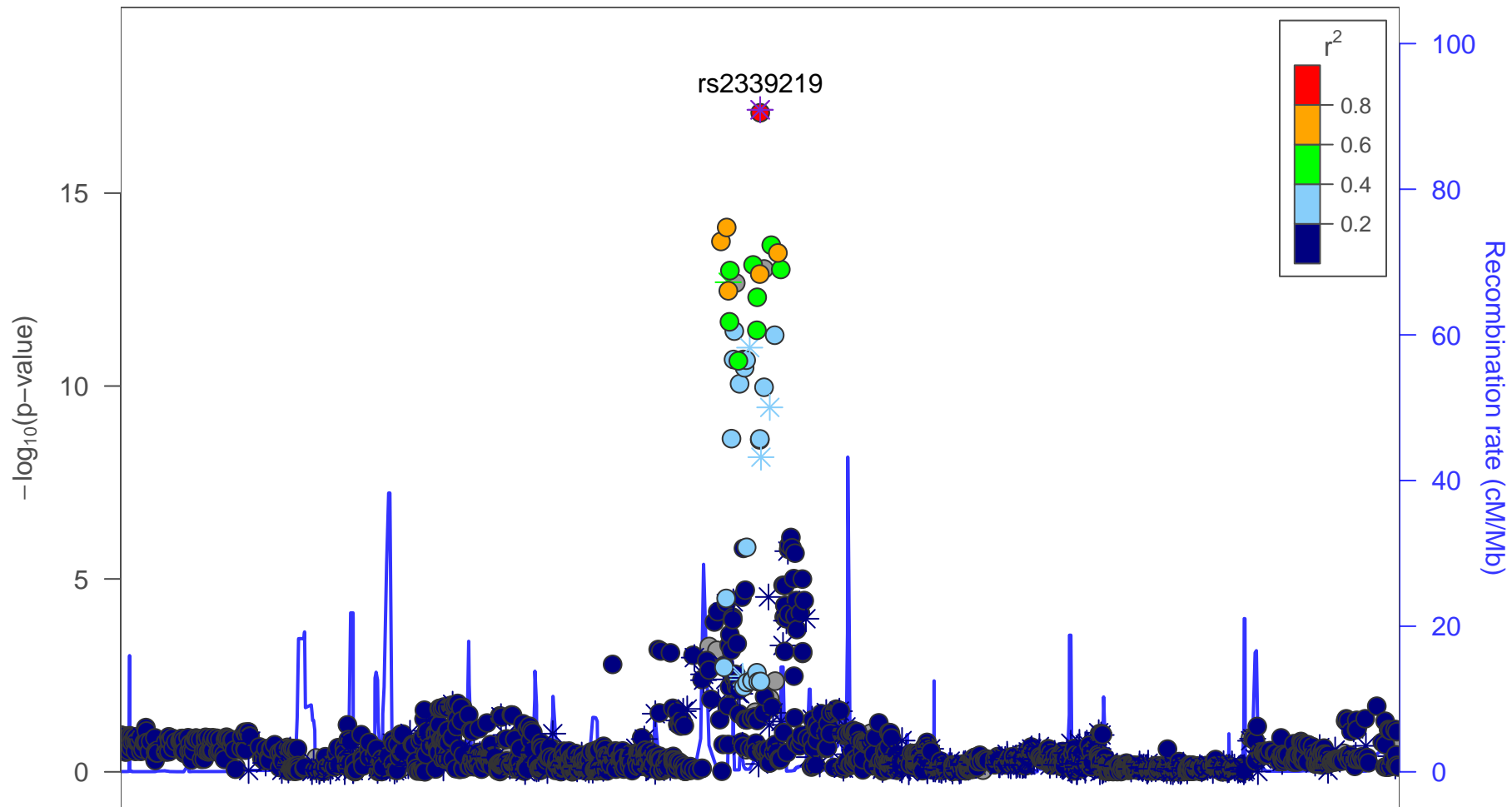


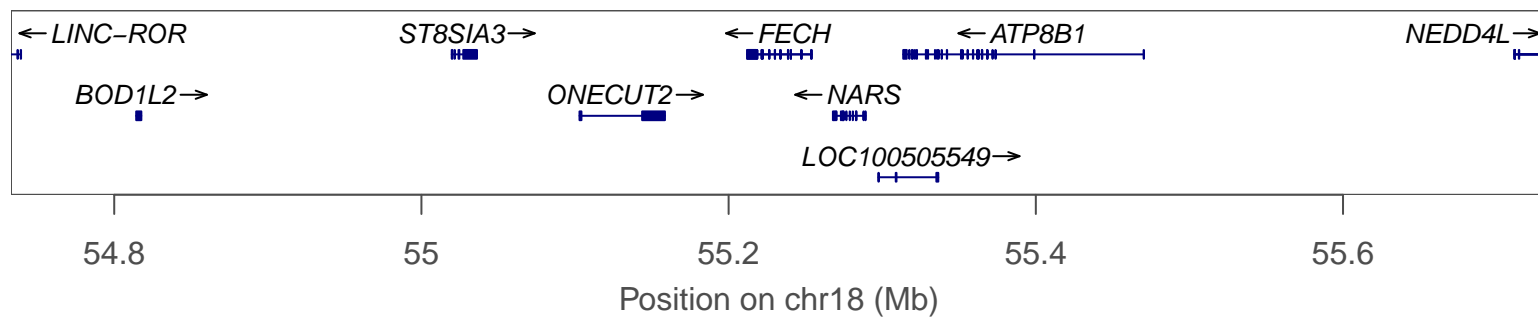
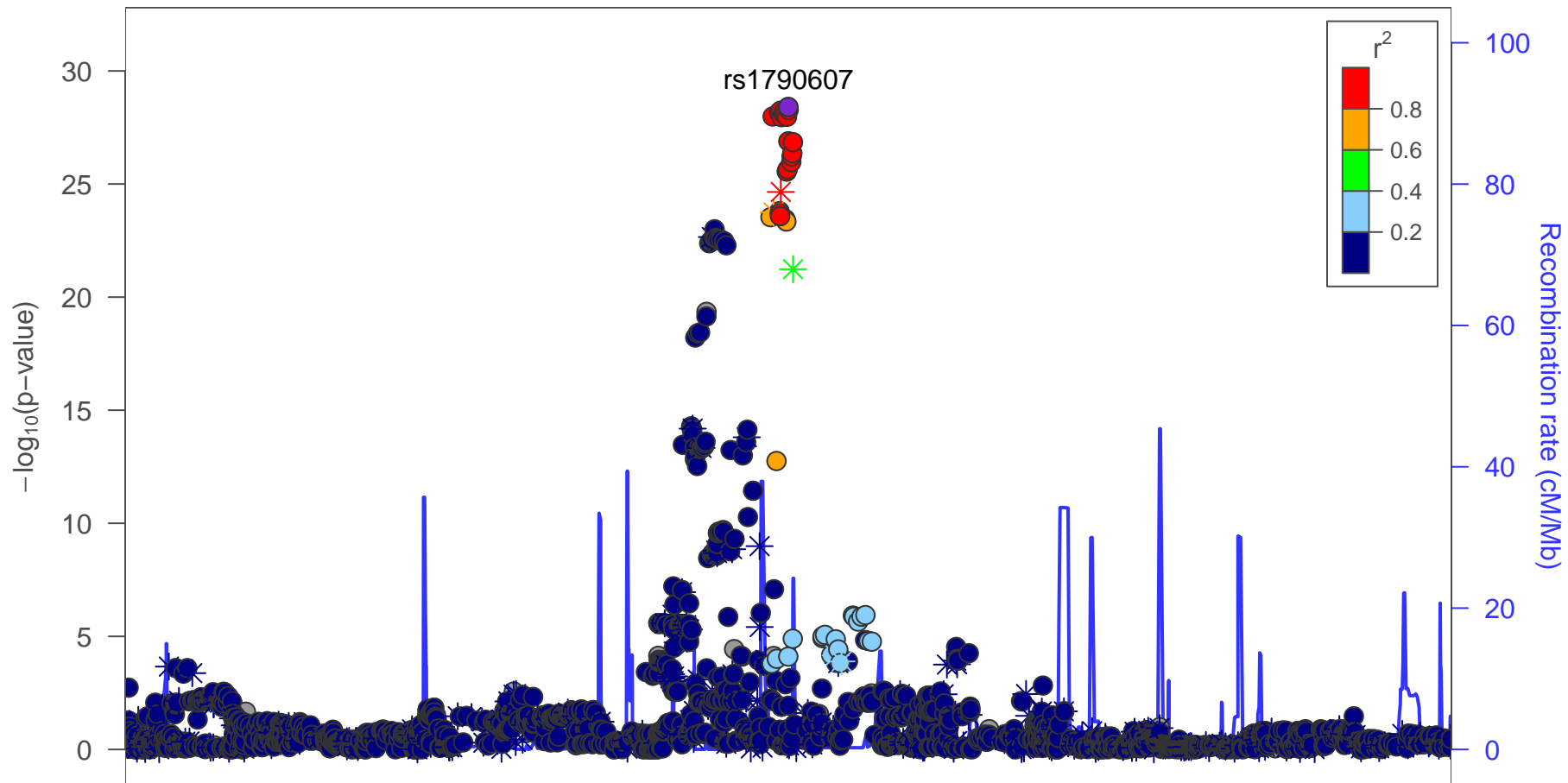


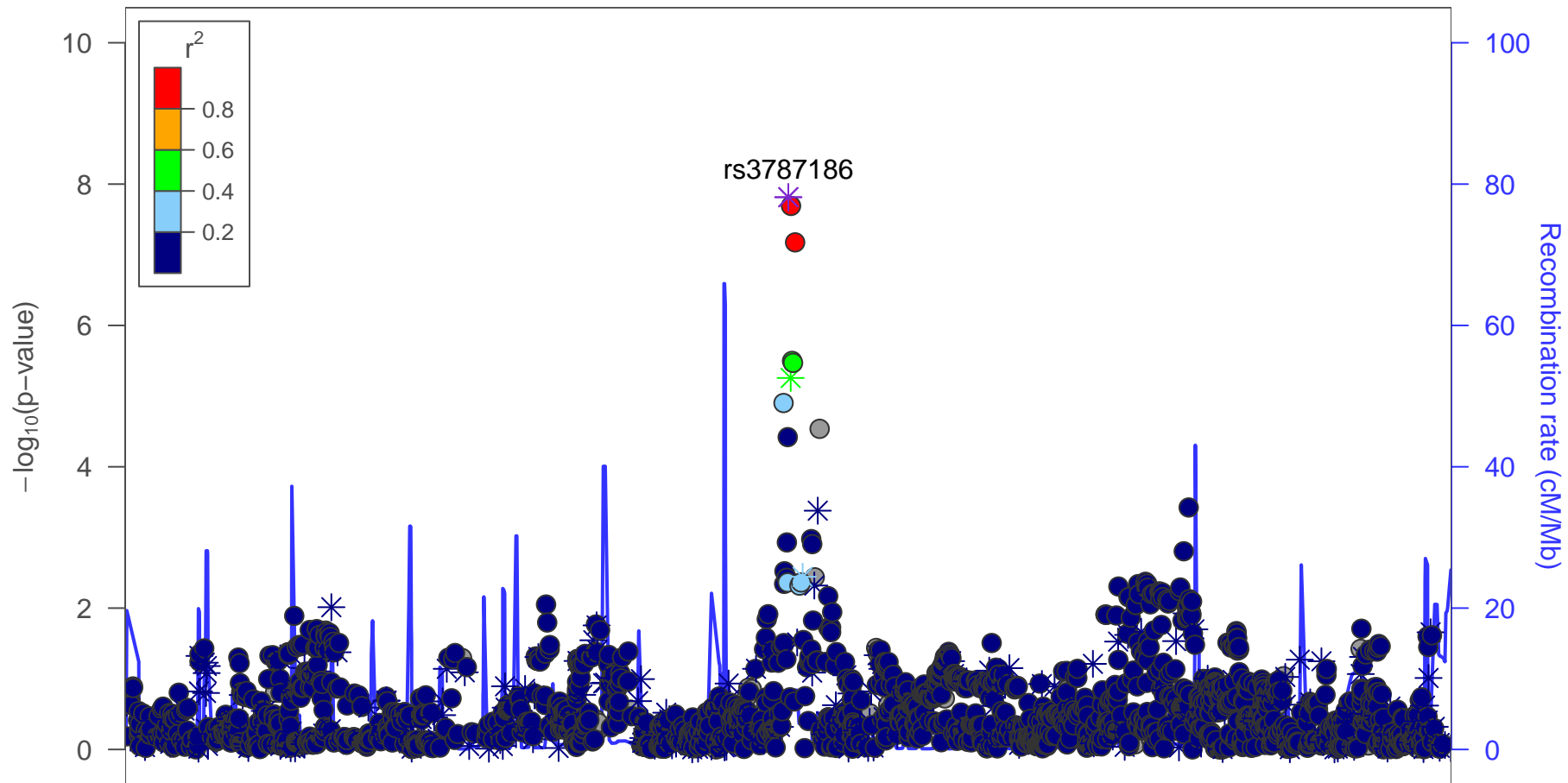












49.8

50

50.2

50.4

50.6

Position on chr20 (Mb)

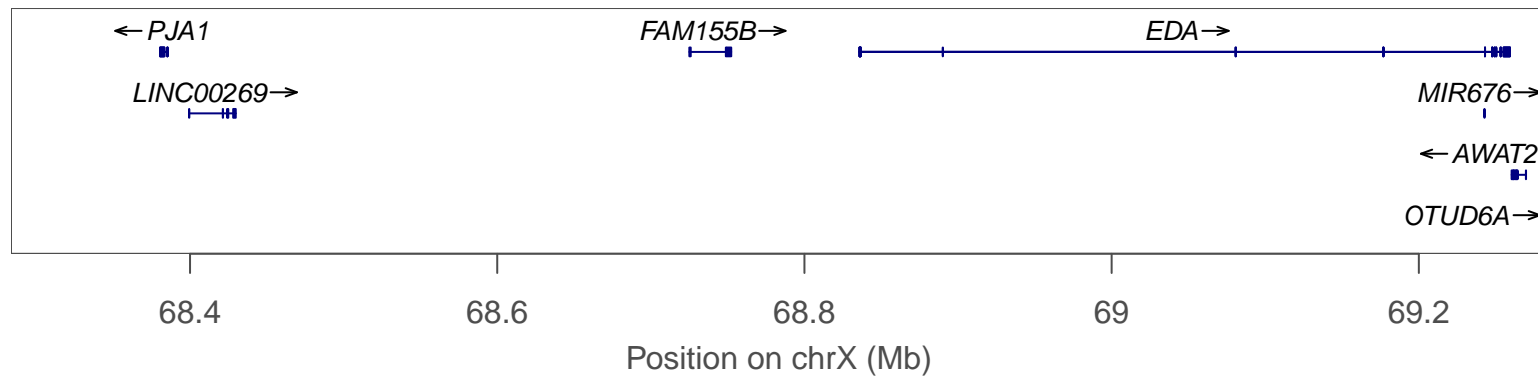
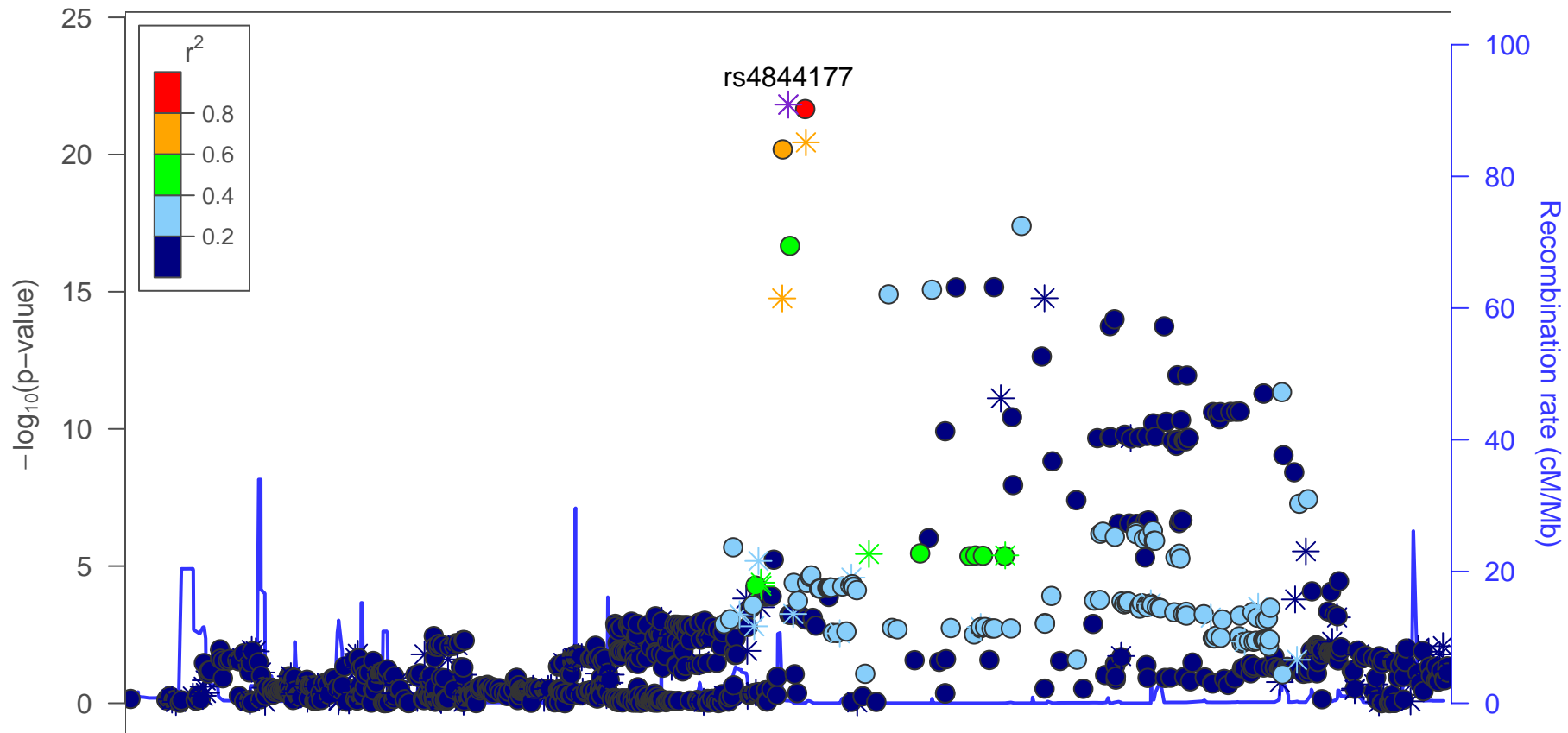
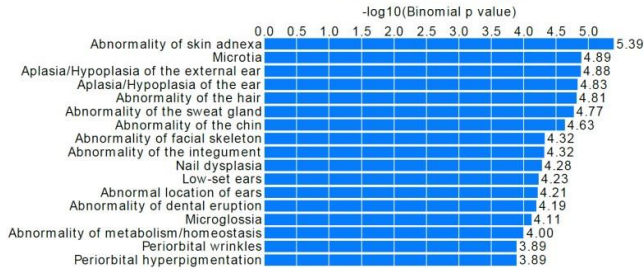
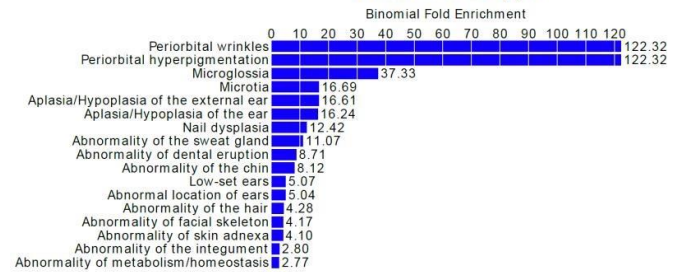


Figure S2: Regional association plots showing significant associations observed in the meta-analysis across all four cohorts.

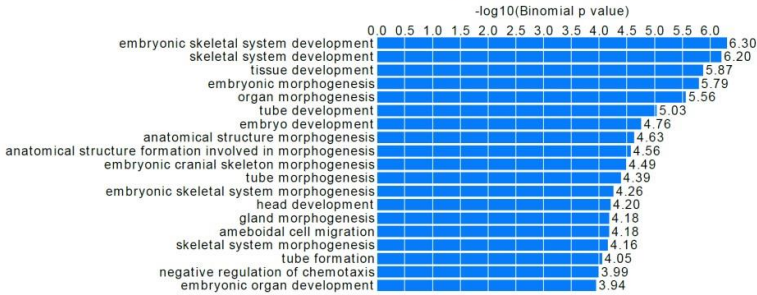
Human Phenotype



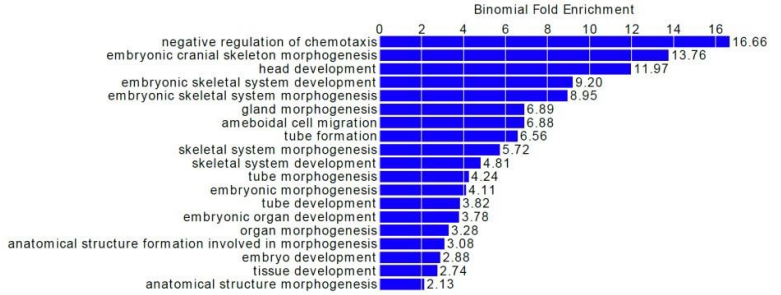
Human Phenotype



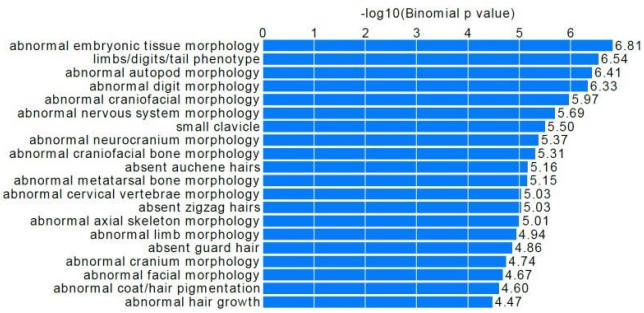
GO Biological Process



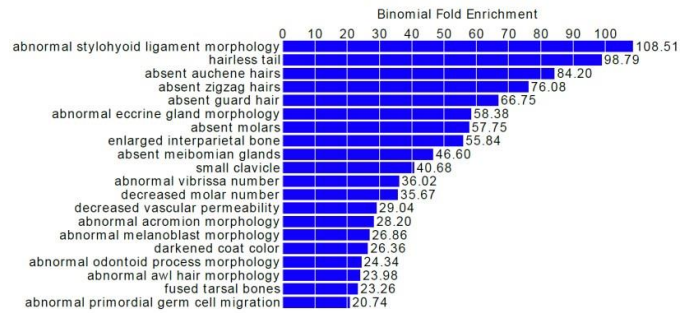
GO Biological Process



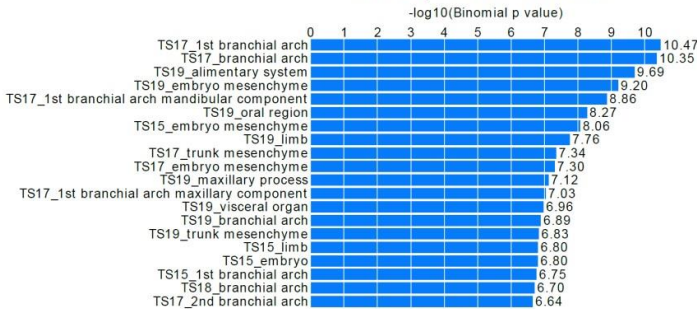
Mouse Phenotype



Mouse Phenotype



MGI Expression: Detected



MGI Expression: Detected

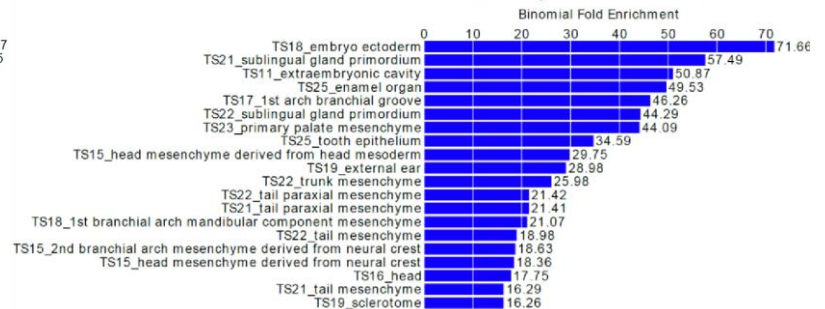


Figure S3: Enrichment for annotation across ontologies for 91 genes (i.e., nearest two genes up to 1000kb from the lead SNP) across 49 loci identified in the meta-analysis across all four cohorts. From top to bottom: human phenotype, Gene Ontology (GO) biological processes, mouse phenotype, and MGI expression ontologies. Left: significant annotations ordered by $-\log_{10}$ -transformed p-values (up to 20 shown). Right: significant annotations order by fold enrichment (up to 20 shown).

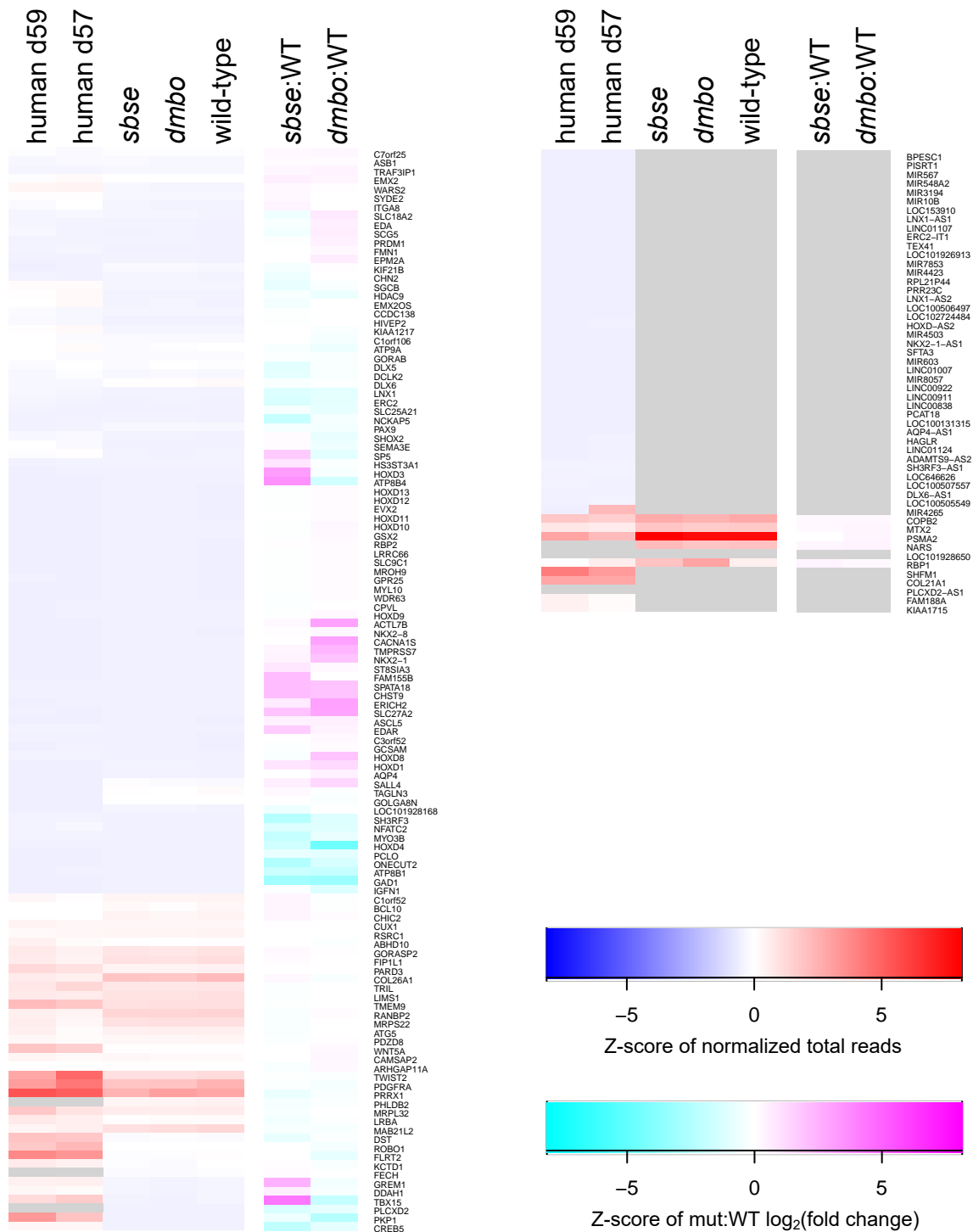


Figure S4: Heatmaps of gene expression in fetal human pinna (at days 59 and 57) and embryonic mouse second branchial arch tissue (in *sbse* mutants, *dmbo* mutants, and wild-type), and fold change of mutant mouse compared to wild-type mouse for 174 genes within 250 kb of the 49 lead SNPs observed in the genetic association meta-analysis across all cohorts. Genes are clustered by expression patterns with the bottom of the heatmap wrapping around to a second column. The shading scale is shown for Z-scores of expression and fold change. Genes for which expression data were not measured are shown in gray.

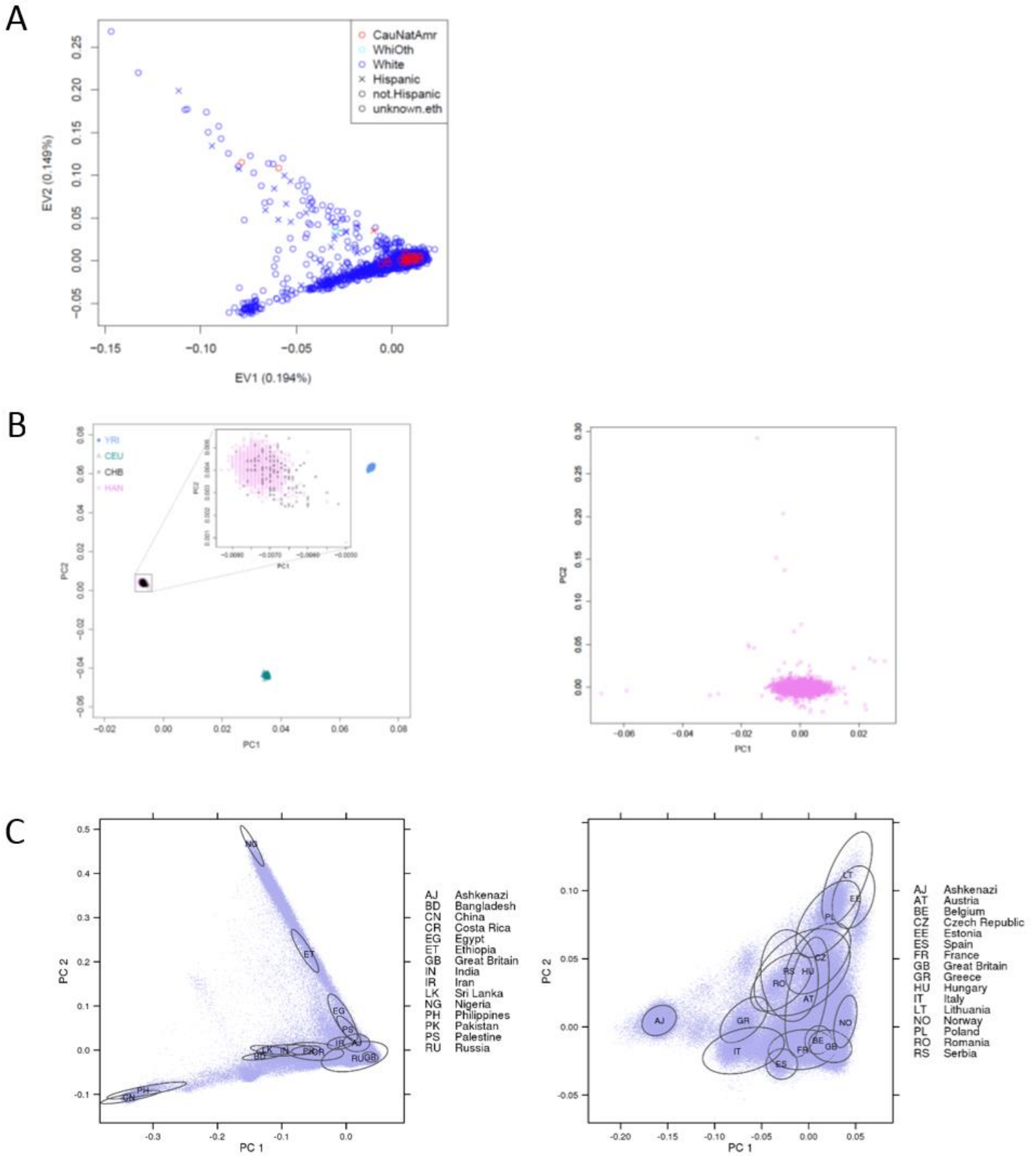


Figure S5: Population structure of cohorts is indicated by principal components of ancestry: (A) European American, (B) Chinese (left, with HapMap controls, right, analysis sample), and (C) 23andMe (left, entire sample, right, European American sample used in the analysis)

Table S1: Characteristics of the cohorts

	European American	Latin American	Chinese	23andMe European ancestry
N	1,791	5062	2,857	64,950
sex (% female)	59.46	52.7	63.67	51.01
age (mean [range])	22.35 [3-40]	24.4 [18 – 43]	56.14 [31-87]	
earlobe attachment (%)				
free	82.52	11.7	3.12	73.75
partially attached	9.94	75.6	31.92	
attached	7.54	12.7	64.96	26.25
ancestry	European	Admixed	East Asian	European
genotyping platform	Illumina HumanOmni Express+Exome	Illumina HumanOmni Express	Illumina HumanOmni Zhonghua-8	Illumina HumanHap550, HumanOmni Express, or custom array
# genotyped SNPs passing QC	653,629	671,772	795,597	1,044,759
# imputed SNPs	10,156,807	9,143,600	7,383,741	15,573,758
regression analysis	linear	linear	linear	logistic
covariates	Age, sex	age, sex, height, BMI	age, sex, BMI	age, sex, genotyping platform
principal components of ancestry included	3	5	0 ^a	5
genomic inflation factor	0.994	1.02	1.037	1.254 ^b

^a Principal components analysis indicated negligible population structure, therefore no adjustments were made for principal components of ancestry in the Chinese sample.

^b Genomic inflation for a subset of 1000 individuals with attached vs. 1000 with unattached earlobes was 1.002.

Table S3: Genetic data quality control filtering criteria

Filter	SNPs omitted	SNPs cumulatively retained
European American		
None (all SNPs)		968,515
Technical filters	8,470	960,045
Missing call rate $\geq 2\%$	9,675	950,370
>1 discordant calls in 69 duplicates ^a	26	950,344
>1 Mendelian error across 8 HapMap trios	122	950,222
HWE p-value $< 10^{-4}$	2,038	948,184
Allele frequency difference ≥ 0.2 between sexes ^b	274	947,910
Heterozygosity difference ≥ 0.3 between sexes ^b	41	947,869
Positional duplicates	19,597	928,272
Monomorphic (MAF = 0)	108,485	819,787
MAF < 0.01	164,959	654,828
non-autosomal or X	1,199	653,629
Latin American		
None (all SNPs)		730,525
Chromosome 0 (SNPs without assigned location)	1230	729,295
Missing call rate $\geq 5\%$	3550	725,745
MAF < 0.01 (includes monomorphic)	53973	671,772
Chinese cohort		
None (all SNPs)		887,270
Technical filters	278	886,992
Missing call rate $\geq 2\%$	18,628	868,364
Minor Allele Frequency < 0.01	67,873	800,491
HWE p-value $< 10^{-3}$	2,588	797,903
Positional duplicates	2,306	795,597

^a one duplicate was removed from QC filters due to a chromosomal anomaly

^b filter applied to SNPs on autosomes and XY pseudo-autosomal region

Table S4: Summary of imputation methods and quality control filters

	European American	Latin American	Chinese	23andMe
Reference data source	1000 Genomes Project Phase 3	1000 Genomes Project Phase 1	1000 Genomes Project Phase 3	1000 Genomes Project Phase 1 custom tool
Pre-phasing	SHAPEIT2	SHAPEIT2	SHAPEIT2	based on BEAGLE
Imputation	IMPUTE2 masked variant analysis	IMPUTE2 masked variant analysis ⁰	IMPUTE2	Minimac2
Other quality control procedures			-	-
Imputation filters				
genotype probability (per SNP per person)	> 0.9	> 0.8	> 0.9	-
INFO score (per SNP)	> 0.5	> 0.4	> 0.6	-
MAF	> 0.01	> 0.01	> 0.02	-
Missing rate	-	-	< 0.05	-
HWE	-	-	≥ 10e-6	-
concordance (concord_type ⁰) for masked variant analysis	-	> 0.7	-	-
info_type ⁰ -concord_type ⁰ (chip genotype quality measure)	-	> 0.1	-	-