

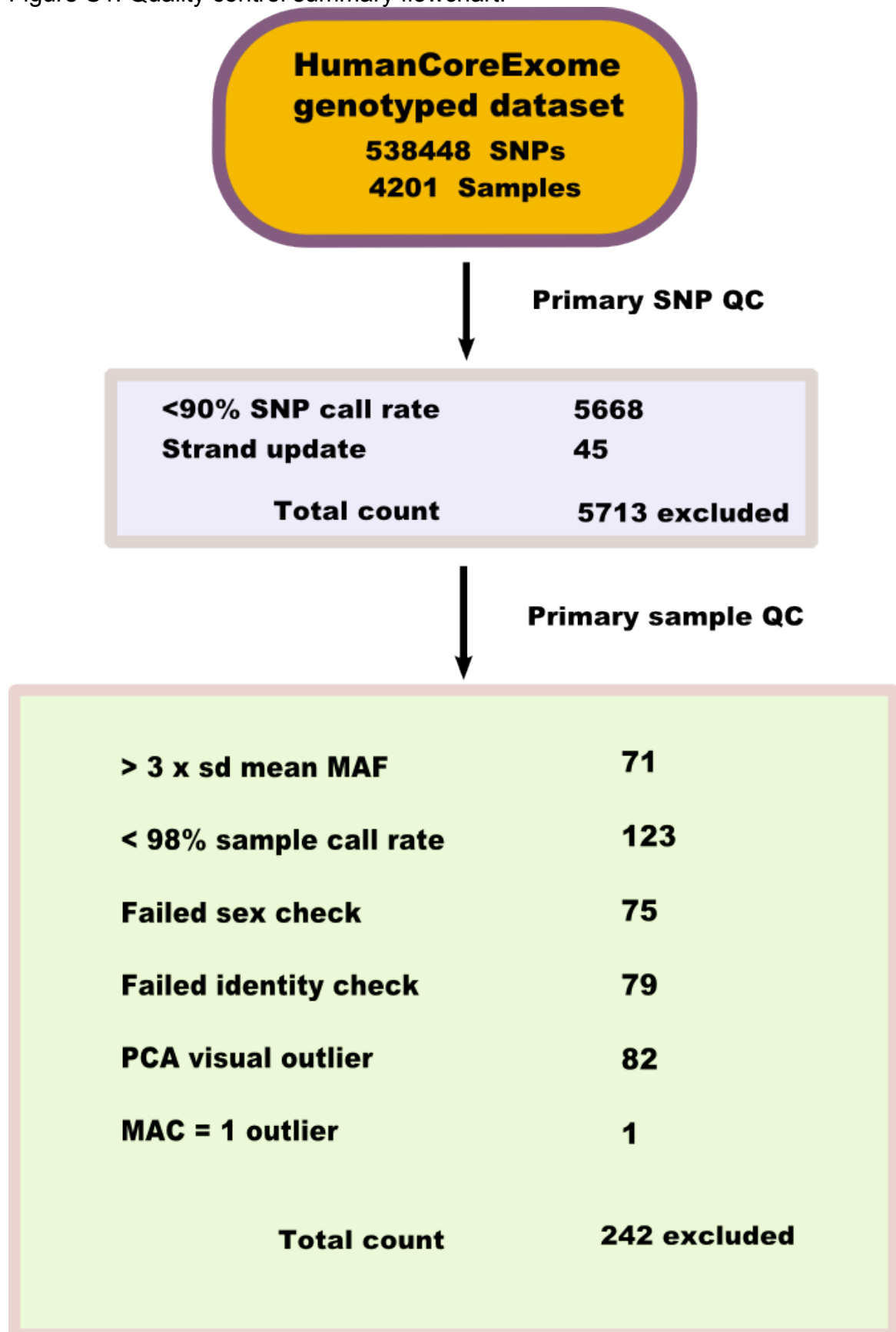
**The American Journal of Human Genetics, Volume 101**

## **Supplemental Data**

### **A Genome-wide Association Study of Dupuytren Disease Reveals 17 Additional Variants Implicated in Fibrosis**

**Michael Ng, Dipti Thakkar, Lorraine Southam, Paul Werker, Roel Ophoff, Kerstin Becker, Michael Nothnagel, Andre Franke, Peter Nürnberg, Ana Isabel Espirito-Santo, David Izadi, Hans Christian Hennies, Jagdeep Nanchahal, Eleftheria Zeggini, and Dominic Furniss**

Figure S1. Quality control summary flowchart.



**Secondary SNP QC**

<b>&lt;98% call rate</b>	<b>13080</b>
<b>HWE <math>p &lt; 0.0001</math></b>	<b>1813</b>
<b>Cluster separation score <math>&lt; 0.4</math></b>	<b>1798</b>
<b>X chr. heterozygous haploid</b>	<b>967</b>
<b>Not autosomal</b>	<b>13289</b>
<b>Duplication</b>	<b>6727</b>
<b>Total count</b>	<b>35705 excluded</b>

**Merged with UKHLS dataset  
4891 samples**

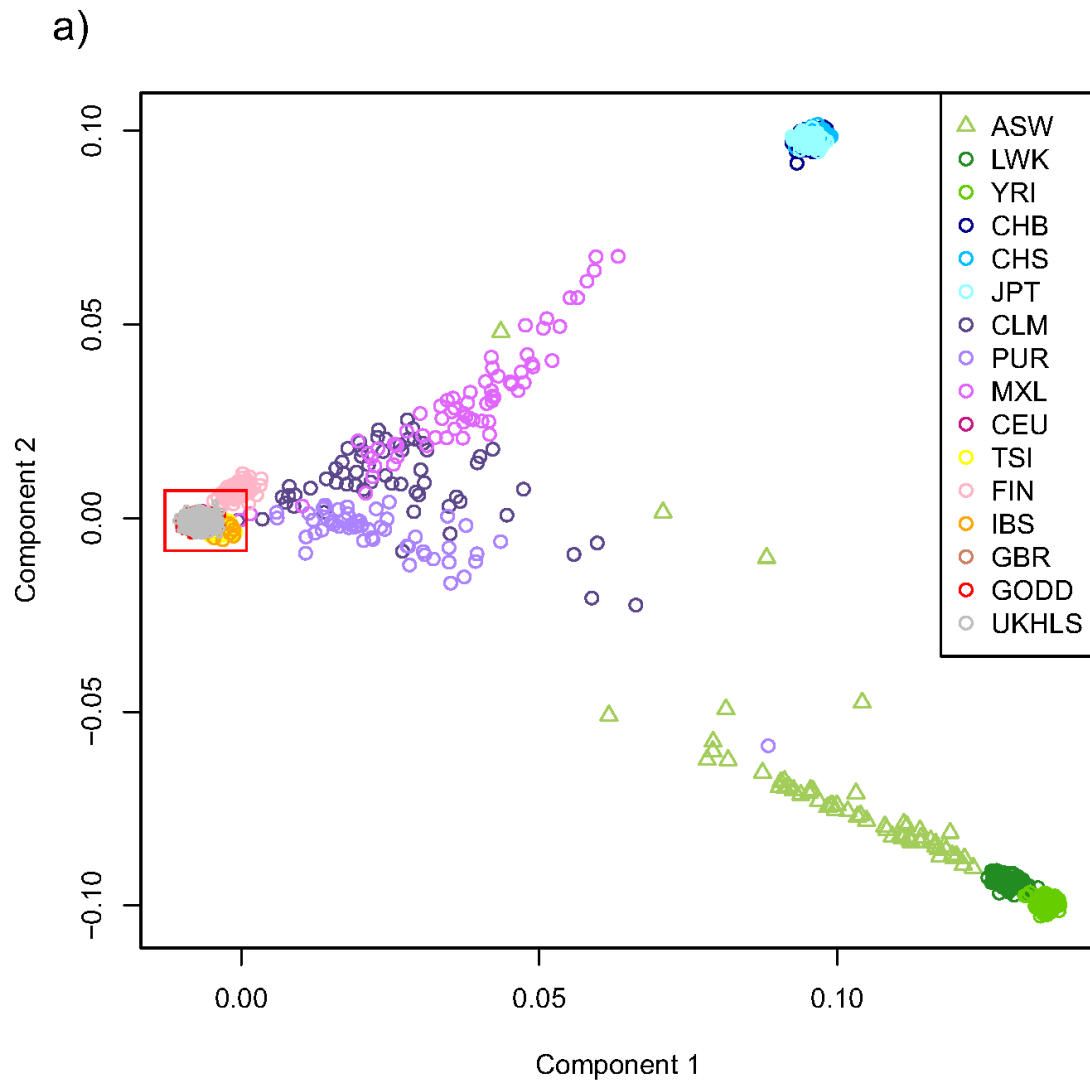
**Secondary sample QC**

<b>Relatedness PIHAT <math>\geq 0.185</math></b>	<b>288</b>
<b>Spurious genotype calling</b>	<b>1</b>
<b>PCA visual outlier</b>	<b>4</b>
<b>Total count</b>	<b>293 excluded</b>

**QC complete**

**Post QC dataset  
494982 SNPs  
3871 cases  
4686 controls**

Figure S2. Principal component analysis plots. a) Principle components 1 and 2 plotted for cases and controls and the 1000 genomes project samples. b) Close-up view of the area highlighted in (a) demonstrating overlap of BSSH-GODD (case) and UKHLS (control) samples. Three-letter abbreviations are those used in the 1000 Genomes Project. GODD – cases from the BSSH-GODD discovery cohort. UKHLS – Controls from the UKHLS cohort.



b)

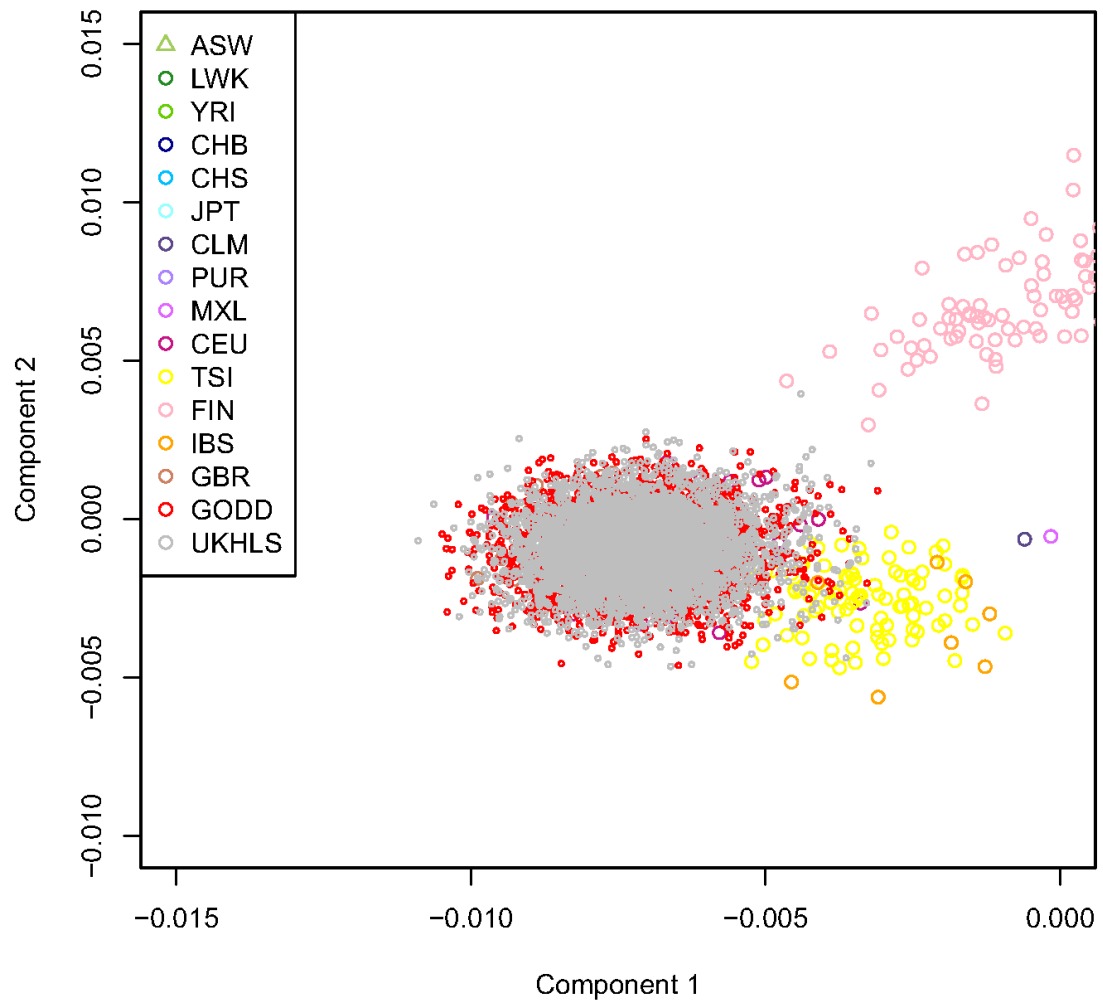


Figure S3. Q-Q plot. Observed versus expected p-values of association in the discovery cohort. The unadjusted genomic control inflation factor ( $\lambda_{GC}$ ) was 1.104. The genomic control inflation factor adjusted to a case control sample size of 1000 ( $\lambda_{1000}$ )<sup>2</sup> was 1.025

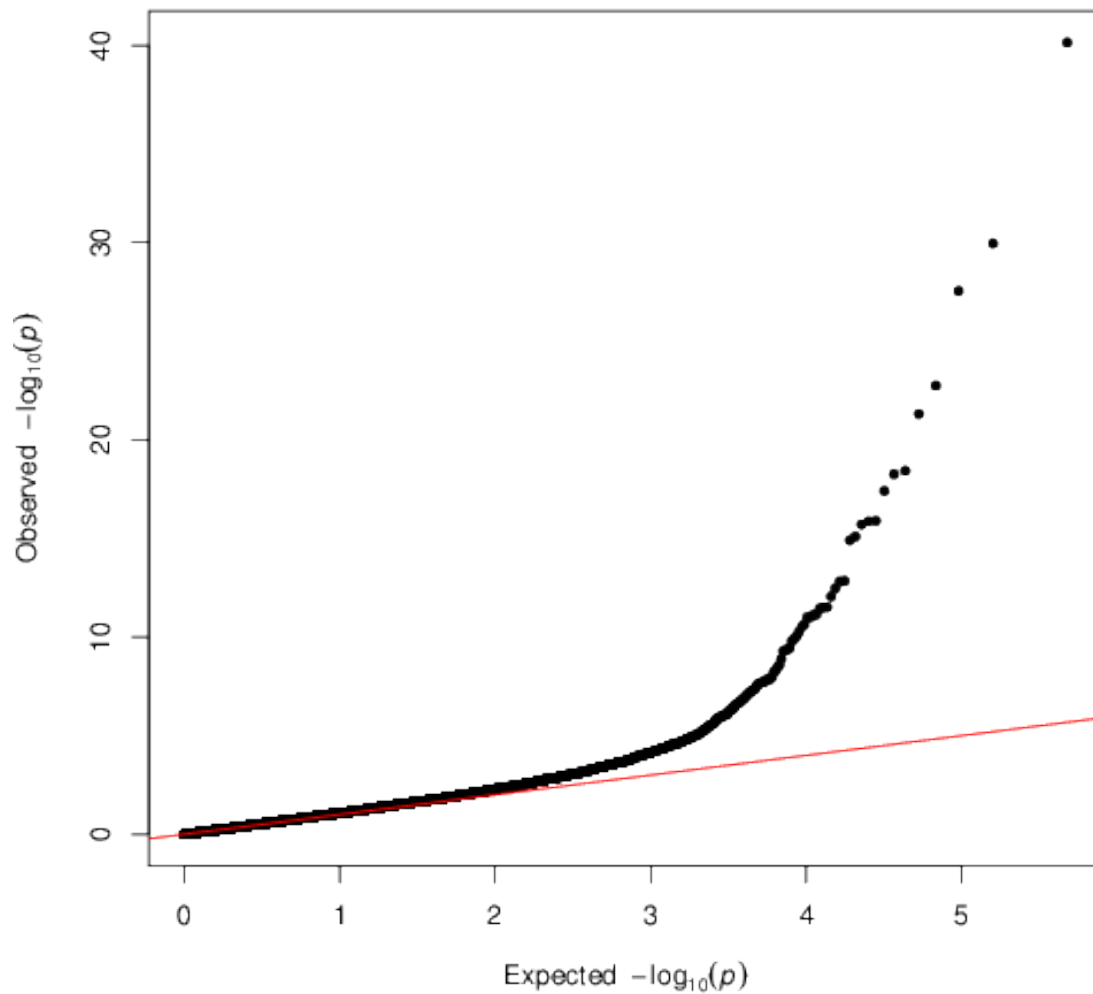
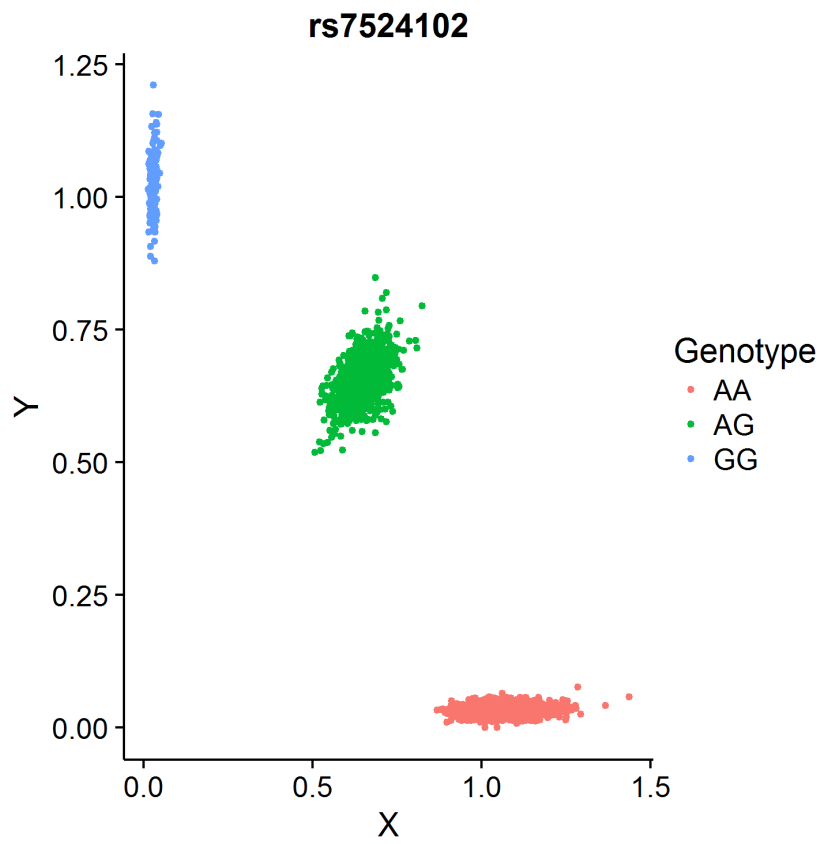
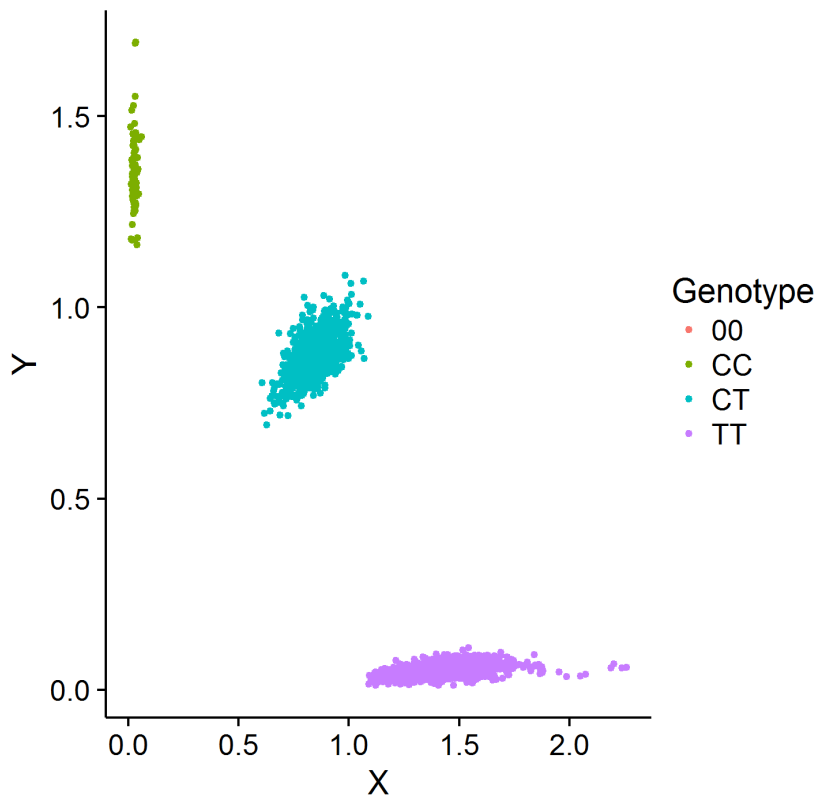


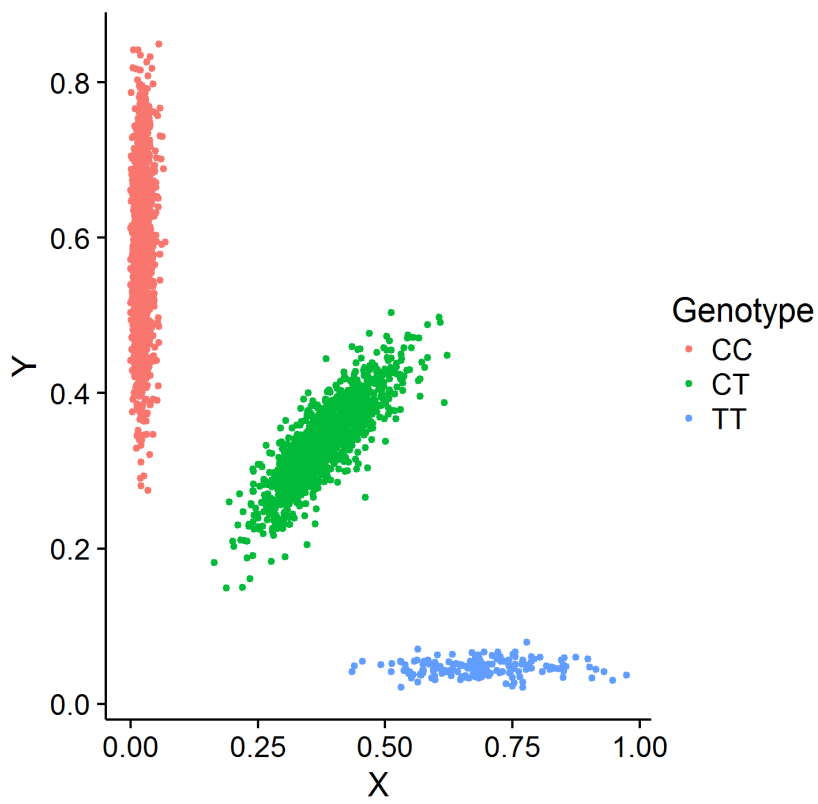
Figure S4. Raw intensity scatter plots of all genome wide significant associated variants generated using R. 00 denotes missing genotype.



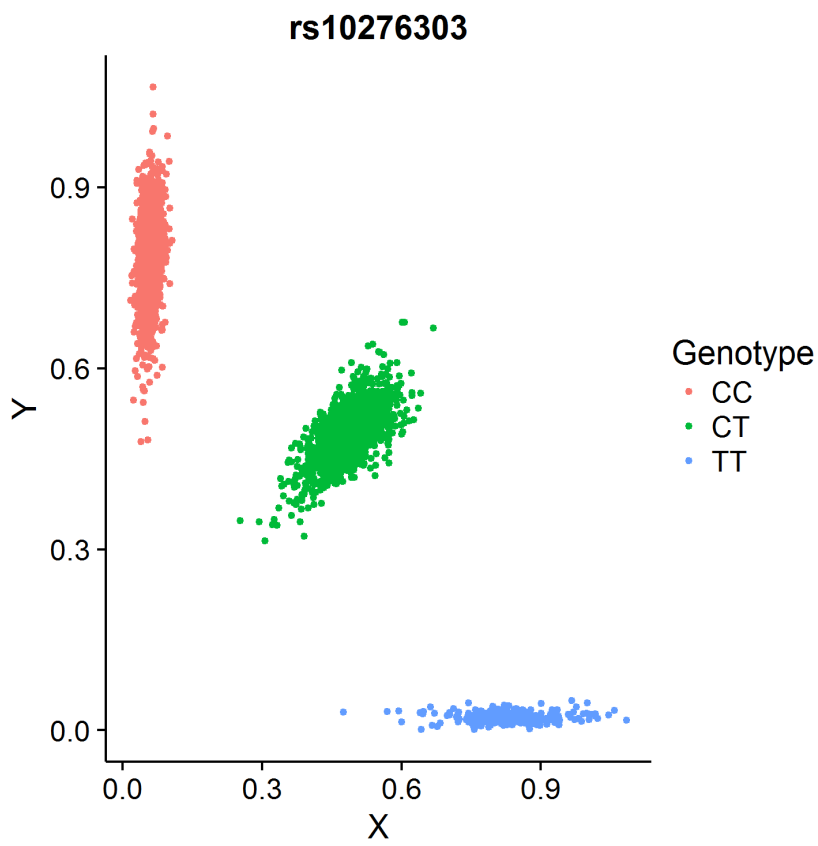
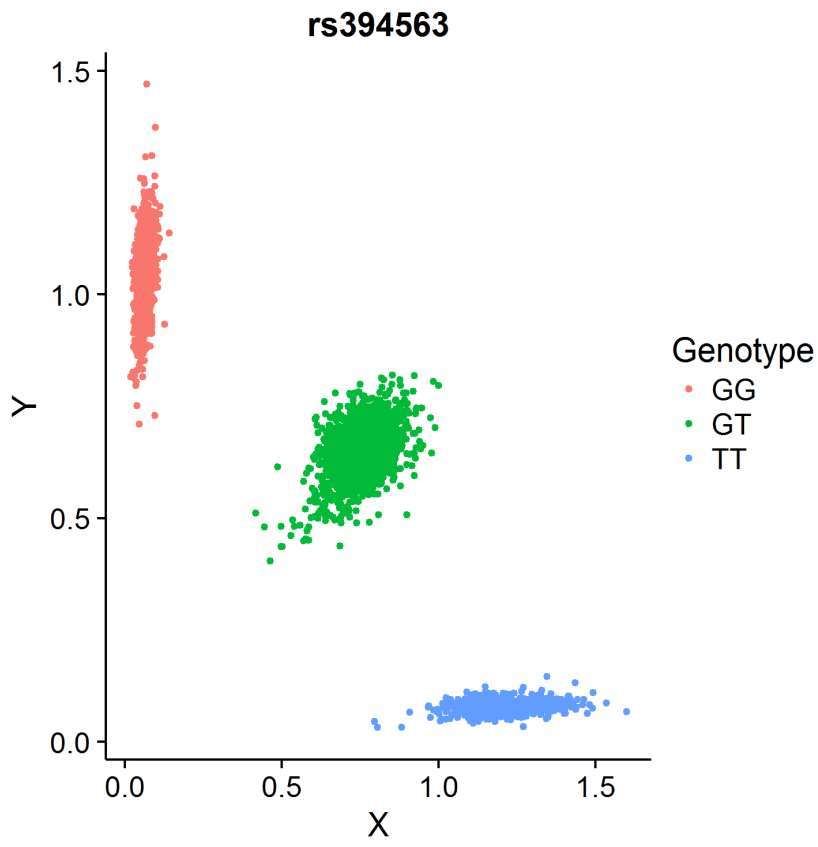
**rs17433710**

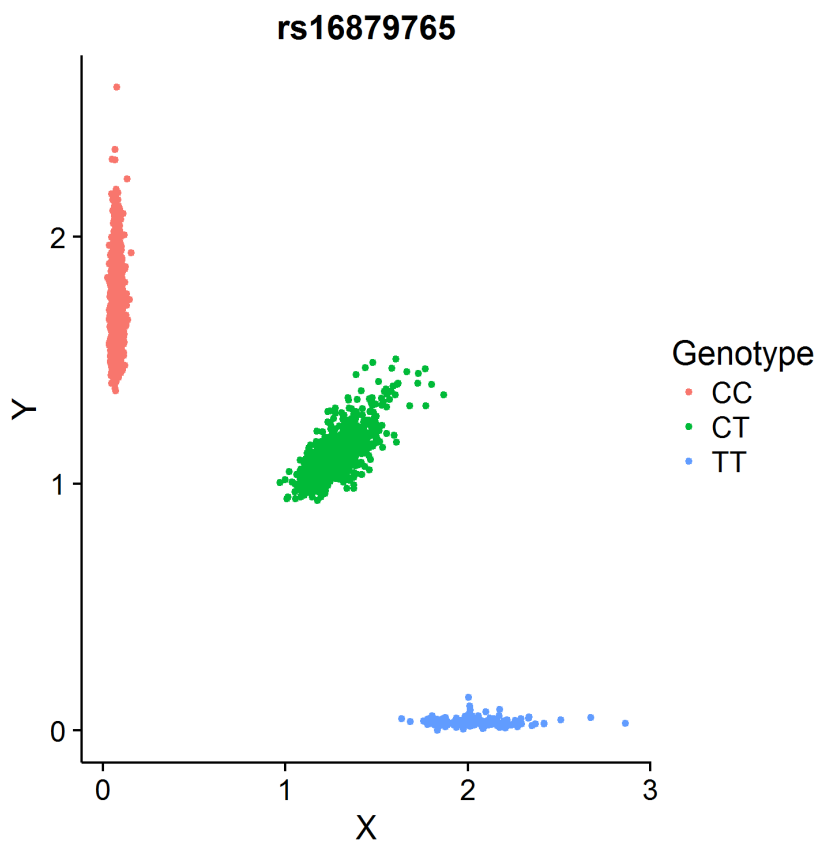
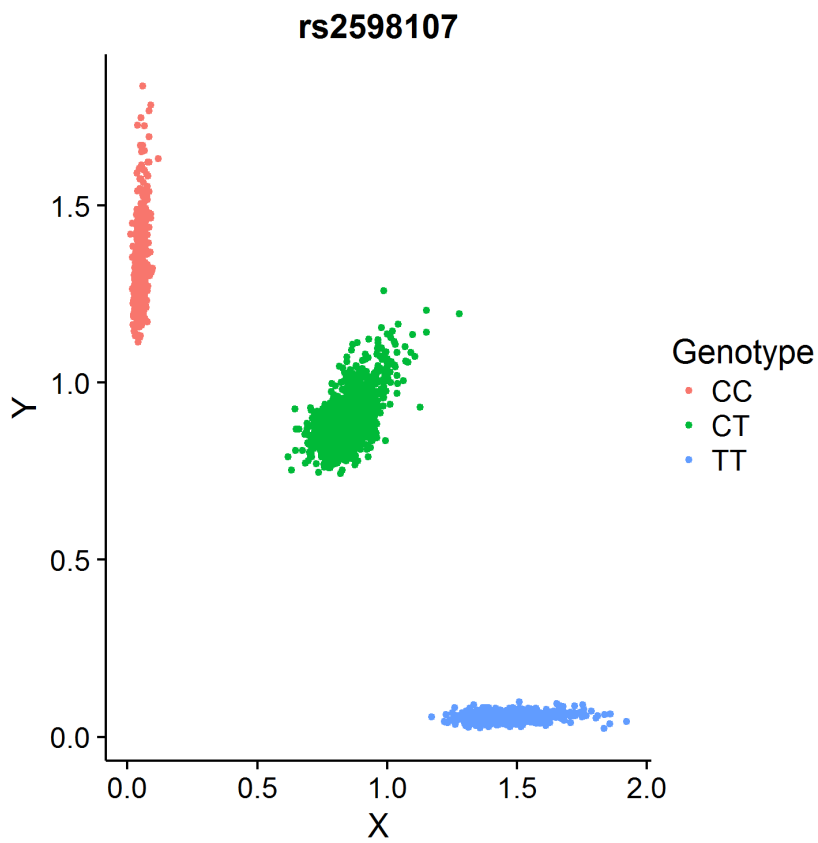


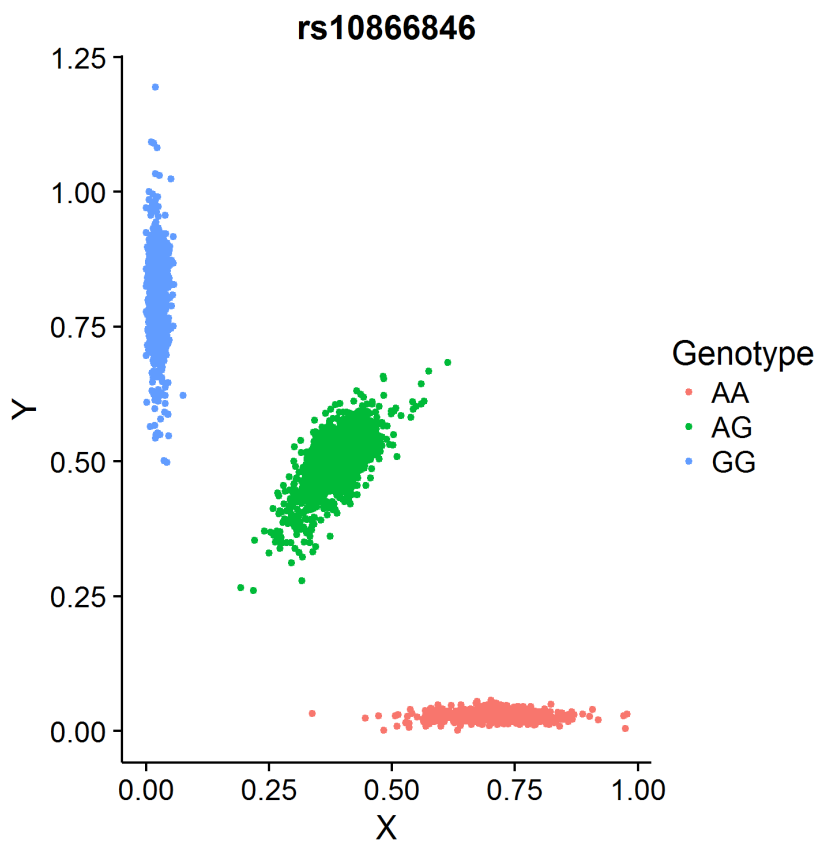
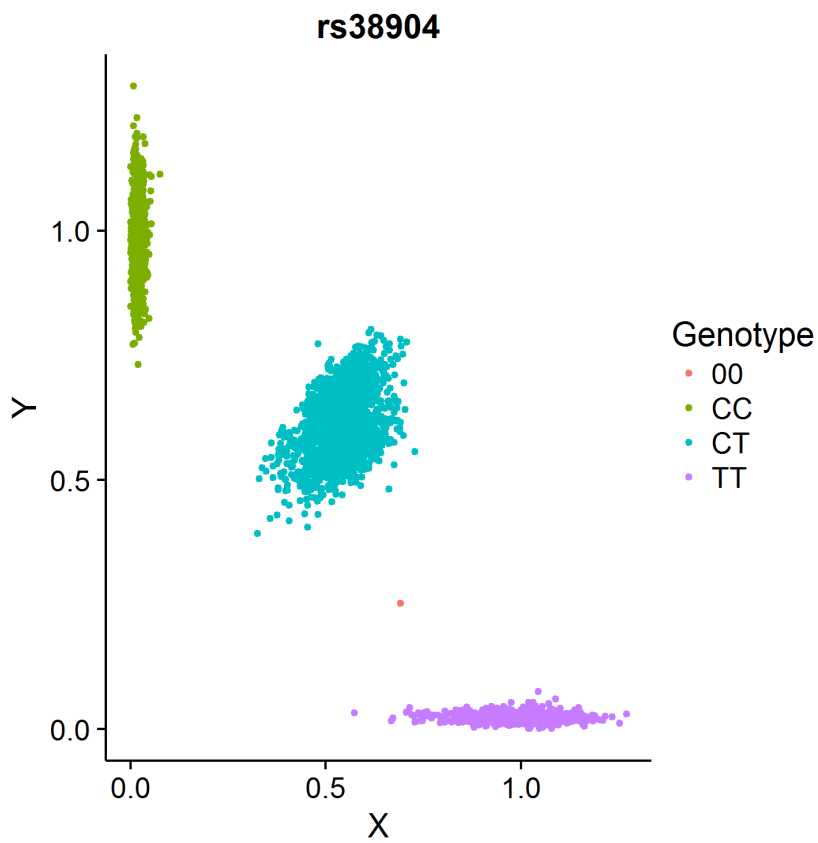
**rs246105**

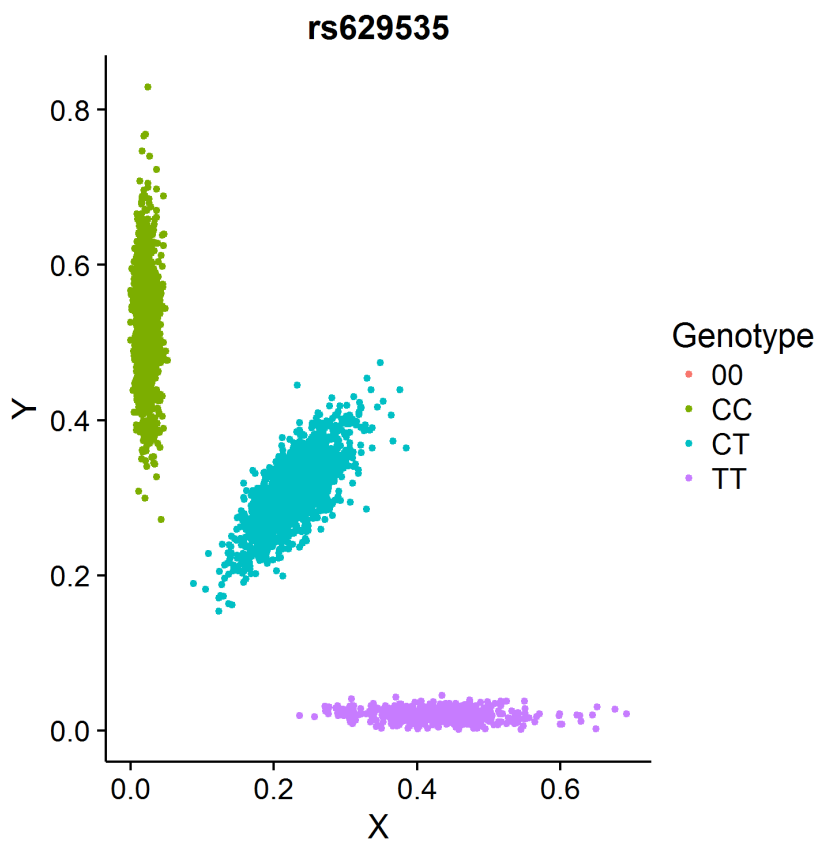
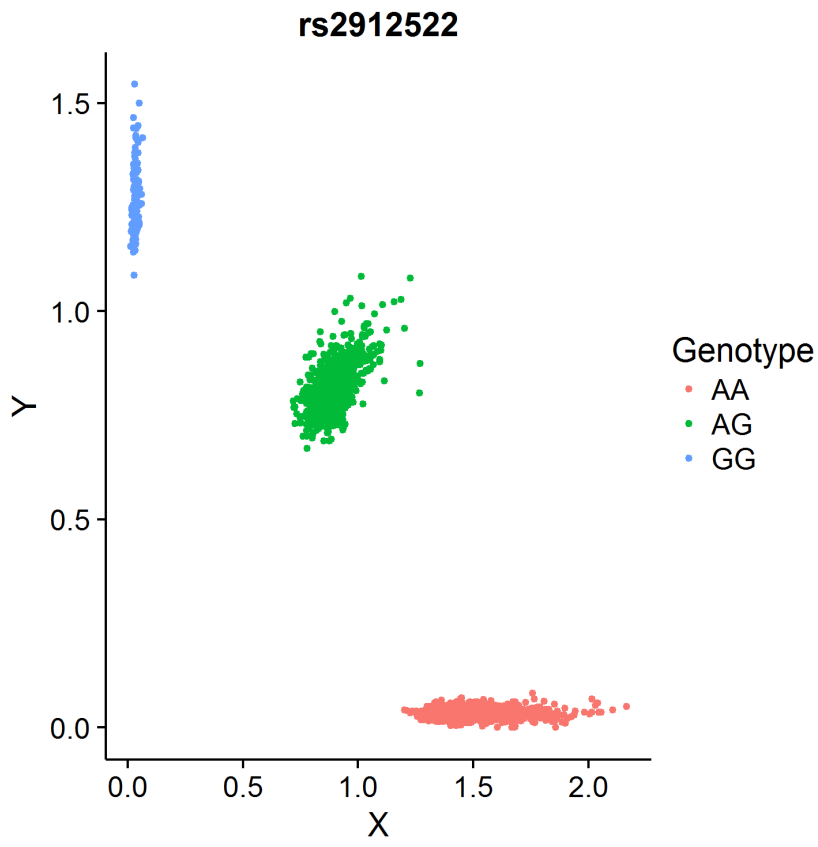


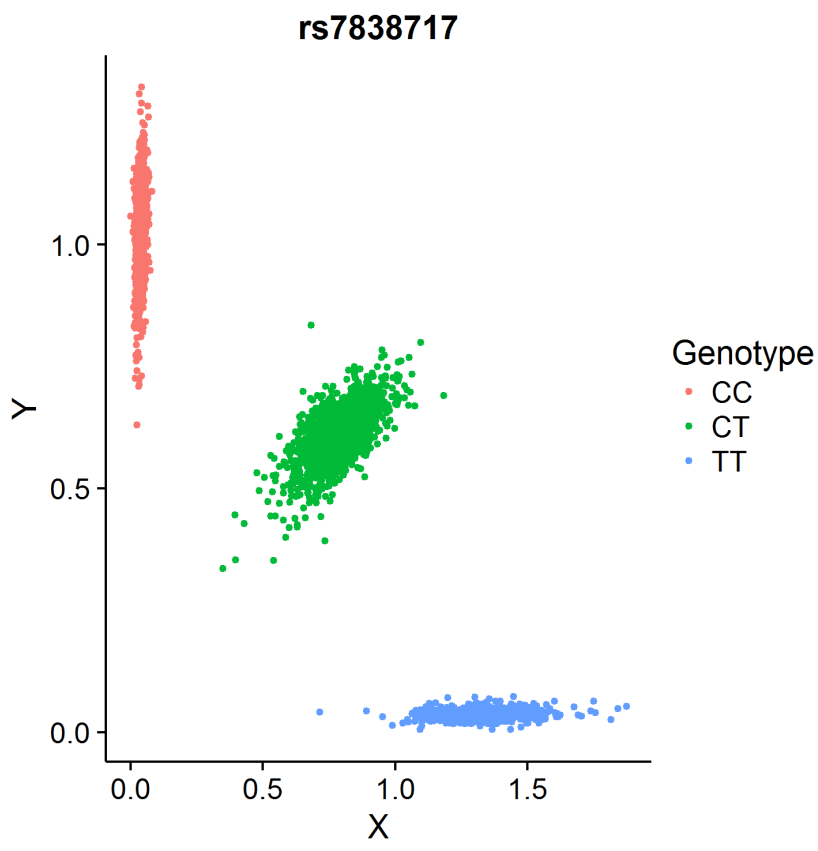
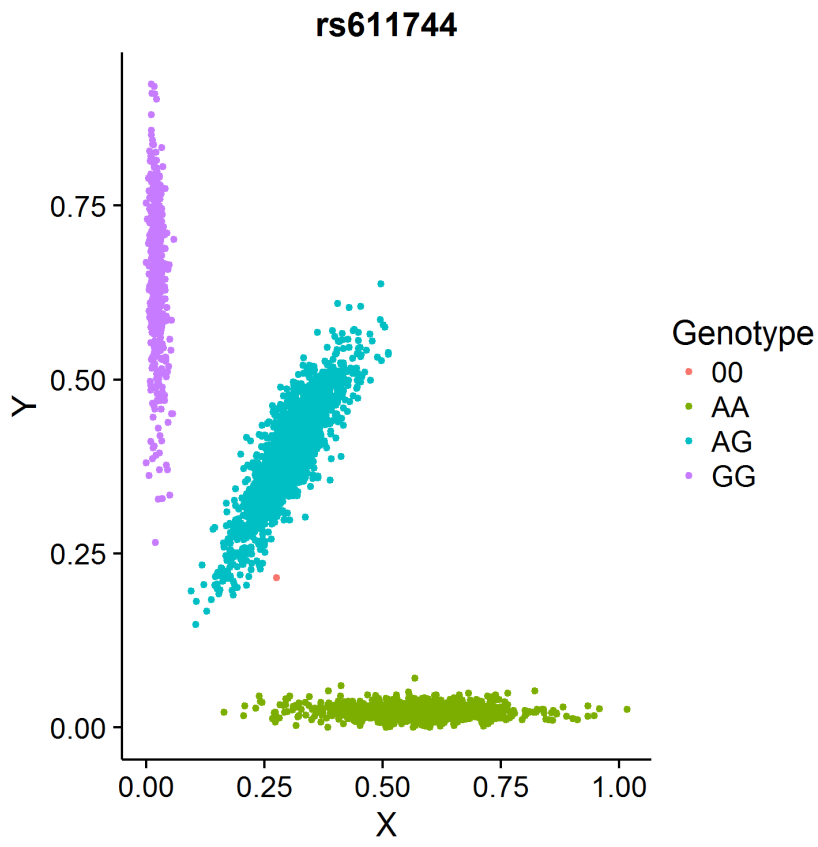


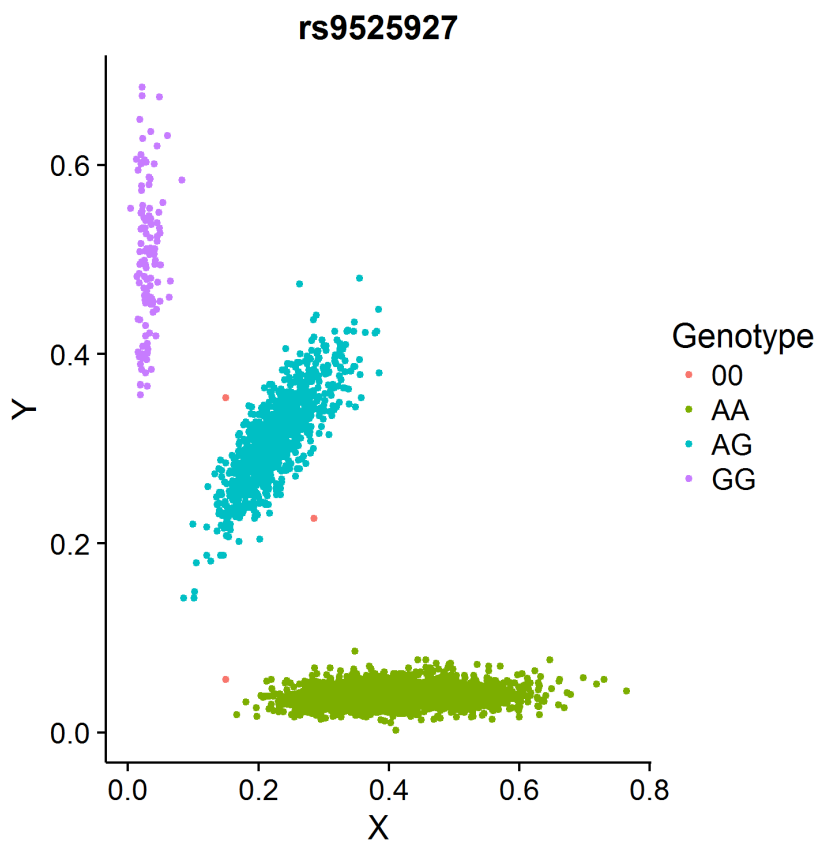
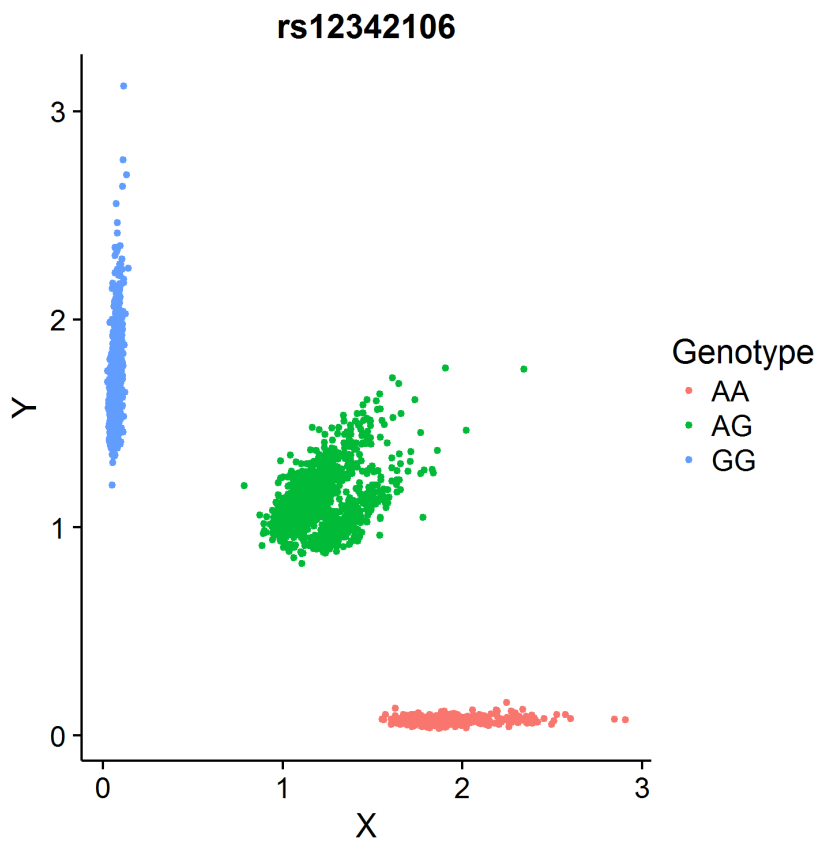


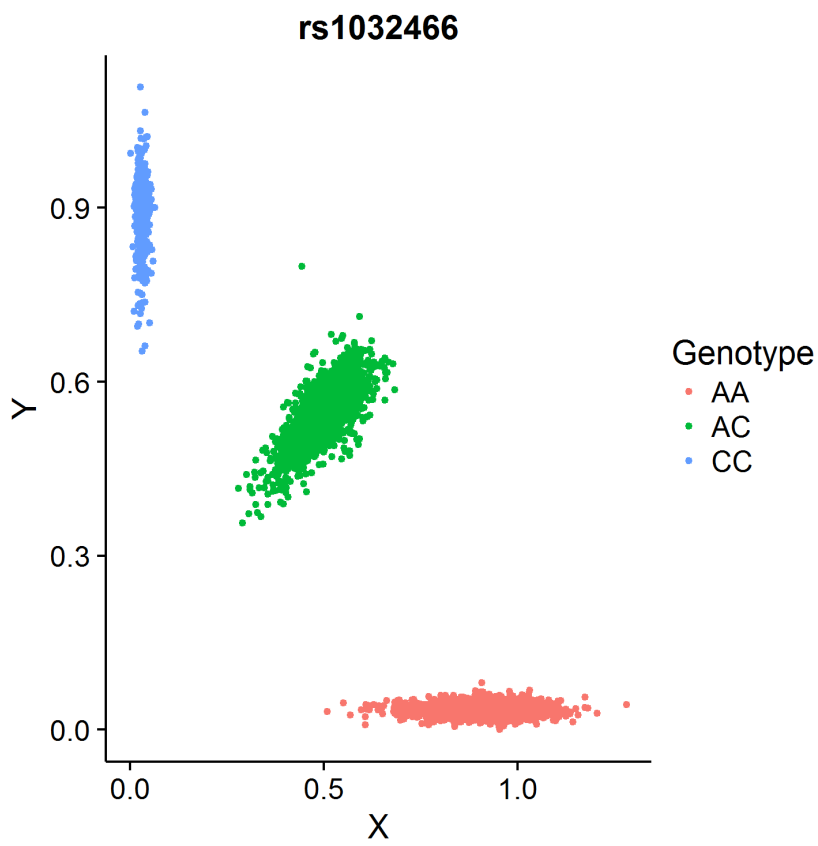
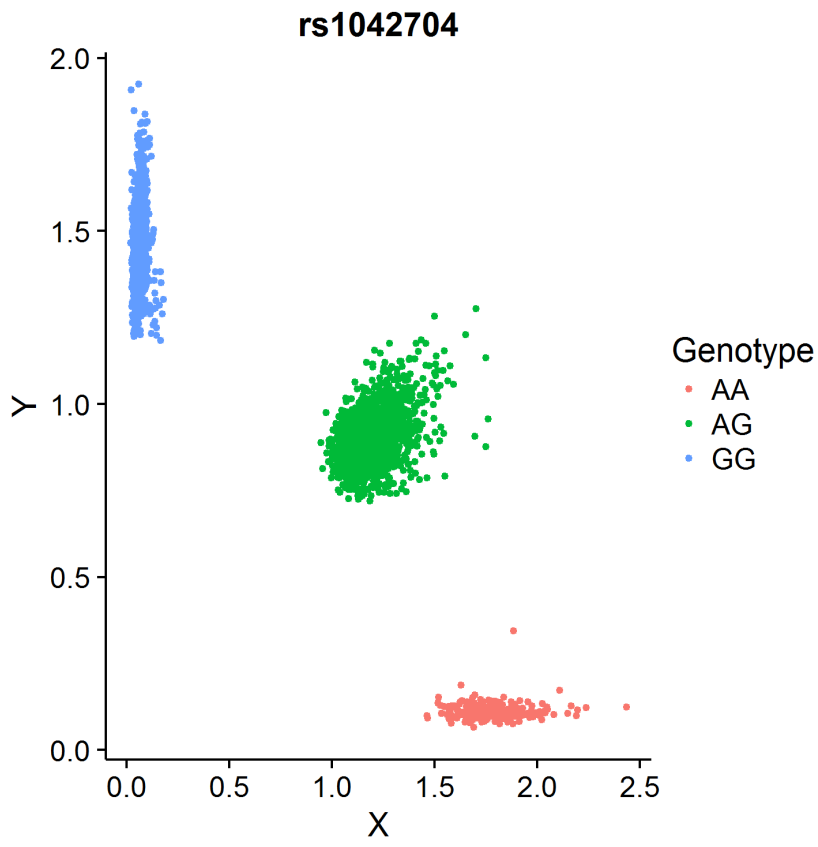


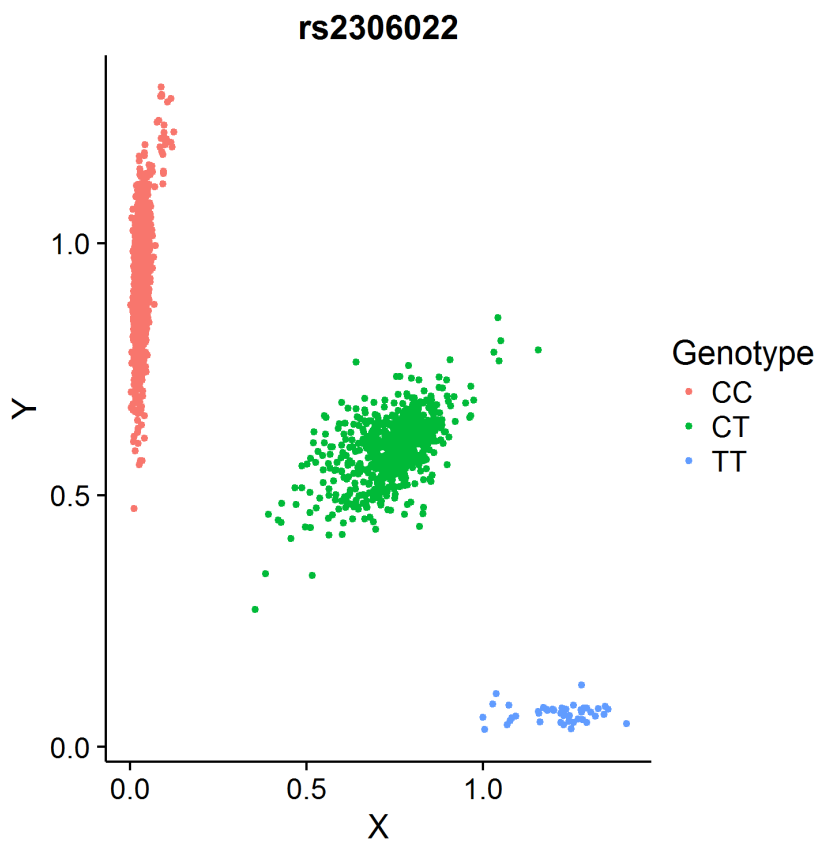
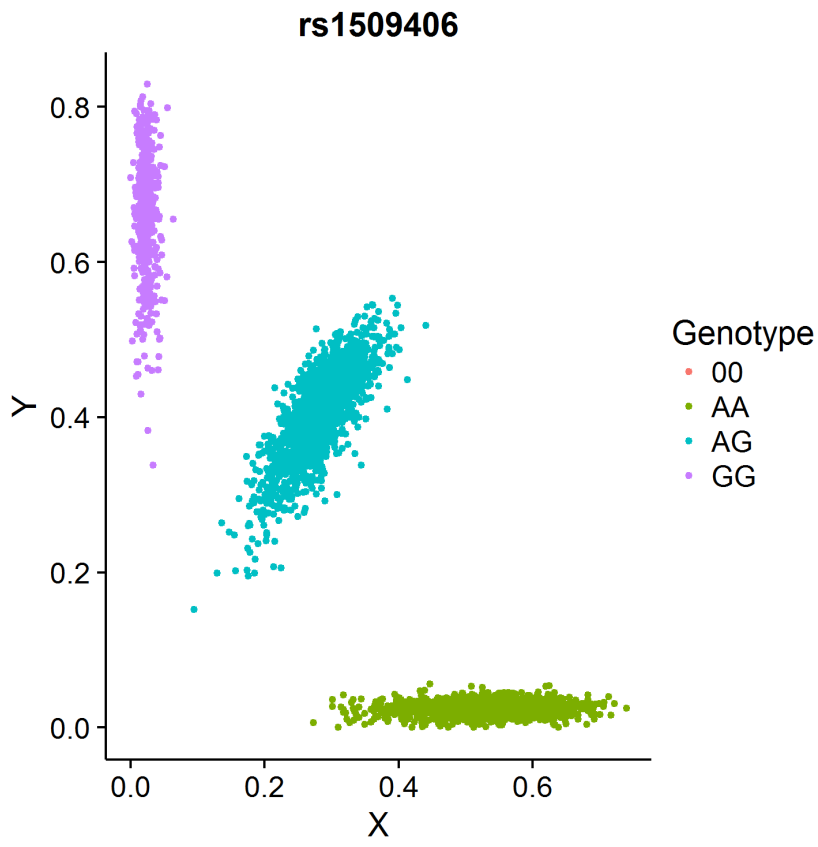




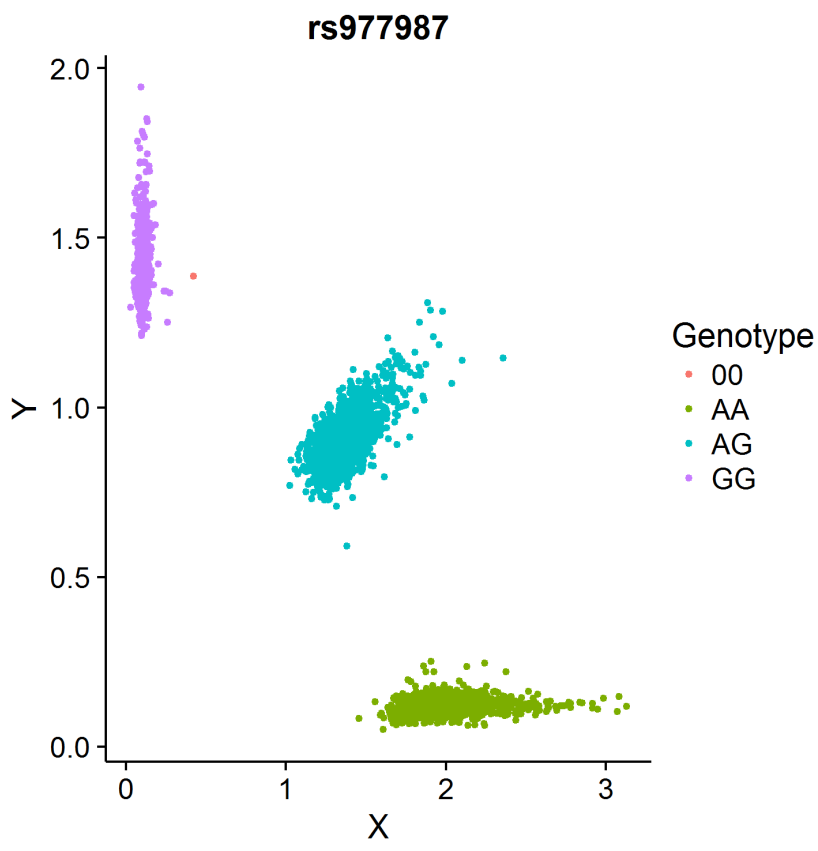
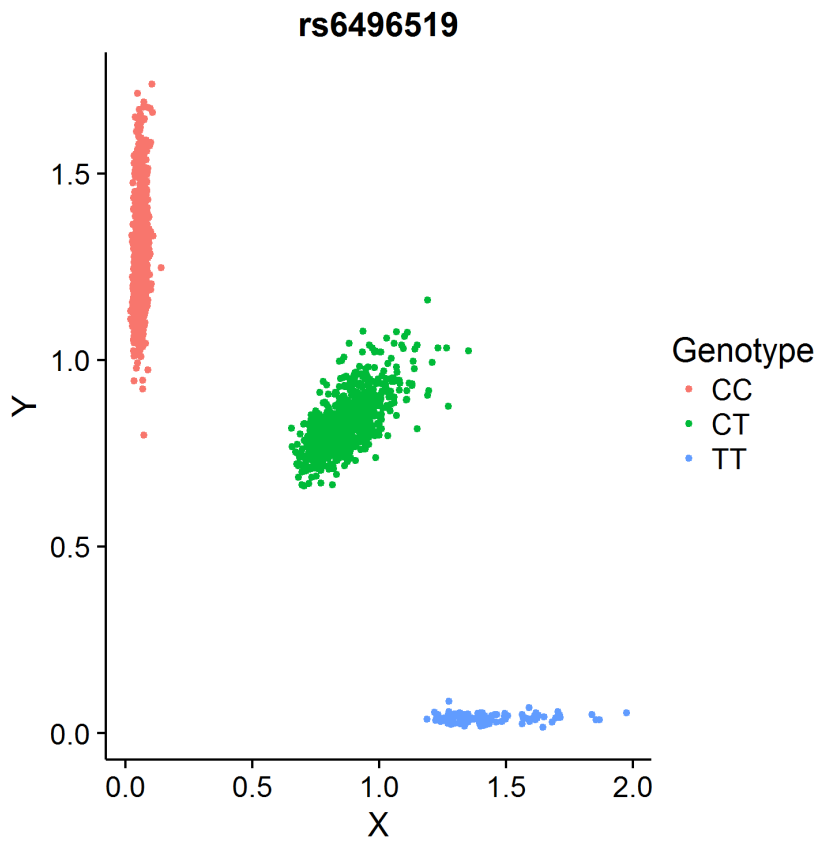


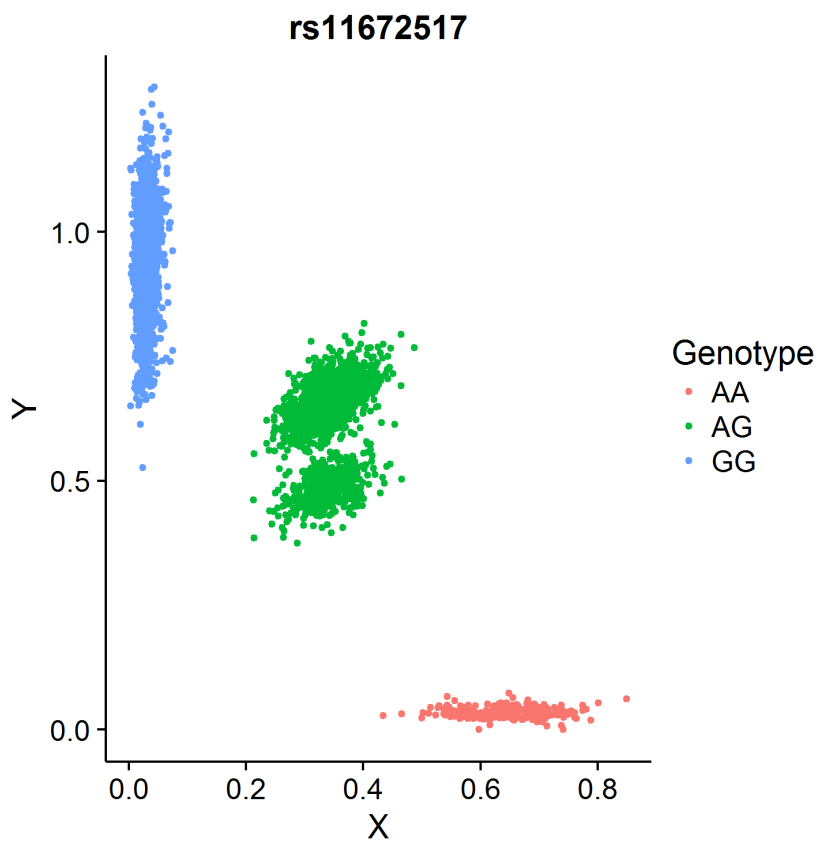
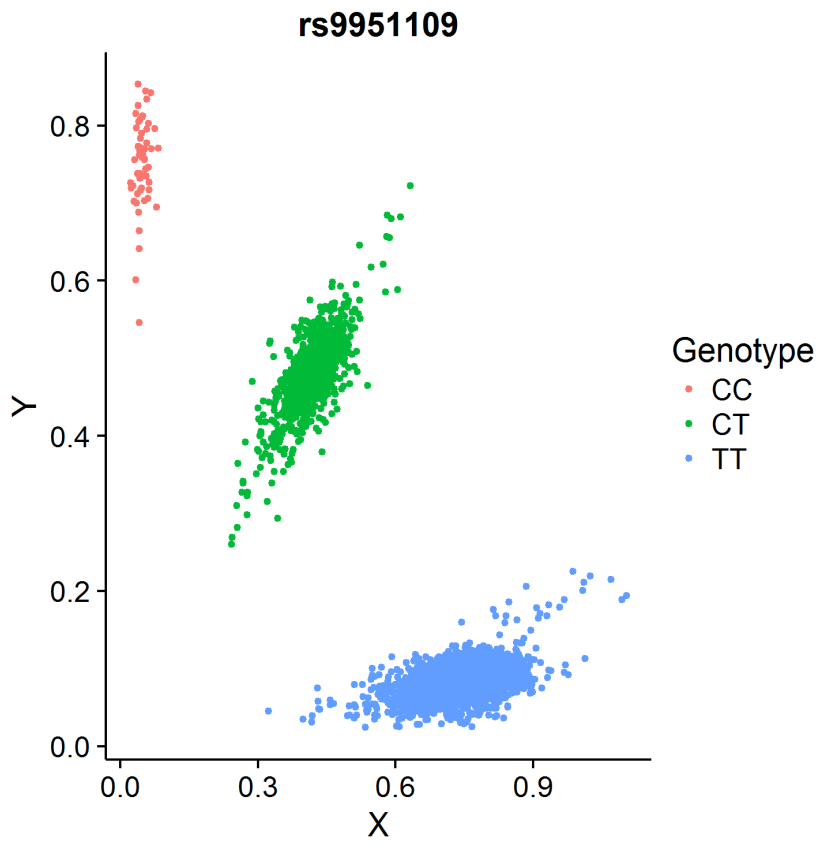


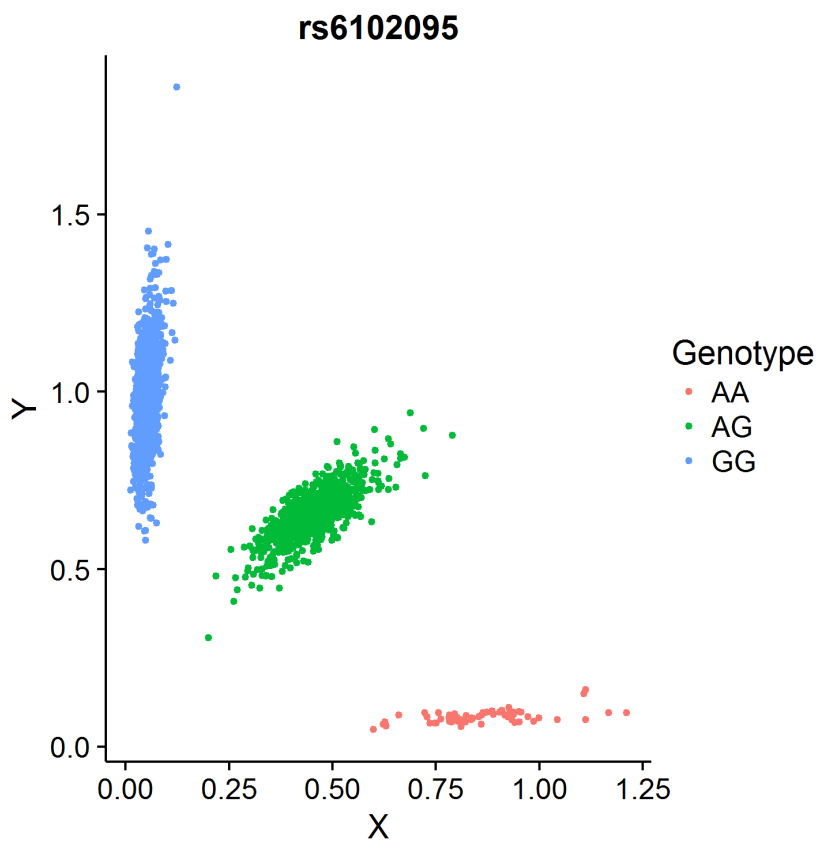
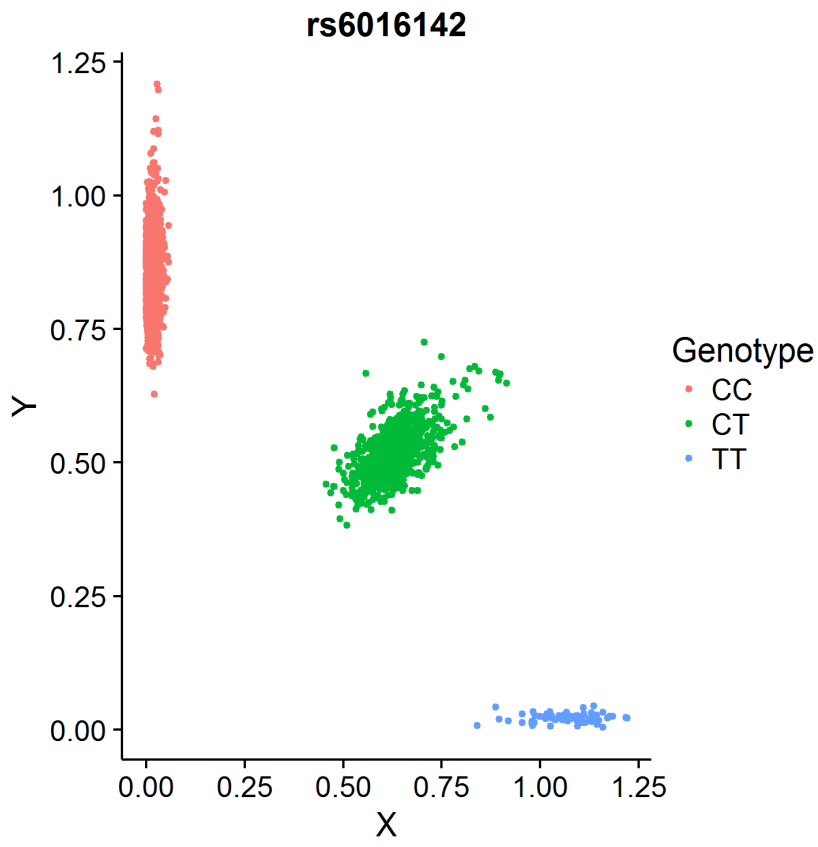












rs7291412

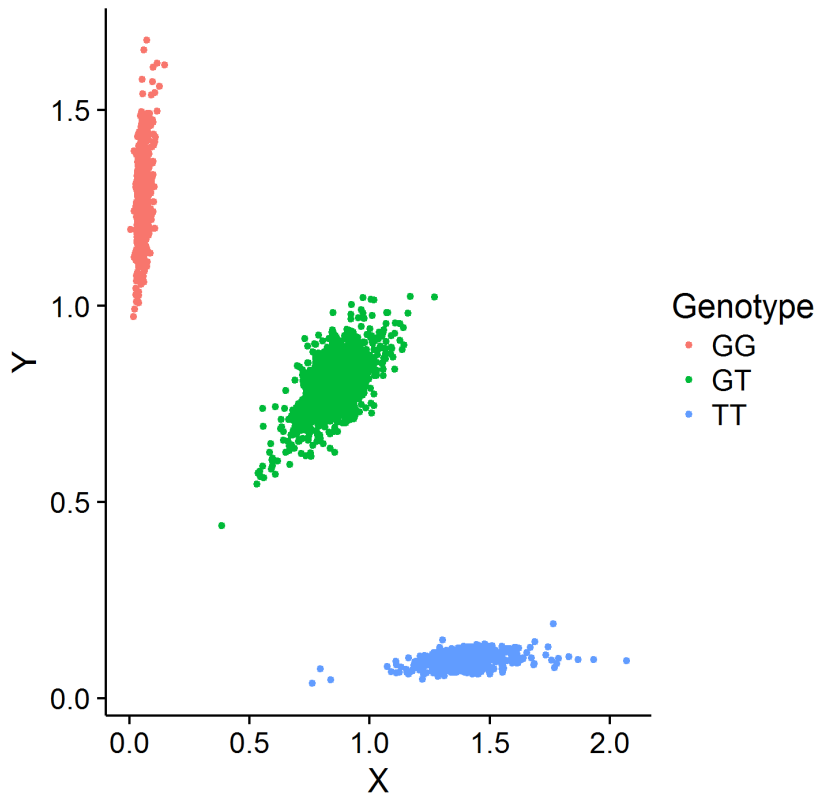
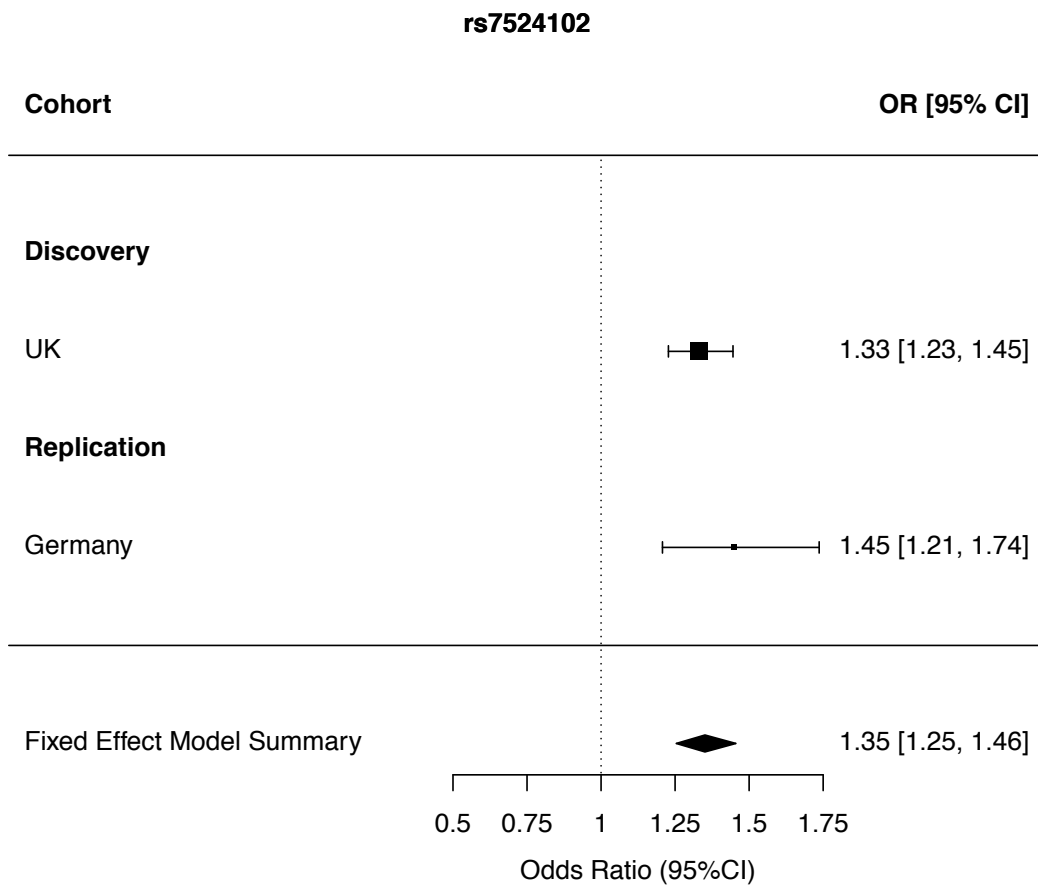
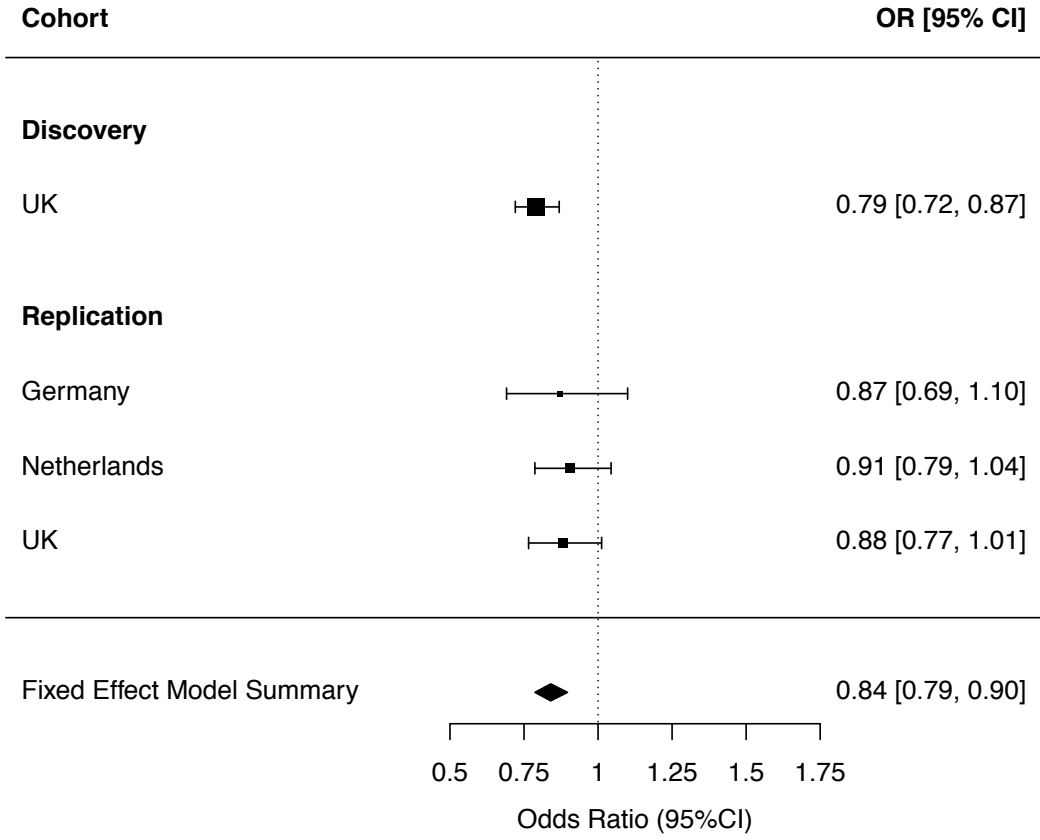


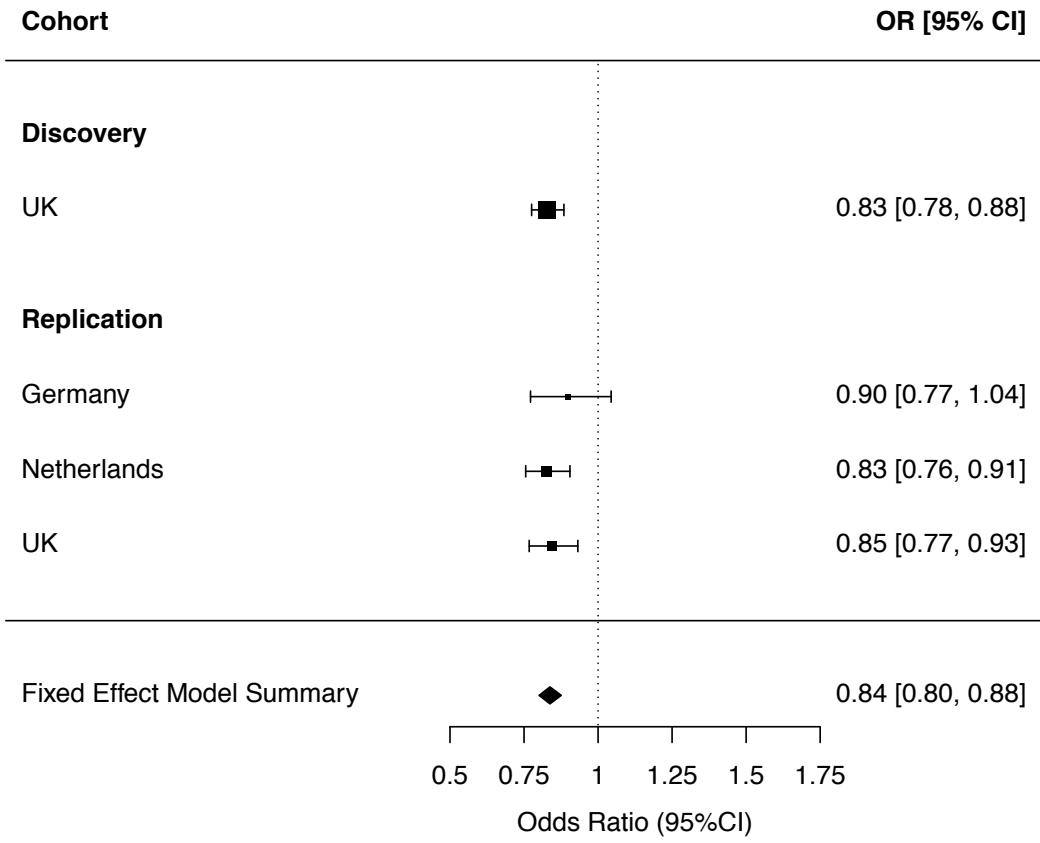
Figure S5. Forest plots showing the magnitude and direction of odds ratio for the risk allele at all associated loci, across the different cohorts in which it was genotyped. The size of the box is proportional to the weight (size) of the dataset in the fixed-effects meta-analysis. The whiskers represent the 95% confidence intervals.



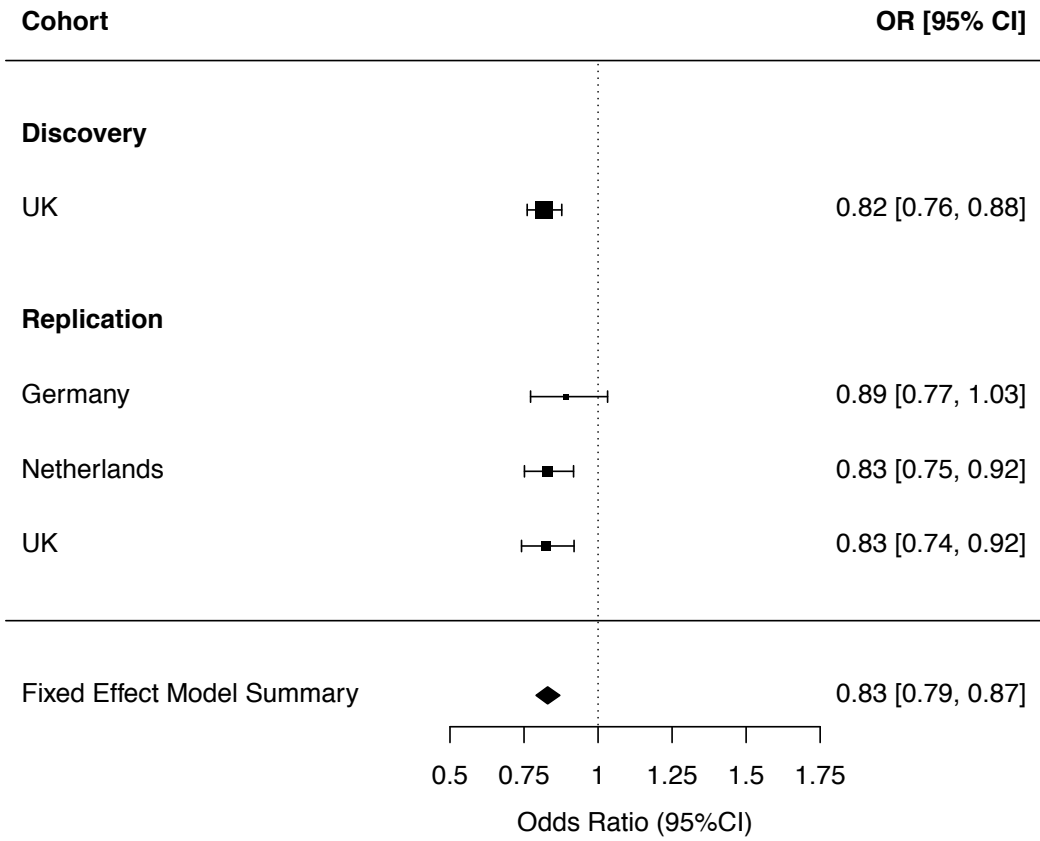
**rs17433710**



**rs394563**

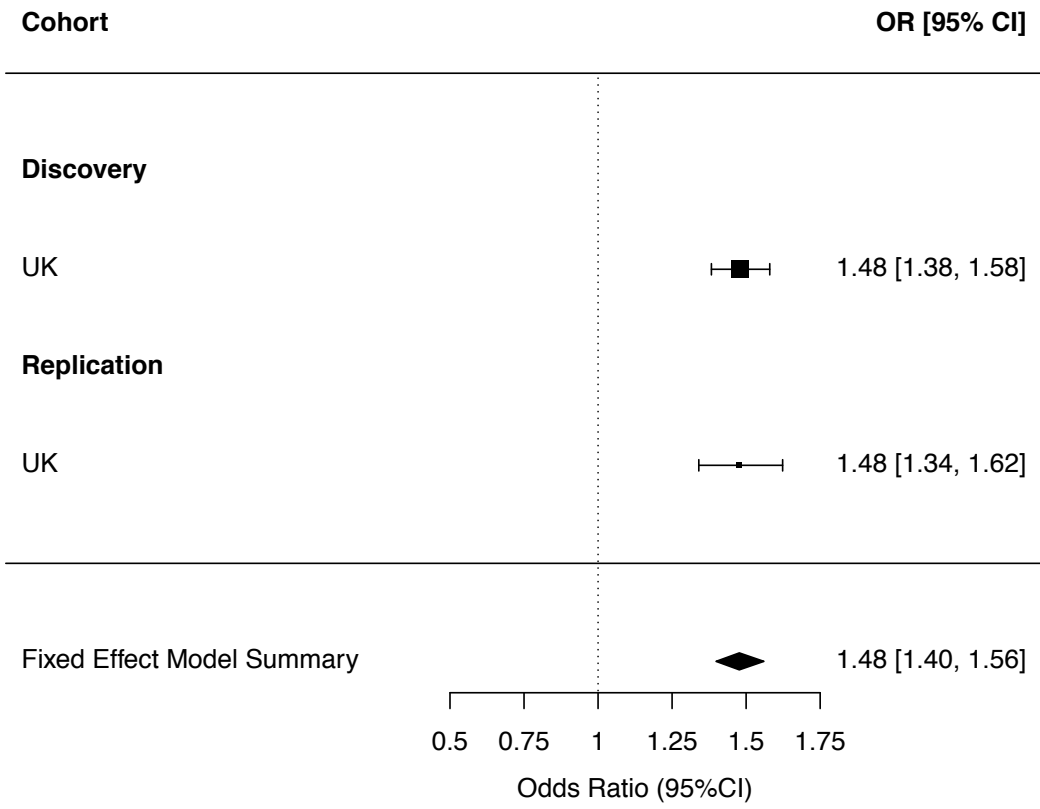


**rs10276303**

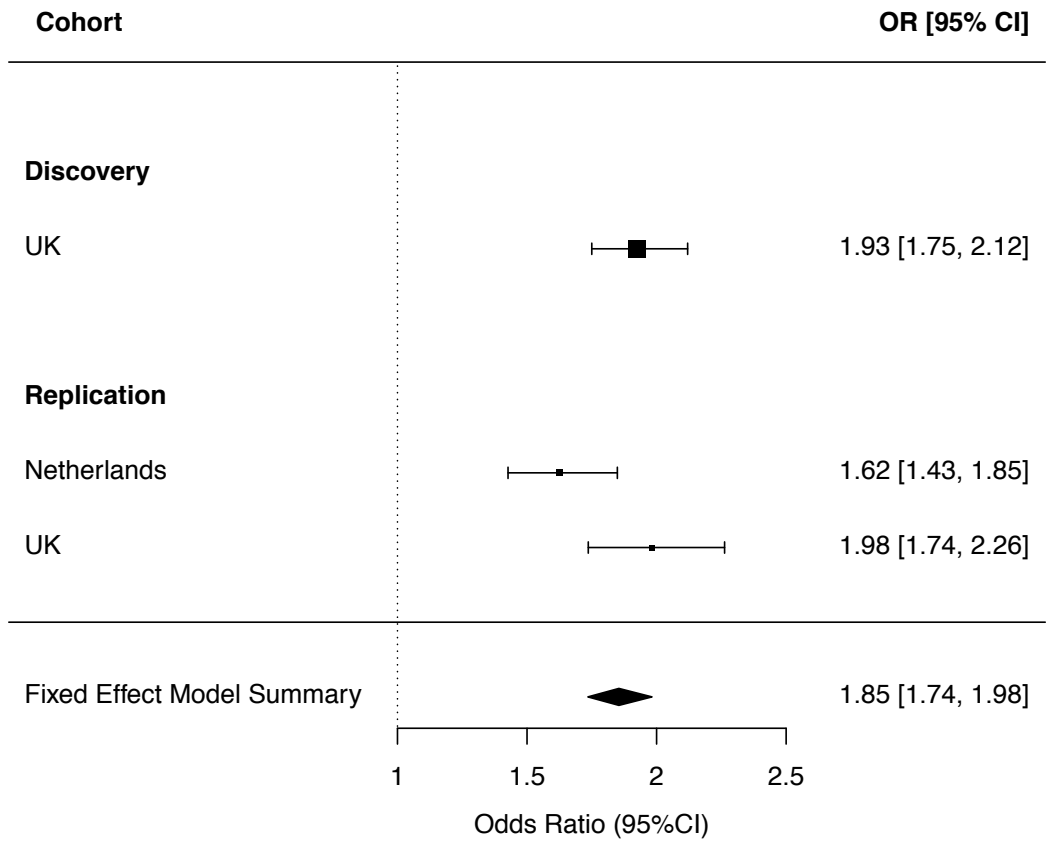




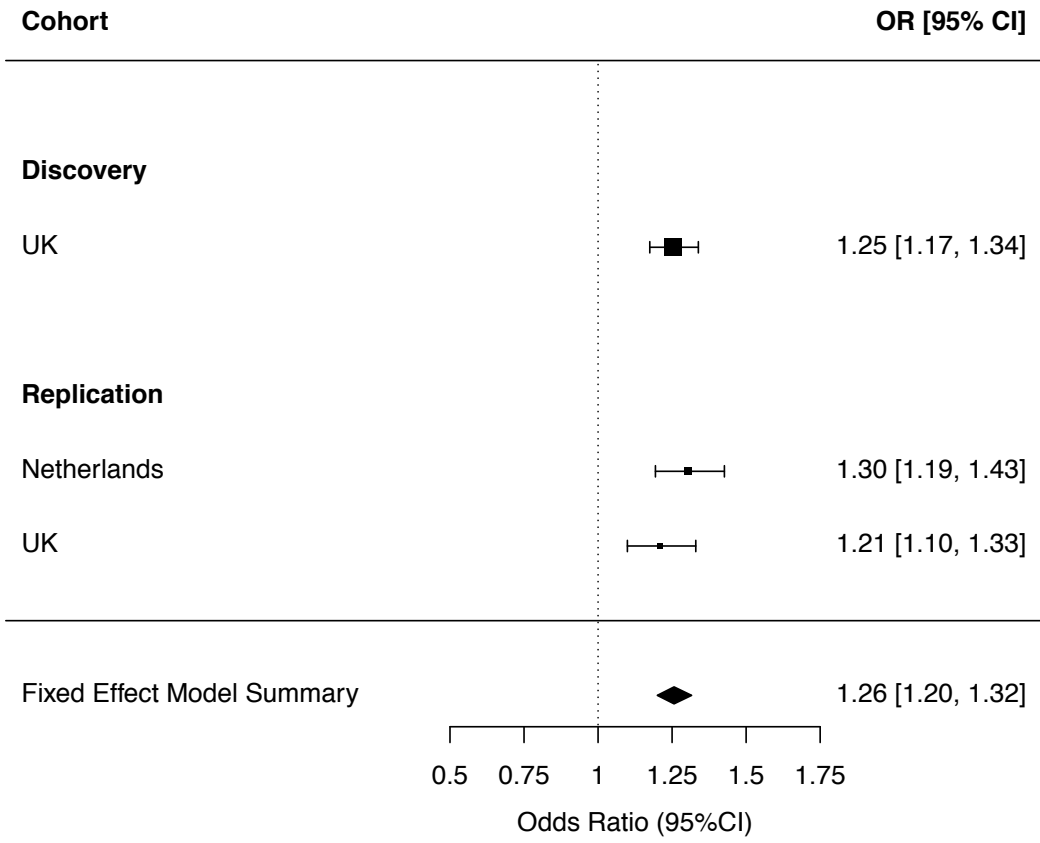
rs2598107



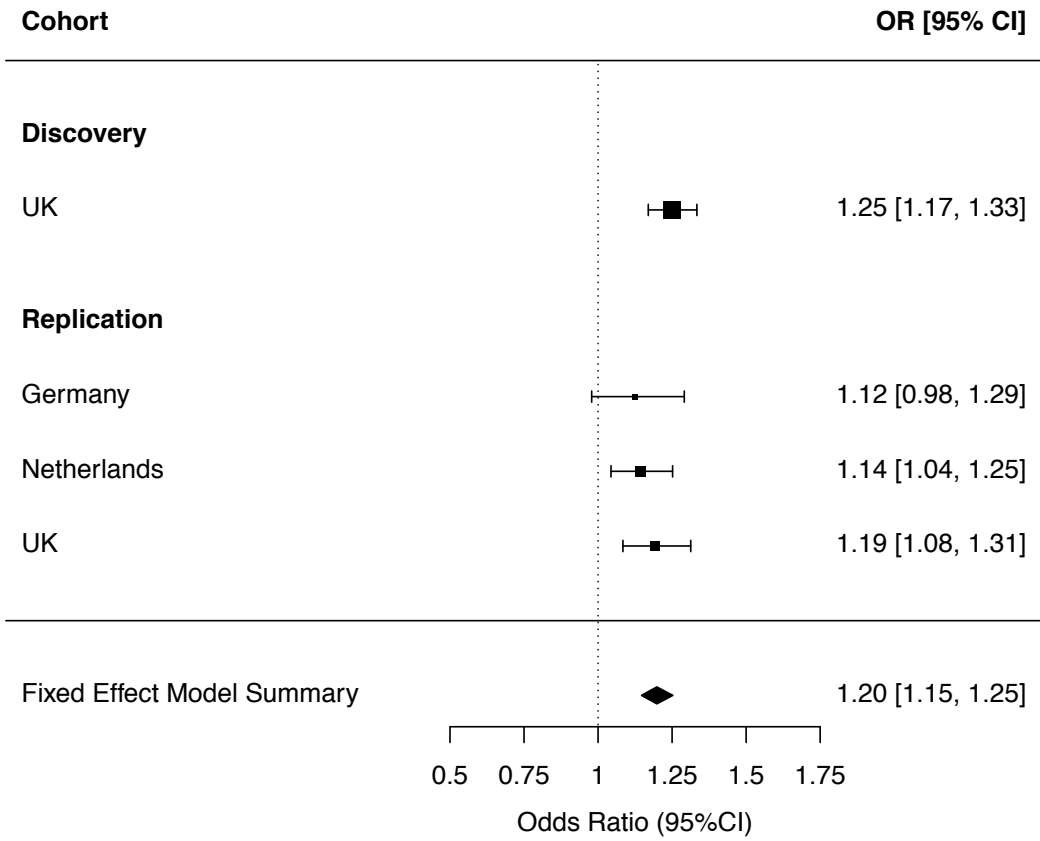
rs16879765



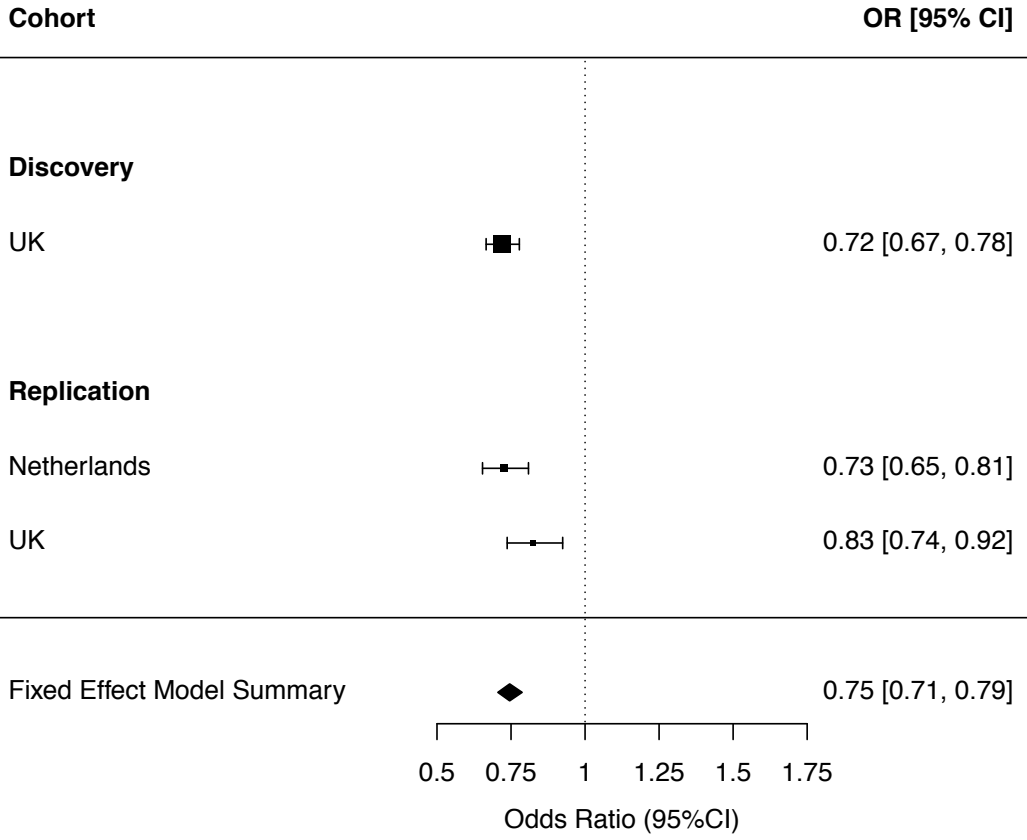
rs38904



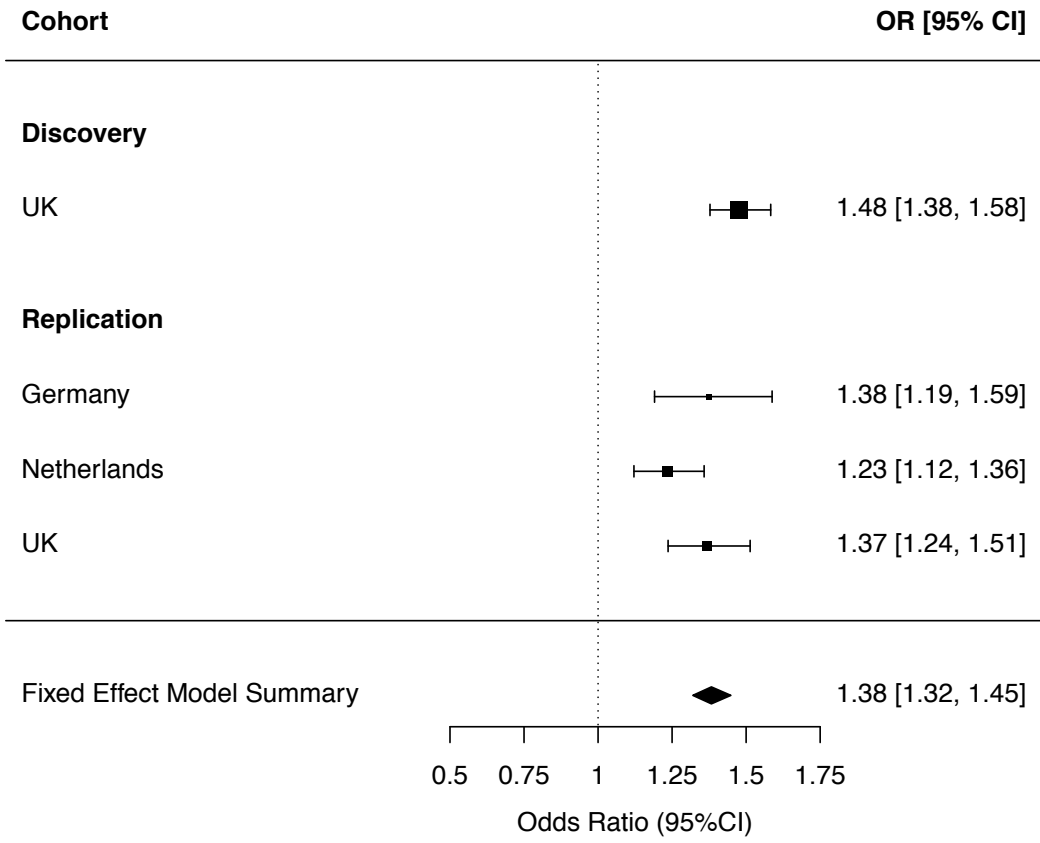
rs10866846



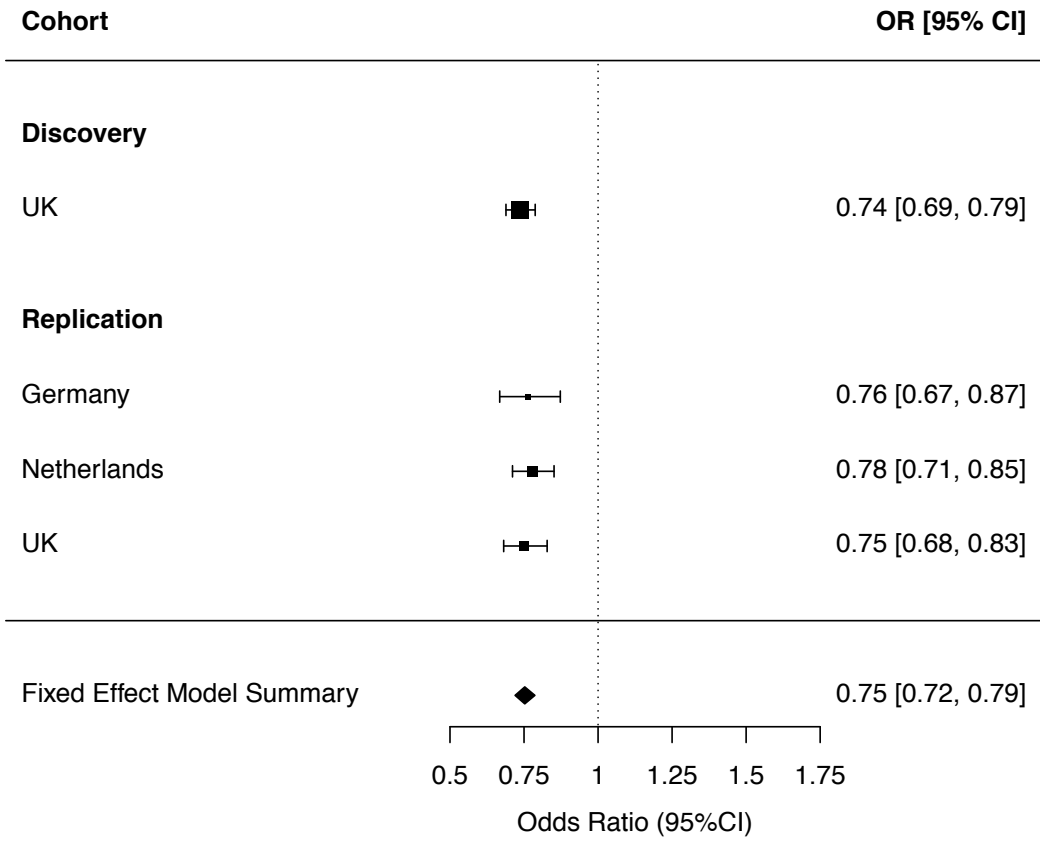
rs2912522



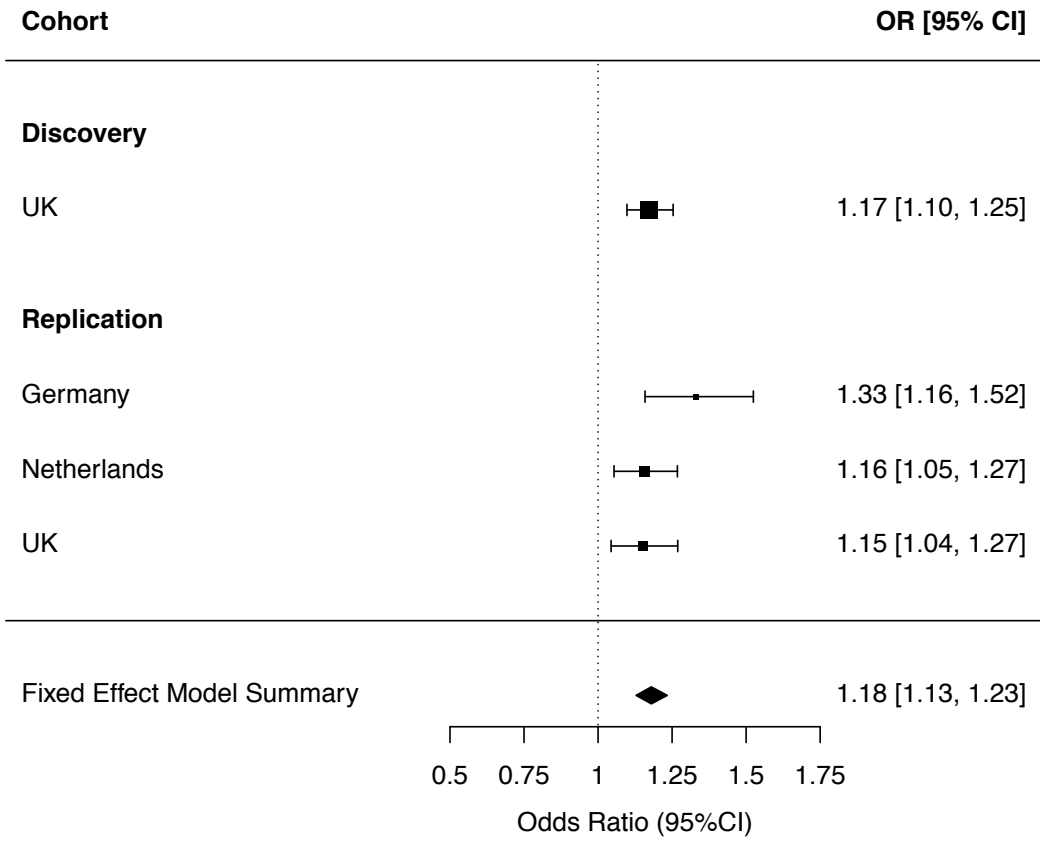
rs629535



rs611744

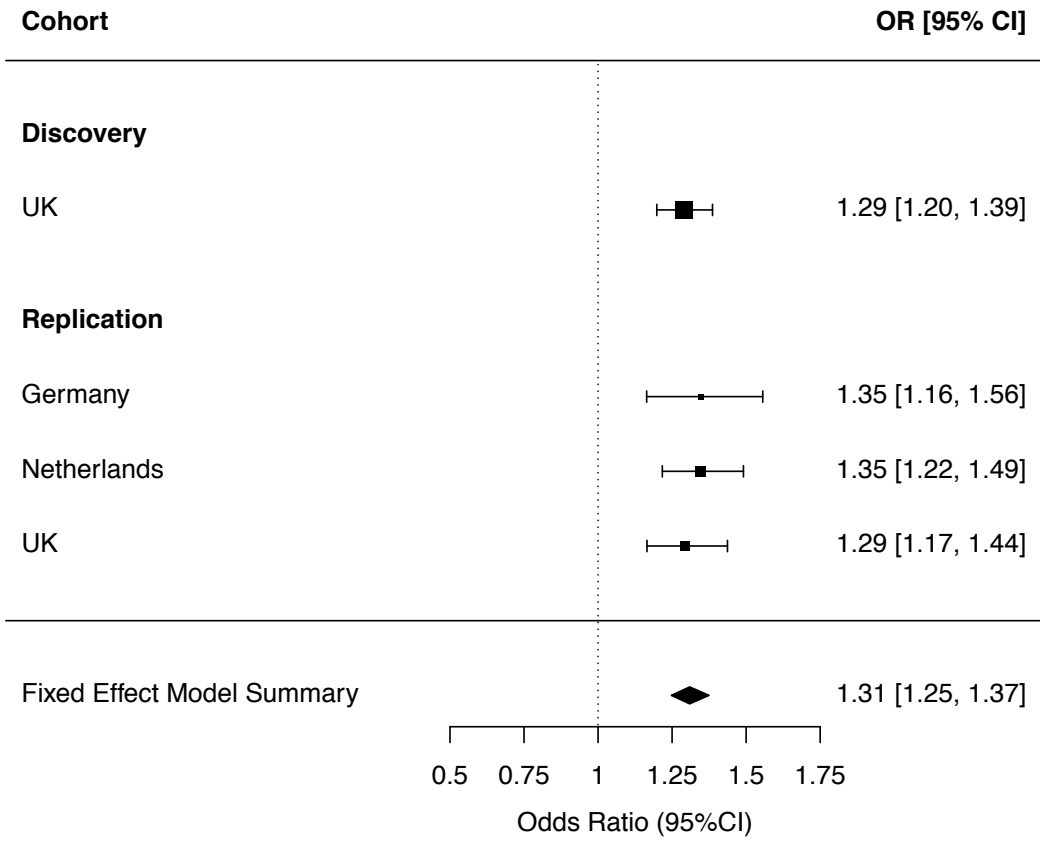


rs7838717

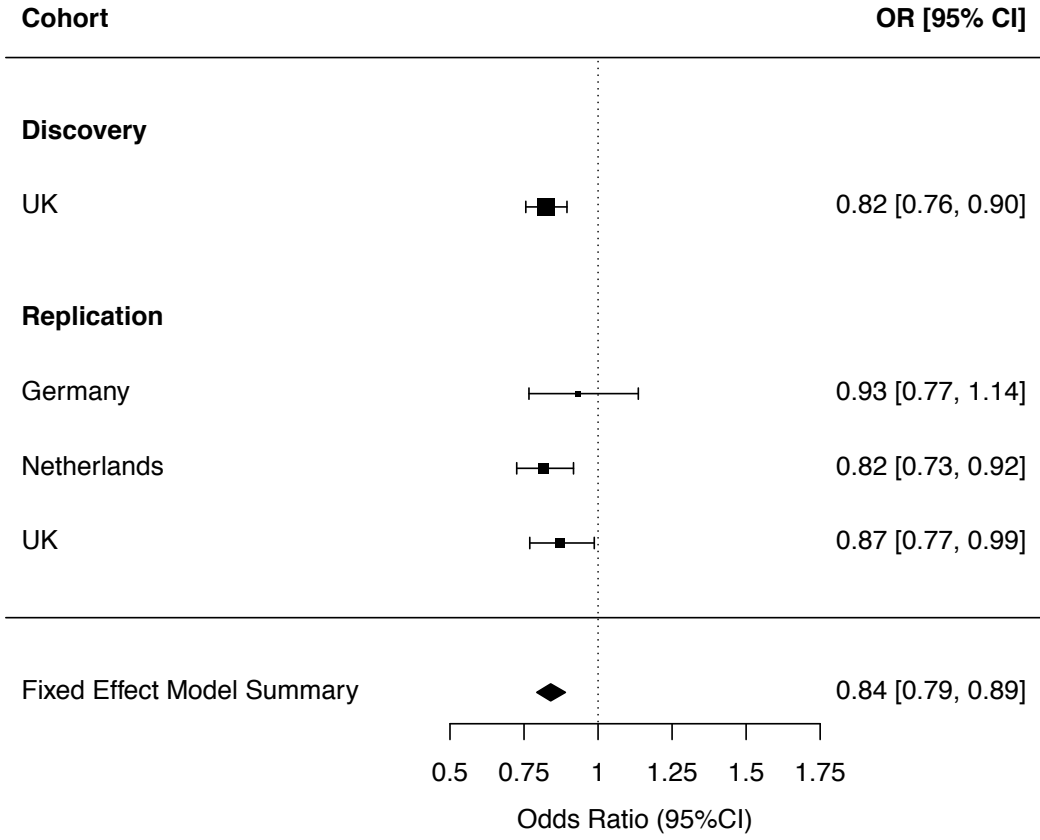




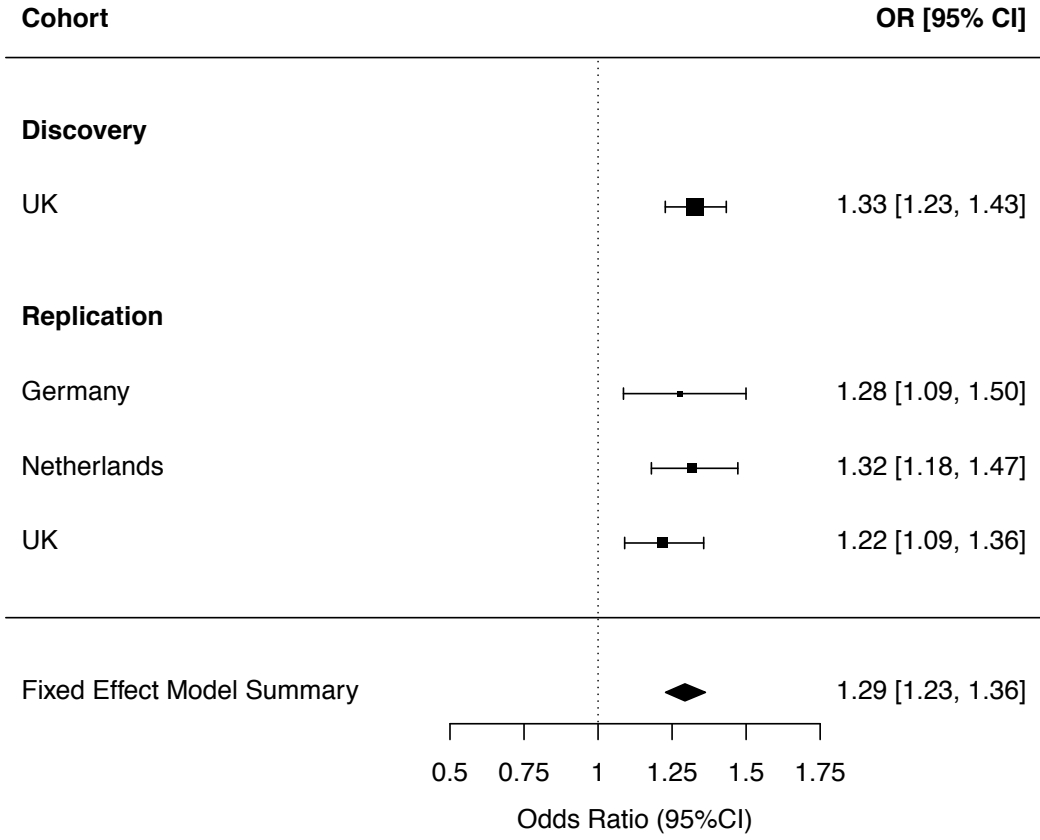
**rs12342106**



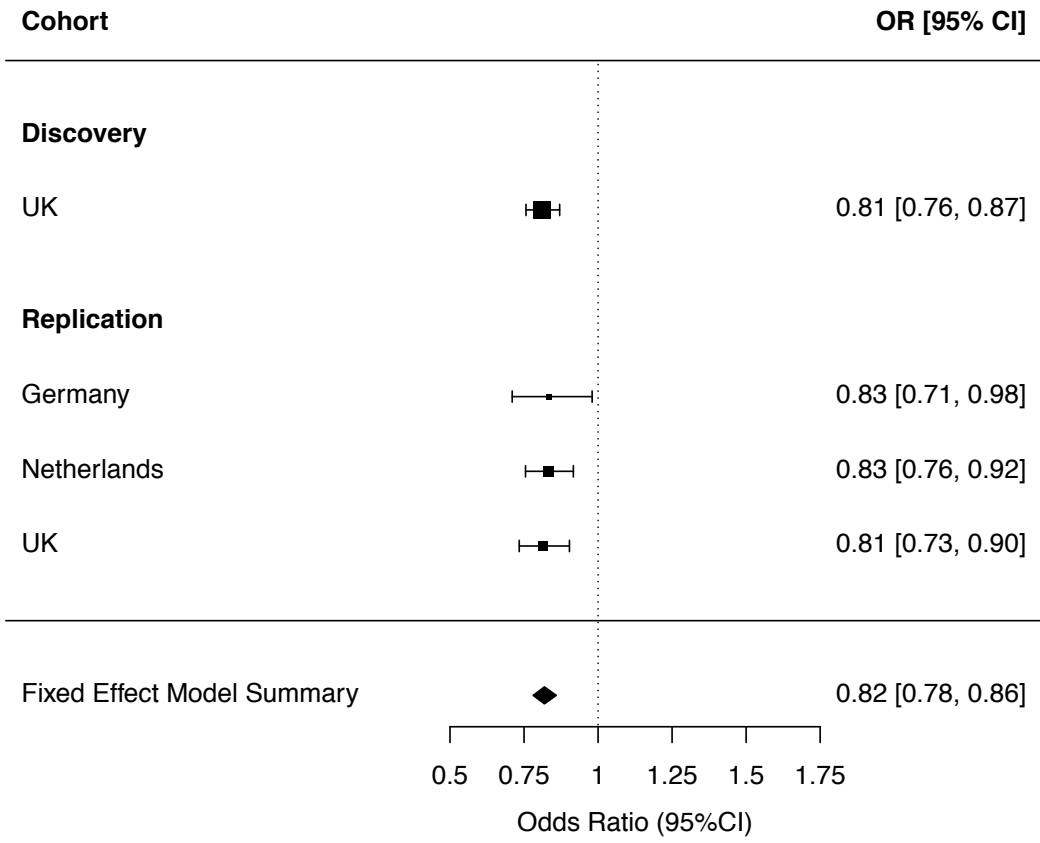
rs9525927



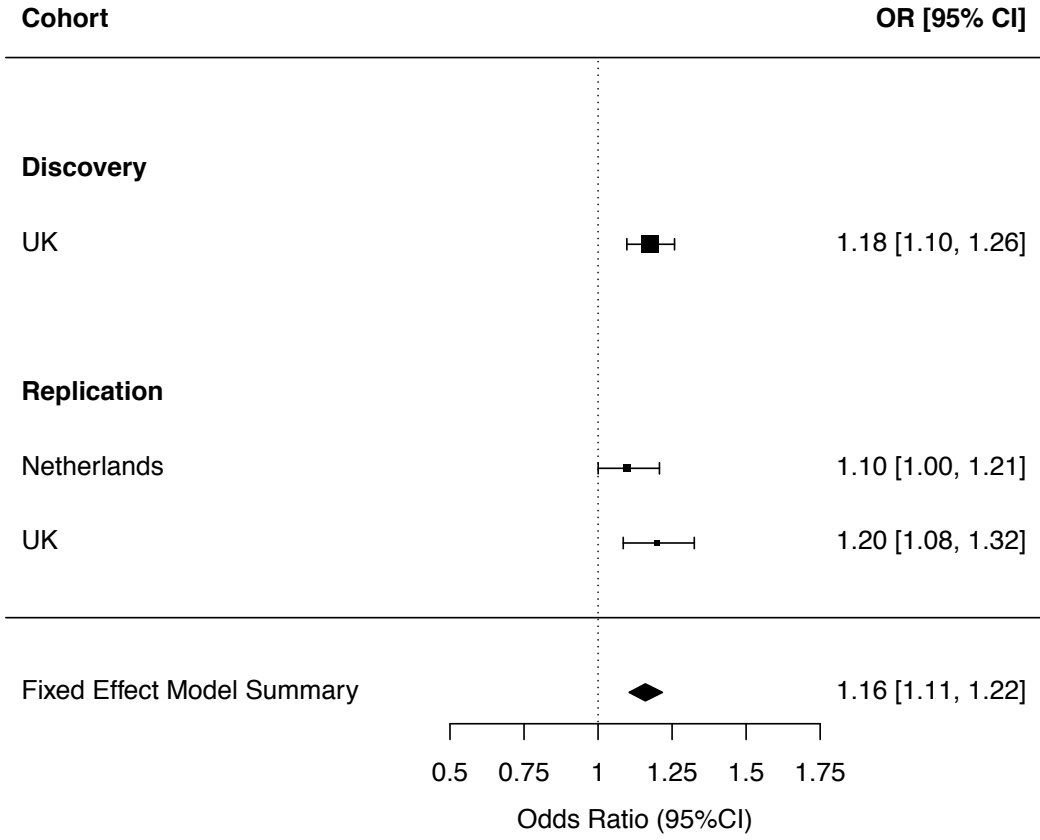
**rs1042704**



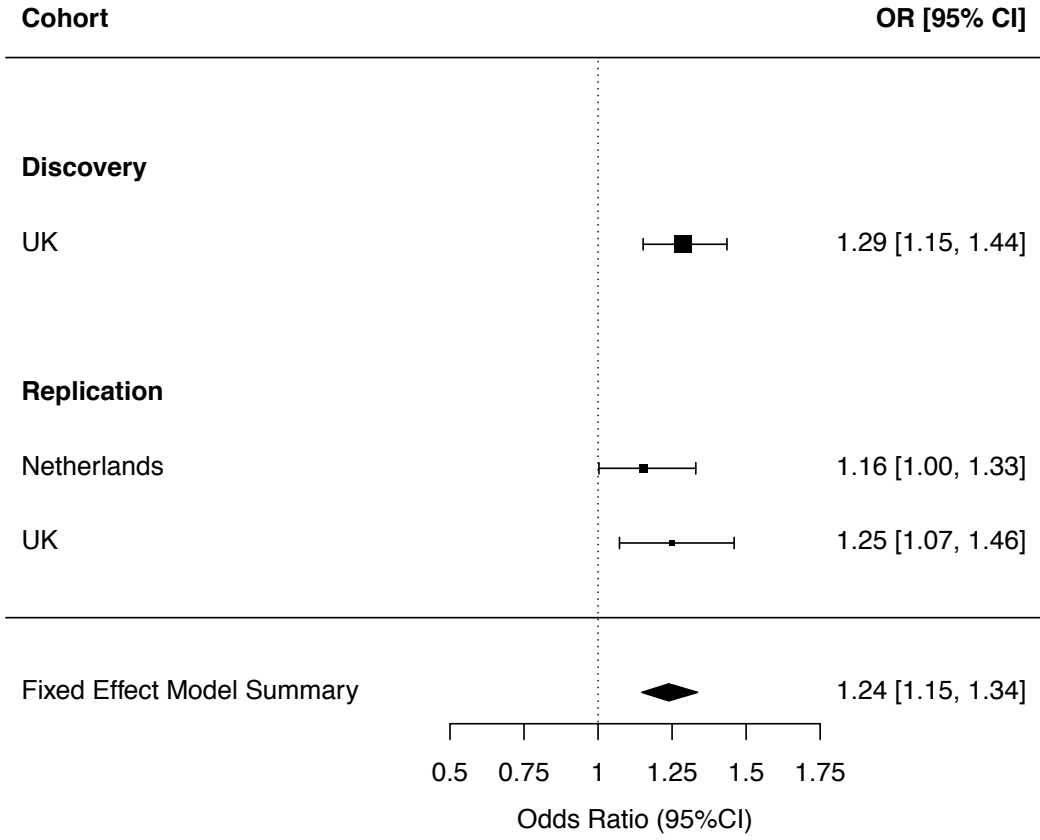
**rs1032466**



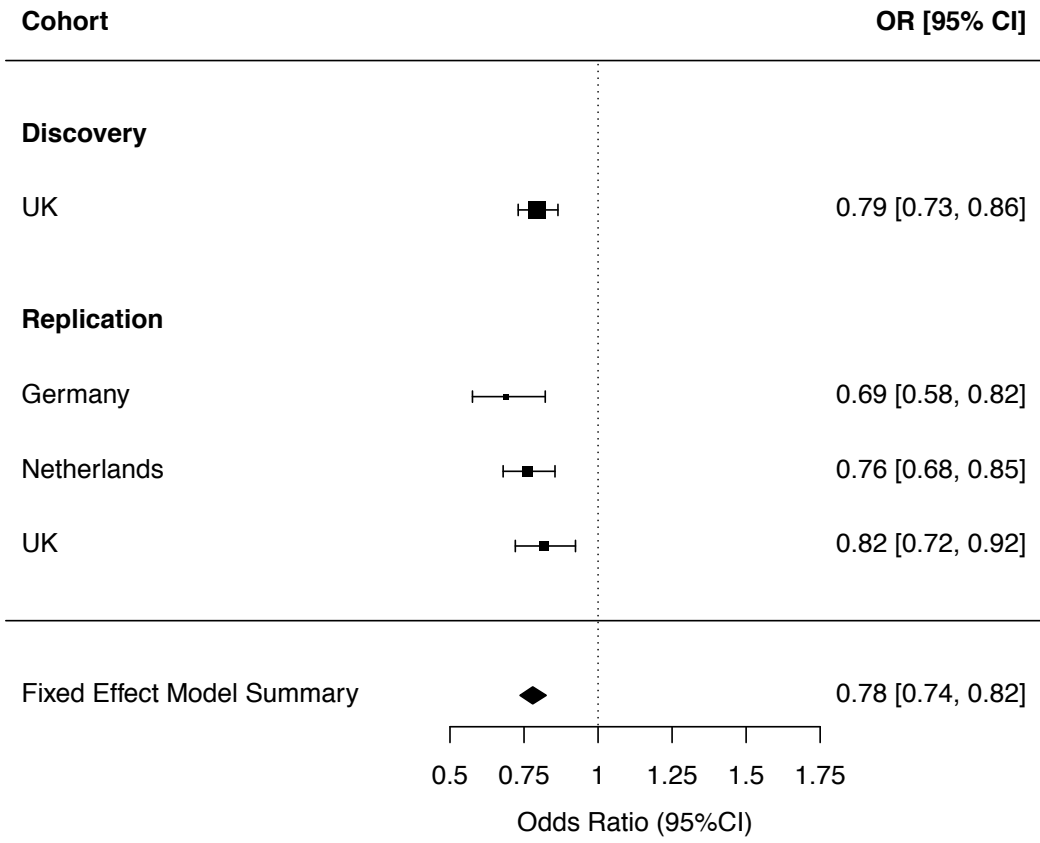
rs1509406



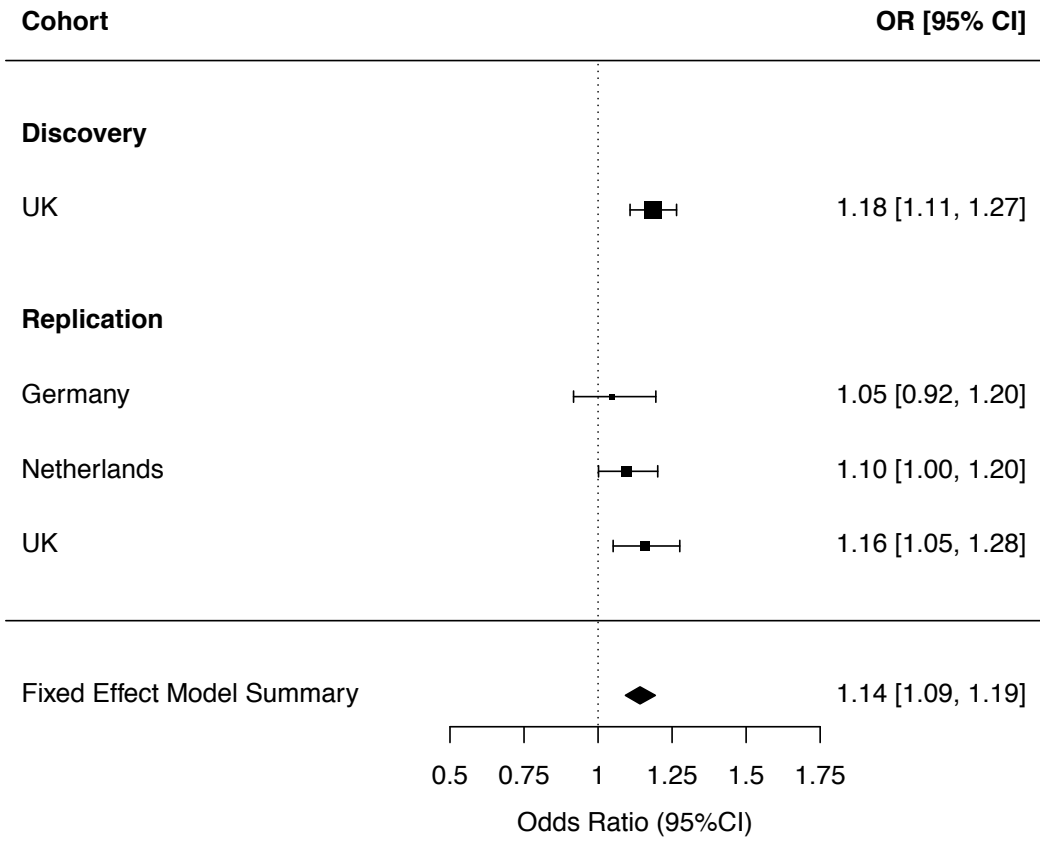
rs2306022



**rs6496519**

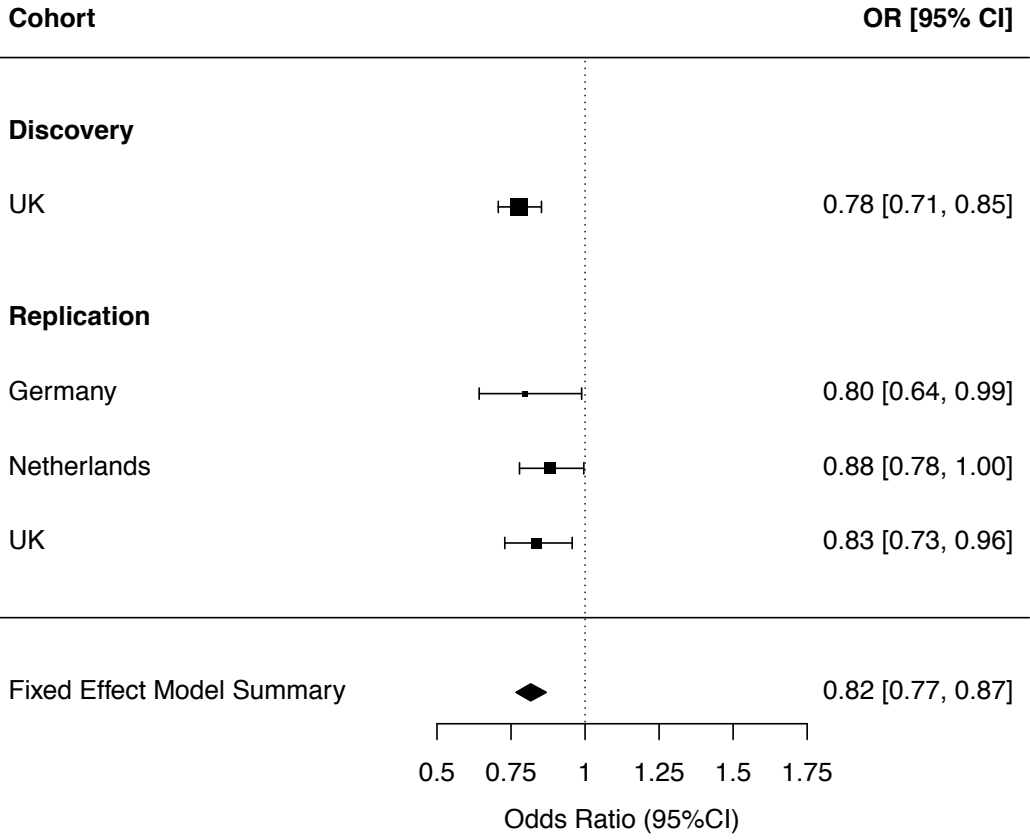


**rs977987**

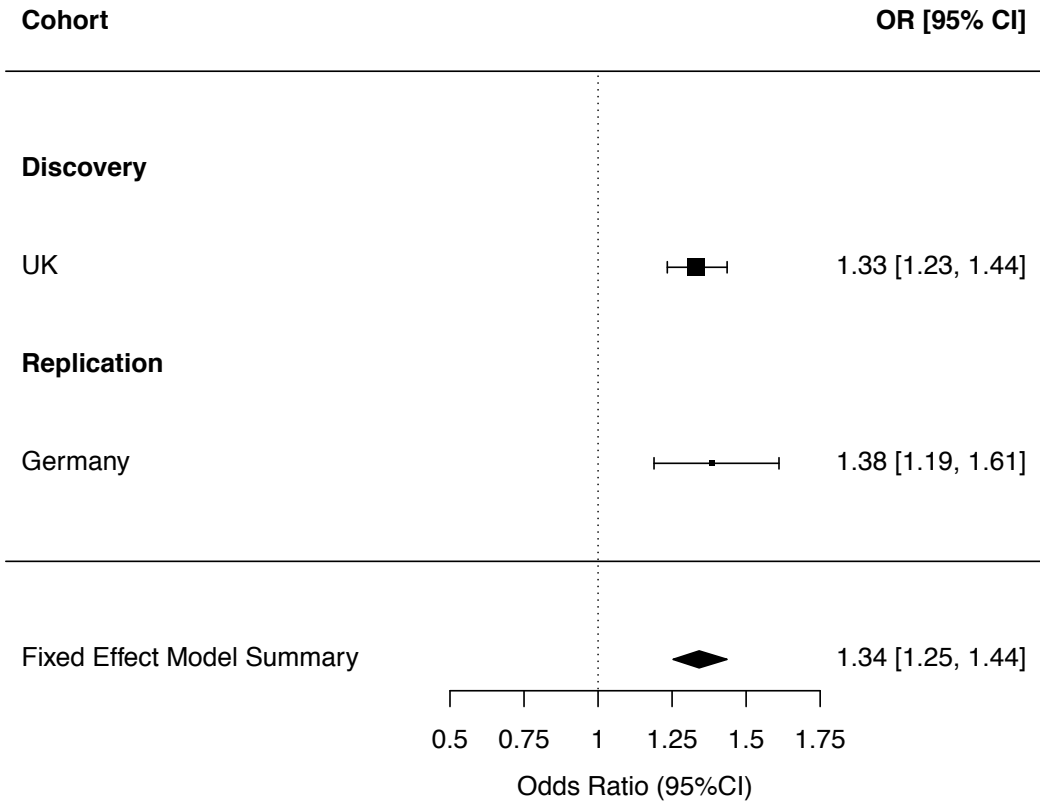




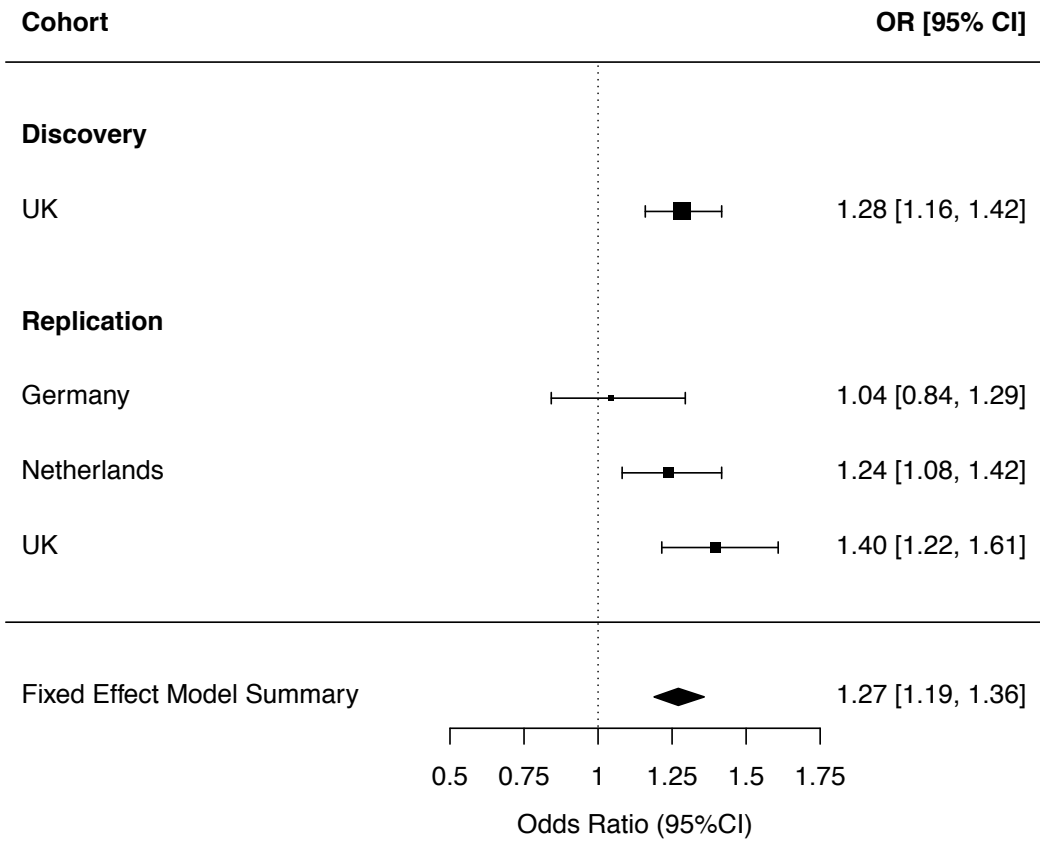
rs9951109



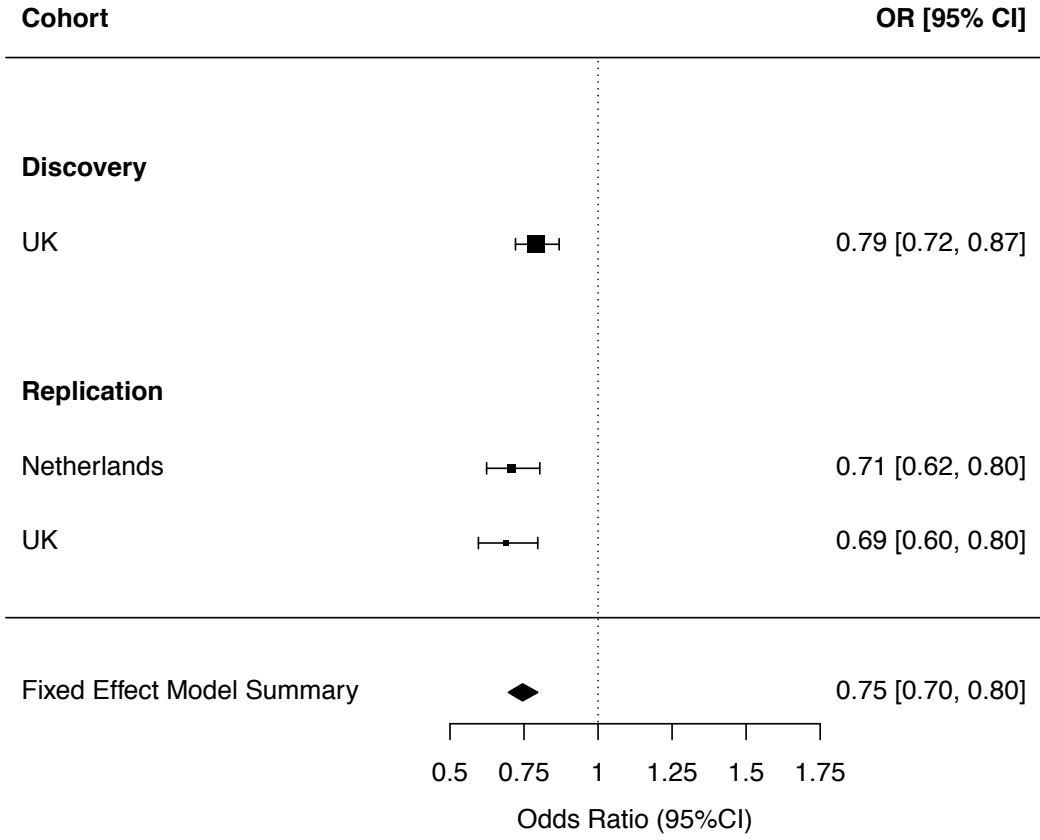
**rs11672517**



**rs6016142**



**rs6102095**



rs7291412

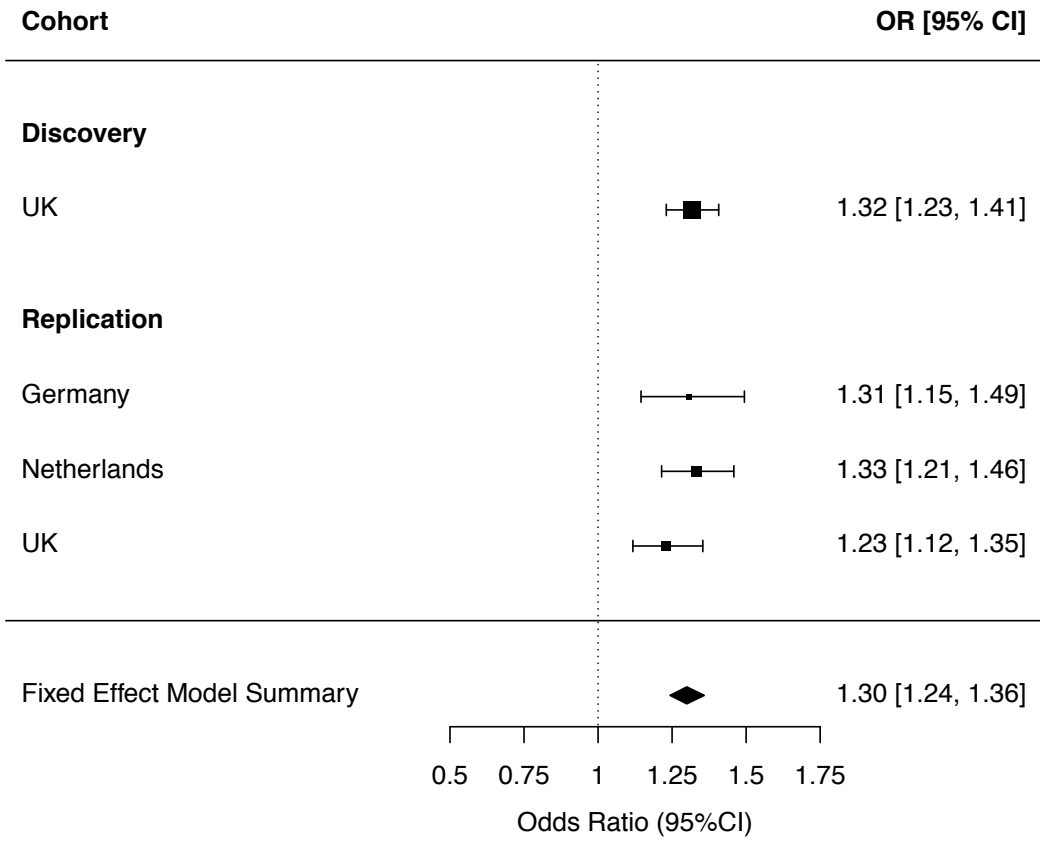
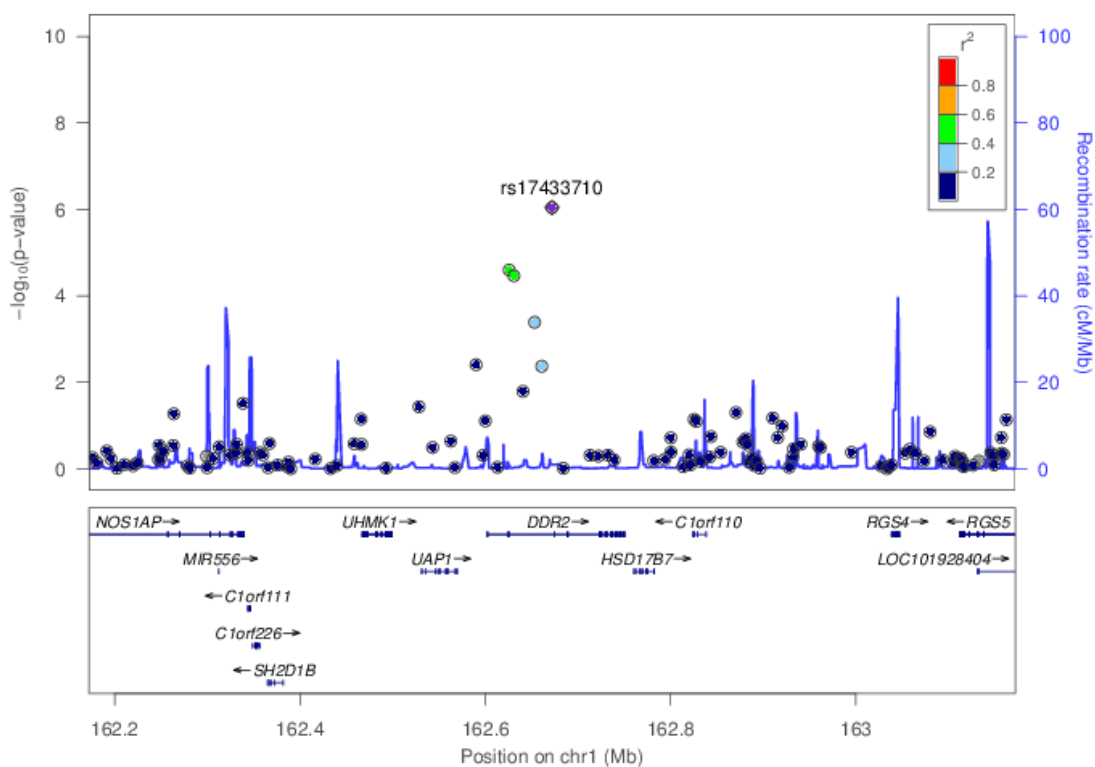
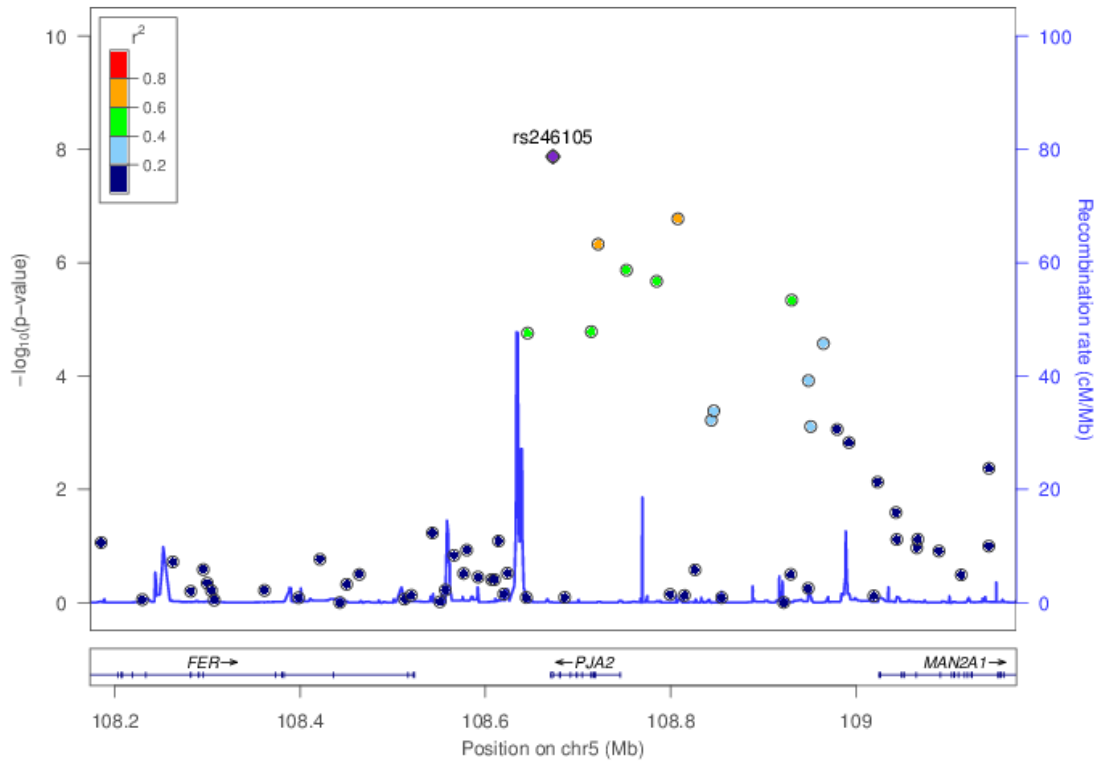
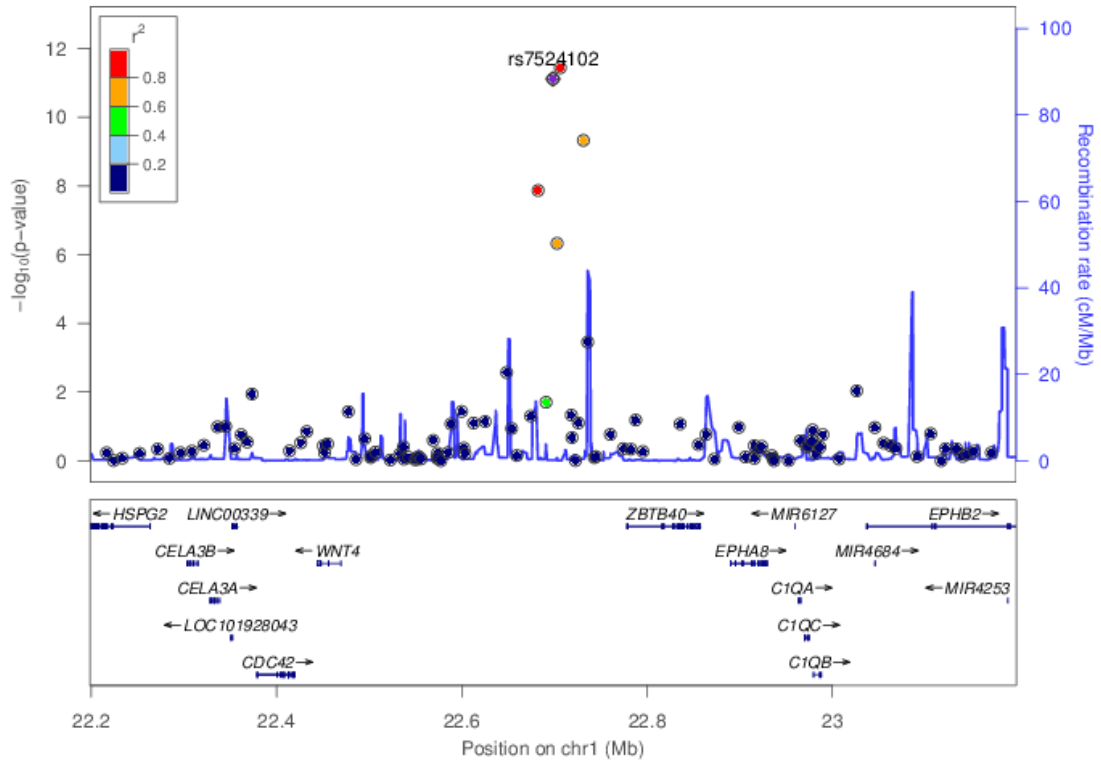
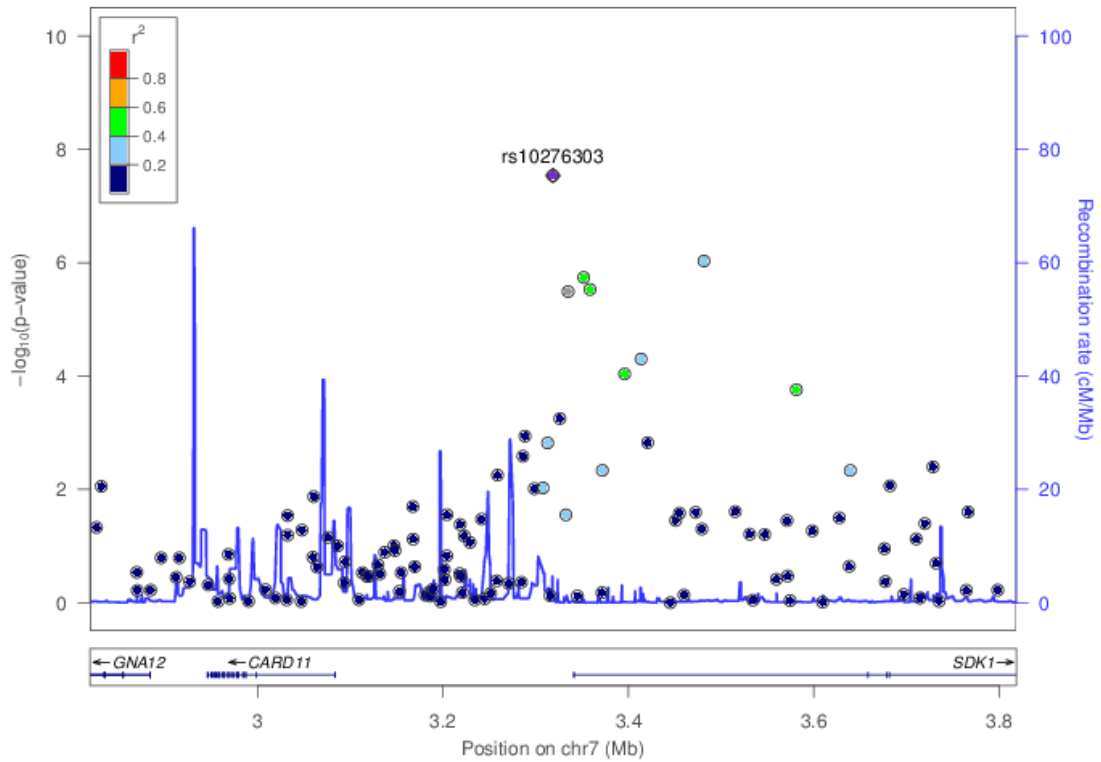
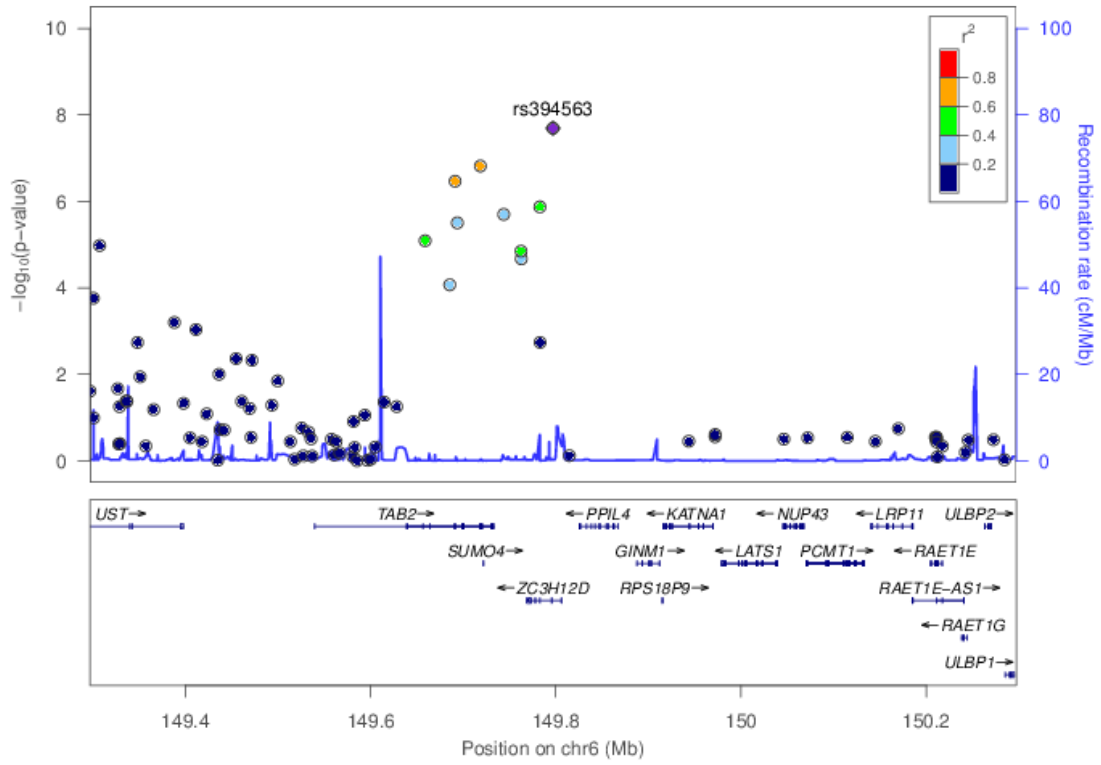


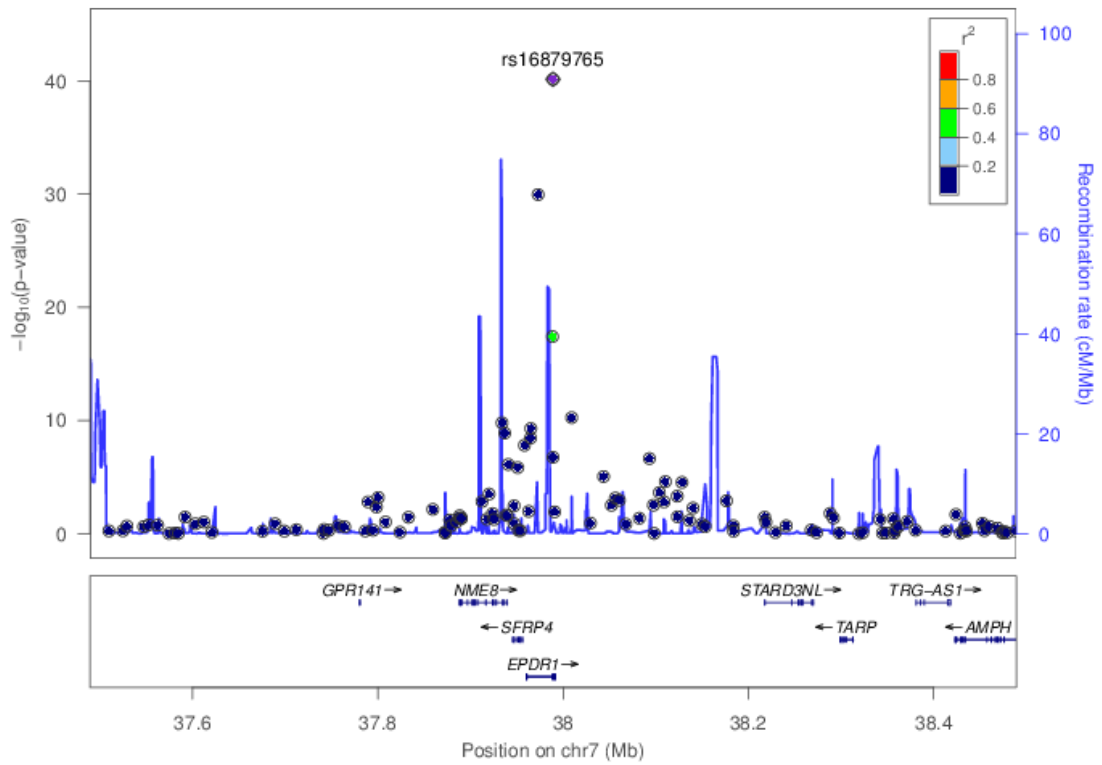
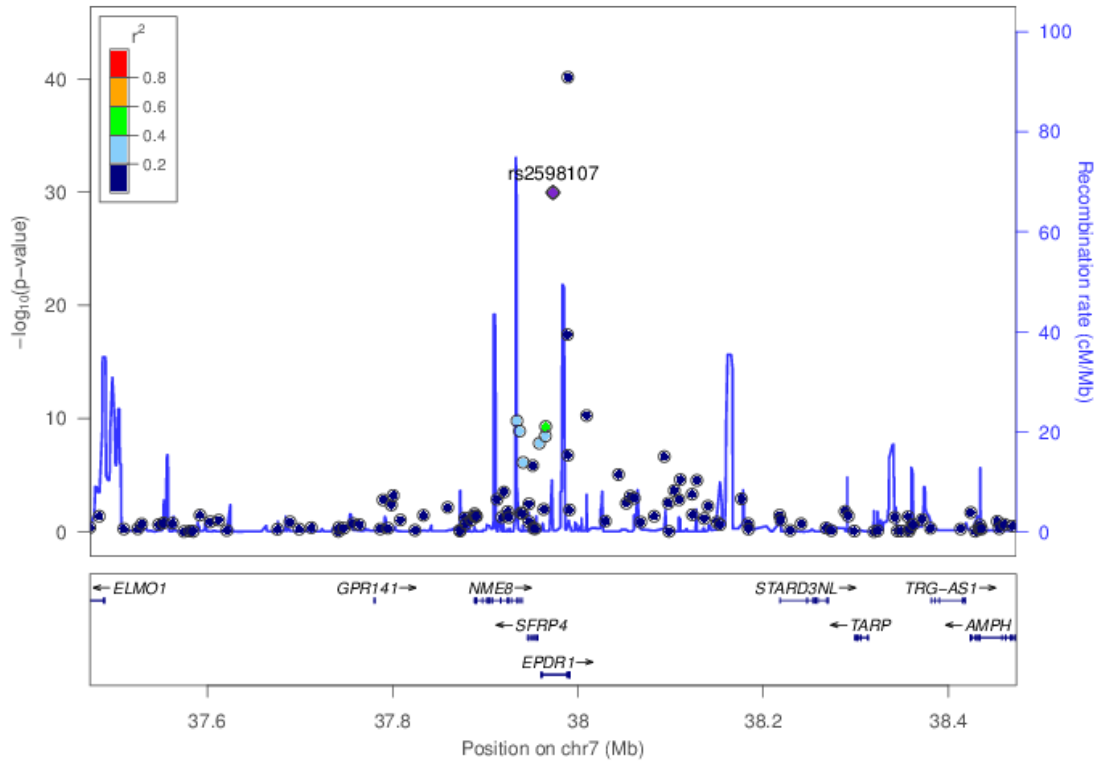
Figure S6. Annotated regional association plots for all 26 variants significantly associated with Dupuytren's disease ( $p \leq 5 \times 10^{-8}$ ) after meta-analysis of discovery and replication cohorts, plotted with LocusZoom<sup>1</sup>. Recombination rate was calculated within the software using data from HapMap genotypes. In the lower panel of each figure, genes within 500kb of the index SNP are shown. The position on each chromosome is shown in relation to Human Genome build hg19.

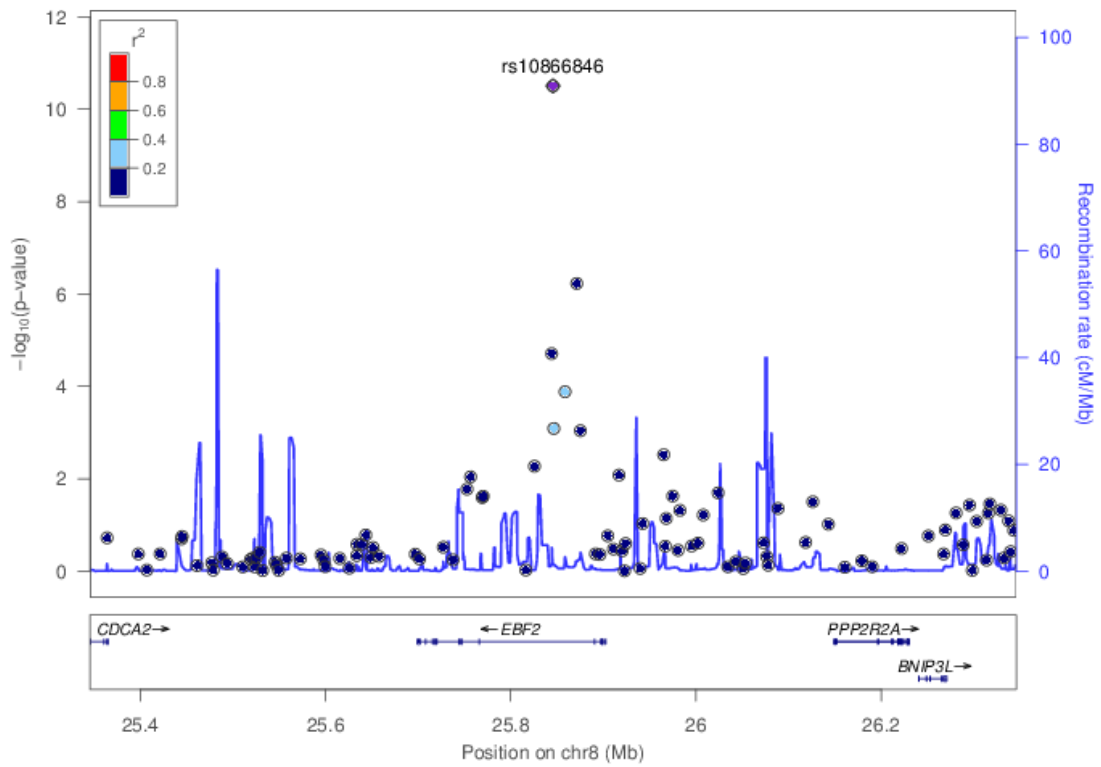
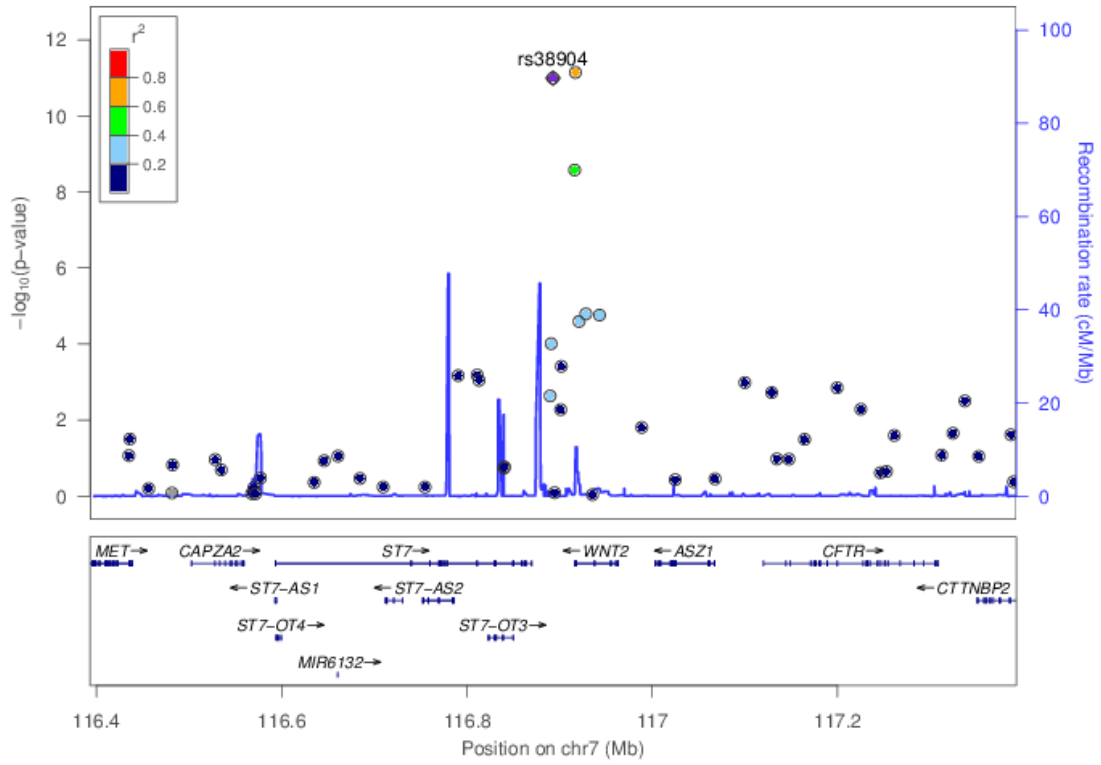


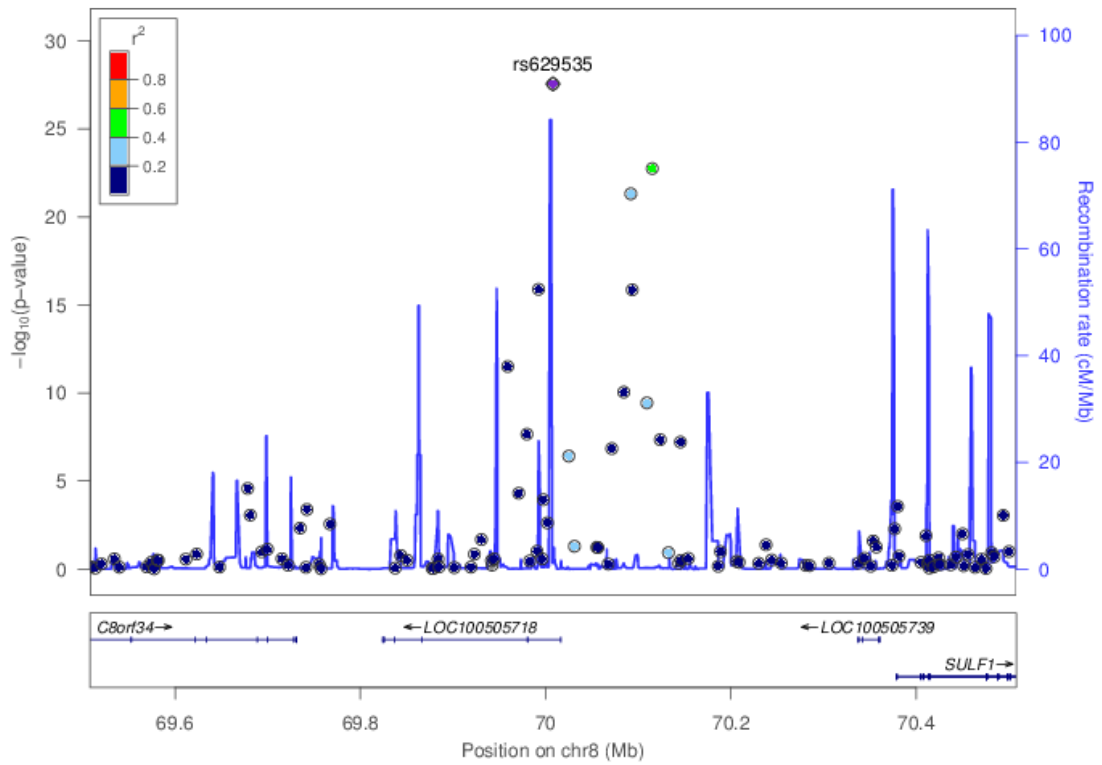
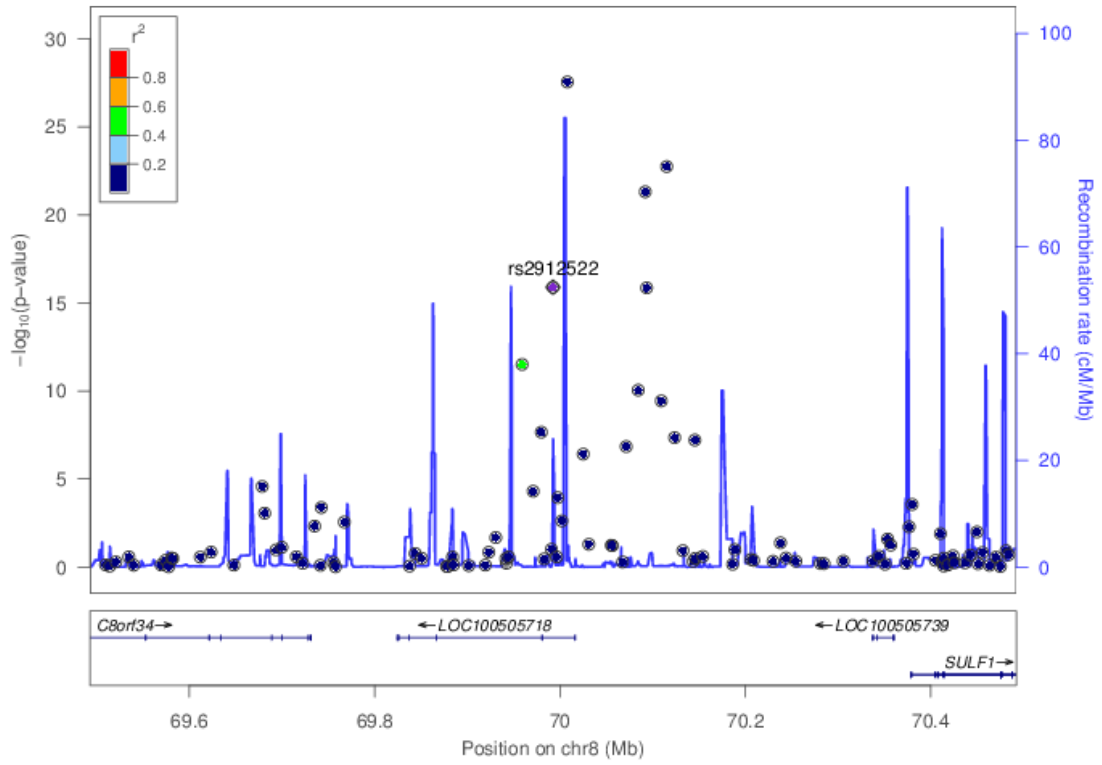


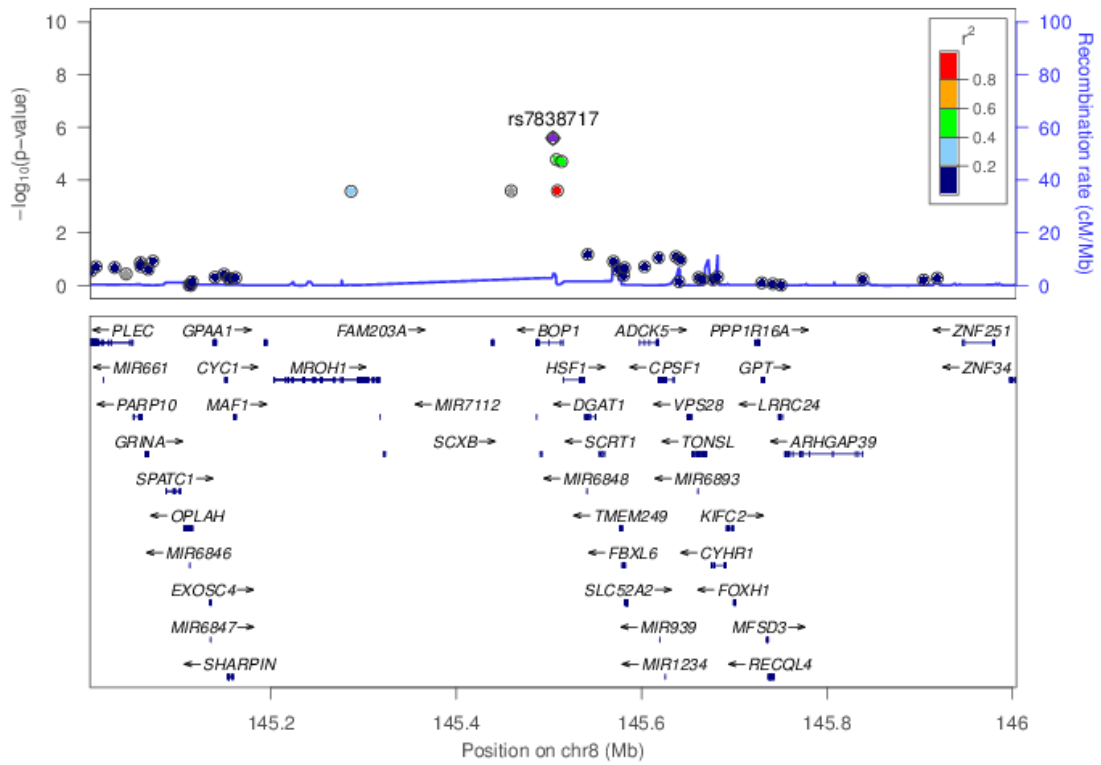
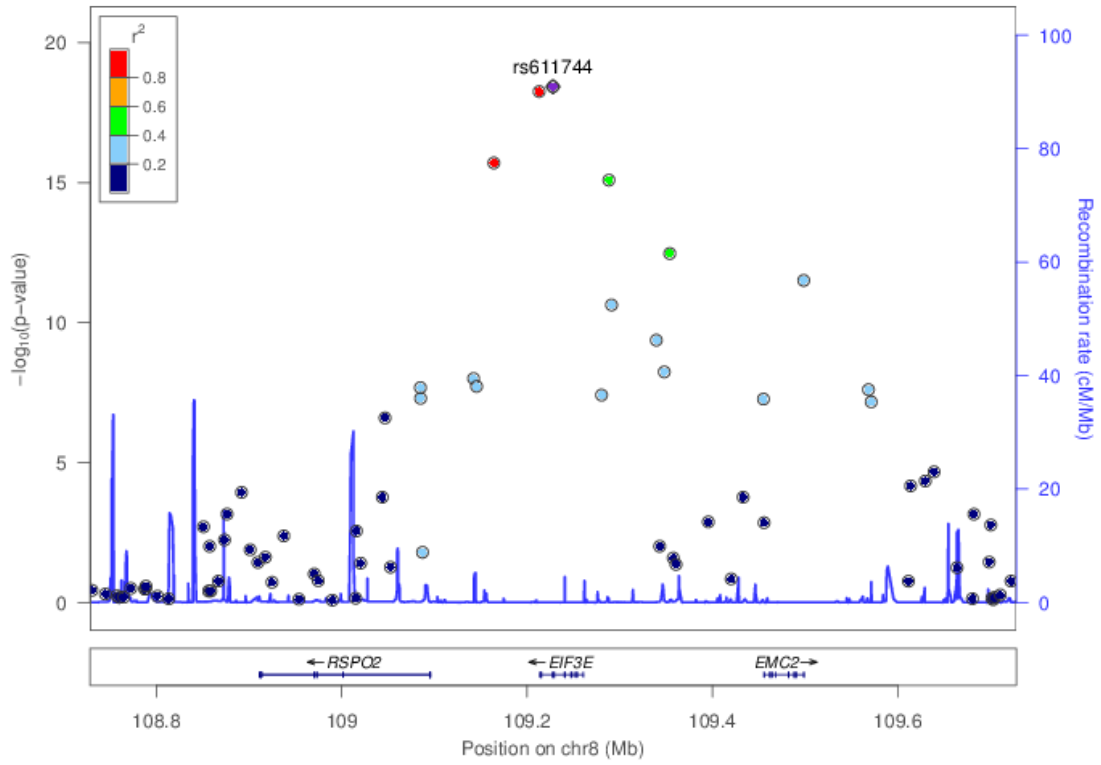


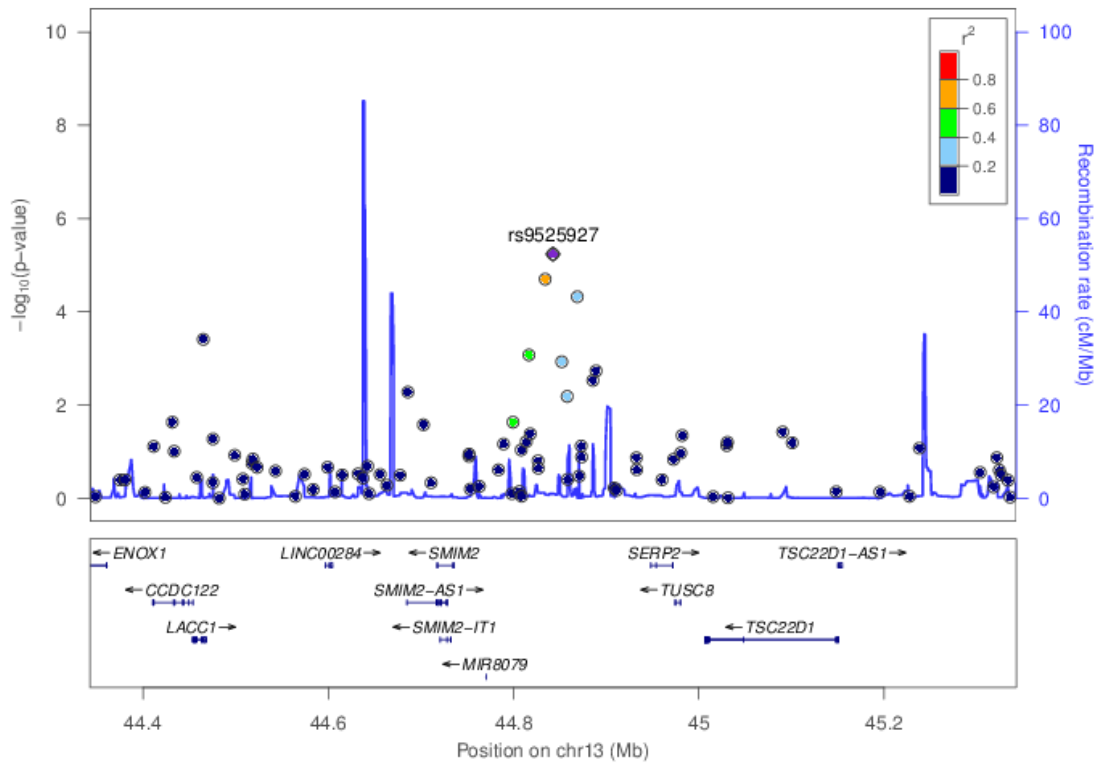
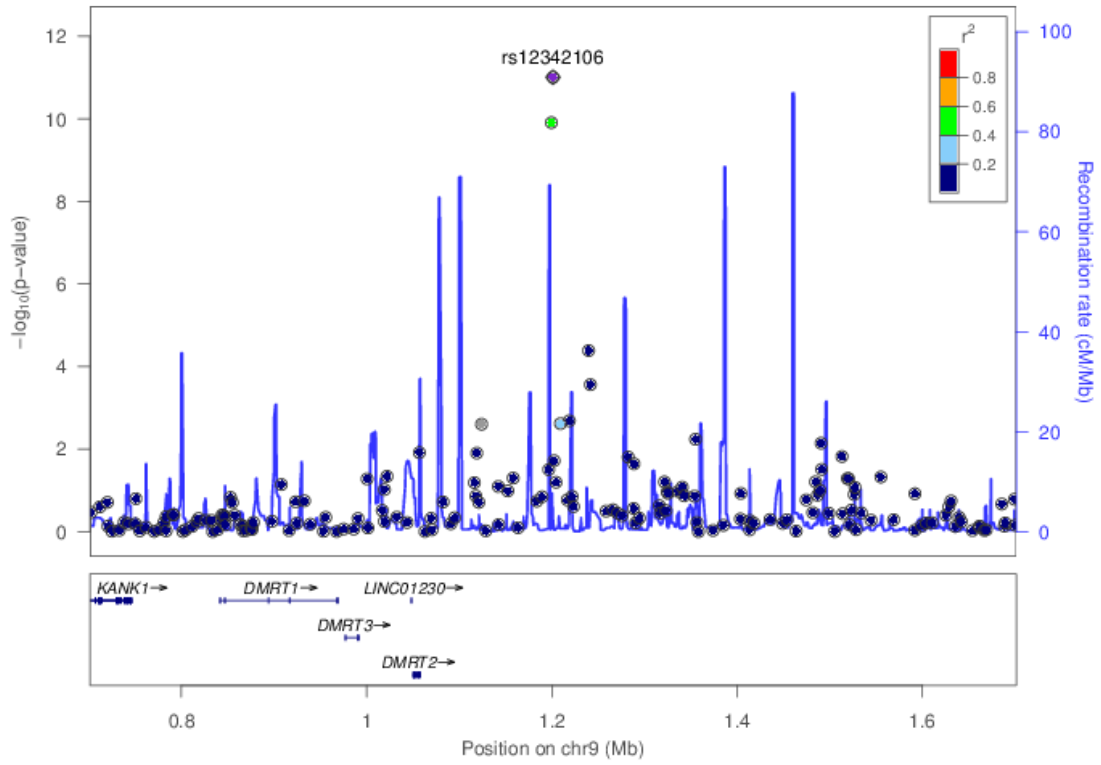


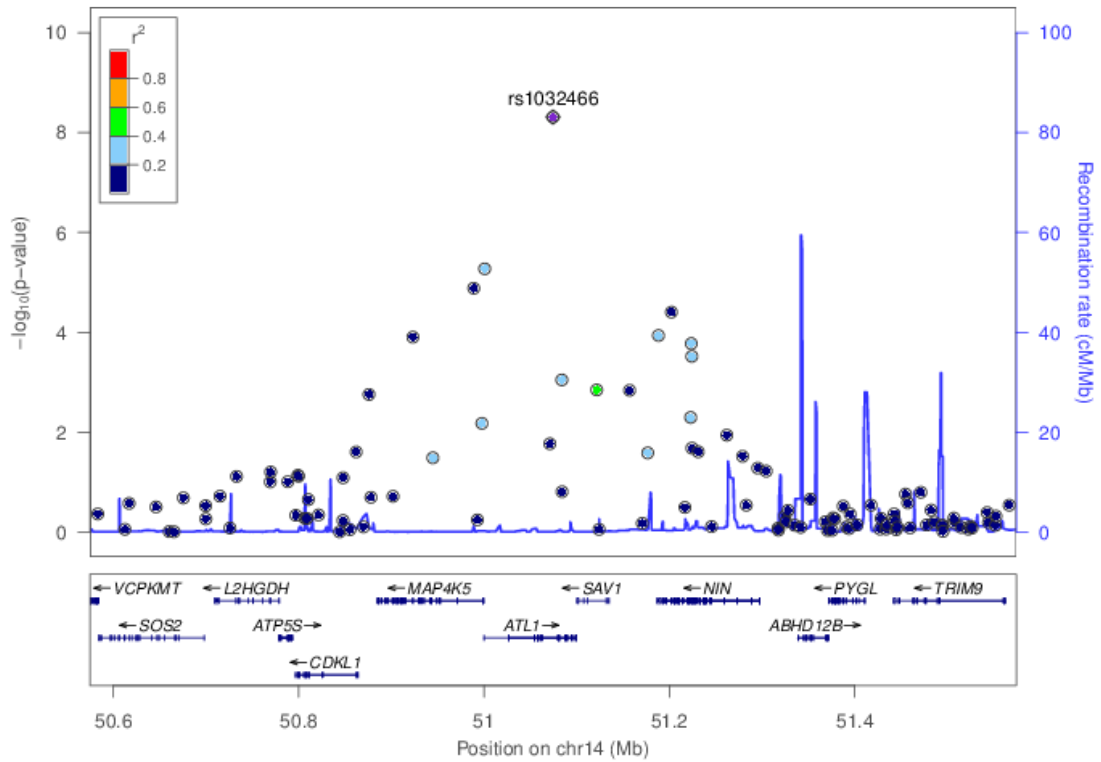
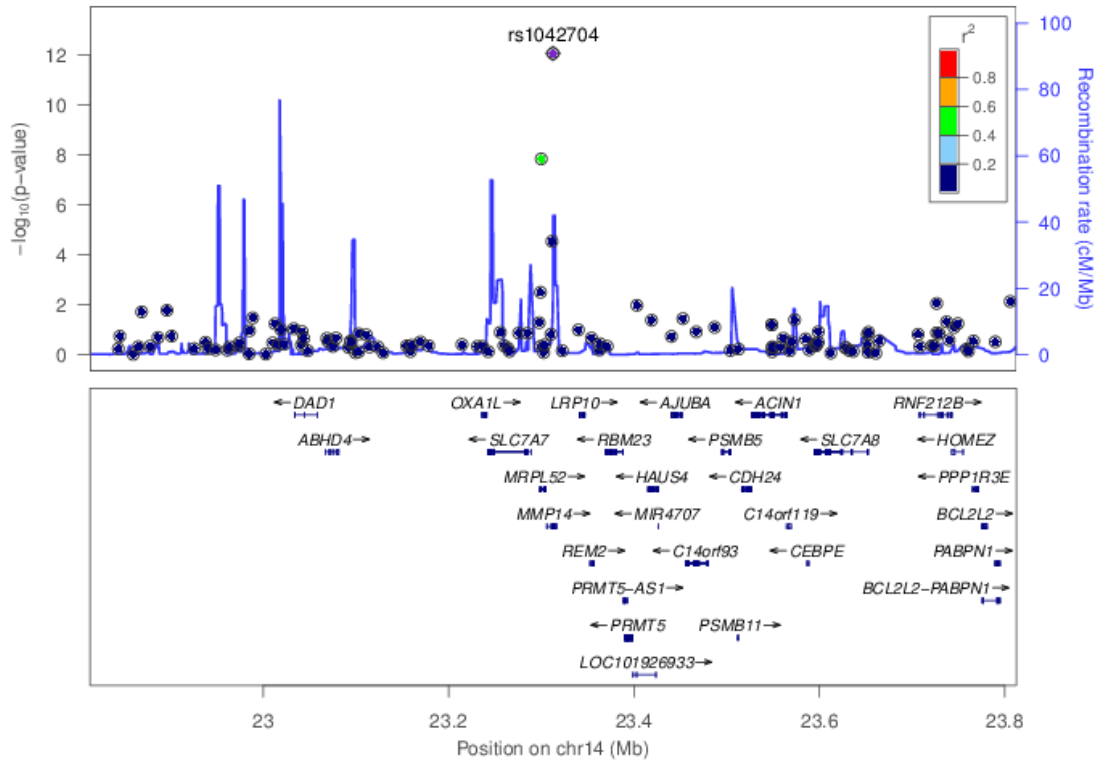


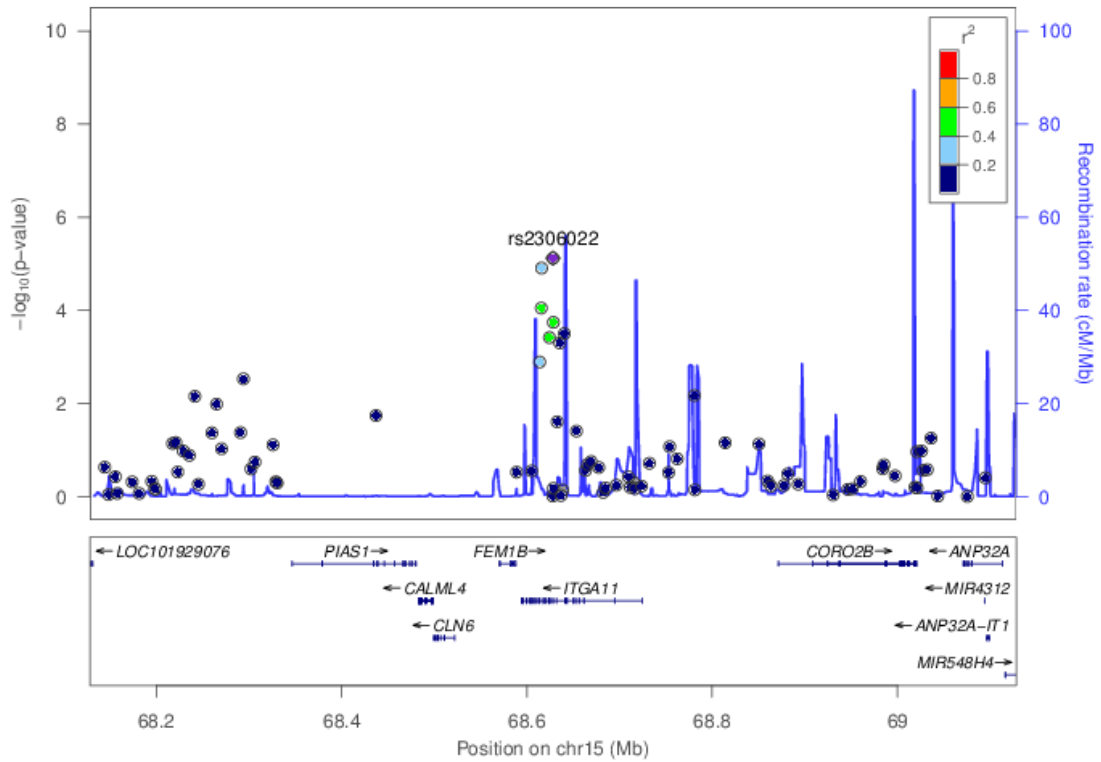
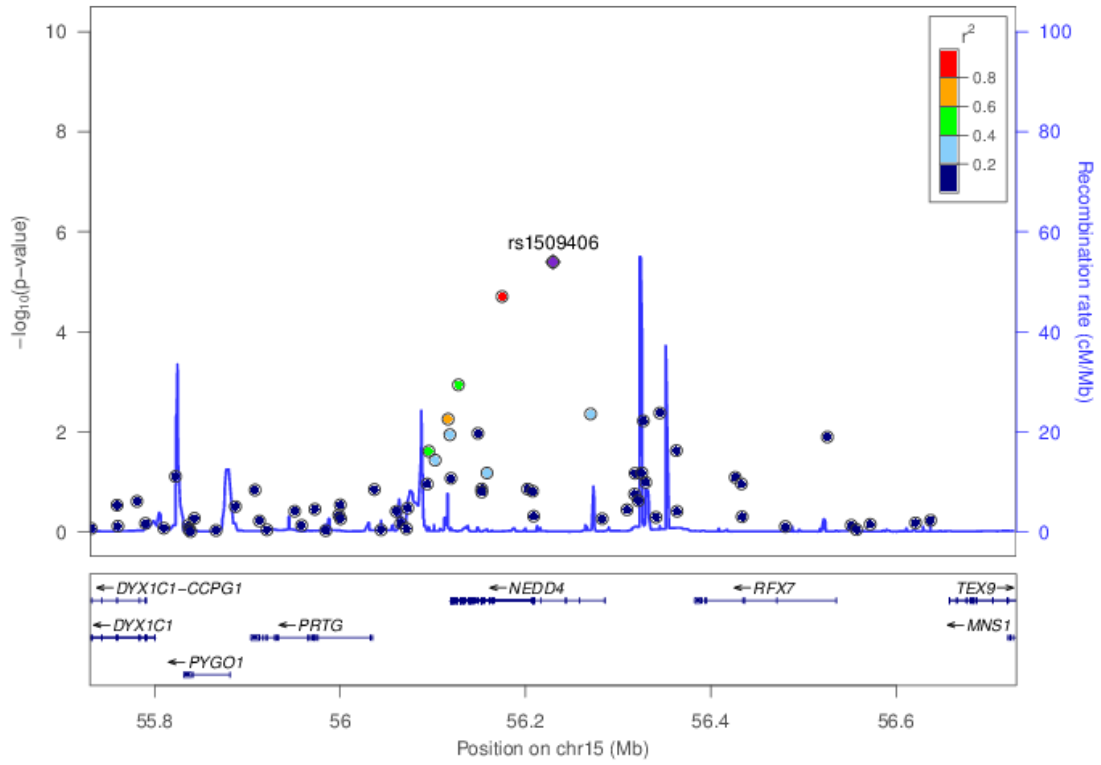


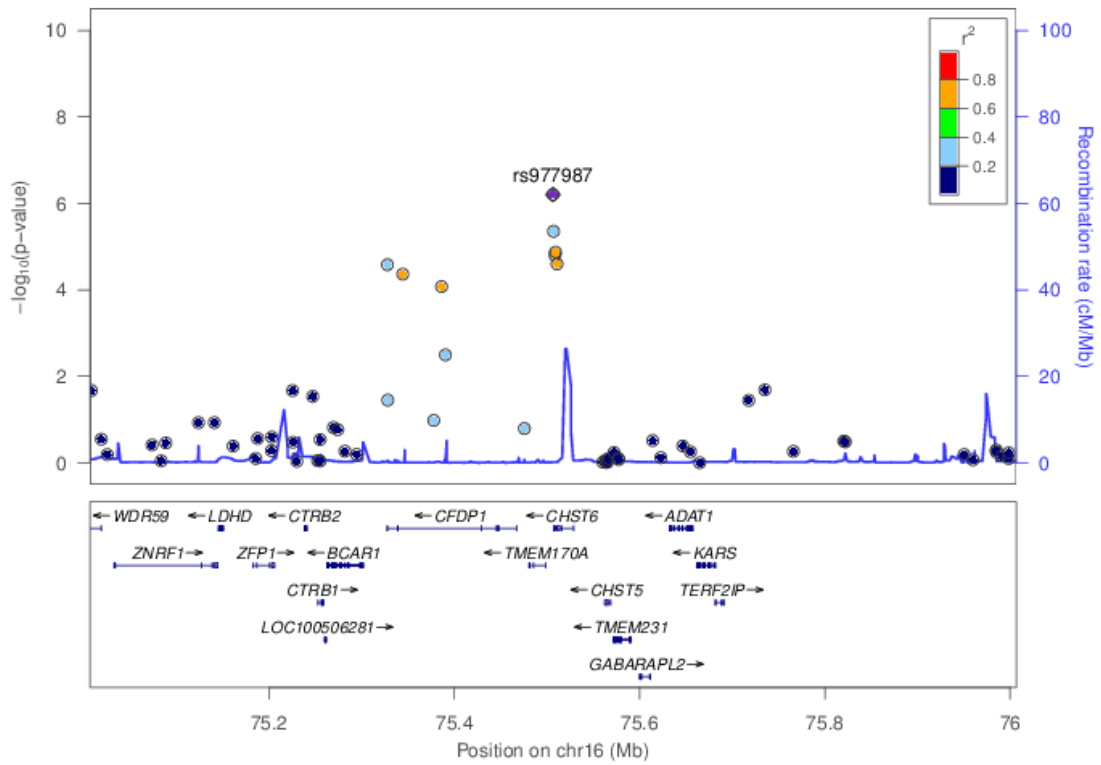
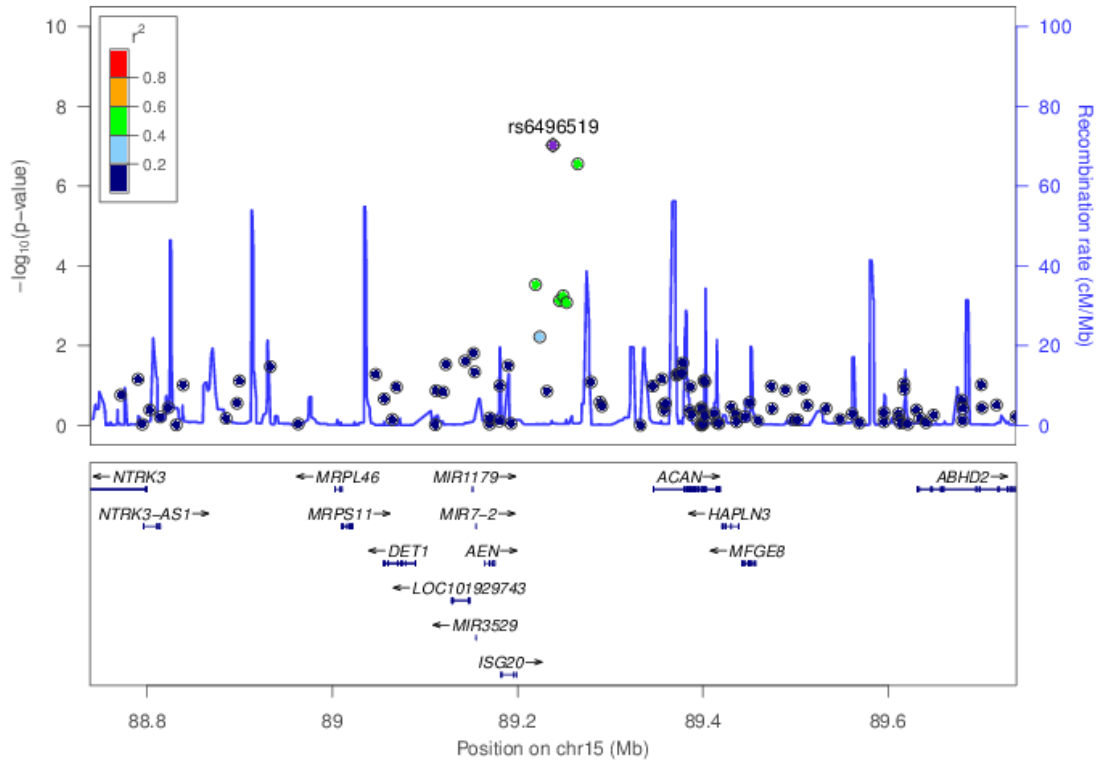




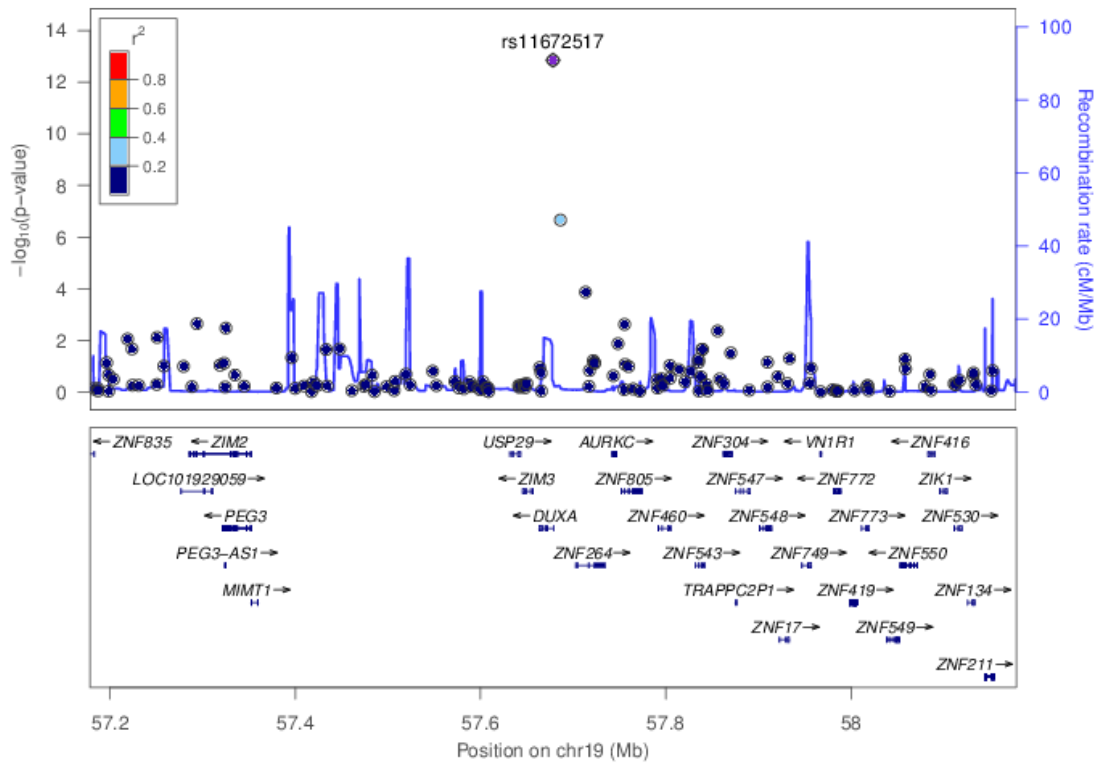
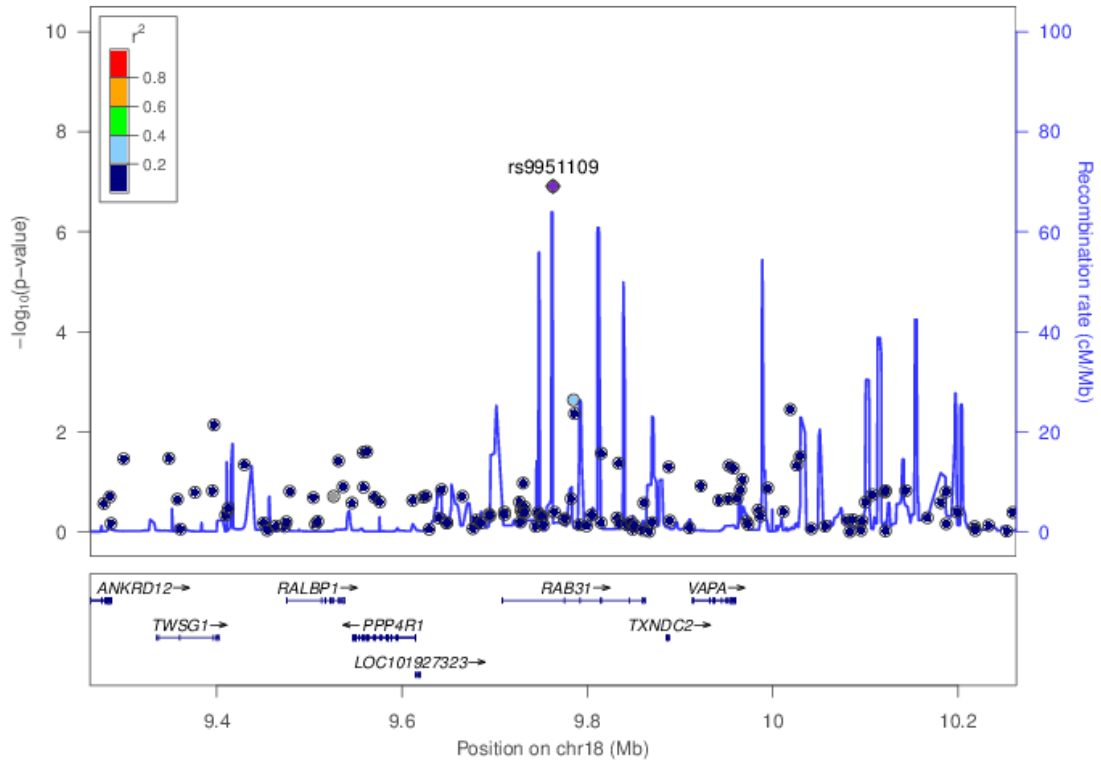


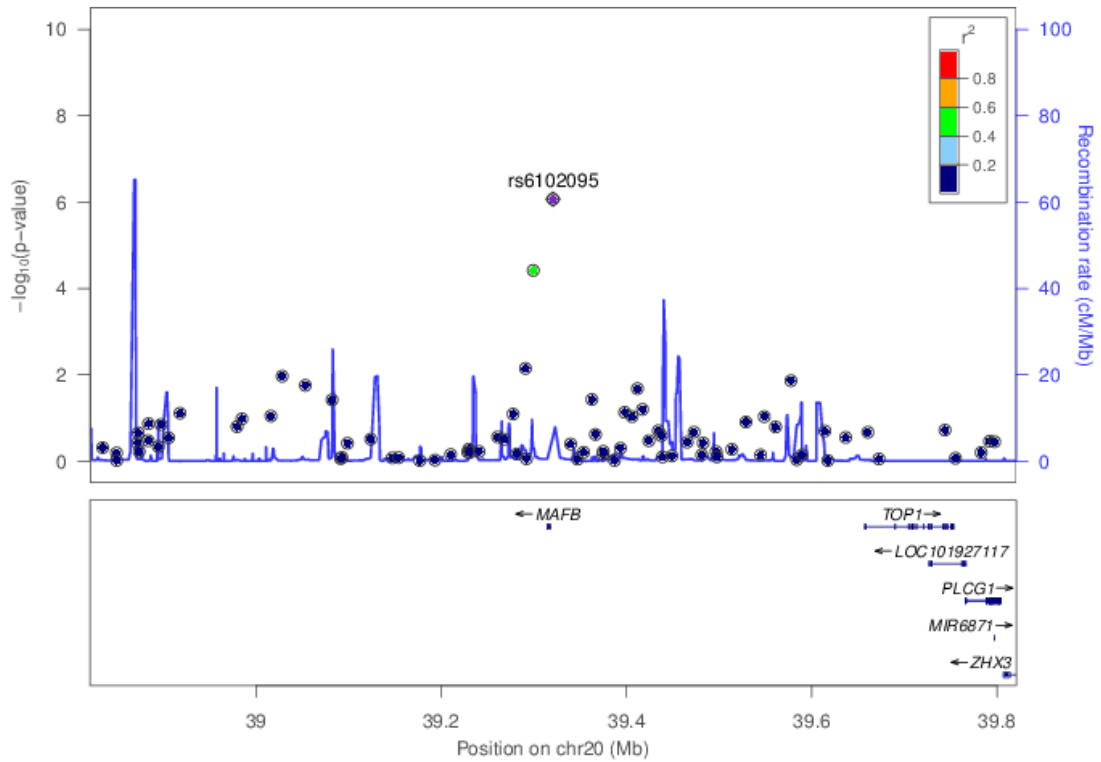
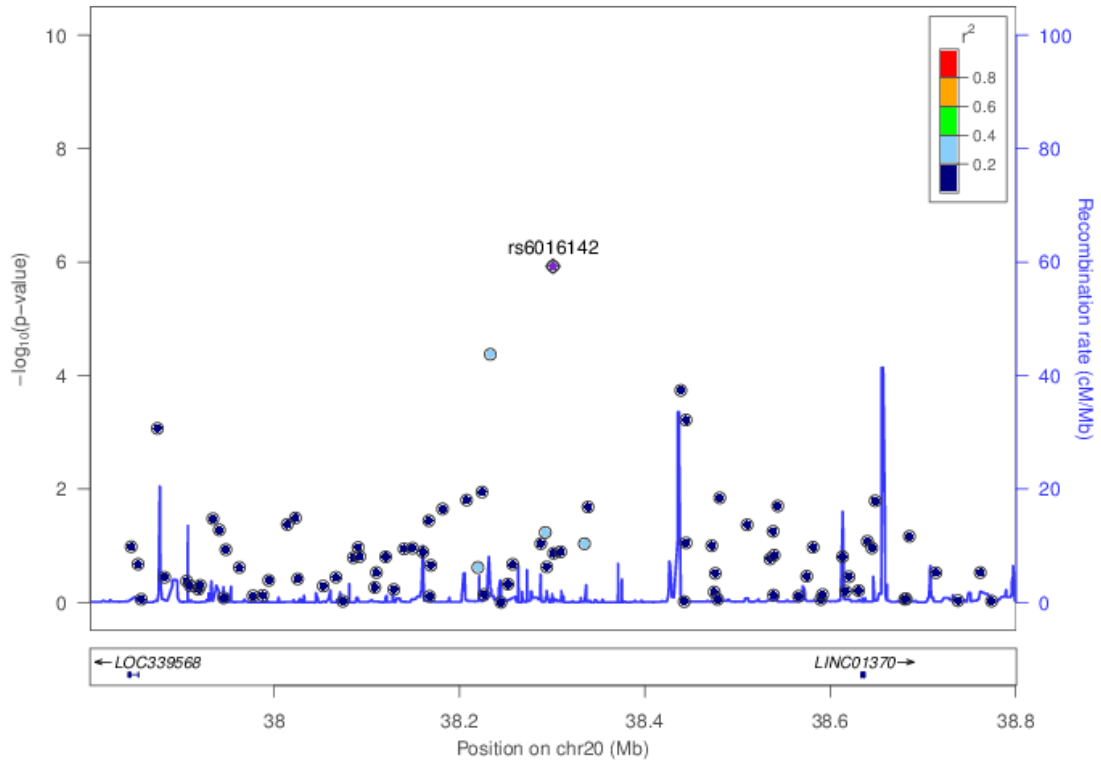












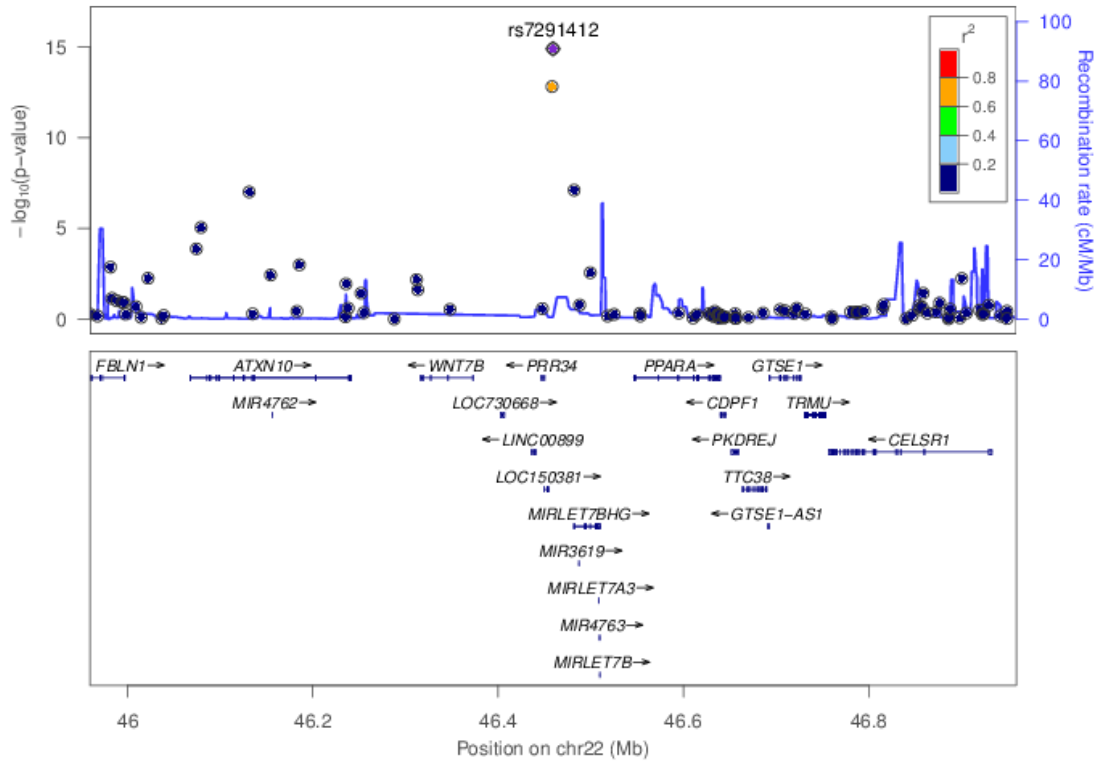
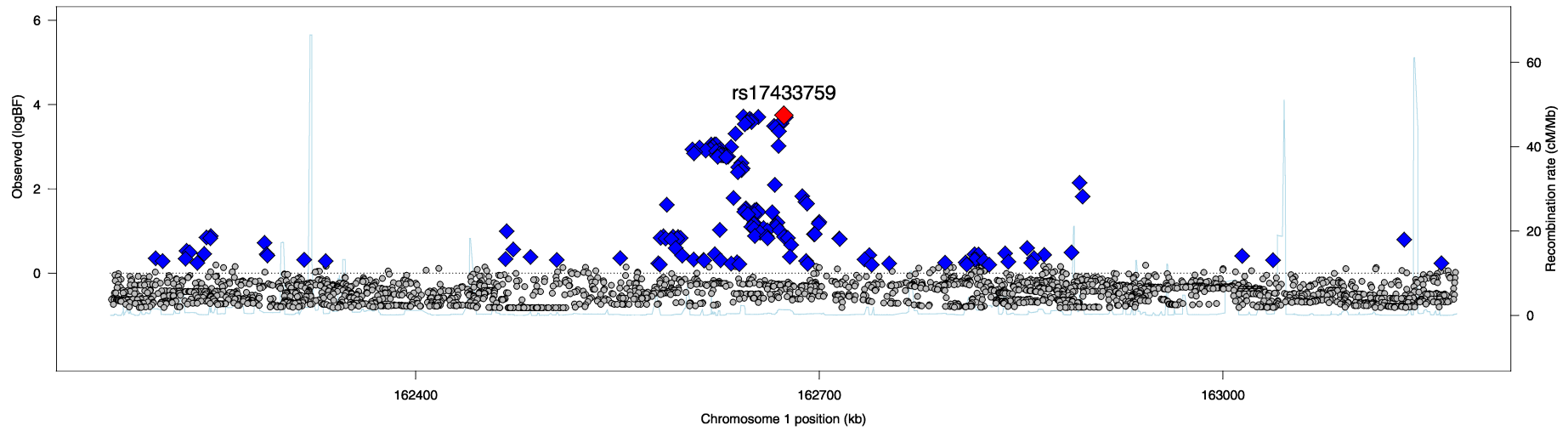
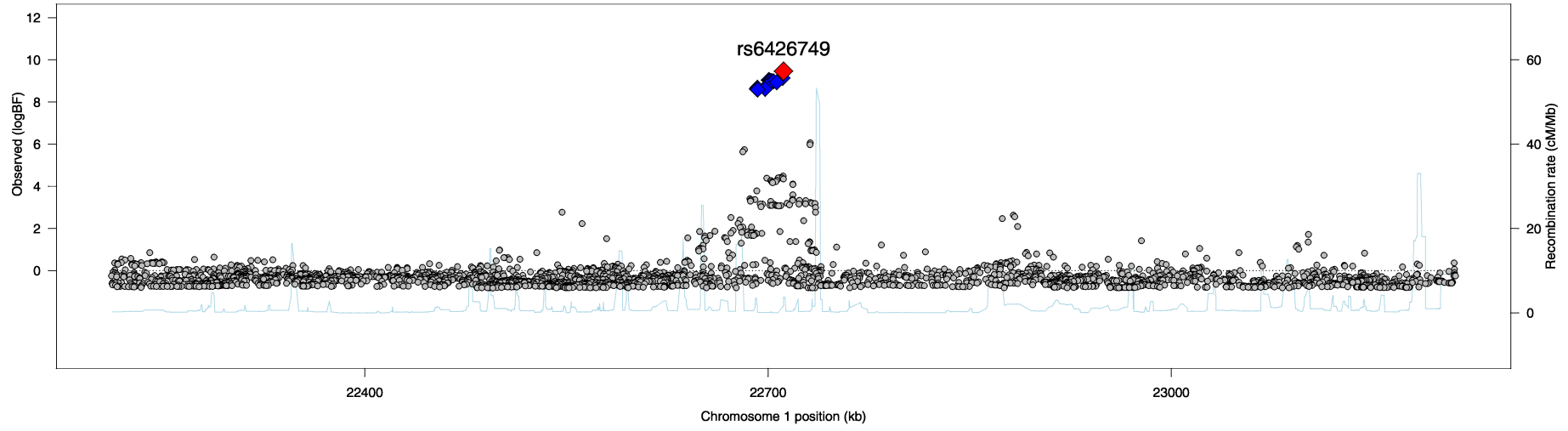
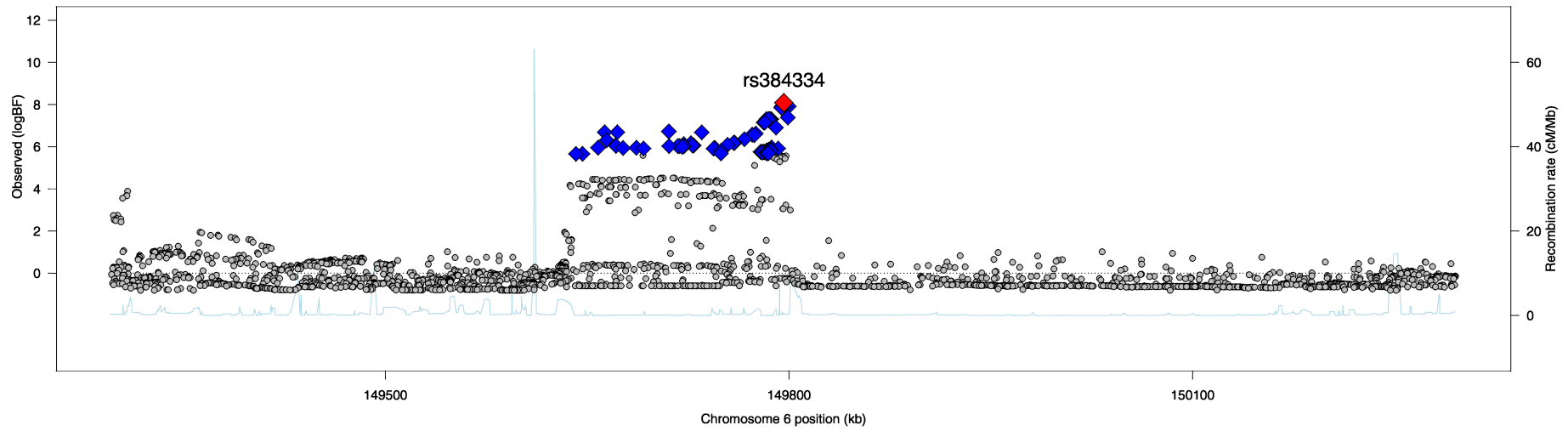
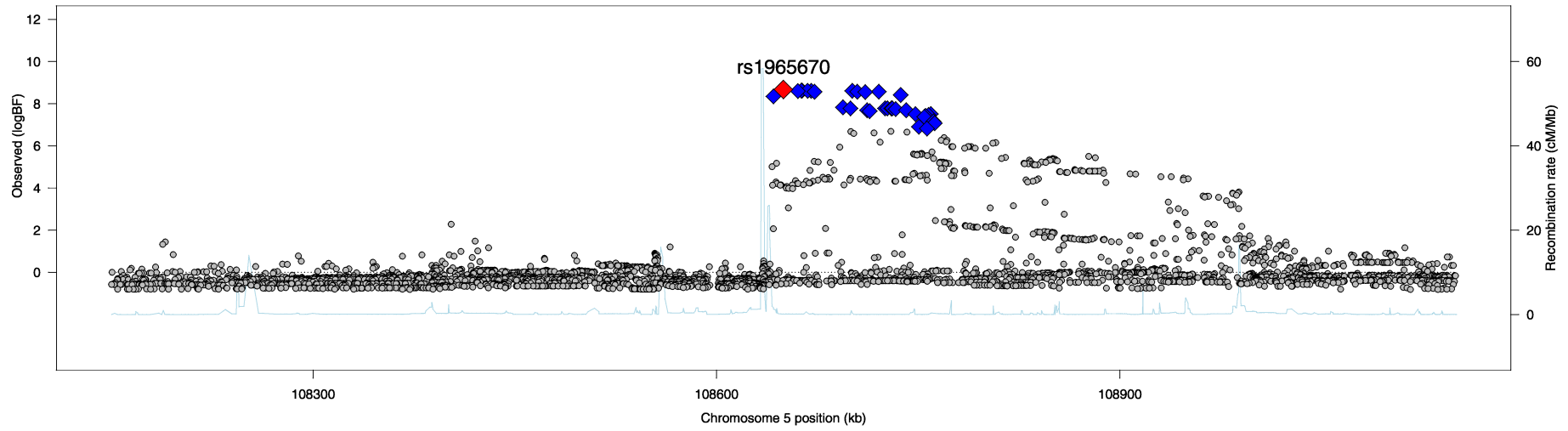
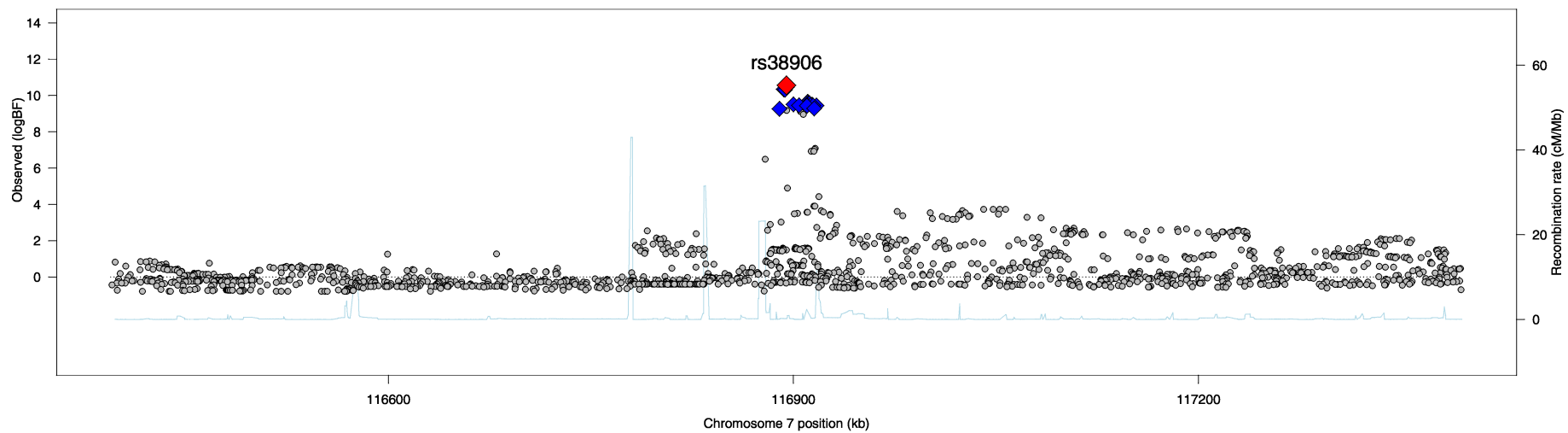
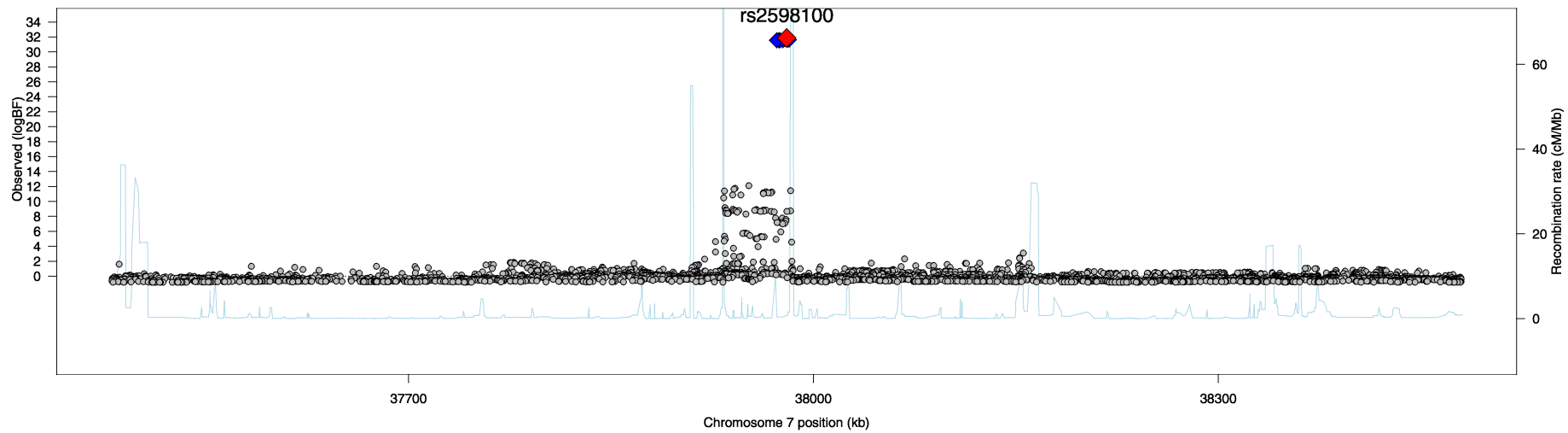
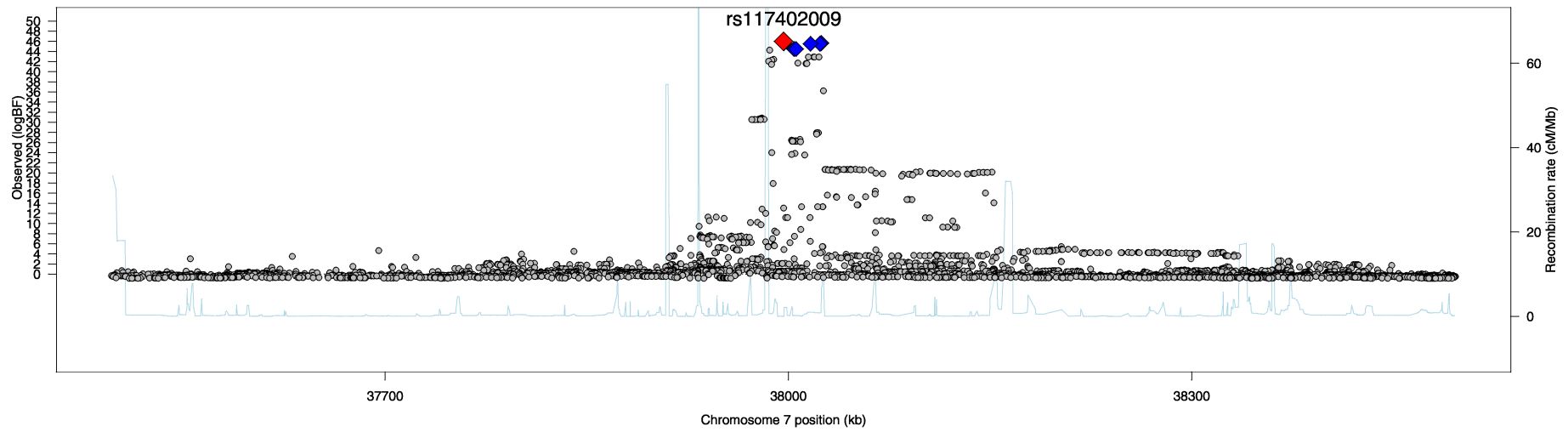
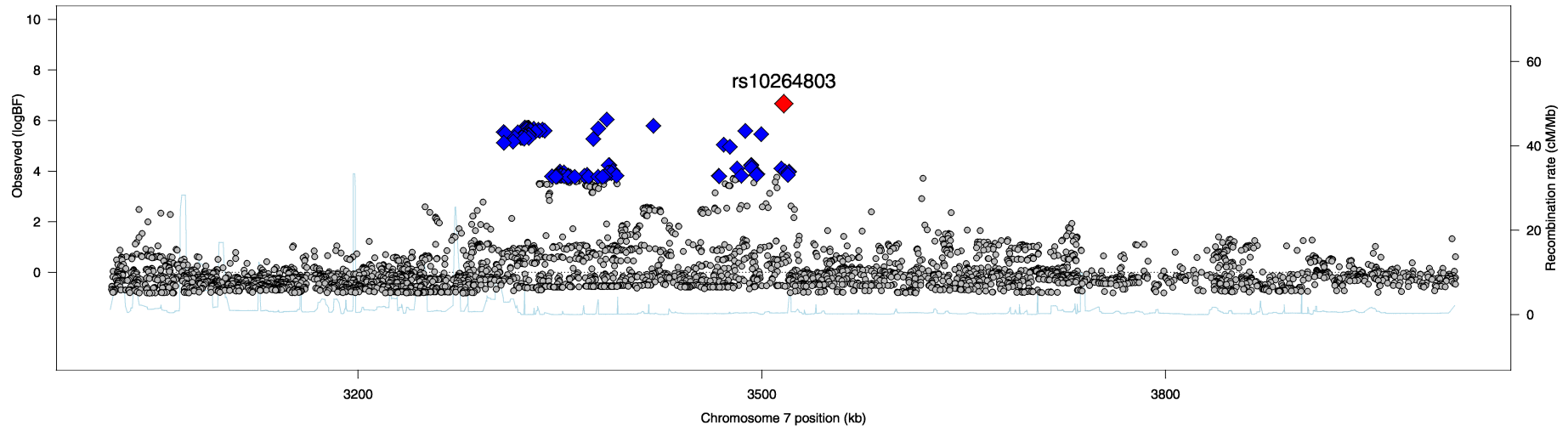


Figure S7. Graphical illustration of the 99% credible sets for all 26 associated regions. For each regional plot, the SNP with the highest posterior probability of association is labeled and depicted with a red diamond. Other SNPs that make up the 99% credible set (Table S2) are depicted with blue diamonds. Imputed SNPs that do not form part of the 99% credible set are depicted with grey circles. Recombination peaks are depicted in light blue and are calculated from HapMap genotypes. The position on each chromosome is shown in relation to Human Genome build hg19.

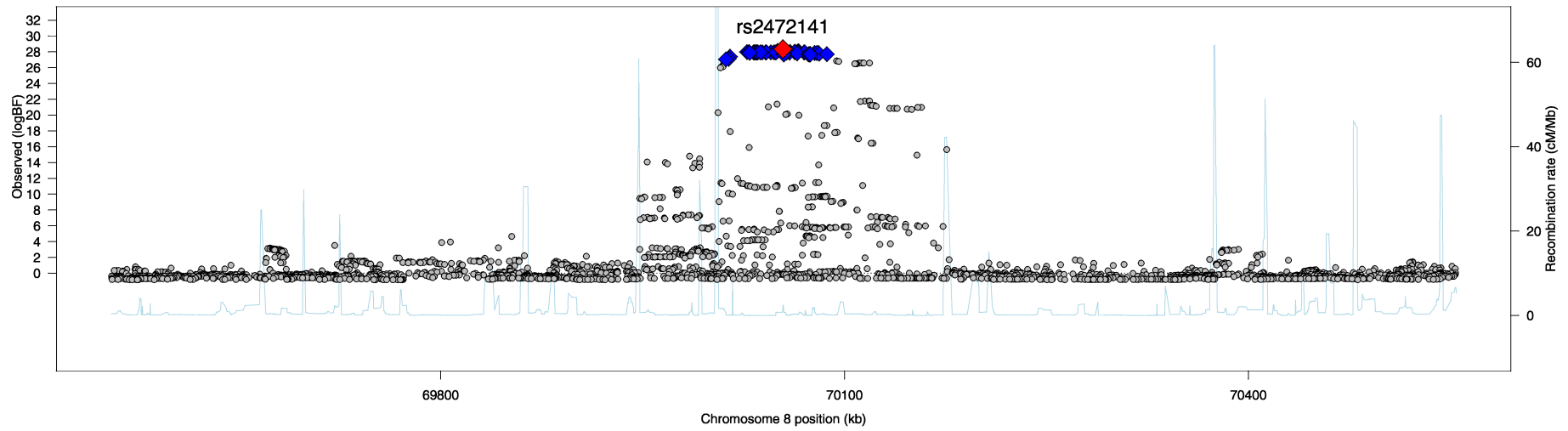
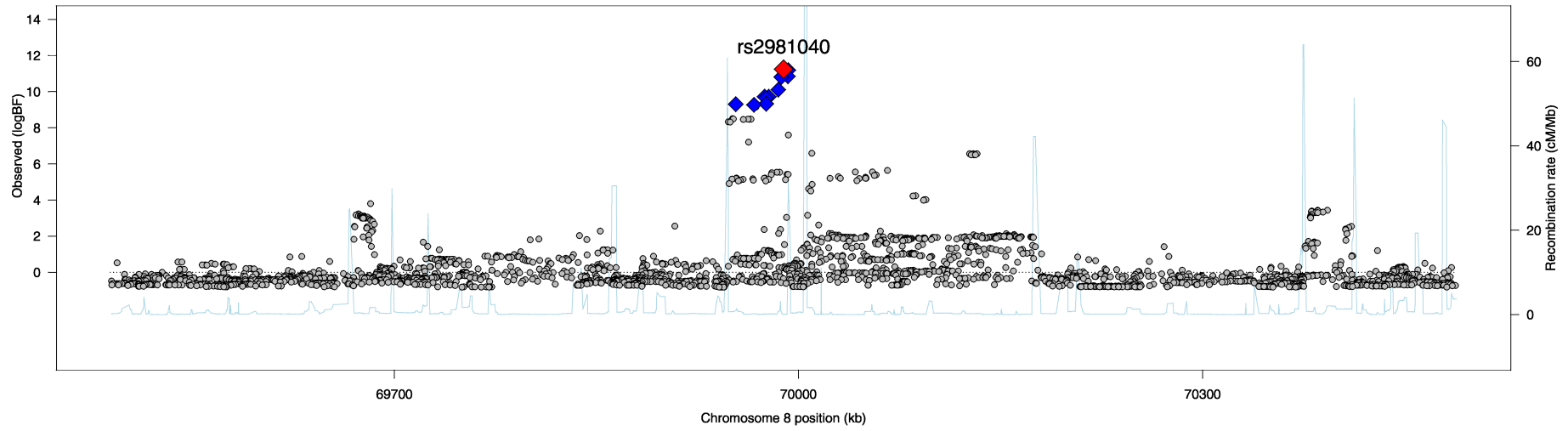


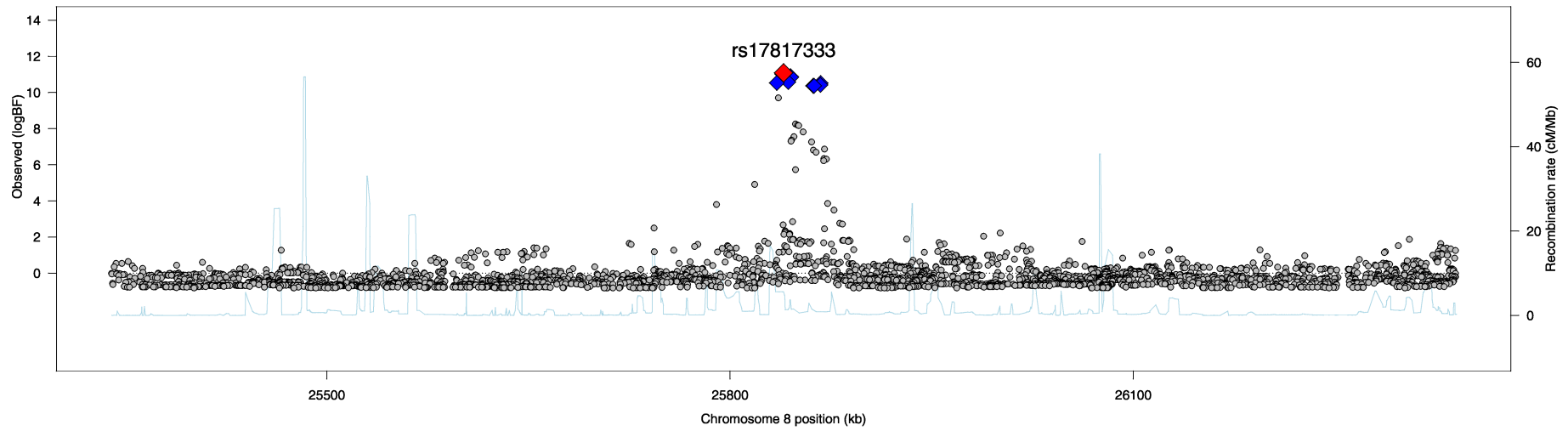
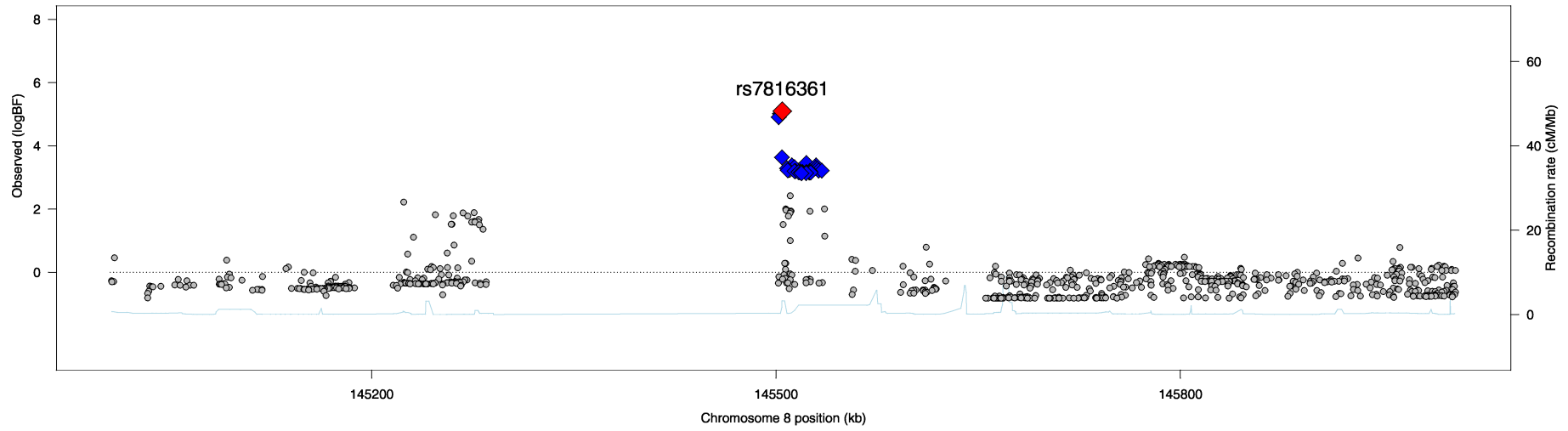


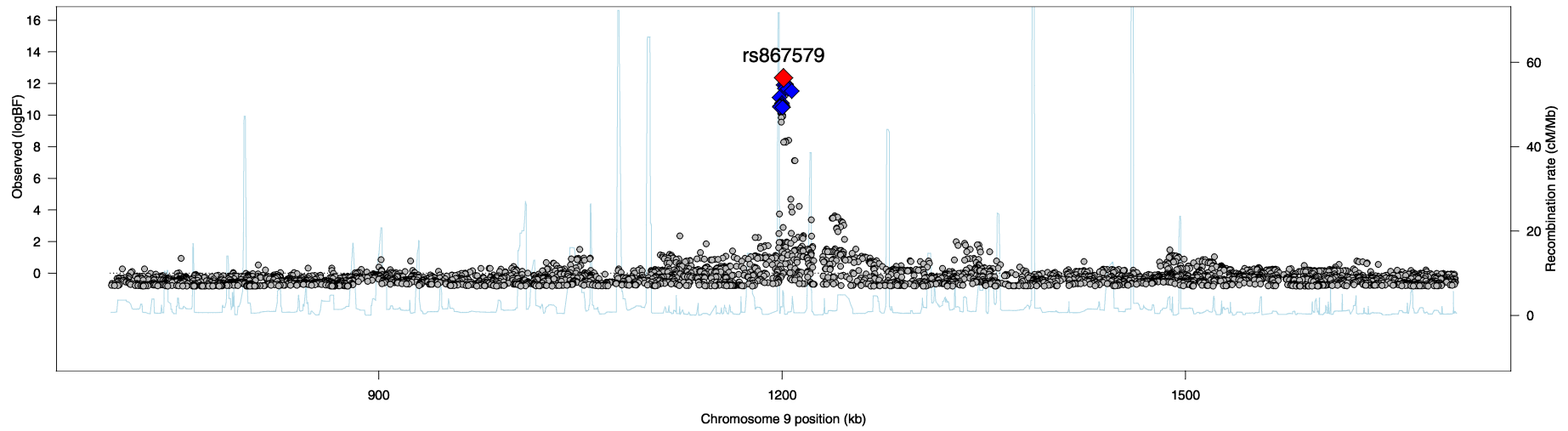
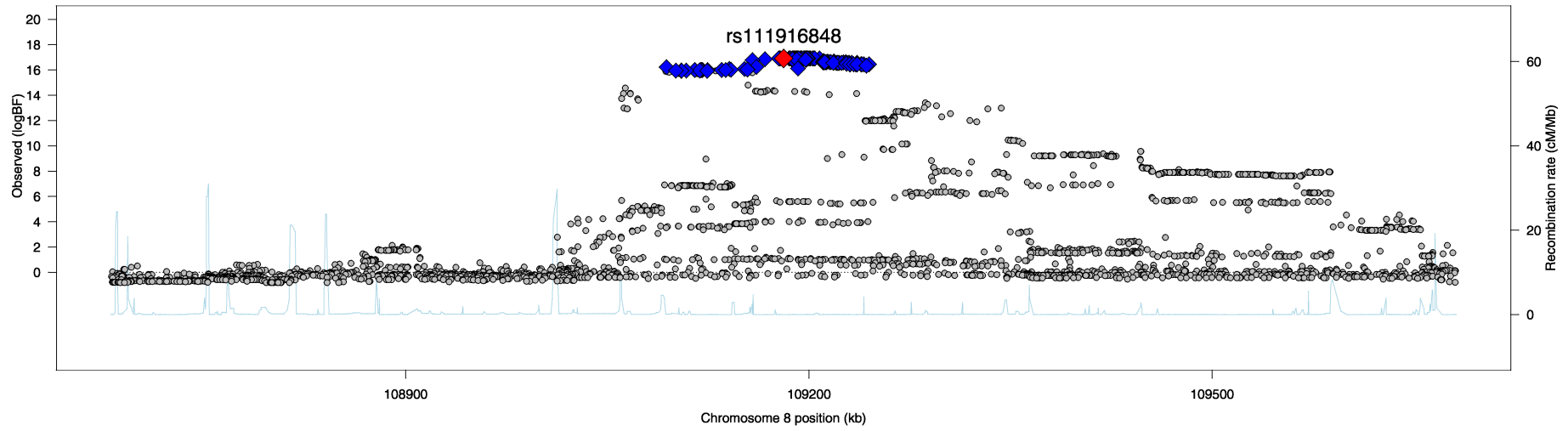


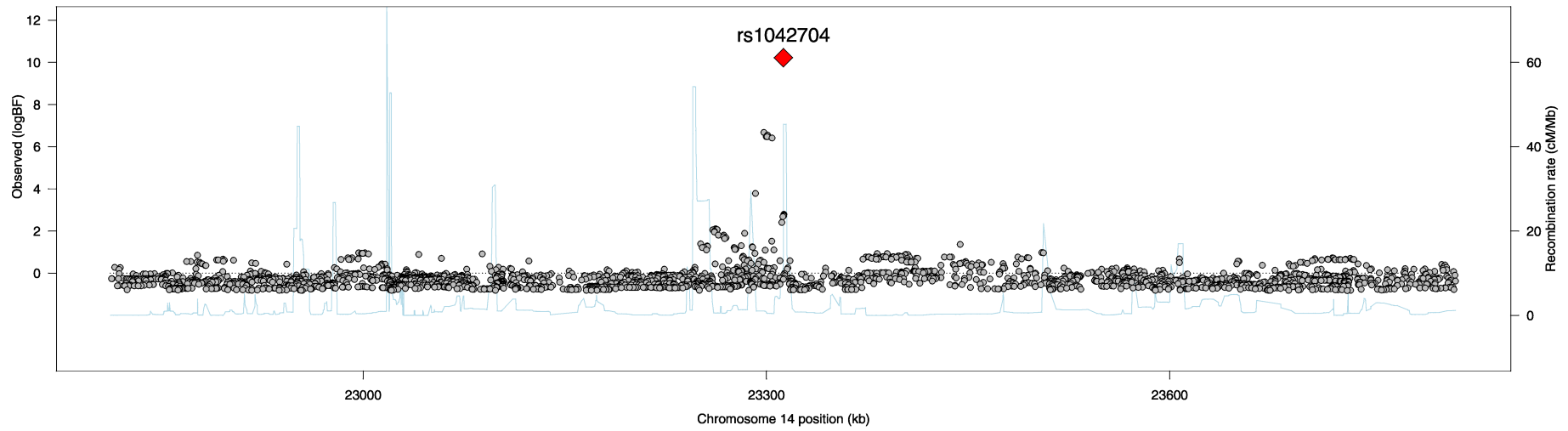
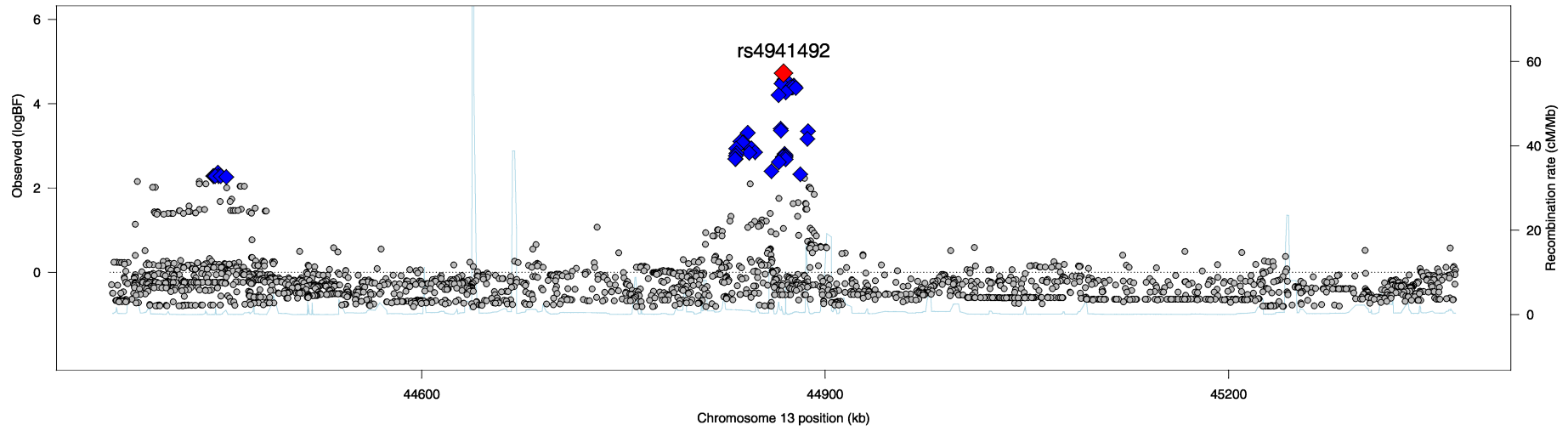


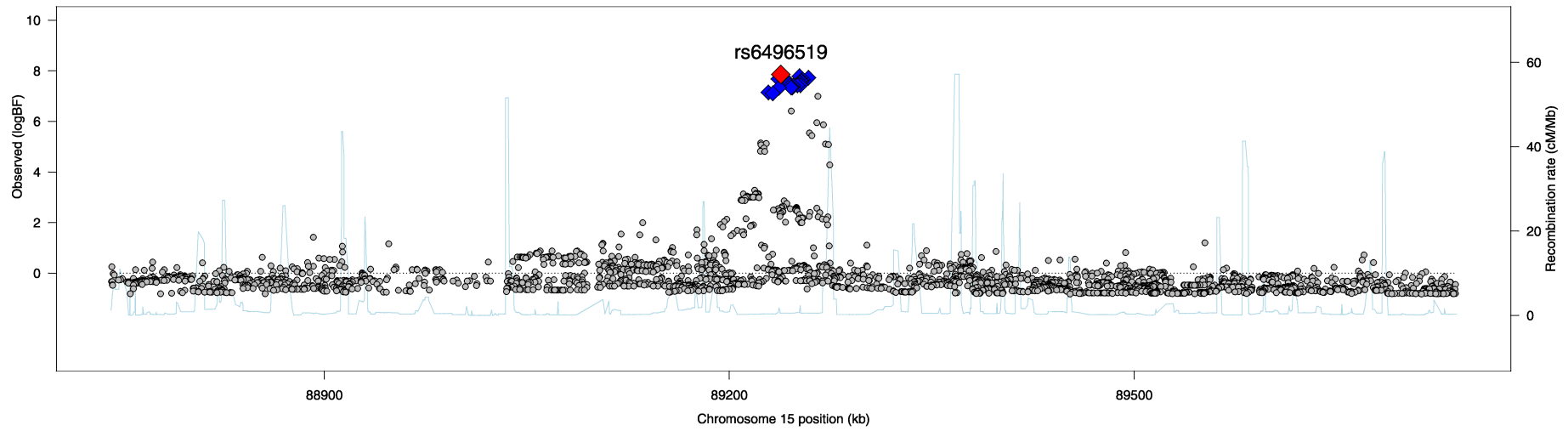
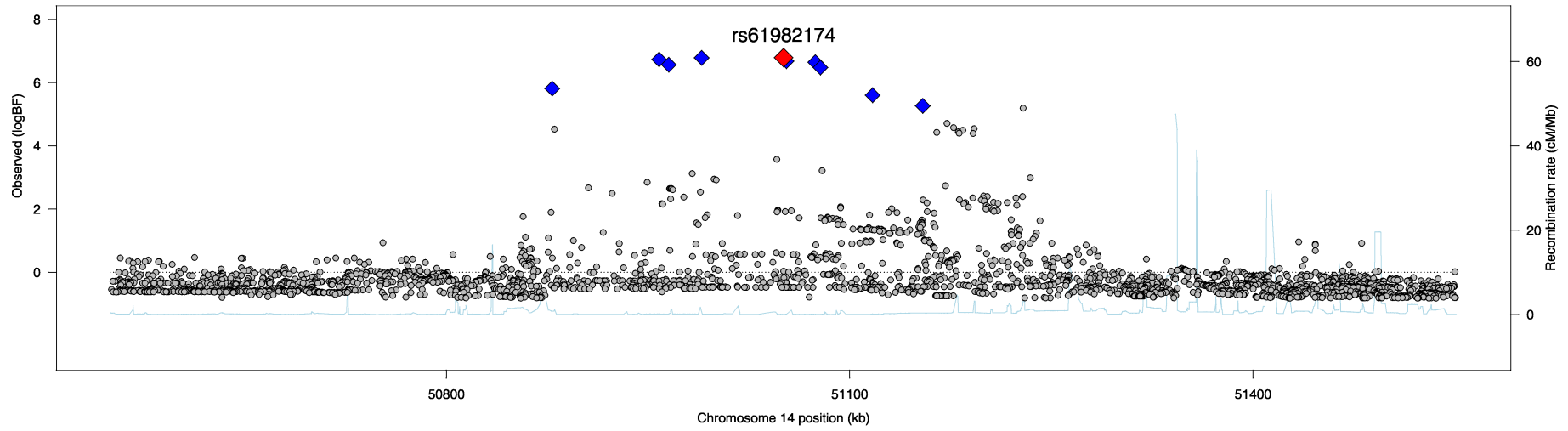


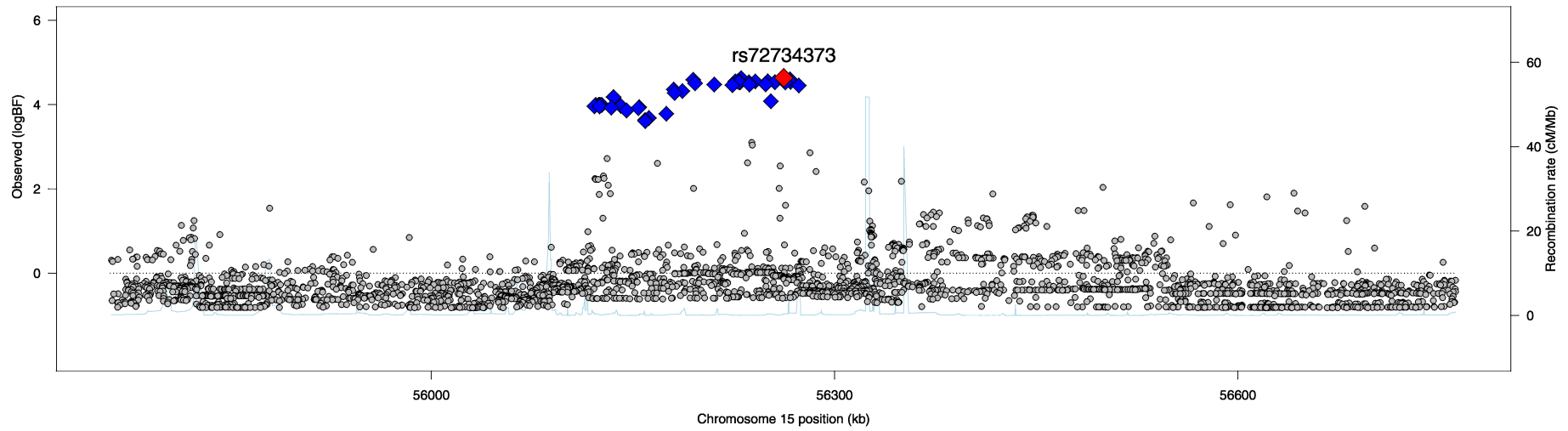
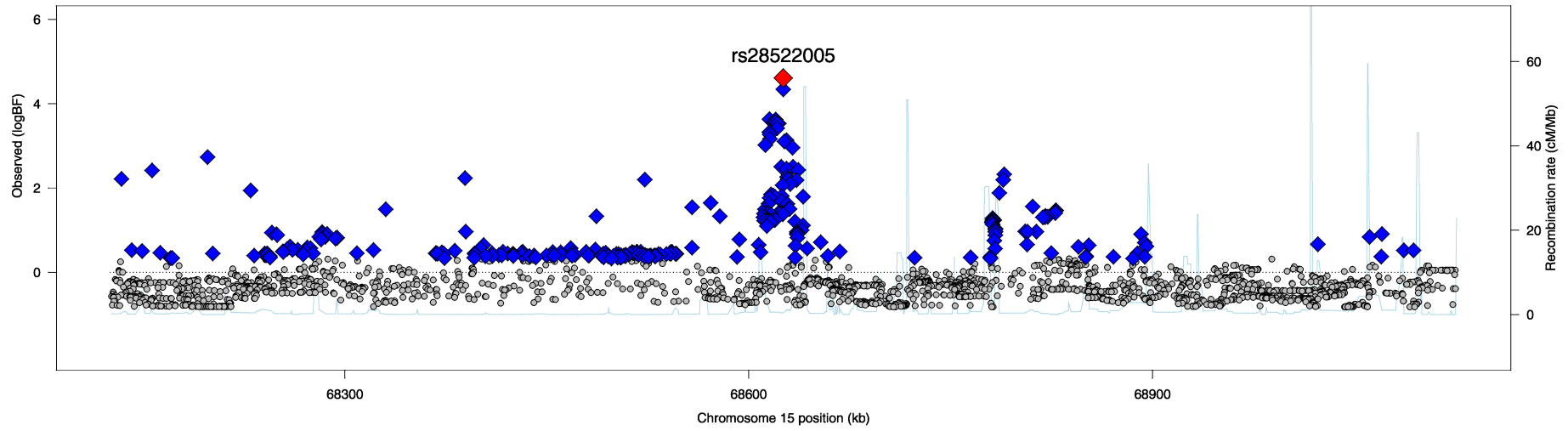


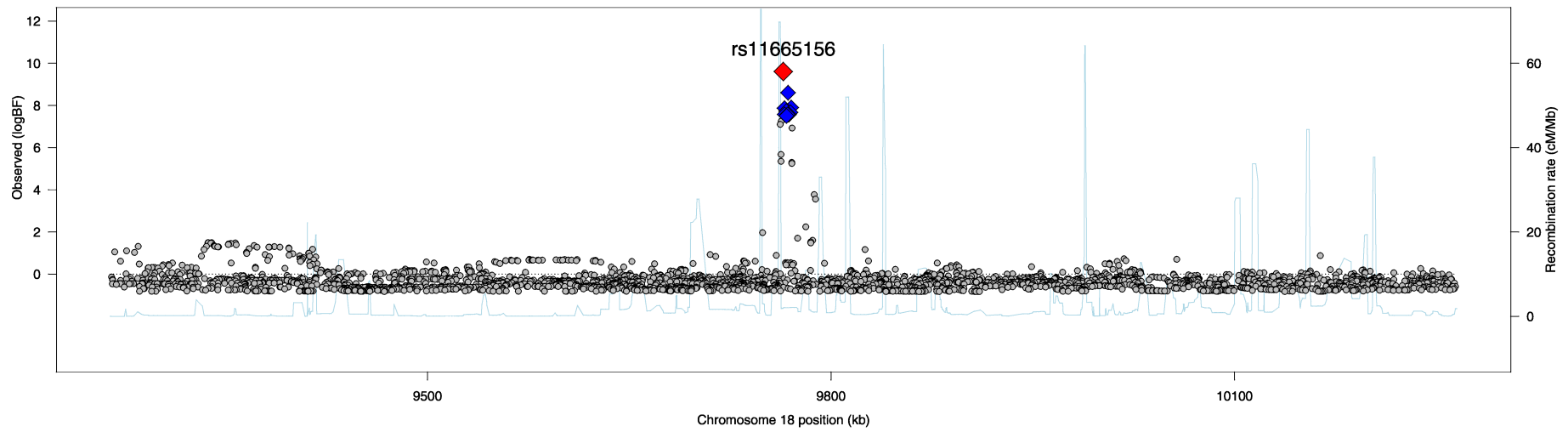
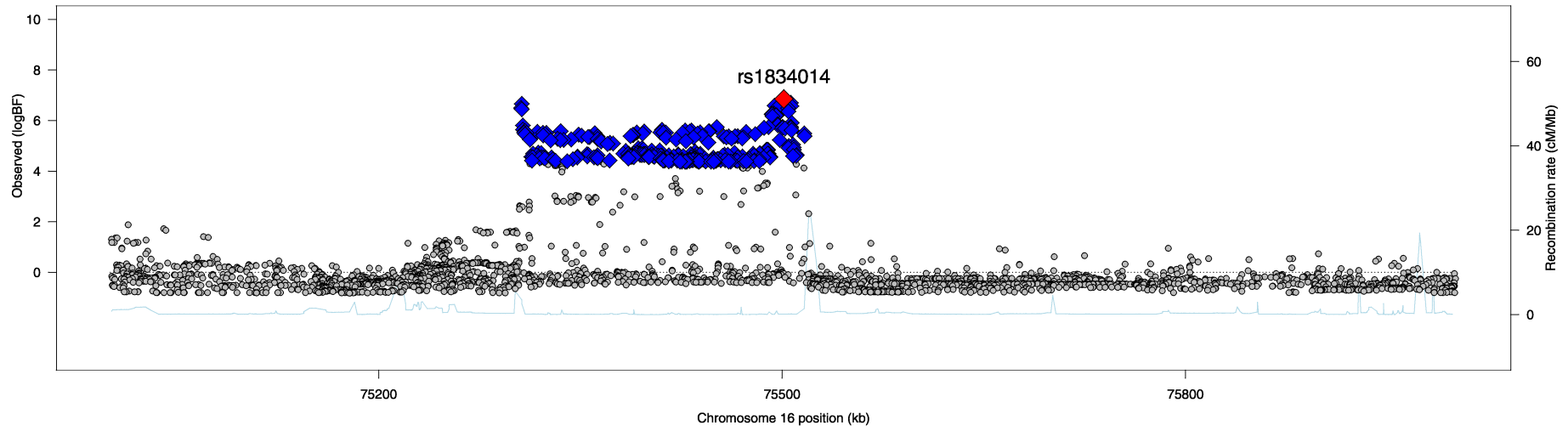


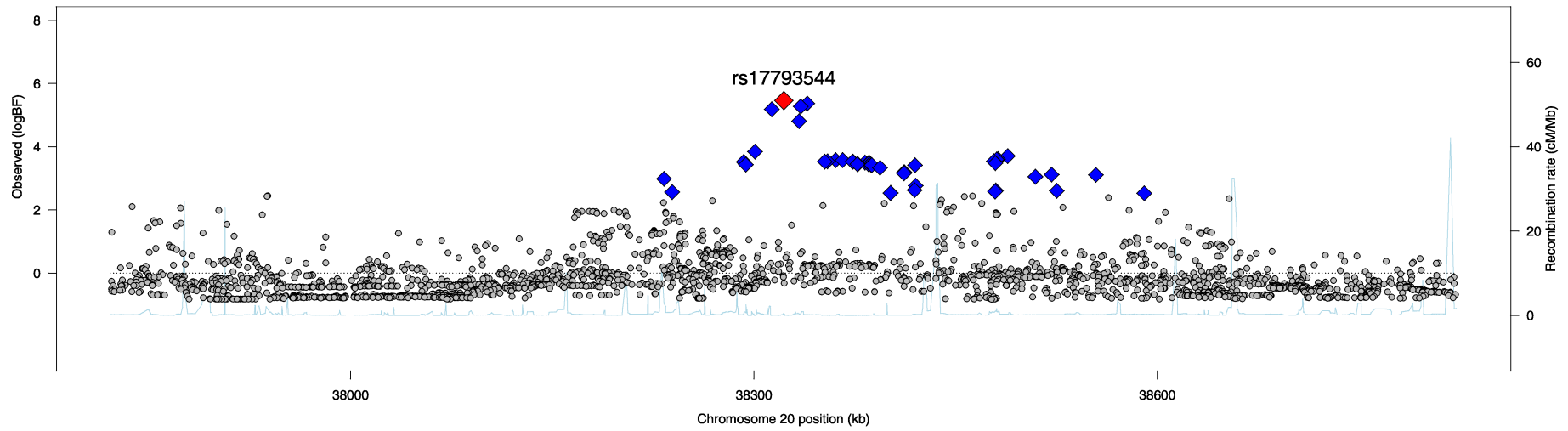
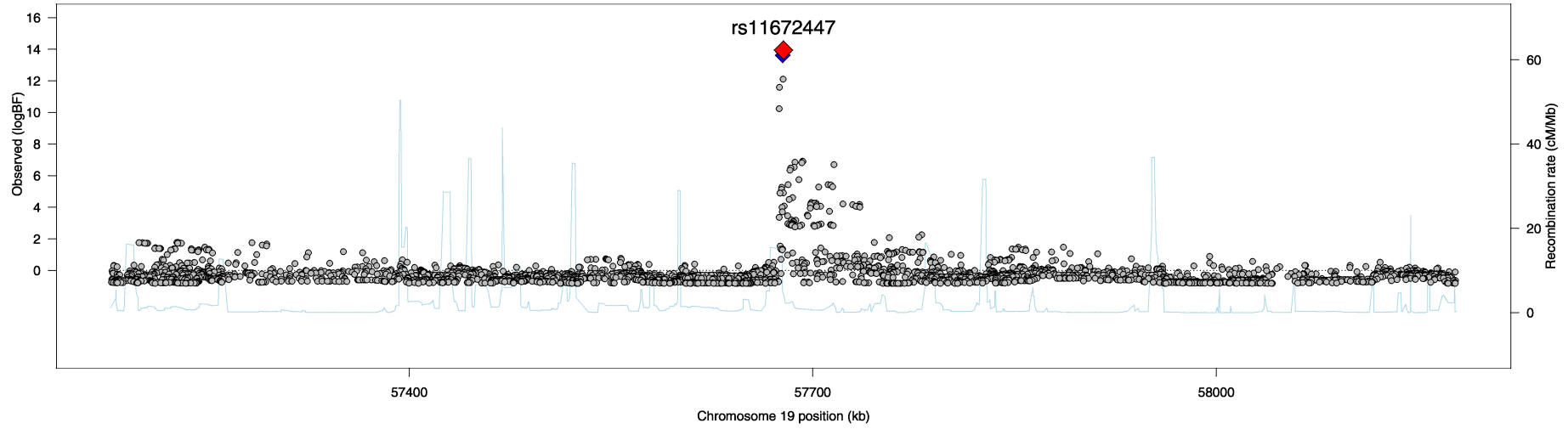




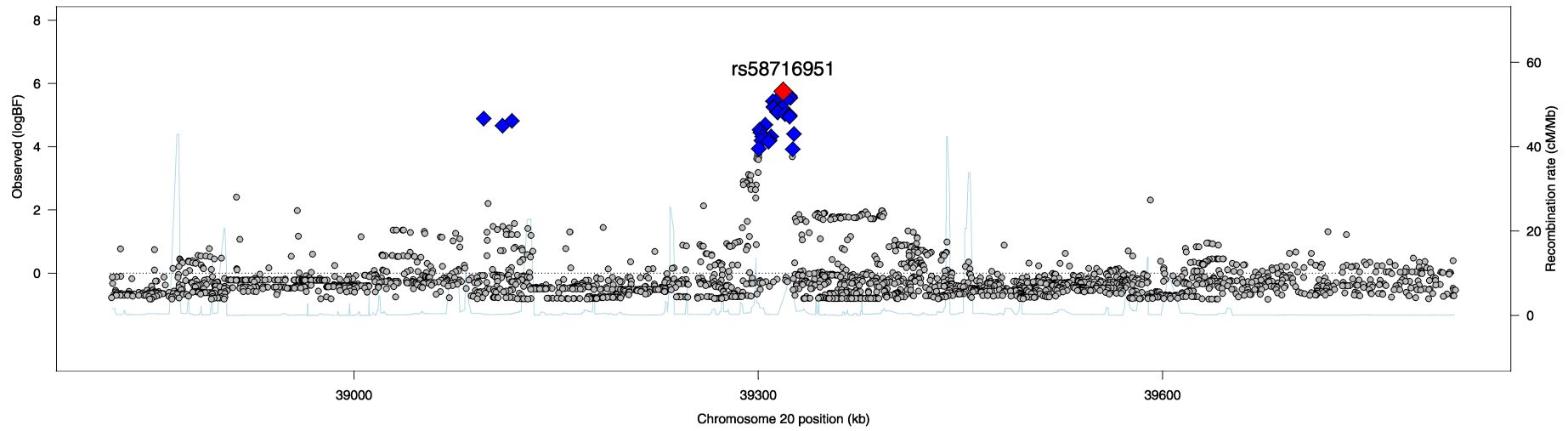
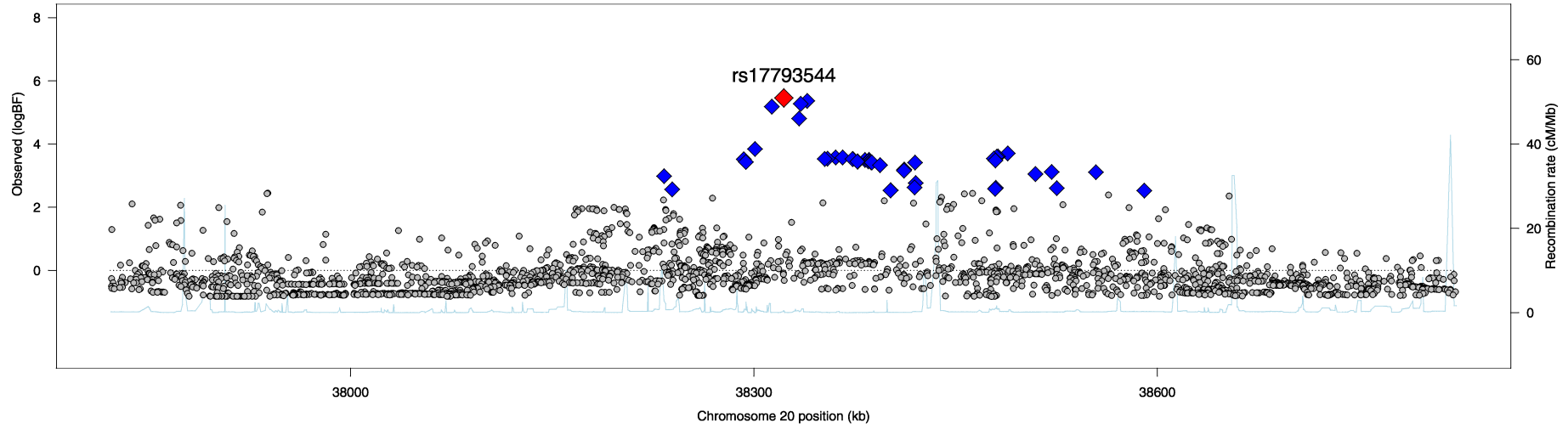












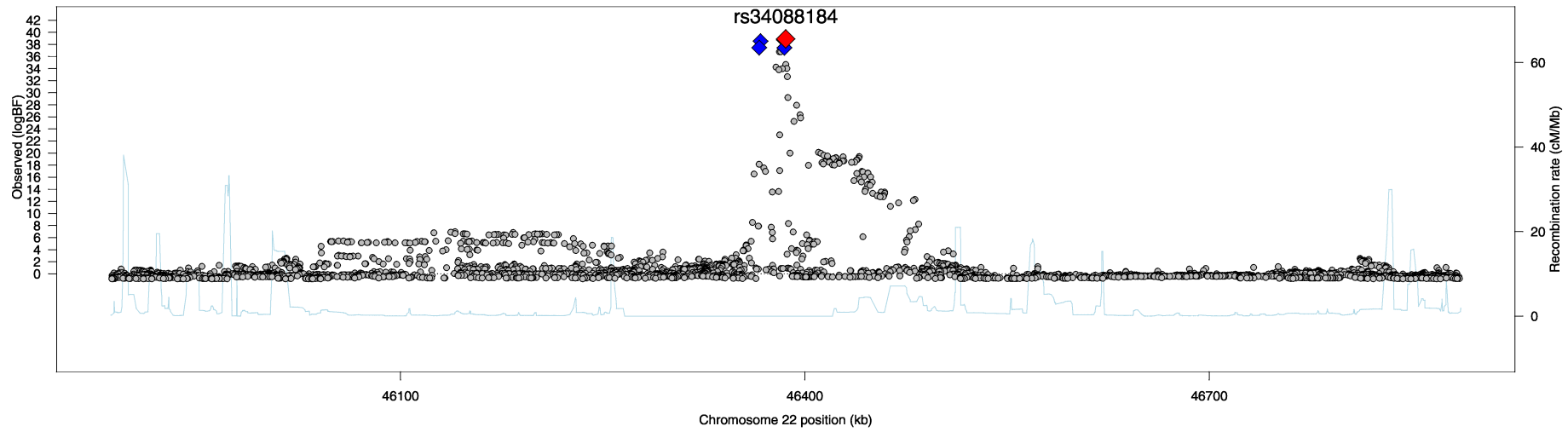


Figure S8. Immunohistochemistry of fixed, surgically resected fibrotic DD tissue confirms WNT3A expression (brown, upper left panel) compared to isotype control (brown, upper right panel), and palm skin (brown, lower left panel). SFRP4 was also not expressed in palm skin (brown, lower right panel). Scale bar = 20 $\mu$ m.

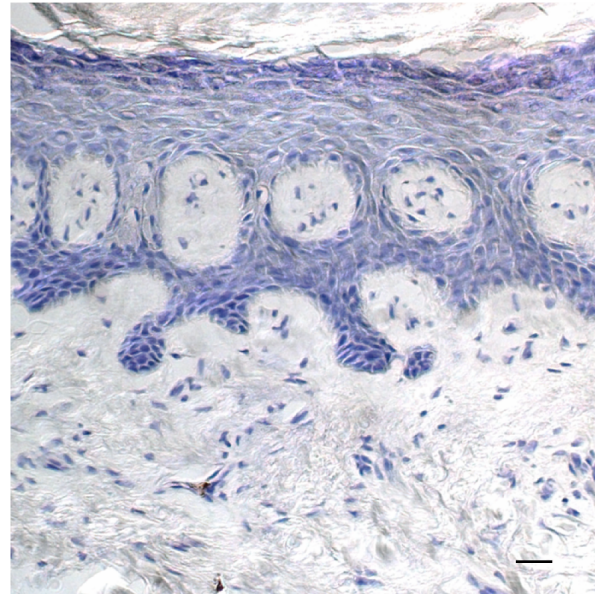
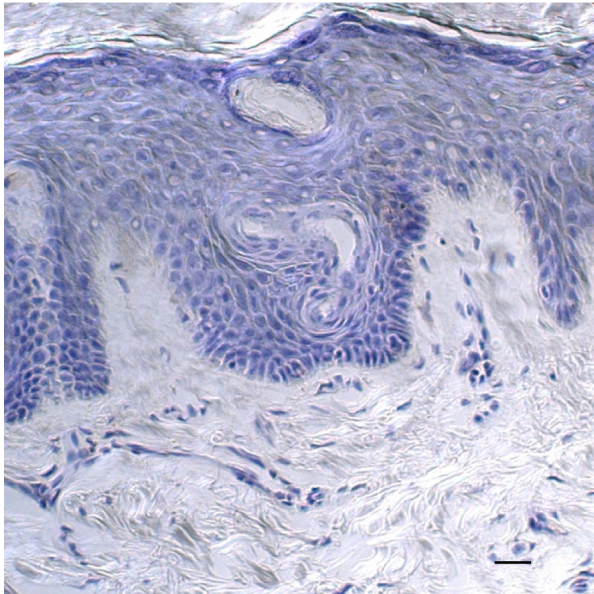
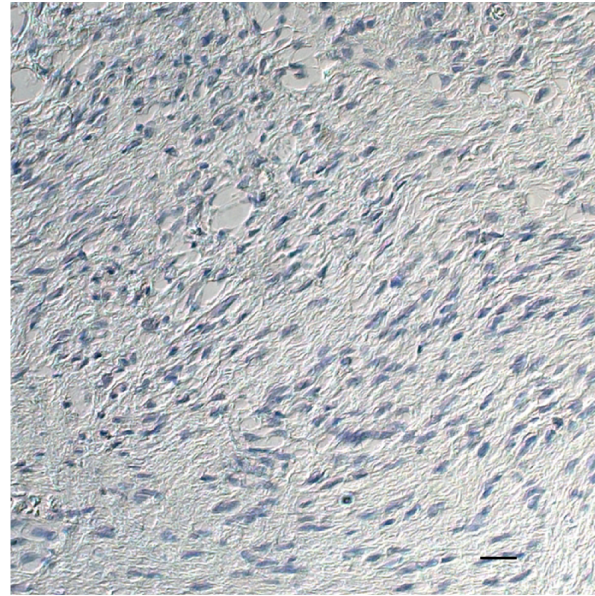
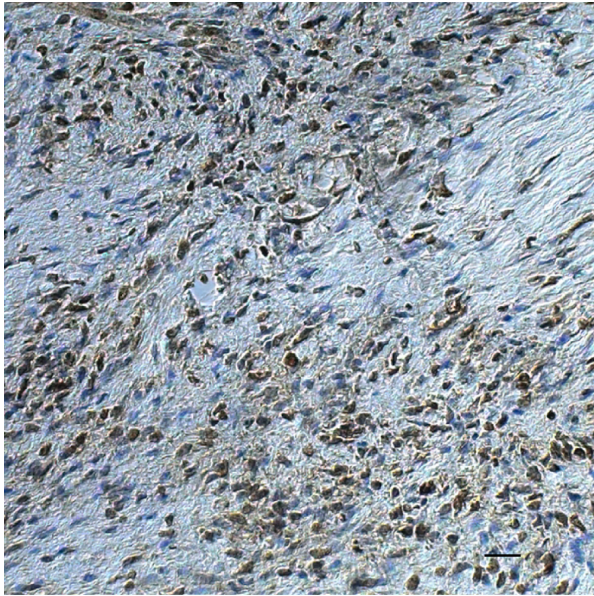
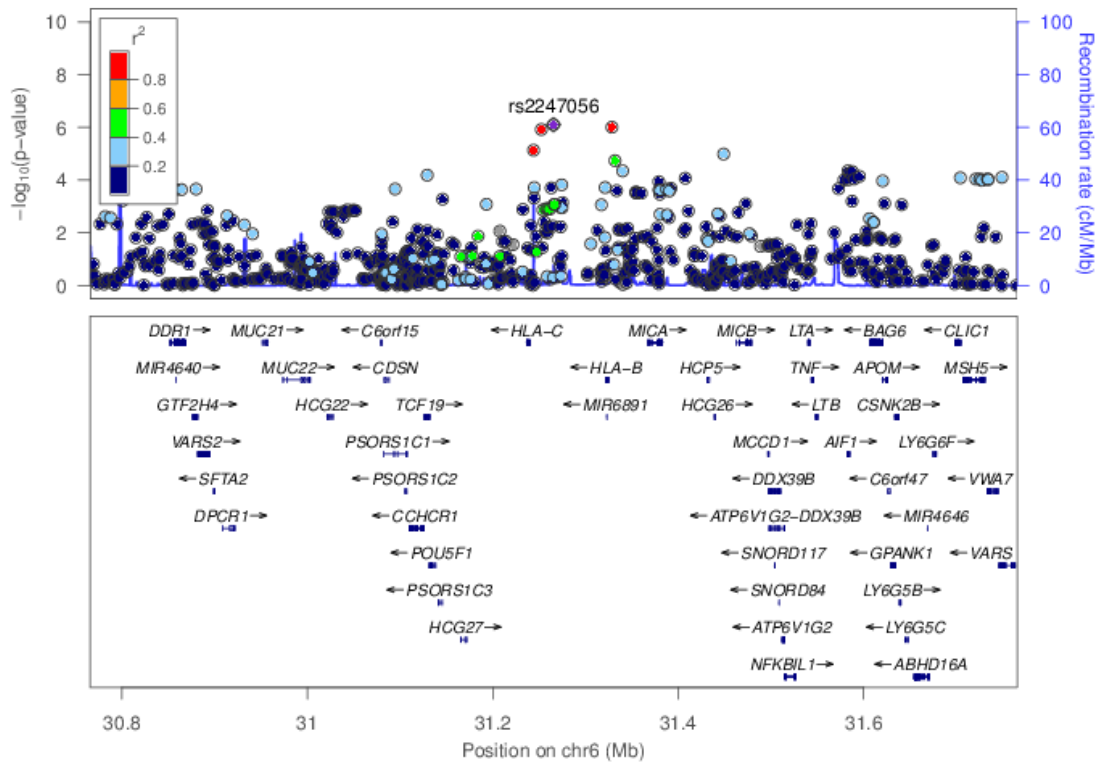


Figure S9. Annotated regional association plots for a) rs12677559 and b) rs2247056 after analysis of the discovery cohort, plotted with LocusZoom<sup>1</sup>. Recombination rate was calculated within the software using data from HapMap. In the lower panel of each figure, genes within 500kb of the index SNP are shown. The position on the chromosomes is shown in relation to Human Genome build hg19.

a)



b)

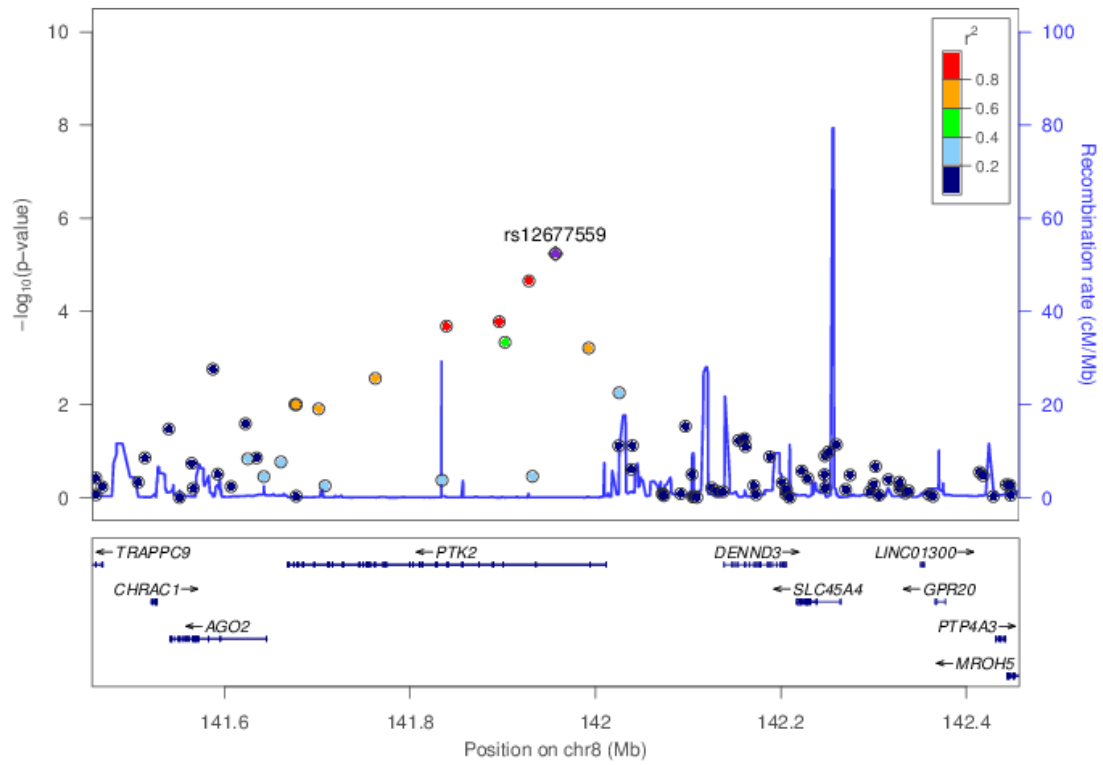


Table S1. Samples and genotyping platforms used in the discovery and replication phases

Phase	Geographical Origin	Genotyping Platform	Pre-QC		Post-QC	
			Cases	Controls	Cases	Controls
Discovery	United Kingdom	HumanCoreExome	4201	4891	3871	4686
		HumanCoreExome				
Replication	Germany	Affymetrix 6.0	768	1353	768	1321
		Affymetrix 6.0				
	Netherlands	Sequenom Custom array	2195	1983	2104	1860
		Sequenom Custom array				
United Kingdom	HumanCoreExome	1207	5070	1169	5070	
	HumanCoreExome					

Table S2. 99% credible sets for associated loci. This table is supplied as an Excel workbook, with a separate worksheet for each individual credible set.

Table S3. Variants carried through to replication phase, if an assay could be designed. Variants that reached genome-wide significance ( $p > 5 \times 10^{-8}$ ) on meta-analysis of discovery and replication cohorts are also reported in Table 1 in the main manuscript. This table is supplied as an Excel workbook.