

Dense genotyping identifies and localizes multiple common and rare variant association signals in celiac disease.

Gosia Trynka^{1,35}, Karen A Hunt^{2,35}, Nicholas A Bockett², Jihane Romanos¹, Vanisha Mistry², Agata Szperl¹, Sjoerd F Bakker³, Maria Teresa Bardella^{4,5}, Leena Bhaw-Rosun⁶, Gemma Castillejo⁷, Emilio G. de la Concha⁸, Rodrigo Coutinho de Almeida¹, Kerith-Rae M Dias⁶, Cleo C. van Diemen¹, Patrick CA Dubois², Richard H. Duerr^{9,10}, Sarah Edkins¹¹, Lude Franke¹, Karin Fransen^{1,12}, Javier Gutierrez¹, Graham AR Heap², Barbara Hrdlickova¹, Sarah Hunt¹¹, Leticia Plaza Izurieta¹³, Valentina Izzo¹⁴, Leo AB Joosten^{15,16}, Cordelia Langford¹¹, Maria Cristina Mazzilli¹⁷, Charles A Mein⁶, Vandana Midah¹⁸, Mitja Mitrovic^{1,19}, Barbara Mora¹⁷, Marinita Morelli¹⁴, Sarah Nutland²⁰, Concepción Núñez⁸, Suna Onengut-Gumuscu²¹, Kerra Pearce²², Mathieu Platteel¹, Isabel Polanco²³, Simon Potter¹¹, Carmen Ribes-Koninckx²⁴, Isis Ricaño-Ponce¹, Stephen S. Rich²¹, Anna Rybak²⁵, José Luis Santiago⁸, Sabyasachi Senapati²⁶, Ajit Sood¹⁸, Hania Szajewska²⁷, Riccardo Troncone²⁸, Jezabel Varadé⁸, Chris Wallace²⁰, Victorien M Wolters²⁹, Alexandra Zhernakova³⁰, CEGEC (Spanish Consortium on the Genetics of Coeliac Disease), PreventCD Study Group, Wellcome Trust Case Control Consortium, B.K. Thelma²⁶, Bozena Cukrowska³¹, Elena Urcelay⁸, Jose Ramon Bilbao¹³, M Luisa Mearin³², Donatella Barisani³³, Jeffrey C Barrett¹¹, Vincent Plagnol³⁴, Panos Deloukas¹¹, Cisca Wijmenga^{1,36}, David A van Heel^{2,36}

Supplementary Note

a) Consortium and Group Memberships

CEGEC (Spanish Consortium for Genetics of Coeliac Disease - Spanish Celiac Disease Society (SEEC))

Montserrat Alsina¹, Maria Esteve Comas¹, Laura Pardo Grau¹, Luis Castaño², Pablo Oliver², Juan Carlos Vitoria², Luis Rodrigo³, Carlos López-Larrea³, Antonio López-Vázquez³, Idoia Hualde⁴, Zuriñe García⁴, Blanca Hernández Charro⁵, Maria Antonia Ramos-Arroyo⁵, Félix Sánchez-Valverde⁵, Carme Farré⁶, Teresa Marqués⁶, Pere Vilar⁶, Eduardo Arranz⁷, José Antonio Garrote⁷

¹CATLAB- Hospital Universitari Mutua Terrassa, Barcelona, Catalonia, Spain

²Hospital Universitario de Cruces, Barakaldo, Basque Country, Spain

³Hospital Universitario Central Asturias, Oviedo, Spain

⁴Hospital de Txagorritxu, Vitoria-Gasteiz, Basque Country, Spain

⁵Hospital Virgen del Camino, Pamplona-Iruña, Spain

⁶Hospital Sant Joan de Déu, Barcelona, Catalonia, Spain

⁷IBGM-Universidad de Valladolid, Valladolid, Spain

PreventCD Study Group

Renata Auricchio¹, Caroline E Hogen Esch², Ania Chmielewska³, Gosia Piescik³, Eva Martinez⁴, Vicente Varea⁵, Francesc Martinez Cerezo⁶, Ester Donat-Aliaga⁷, Joaquin Escribano⁶

Department of Paediatrics and European Laboratory for the Investigation of Food-Induced Diseases, University “Federico II”, Naples, Italy.

²Department of Paediatric Gastroenterology, Leiden University Medical Center, Leiden, the Netherlands.

³Department of Paediatrics, The Medical University of Warsaw, Warsaw, Poland

⁴Pediatric Gastroenterology Department, Hospital La Paz, Madrid, Spain.

⁵Hospital Sant Joan de Déu, Barcelona, Spain.

⁶Hospital Universitari Sant Joan de Reus, Spain.

⁷La Fe University Hospital, Pediatric Gastroenterology, Bulevar Sus 46026 Valencia, Spain.

WTCCC (Wellcome Trust Case Control Consortium)

Please see the website www.wtccc.org.uk for further details.

b) Ethics Committee / Institutional Review Board Approvals

UK: Oxford REC B, East London and The City REC 1, SouthEast Multi-Centre Research Ethics Committee.

The Netherlands: Medical Ethical Committee of the University Medical Center Utrecht, Medical Ethical Committee of the Free University Medical Center Amsterdam.

Poland: Bioethical Committee of the Children's Memorial Health Institute, Warsaw.

Spain - CEGEC: Clinical Trials and Ethics Committee of Hospital de Cruces

Spain - Madrid: CEIC- Comité ético de investigación clínica del Hospital Clínico San Carlos.

Italy - Rome, Ethics Committee of Policlinico Umberto I, “La Sapienza” University of Rome

Italy - Naples: Ethical Committee of the Faculty of Medicine of the University of Naples Federico II.

Italy - Milan: Ethical Committee Fondazione IRCCS Ca' Granda – Ospedale Maggiore Policlinico.

India - Punjab: Drug trial ethics committee, Dayanand Medical College and Hospital, Ludhiana, Punjab.

c) Summary of additional case/control resequencing content on ImmunoChip

We provide a brief summary of regional resequencing data on ImmunoChip. This is content that was submitted in addition to the 1000Genomes Pilot CEU based fine mapping data for the 183 non-HLA autoimmune loci. All coordinates are hg18/NCBI build 36.

Summary of celiac disease case resequencing.

80 random celiac case DNA samples were selected for 10 blind DNA pools of 8 samples each. A total of 1.26Mb of genomic sequence was selected for a capture array. Genomic regions were selected from previous celiac disease GWAS:

<i>IL2/IL21</i>	chr4:123080000-123820000
<i>IL12A</i>	chr3:161090000-161300000
<i>RGS1</i>	chr1:190530000-190850000

Repetitive sequence prevented capture of 245kb (20%) of sequence, which was not tiled on the capture array. We noted that nearly all known high-quality SNPs (96%, 1512 of 1575 CEU HapMap SNPs) fell within the 1015kb non-repetitive regions (“the PCR accessible genome”).

Each pool was hybridised to a Nimblegen 385k capture array, using 60-100bp probes. Sequencing was performed using 45bp paired end reads on an Illumina GAI sequencer. Alignment was performed using Novoalign, allowing gapped reads. SNP calling was performed in the pools at bases with read-depth>150X (~20X per sample in the pool) to maintain accuracy. On average, 84% of tiled bases had sufficient read-depth in a pool (and 85% of HapMap SNP positions). We calculated approximately 5% false positive and false negative variant discovery rates using Hap300 genotype data (all samples were array genotyped) and Hap300 SNPs mapping to the capture regions. Validation of 43 of 46 randomly selected SNPs by Sanger sequencing estimated a 6.5% false positive rate.

3457 SNPs and 616 small indels were called and all were submitted to both dbSNP (dbSNP Submitter Handle VANHEEL_LAB_LONDON) and for ImmunoChip design.

Summary of WTCCC resequencing.

40 cases (for each disease) and 40 controls were sequenced per genomic region in blind pools of 8 samples. The samples sequenced differed between regions and were chosen to ensure representation of both the known disease risk allele(s) (in case samples) and protective allele(s) (in control samples) as well as the known haplotypic diversity in each region (derived from HapMap2). Genomic regions were selected from the relevant disease GWAS studies.

PCR primers were designed in a triple tile across the targeted region and used to amplify long-range PCR products of between 5 and 10 kb. PCR products were pooled, and sequenced using 37bp paired end reads on an Illumina GAI or GAI sequencer. Alignment was performed using MAQ. Potential SNPs are filtered according to read depth, base quality, the nearness of neighbouring SNPs and het ratio where appropriate.

All variants were submitted to both dbSNP (dbSNP Submitter Handle WTCCC) and for ImmunoChip design.

Type 1 diabetes

<i>IFIH1</i>	chr2:162704739-163072739
<i>CTLA4/ICOS</i>	chr2:204386046-204533060
<i>IL2/IL21</i>	chr4:123155845-123804957
<i>IL2RA</i>	chr10:6068112-6311849
<i>ERBB3</i>	chr12:54632267-55043535
<i>CLEC16A</i>	chr16:10899000-11217000
<i>CD226</i>	chr18:65659000-65725000

Crohn's disease

<i>IL23R</i>	chr1:67426567-67540567
<i>CD244</i>	chr1:159073551-159178551
-	chr5:40323000-40660000
<i>IRGM</i>	chr5:150151000-150318000
<i>IL12B</i>	chr5:158688000-158758500
<i>CDKALI</i>	chr6:20630000-20842000
-	chr6:106514000-106616000
<i>CCR6</i>	chr6:167264579-167466579
-	chr7:50223285-50304285
-	chr8:126594000-126615500
<i>TNFSF8/TNFSF15</i>	chr9:116,572,267-116,740,267
<i>ZNF365</i>	chr10:64049000-64120000
<i>NKX2-3</i>	chr10:101257000-101319000
<i>STAT3</i>	chr17:37733000-37951000
<i>PTPN2</i>	chr18:12728000 -12878000
<i>GPX4, SBNO2</i>	chr19:1049000-1145000
-	chr21:15720000-15764500
<i>ICOSLG</i>	chr21:44428000-44478000

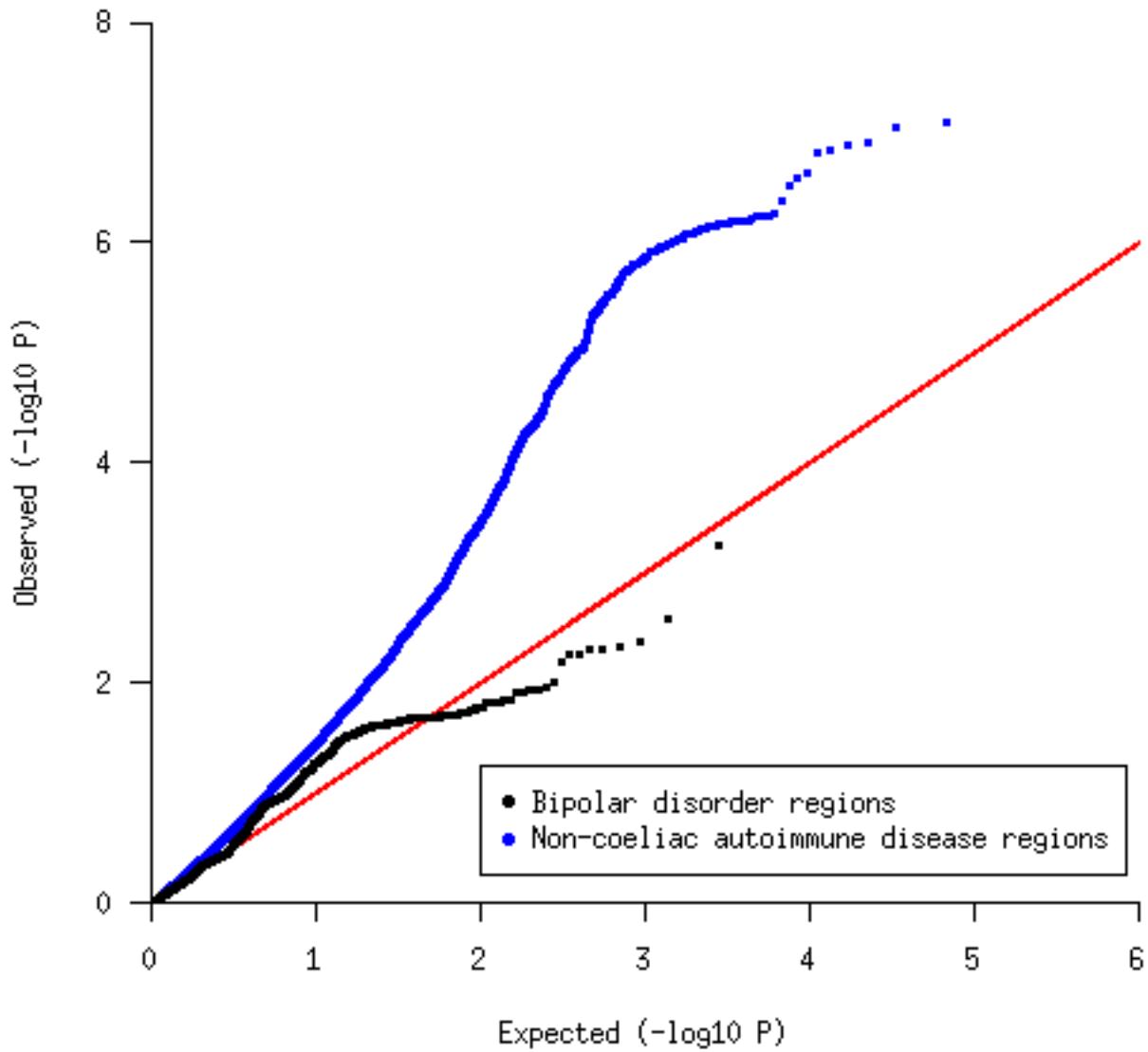
Ankylosing spondylitis

-	chr2:62334538-62462791
<i>IL1R1 IL1R2</i>	chr2:99909427-102280256
<i>CAST, ERAP1</i>	chr5:96053960-96225604
<i>TNFRSF1A</i>	chr12:6284227-6404385
-	chr21:39308249-39406803

Rheumatoid arthritis

<i>CD74</i>	chr5:149663031-149779532
<i>TNFAIP3</i>	chr6:138002656-138290619
<i>PRKCQ</i>	chr10:6427003-6673677
multiple genes	chr12:56140929-56635212
<i>CD40</i>	chr20:44112612-44218338
<i>IL2RB</i>	chr22:35871557-35893351

Supplementary Figure 1 QQ plot of association results at non-coeliac auto-immune, and bipolar disorder, regions.
Observed versus expected celiac disease association test results for immunoChip 1000Genomes dense-genotyped regions. Non-coeliac other immune-mediated disease loci (144 regions excluding HLA 20-40Mb, 67336 variants). Bipolar disorder (3 regions, 2769 variants) shown as a negative control (association statistic inflation factor, lambda, 1.018).

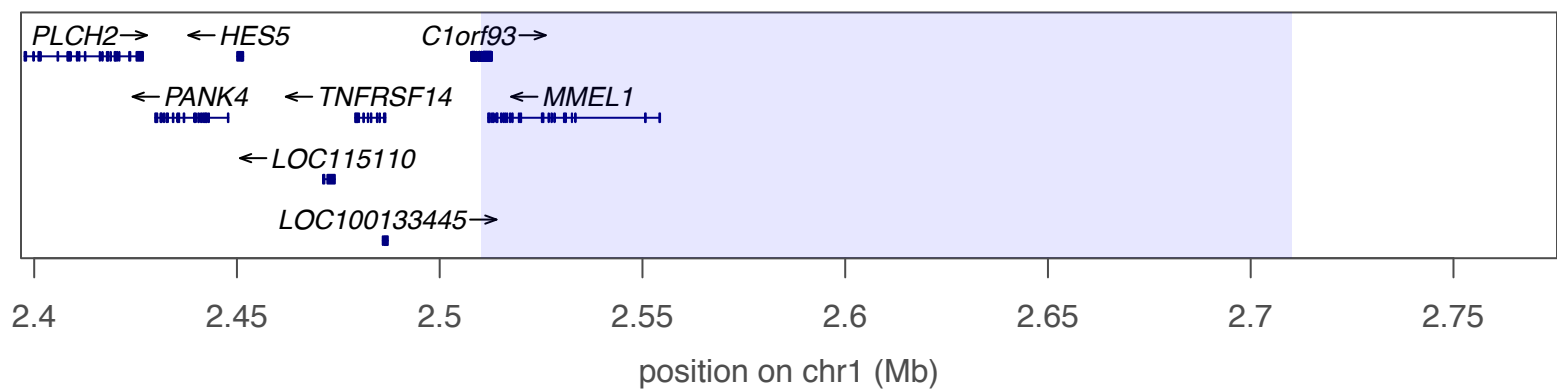
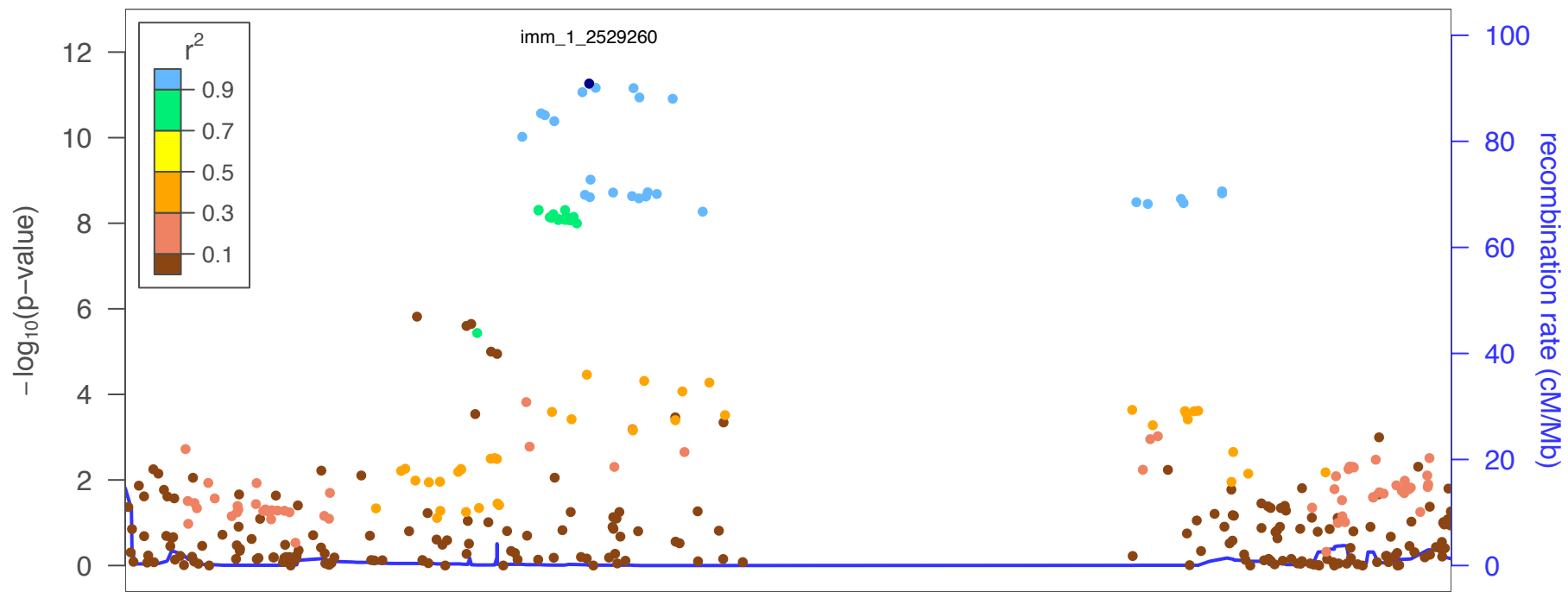


Supplementary Figure 2

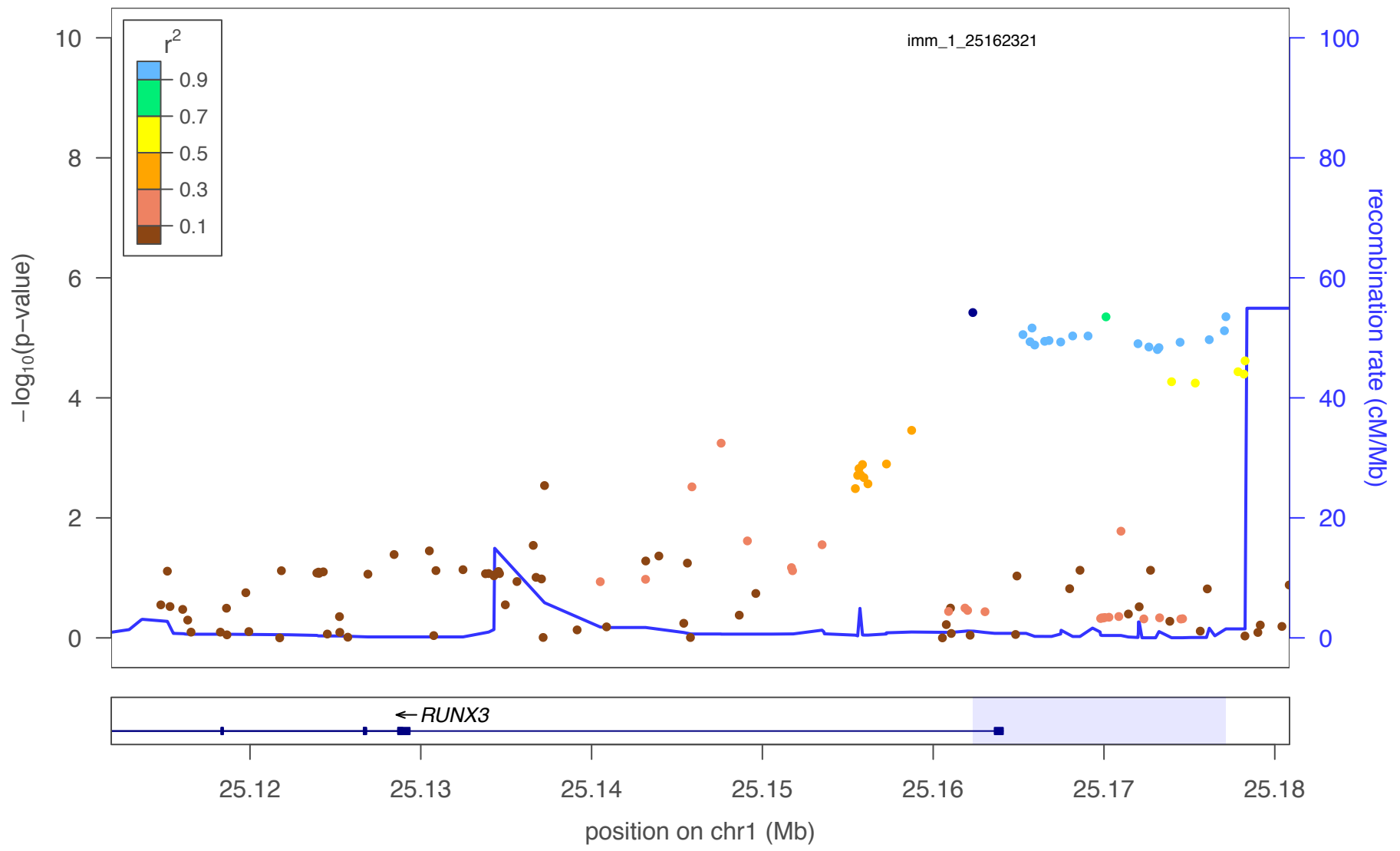
Detailed association and linkage disequilibrium plots for each of 57 independent non-HLA celiac disease risk signals.

ImmunoChip

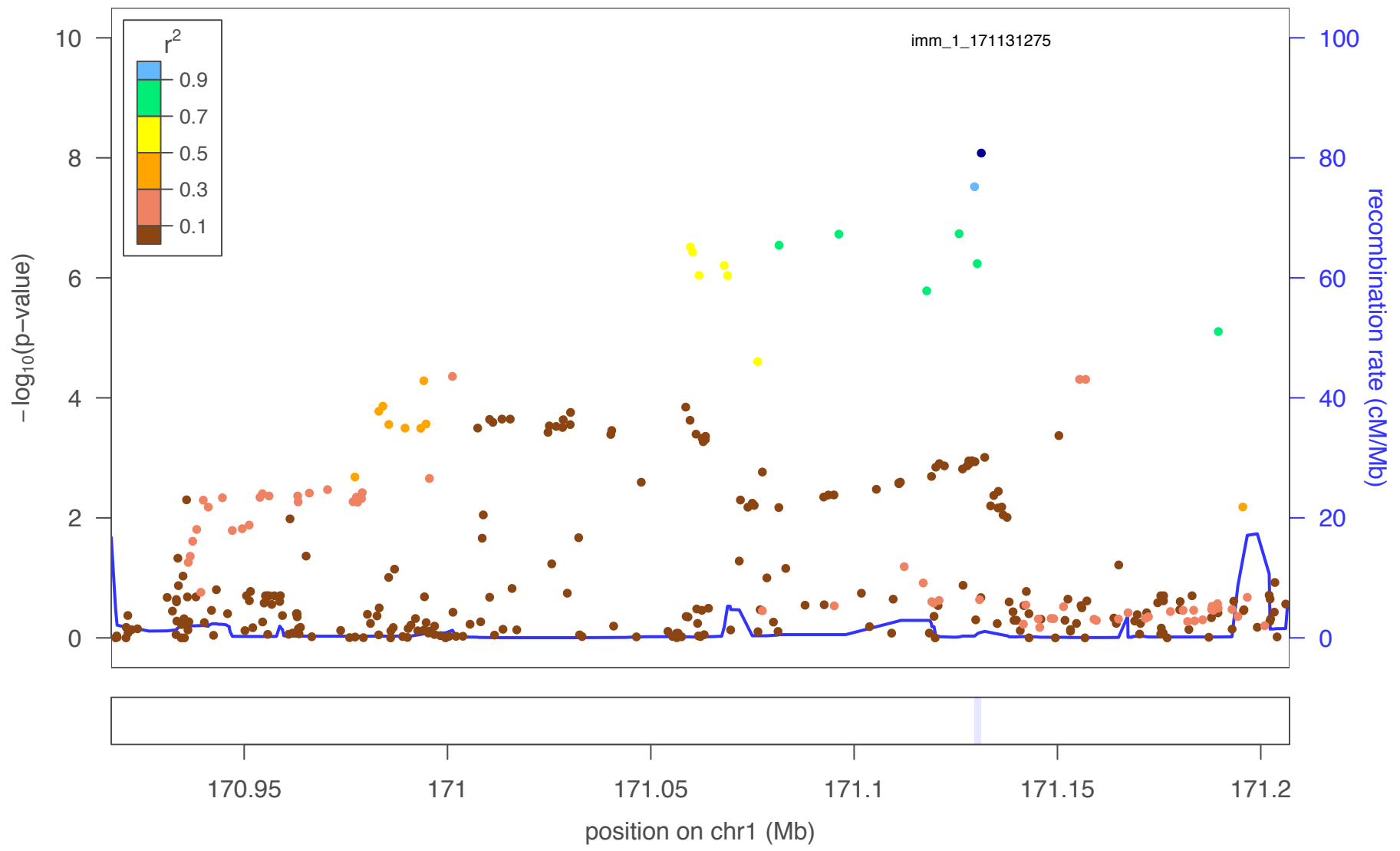
ImmunoChip



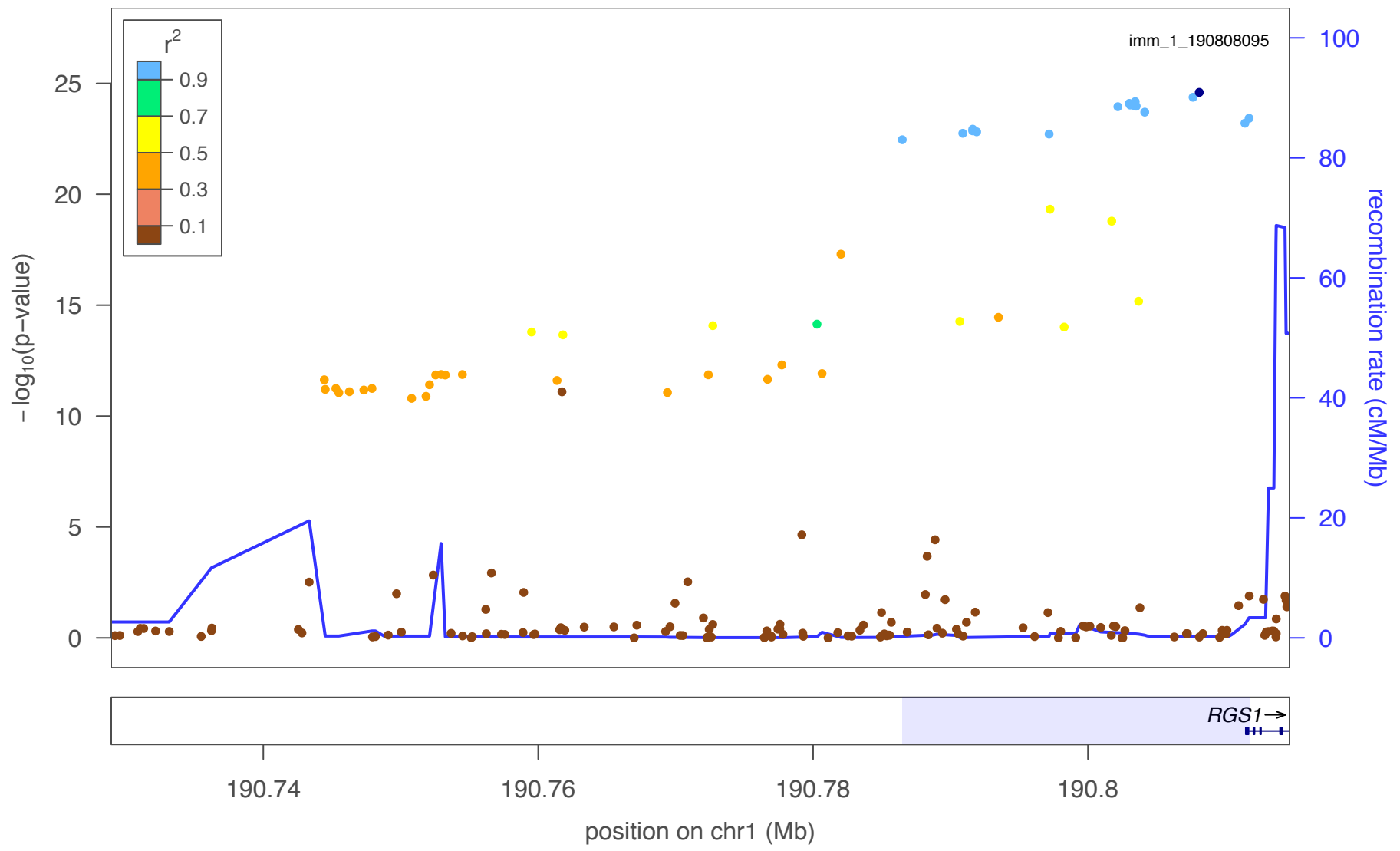
ImmunoChip



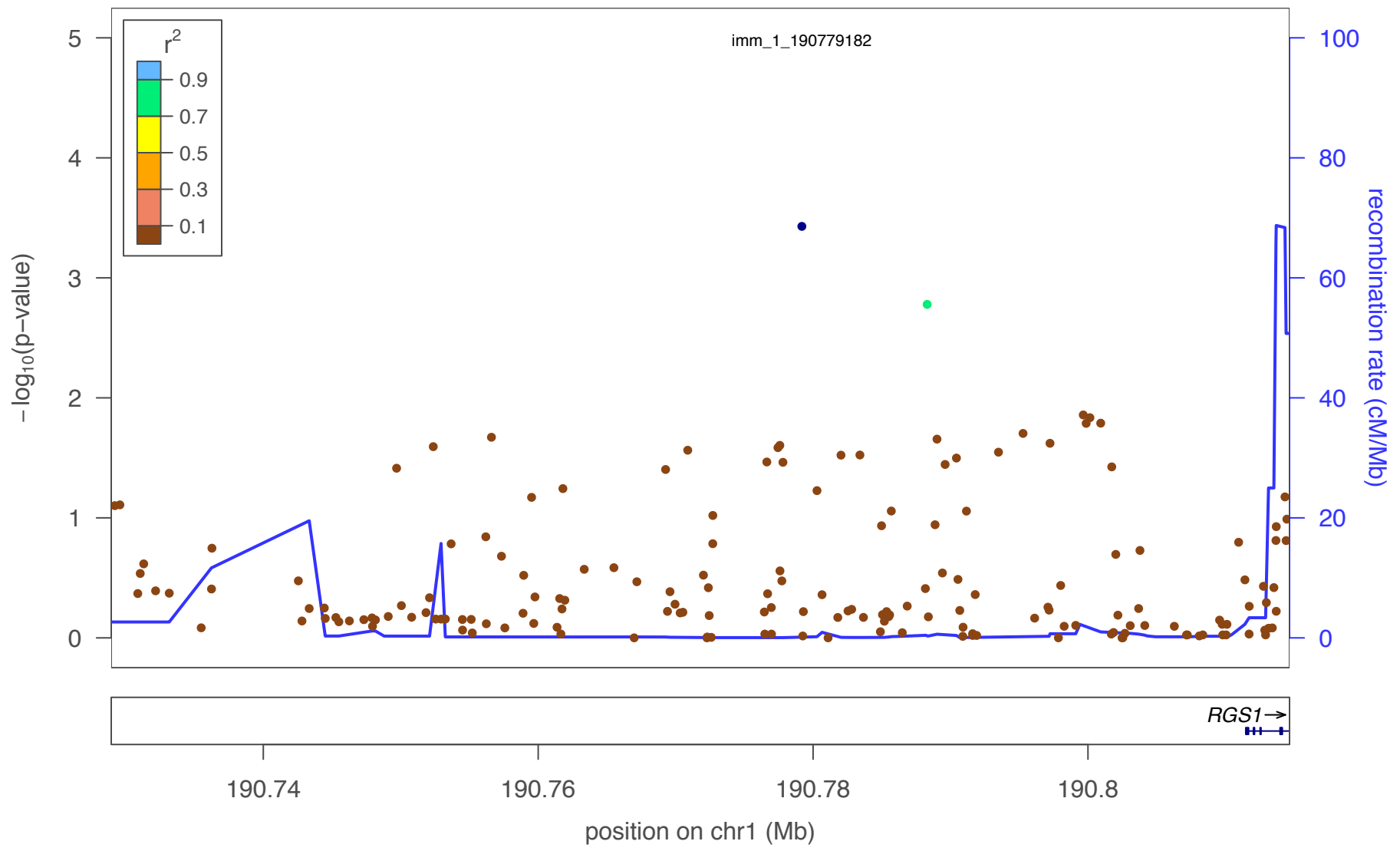
ImmunoChip



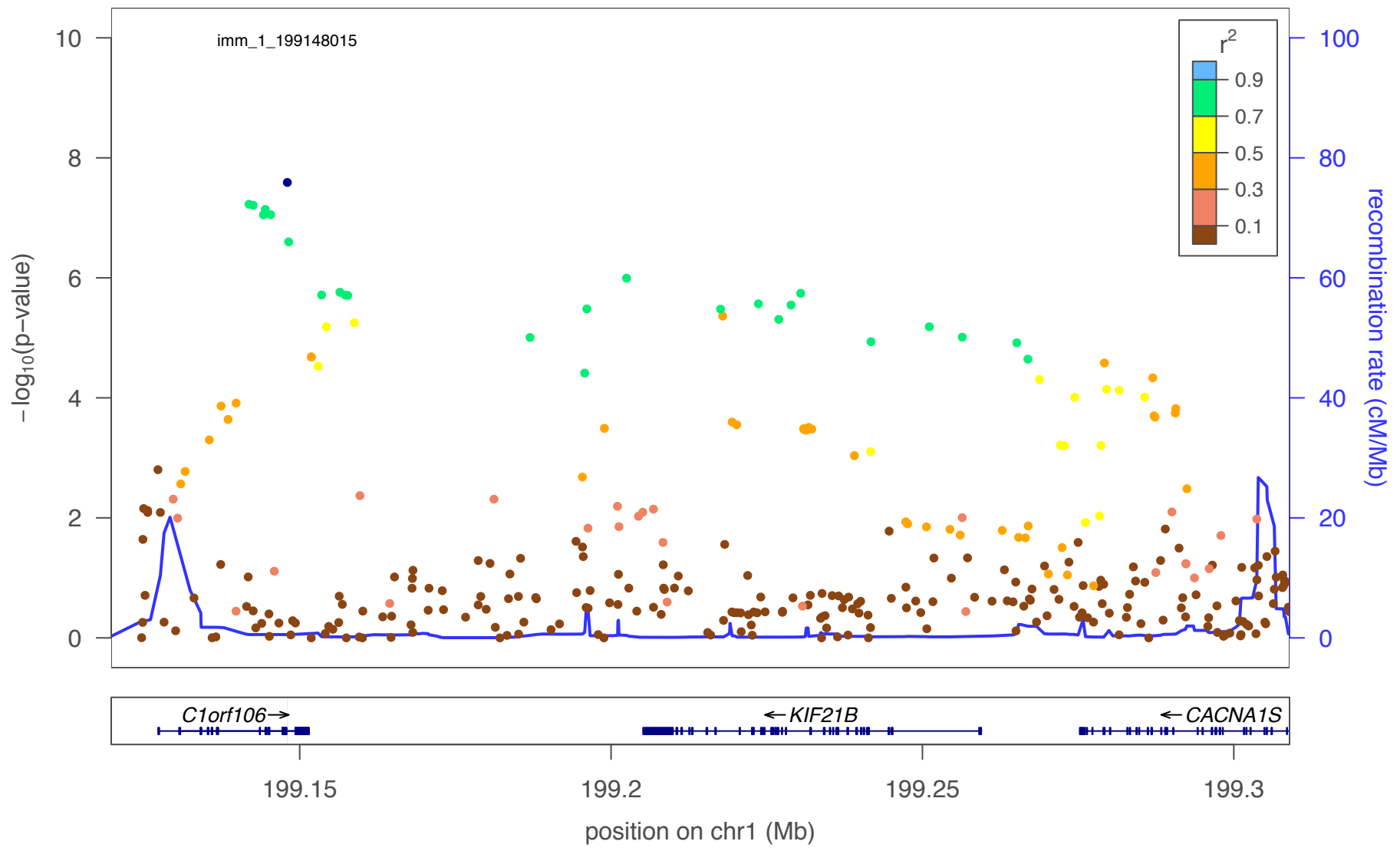
ImmunoChip || ||| || |



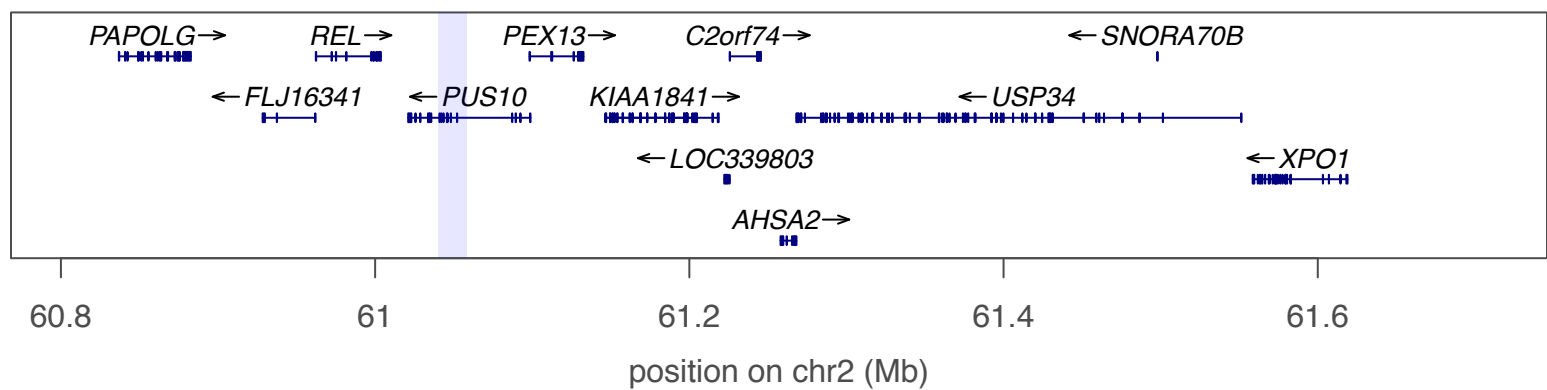
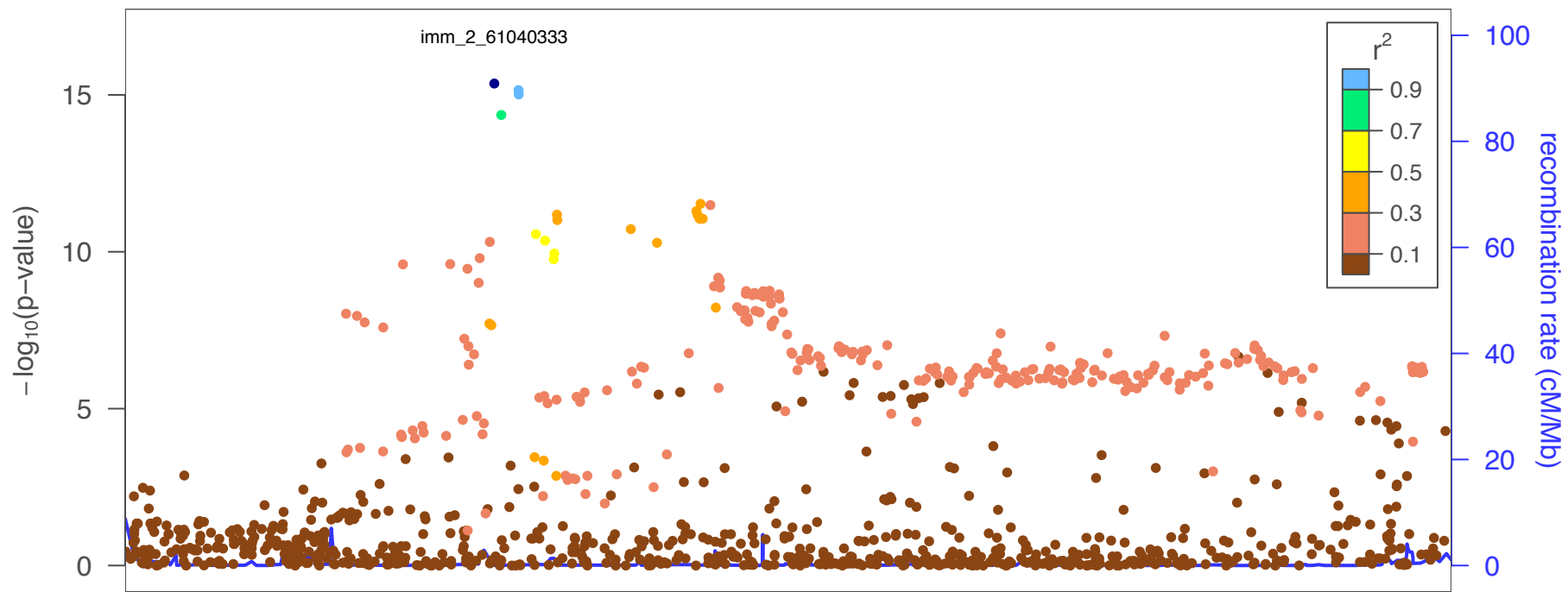
ImmunoChip || ||| || |



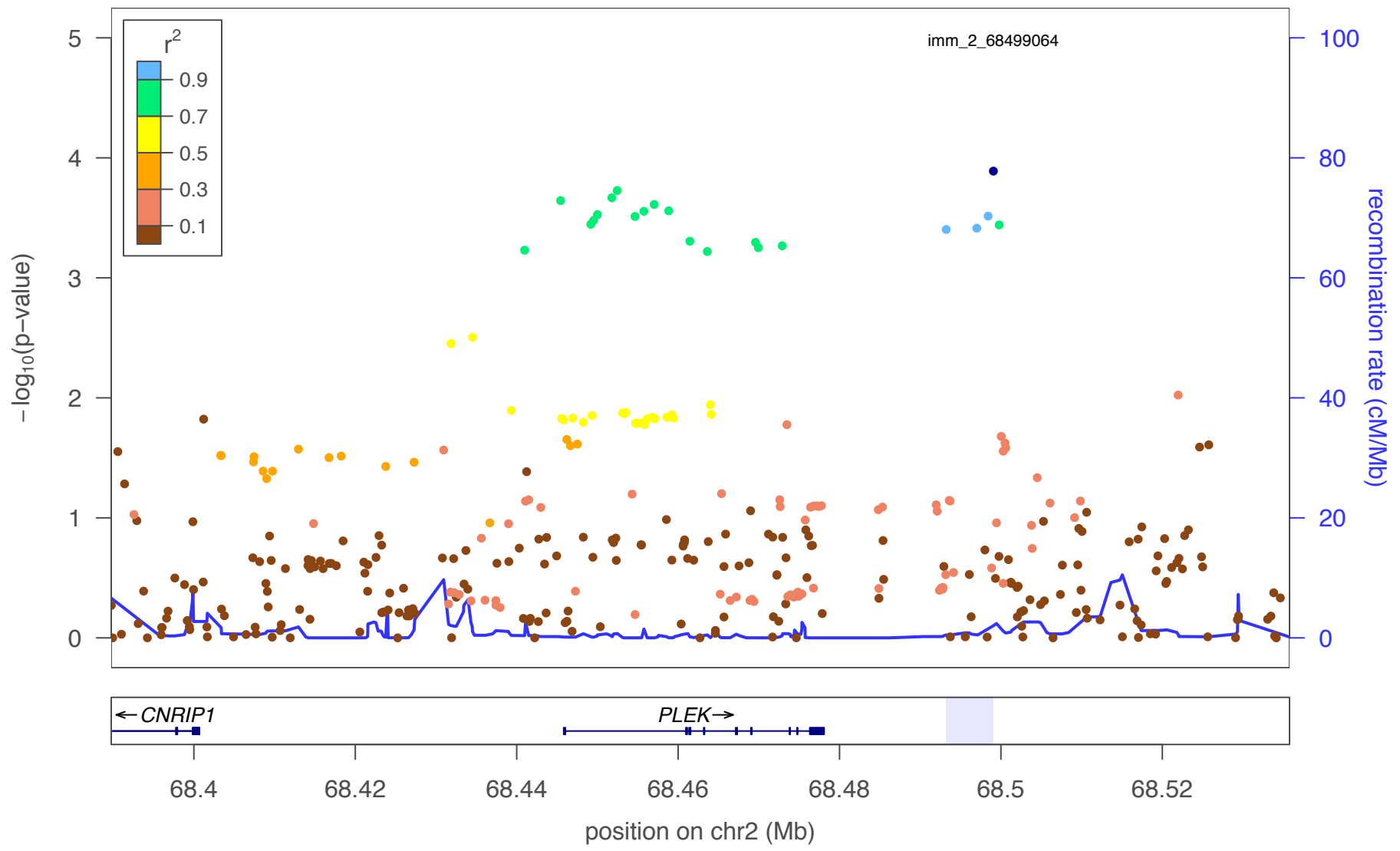
ImmunoChip 



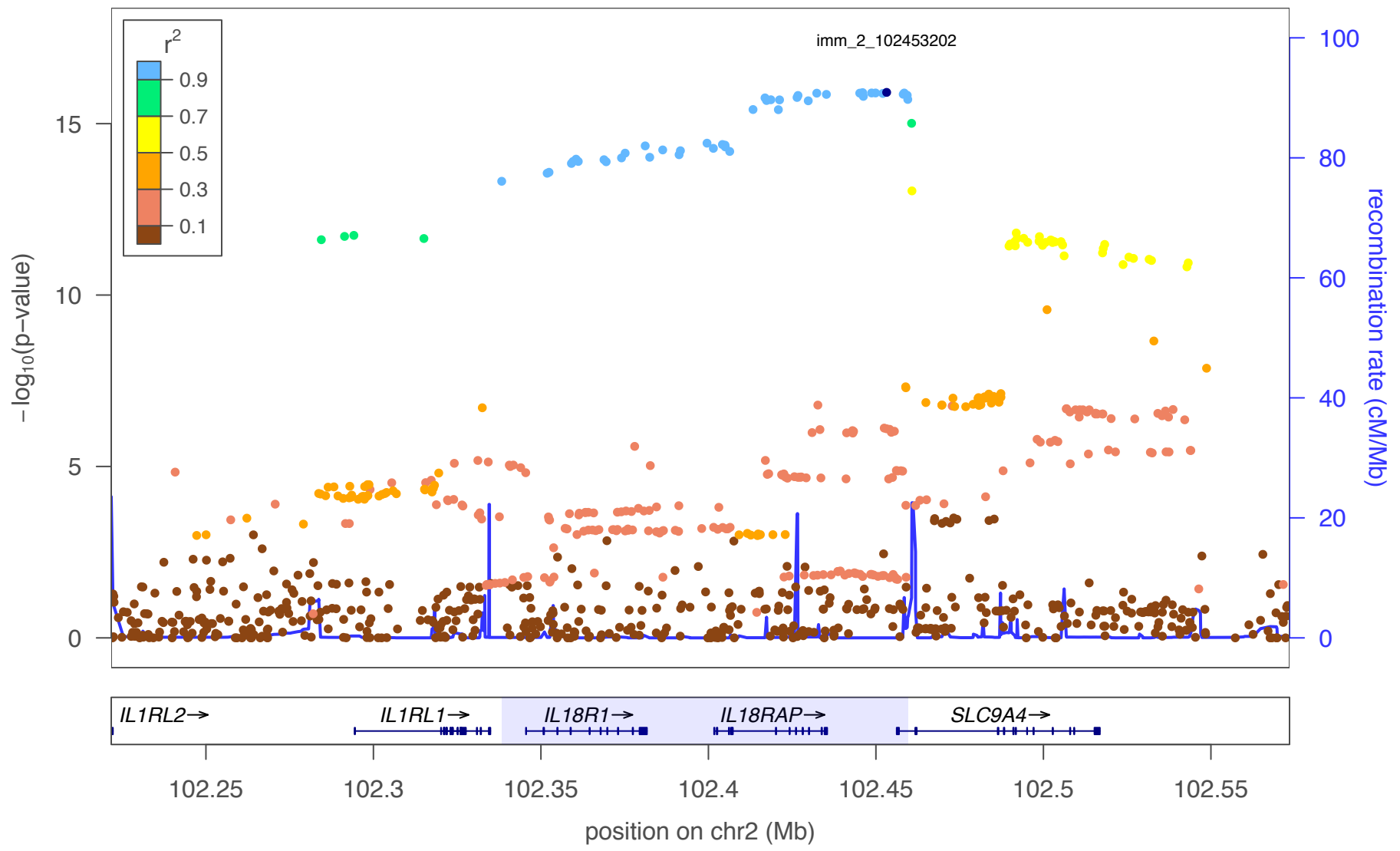
ImmunoChip



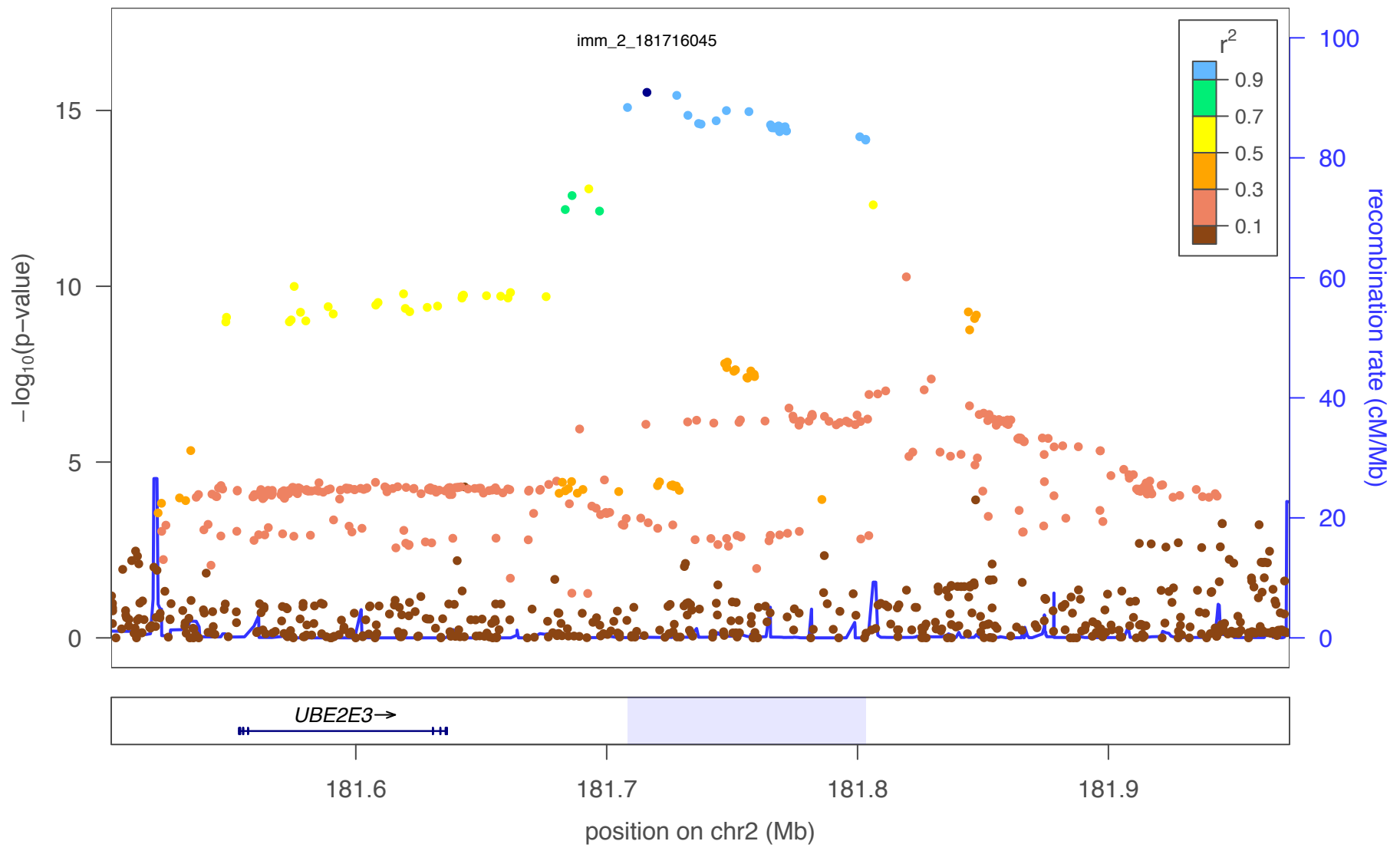
ImmunoChip 



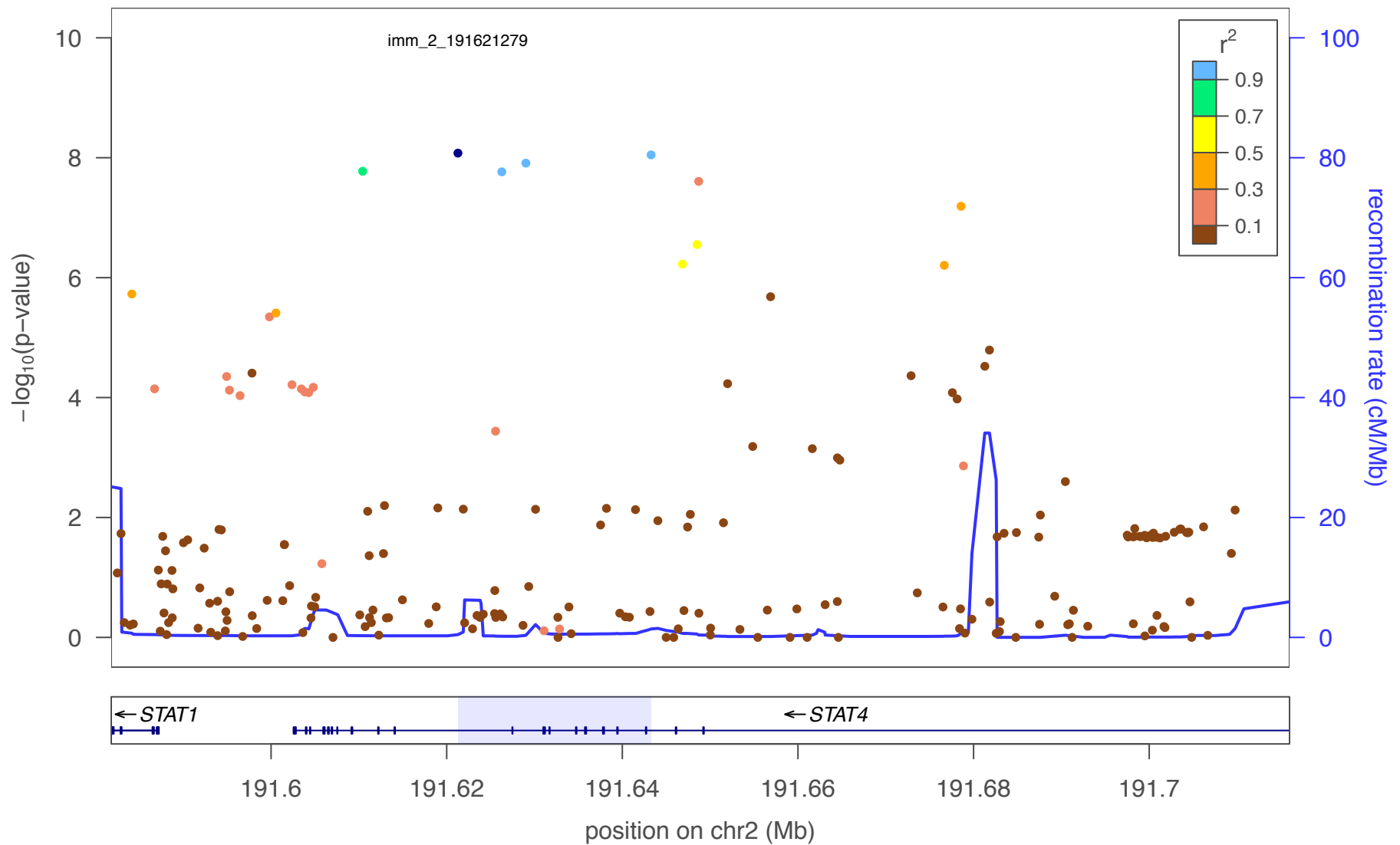
ImmunoChip



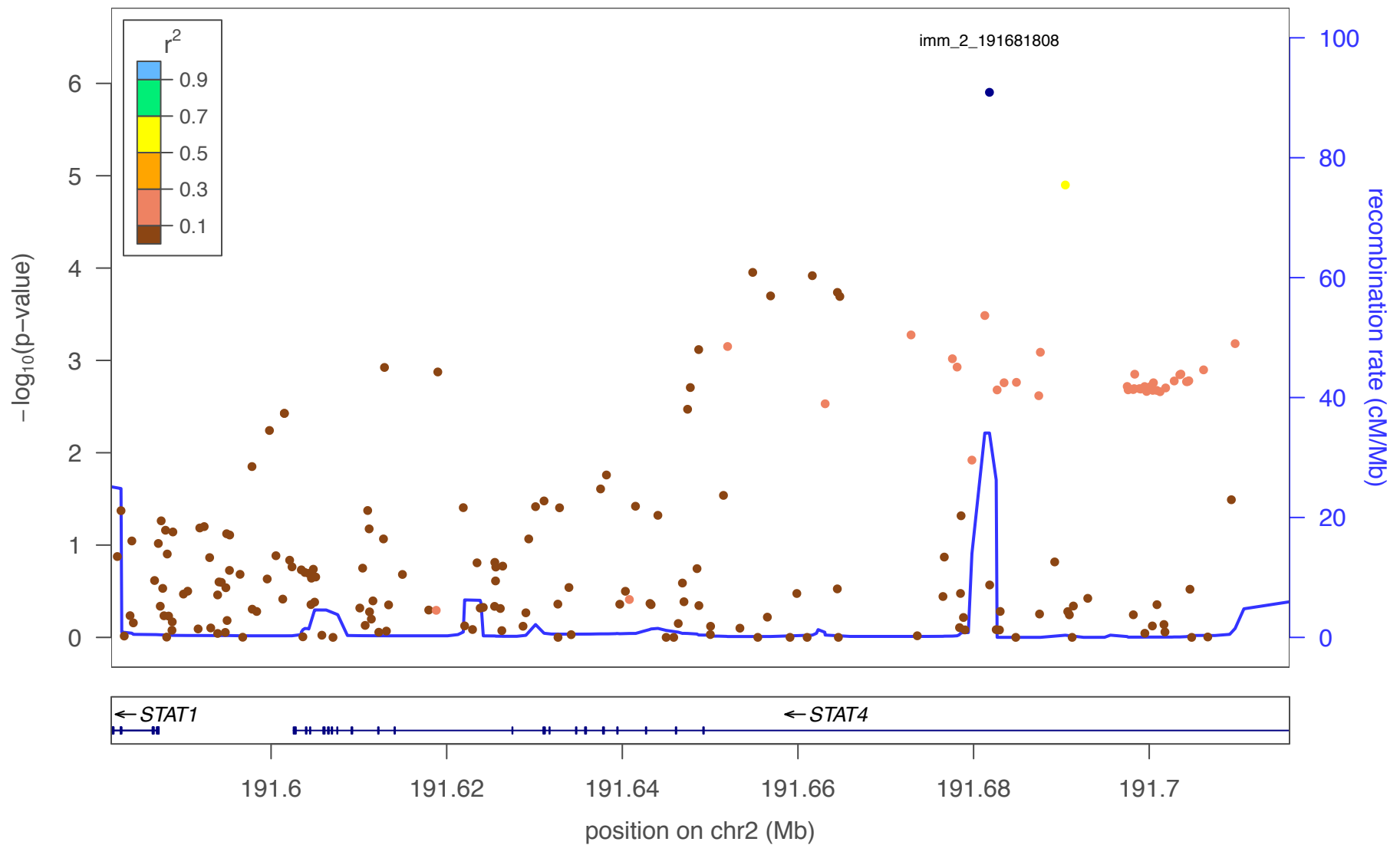
ImmunoChip



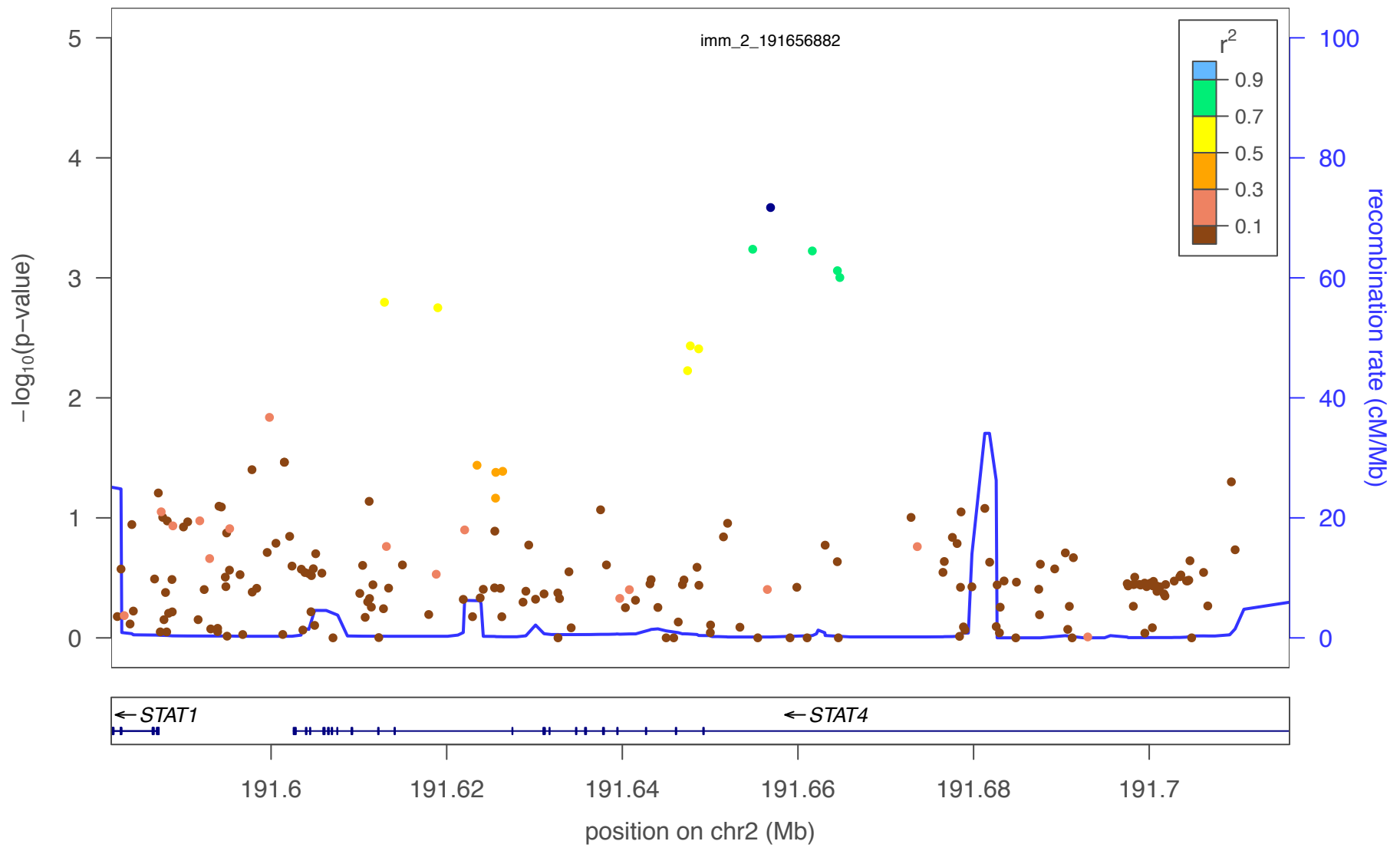
ImmunoChip 



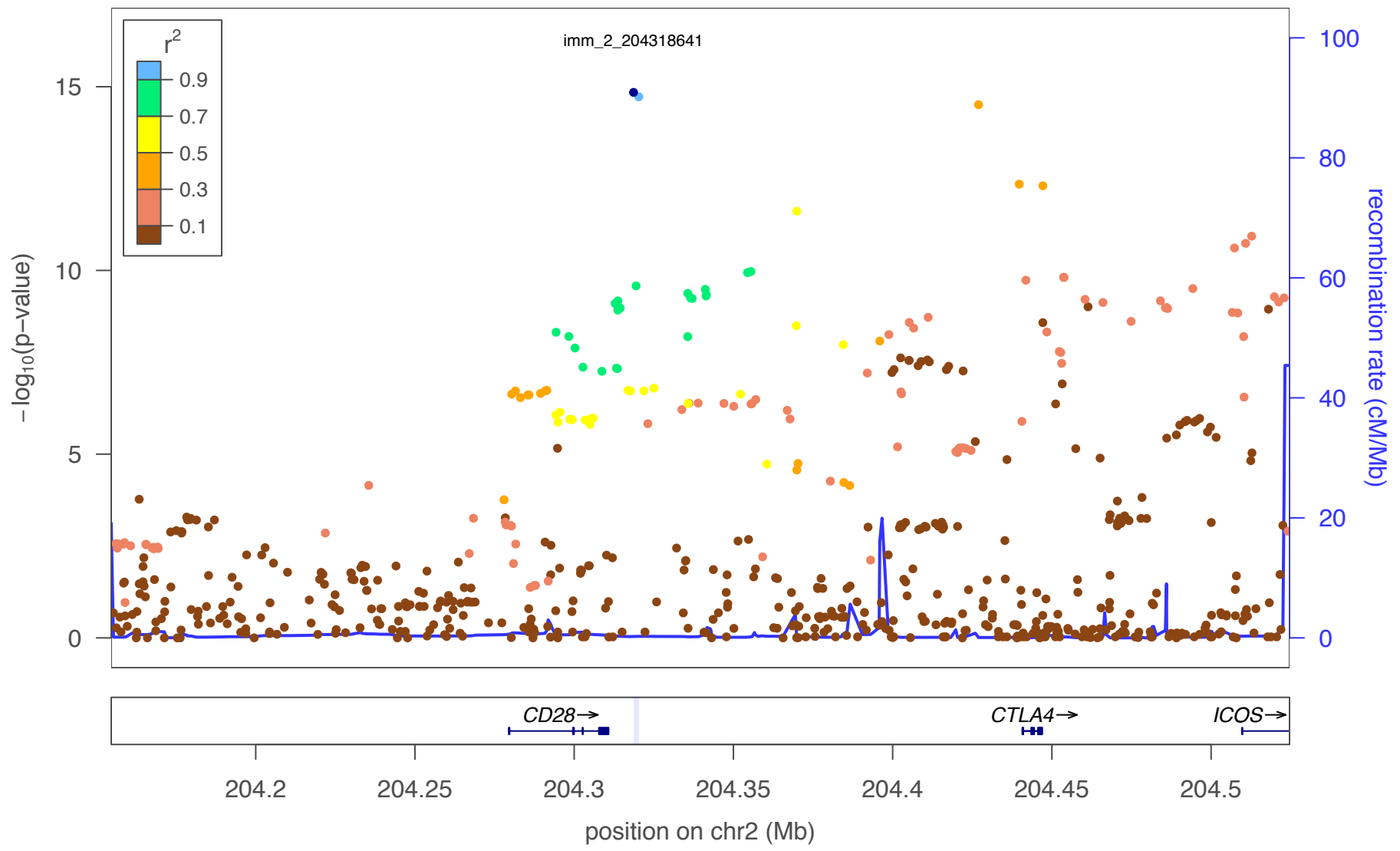
ImmunoChip 



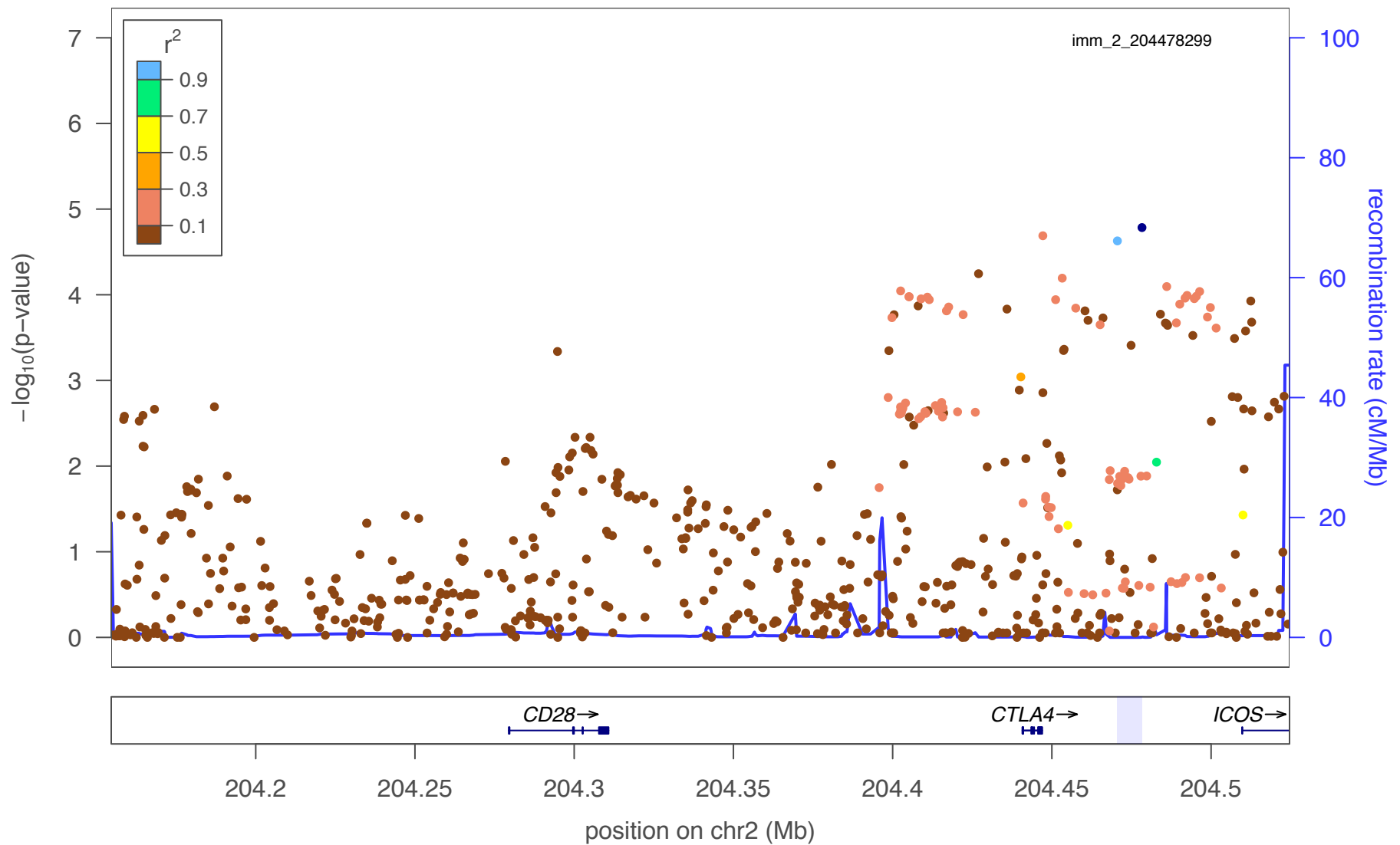
ImmunoChip



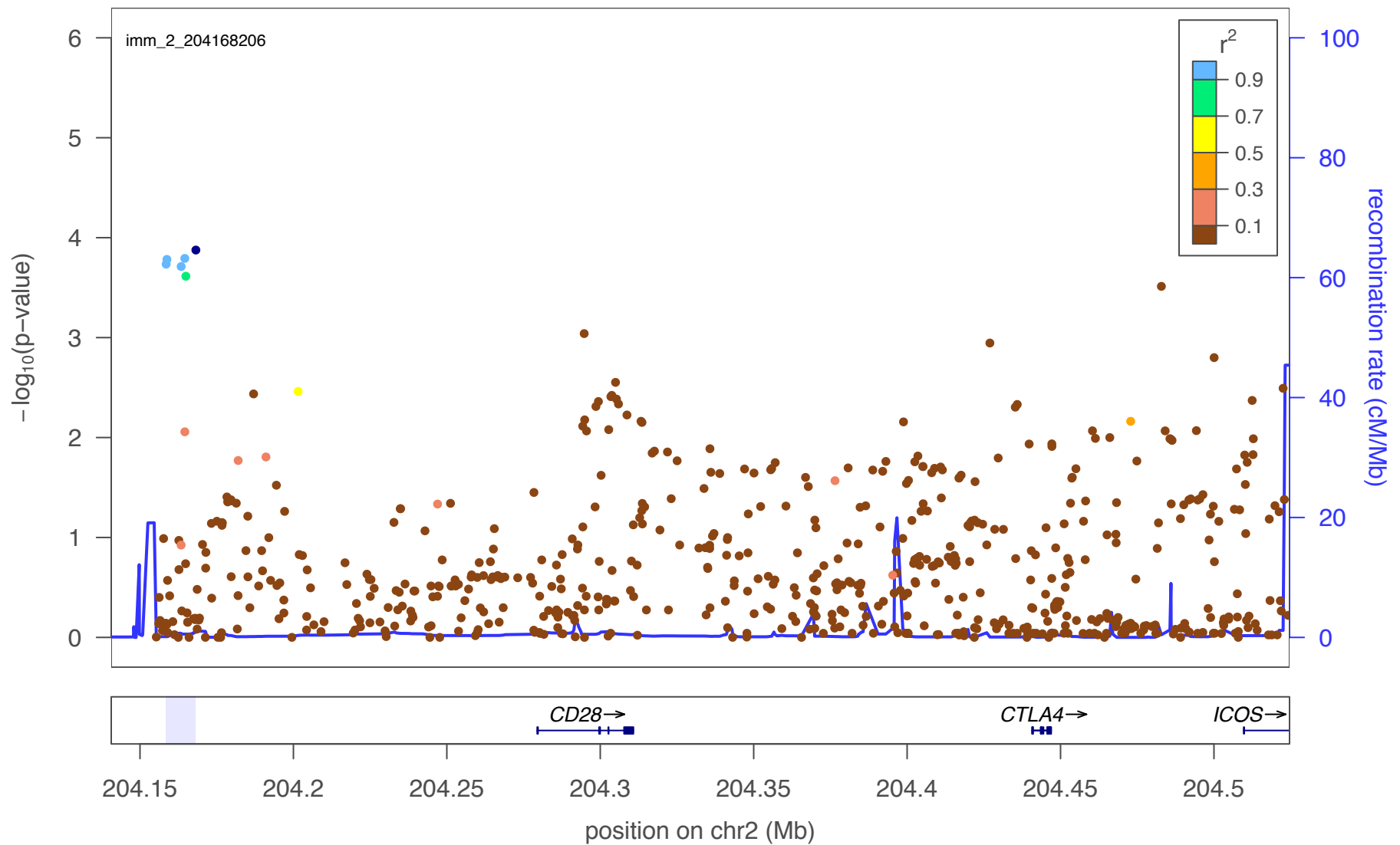
ImmunoChip



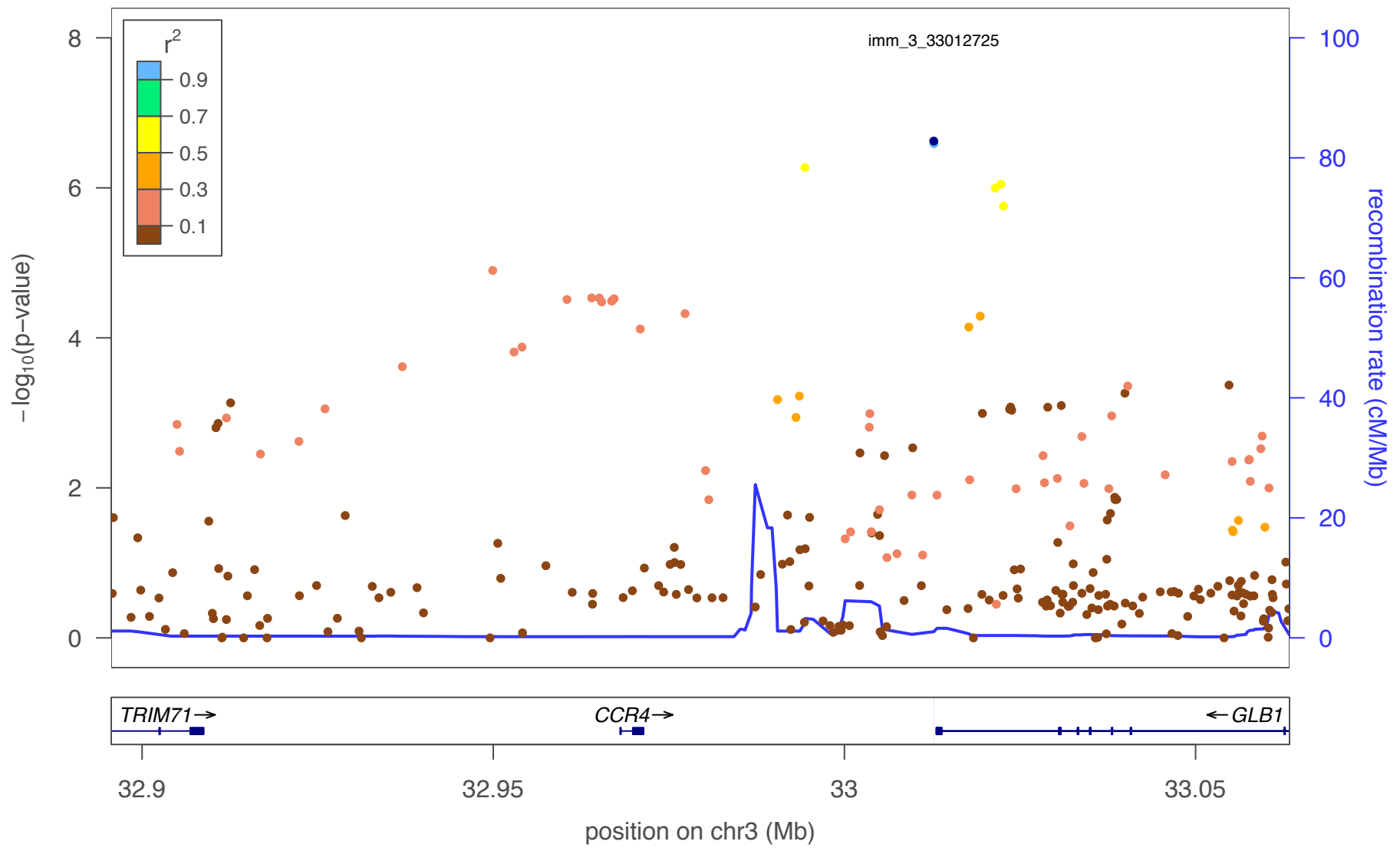
ImmunoChip



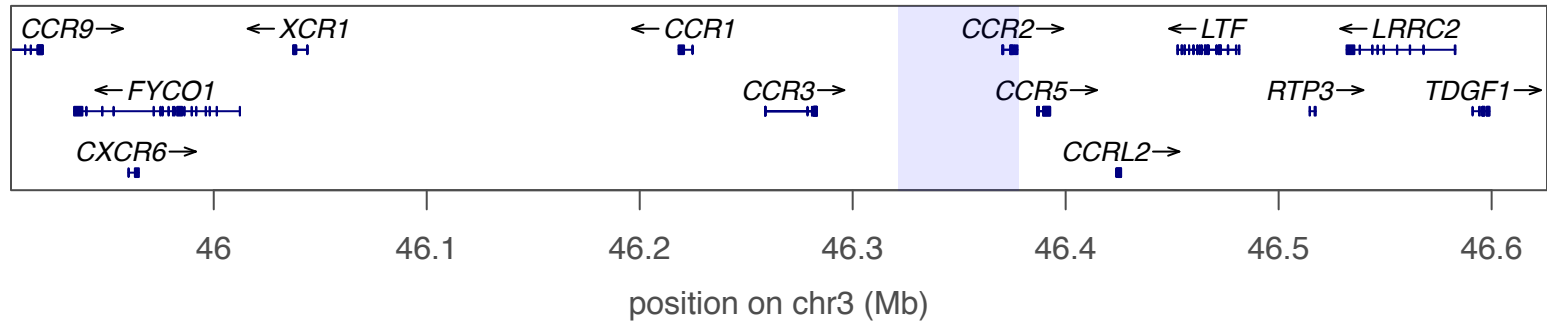
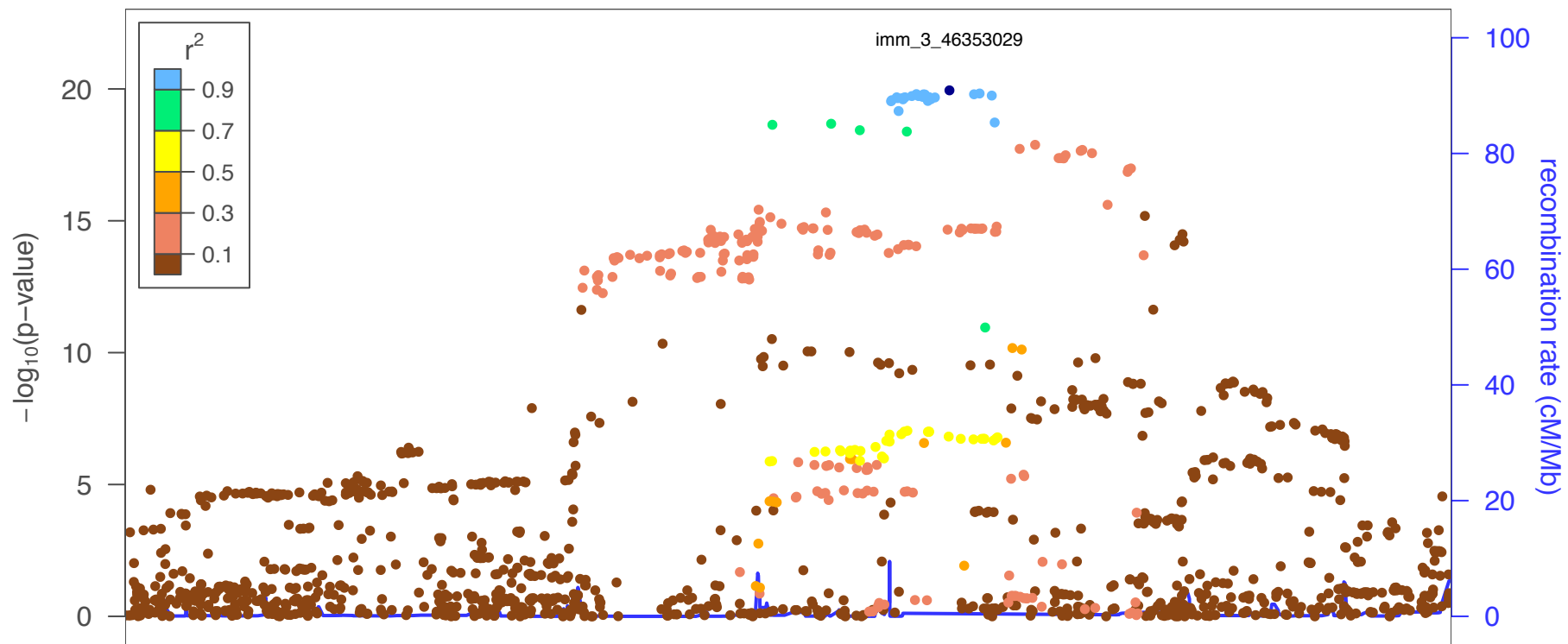
ImmunoChip



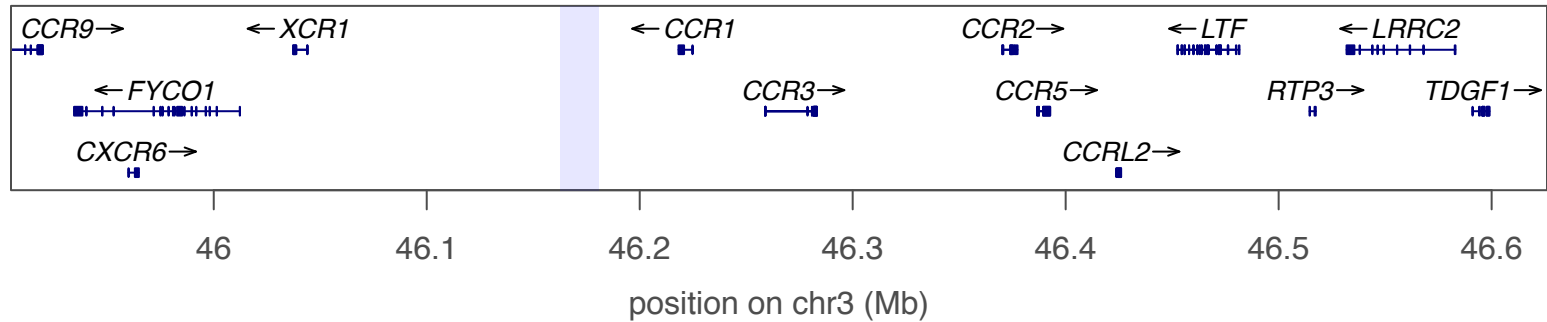
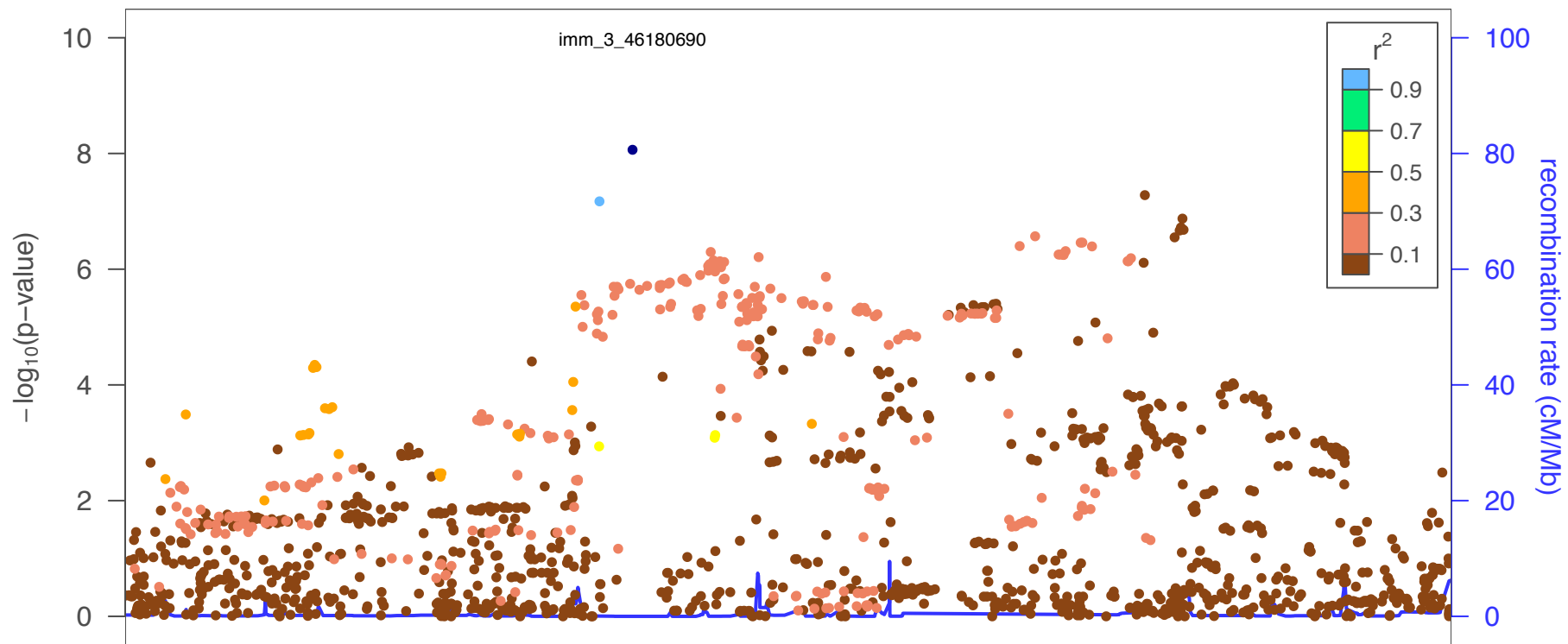
ImmunoChip | 



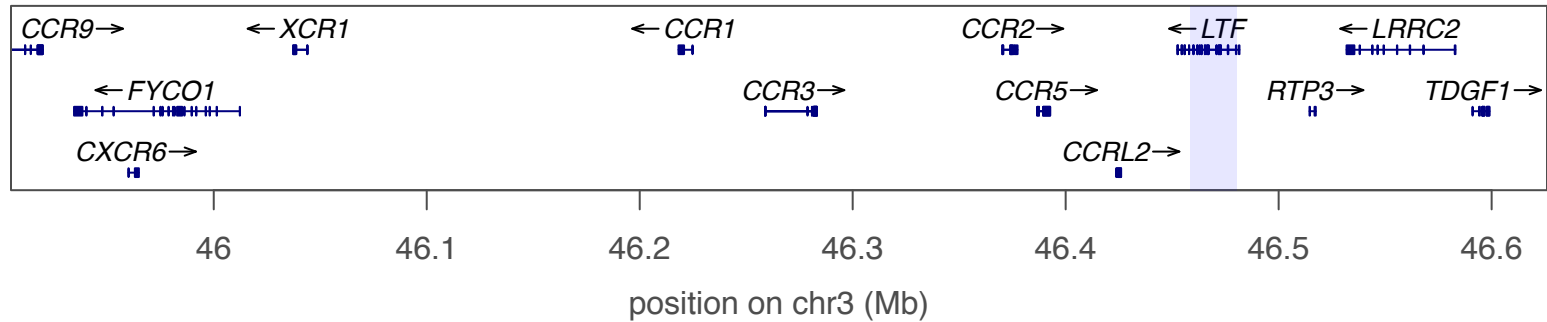
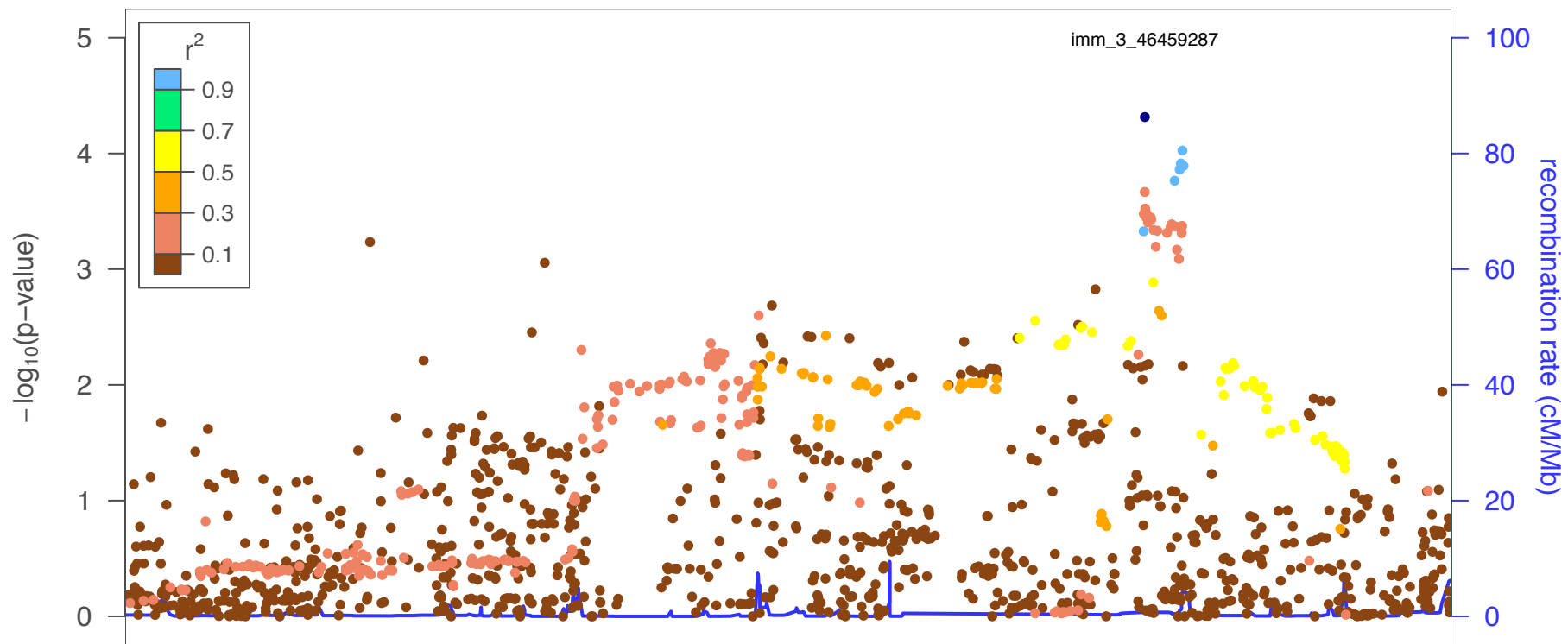
ImmunoChip 



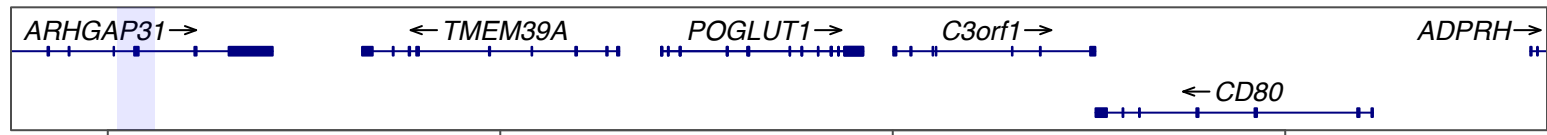
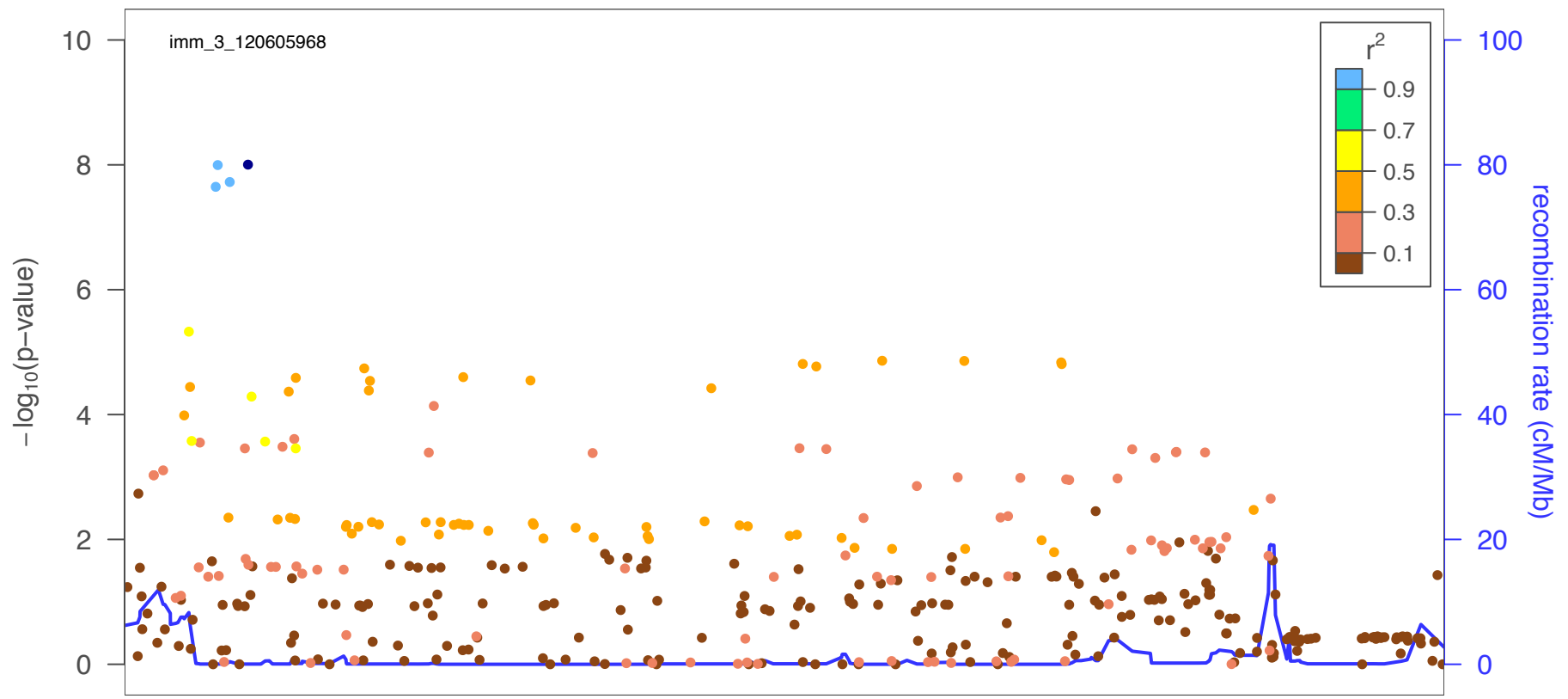
ImmunoChip 



ImmunoChip



ImmunoChip |



120.6

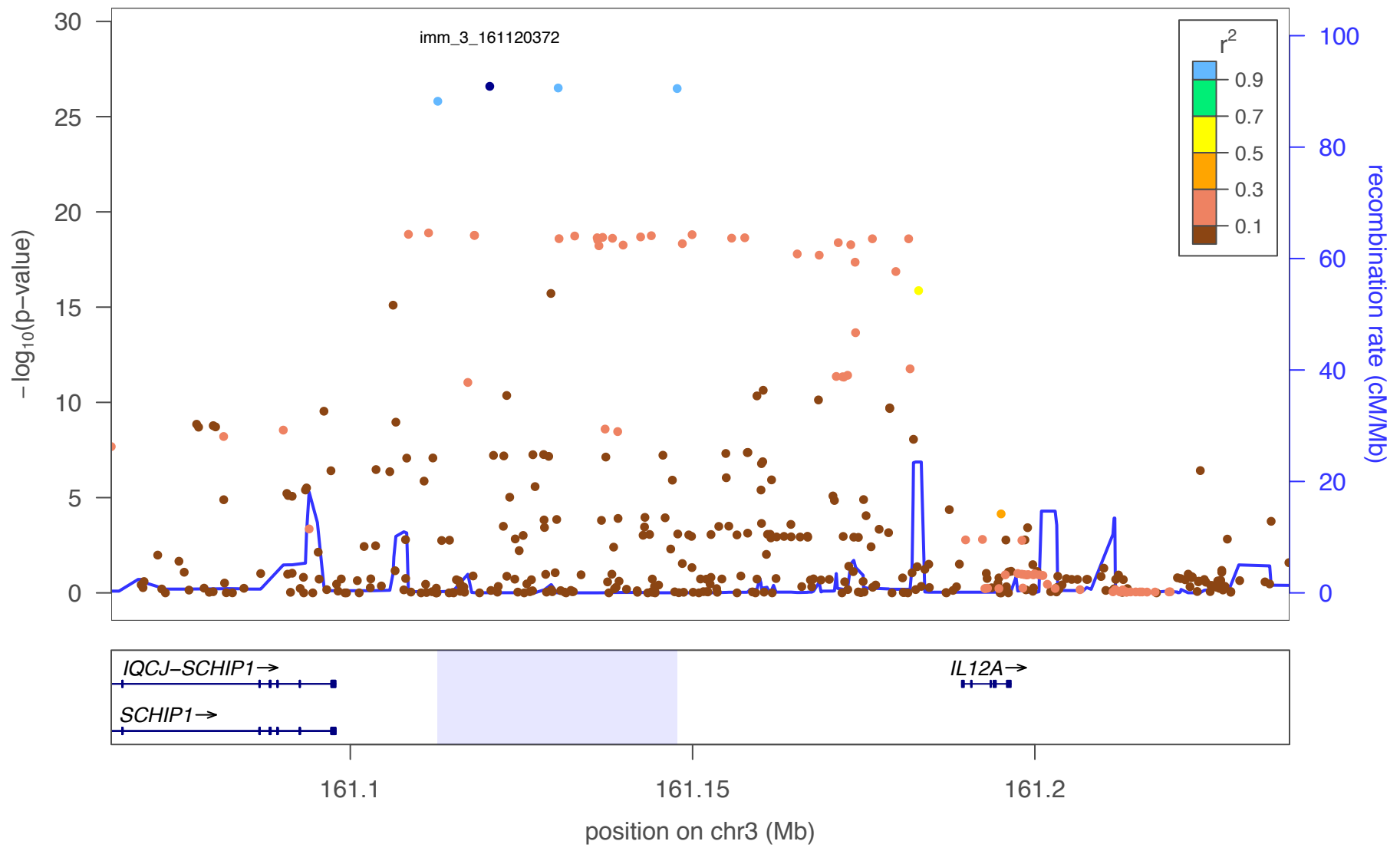
120.65

120.7

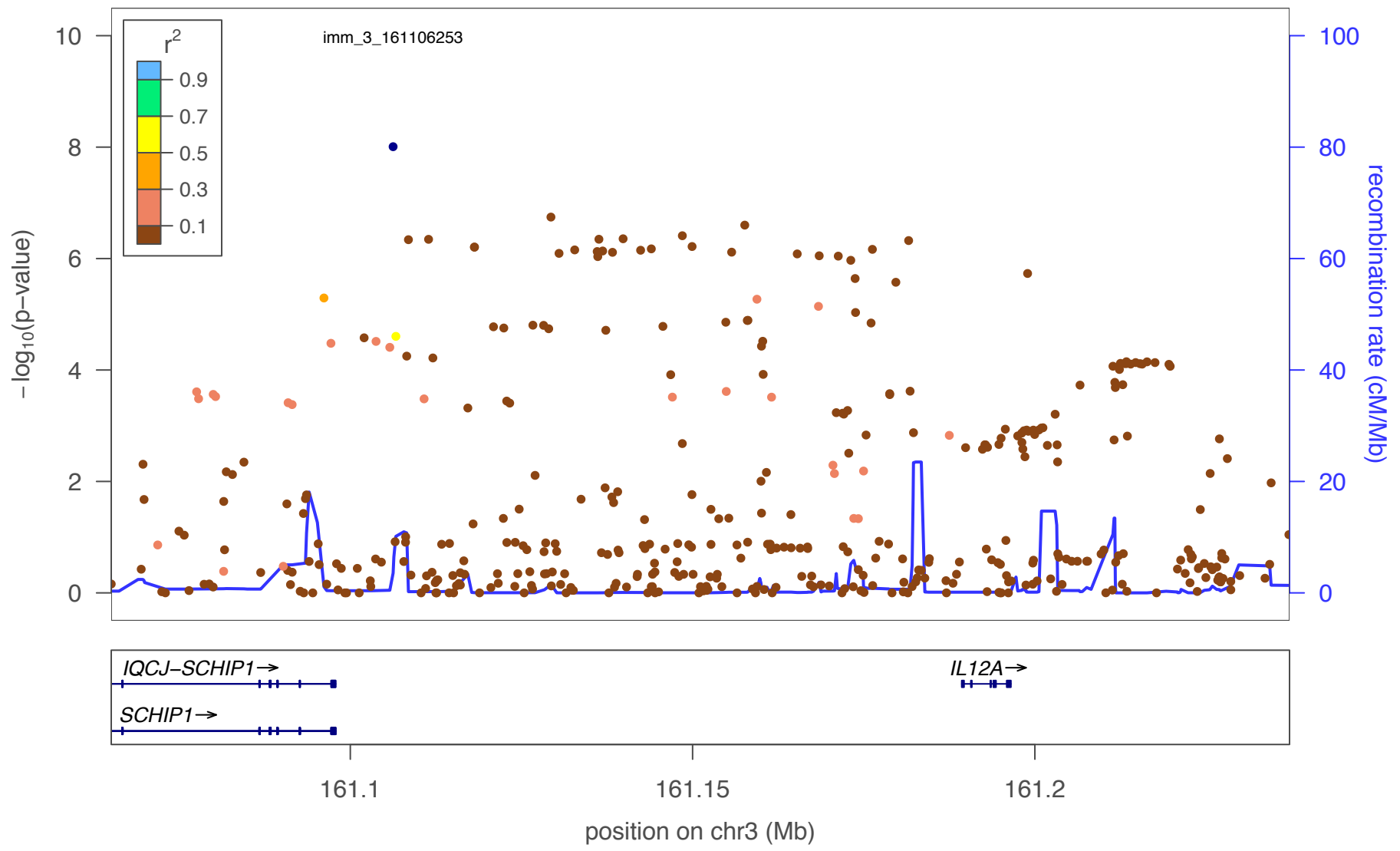
120.75

position on chr3 (Mb)

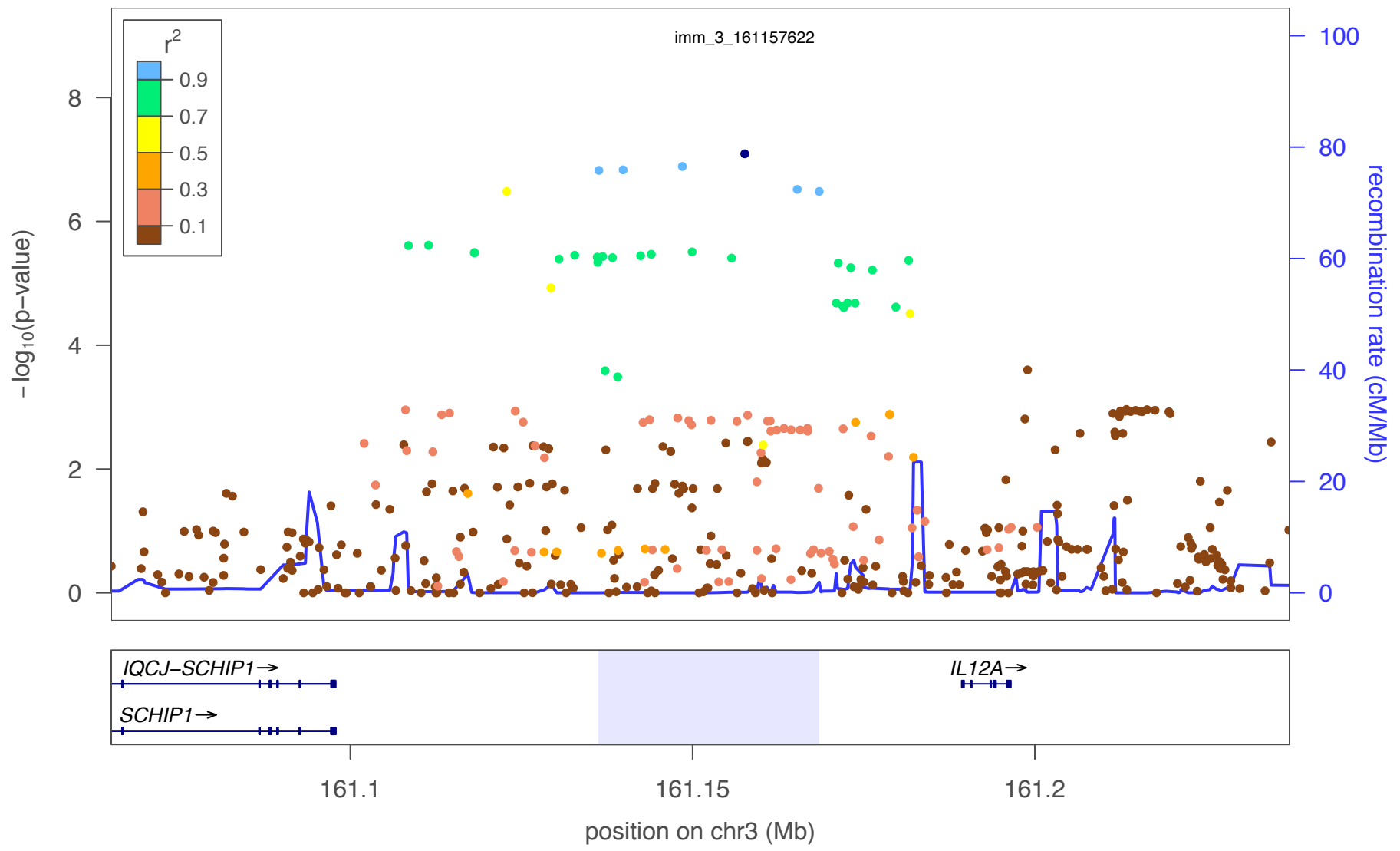
ImmunoChip | 



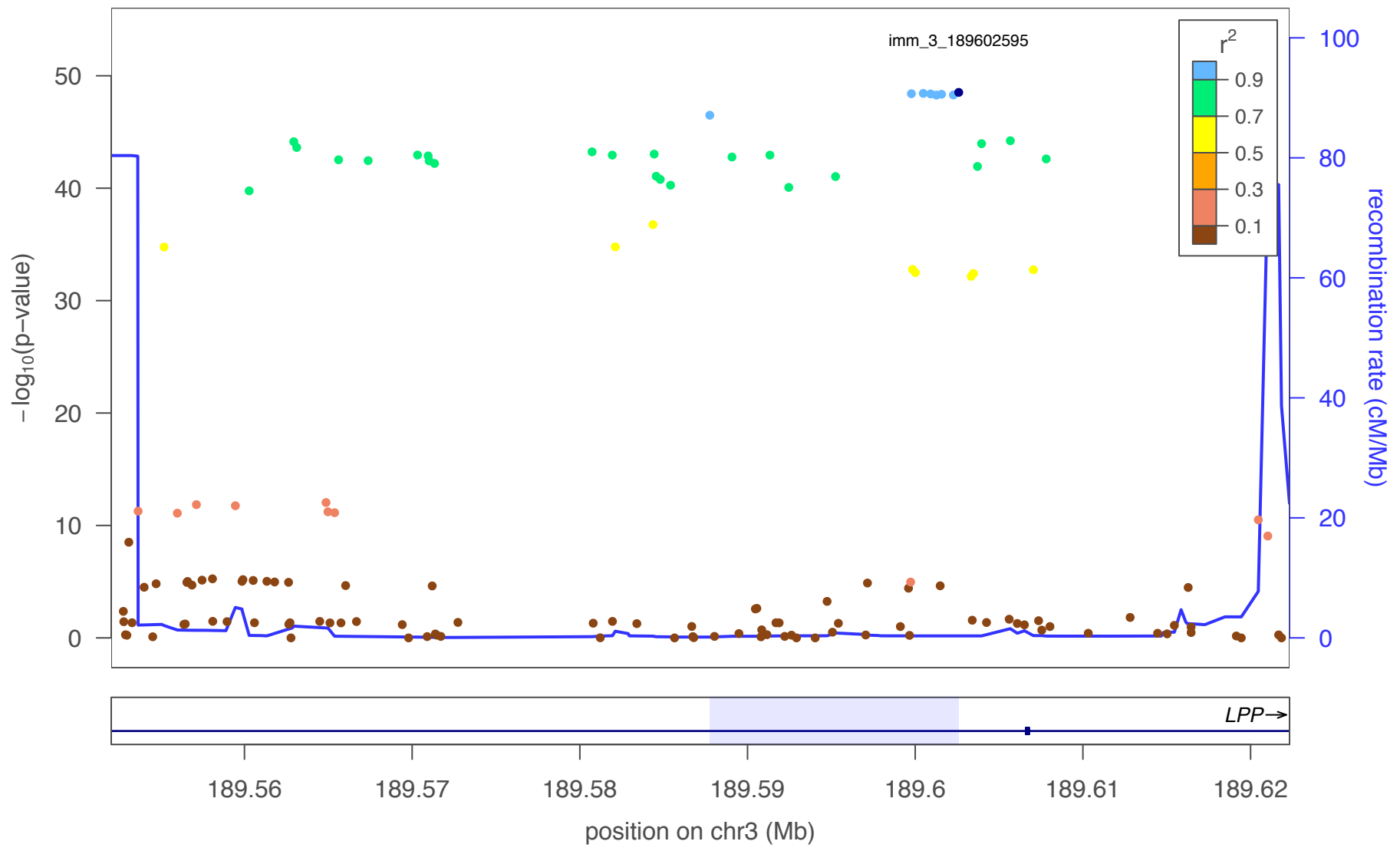
ImmunoChip | 



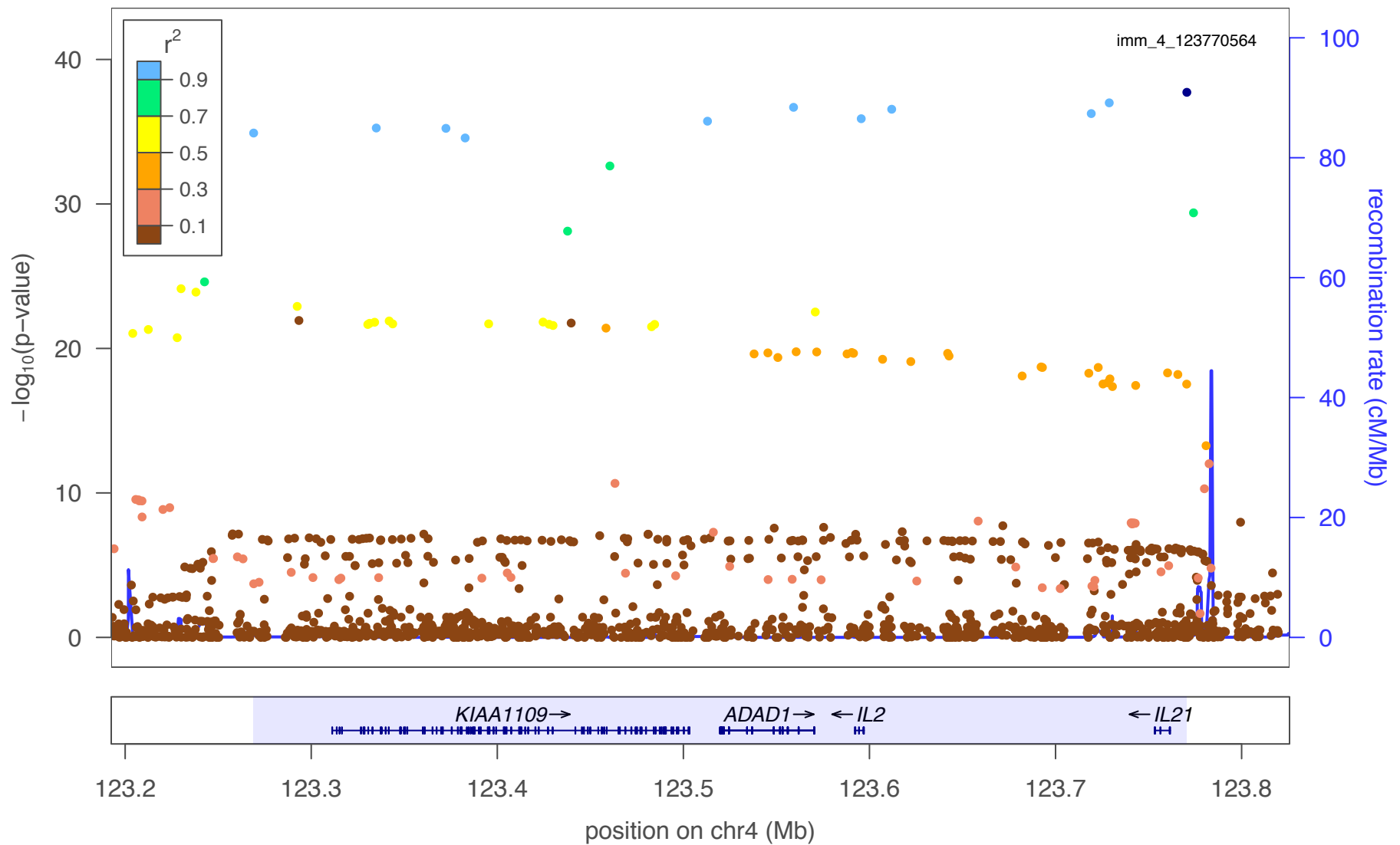
ImmunoChip | 



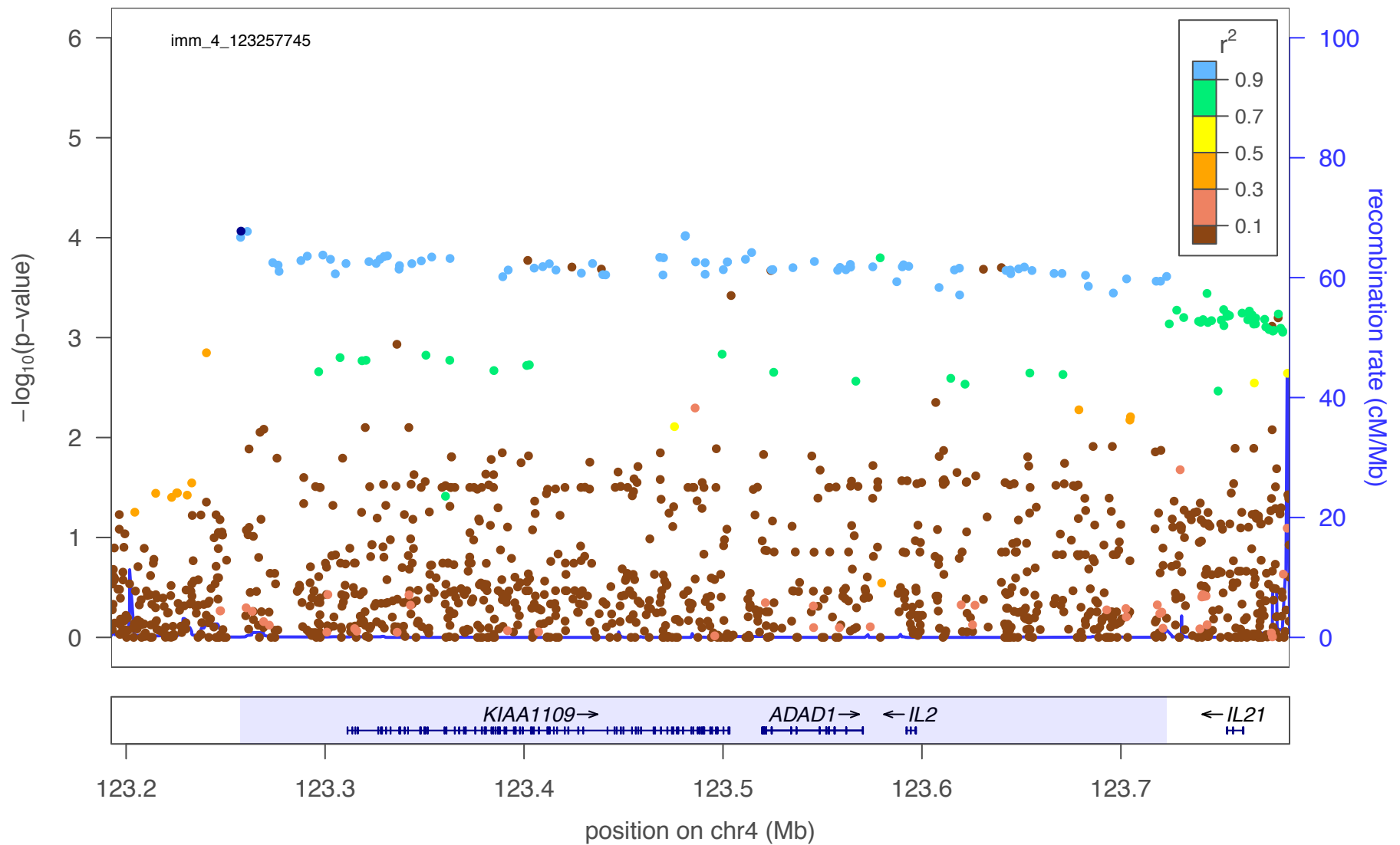
ImmunoChip



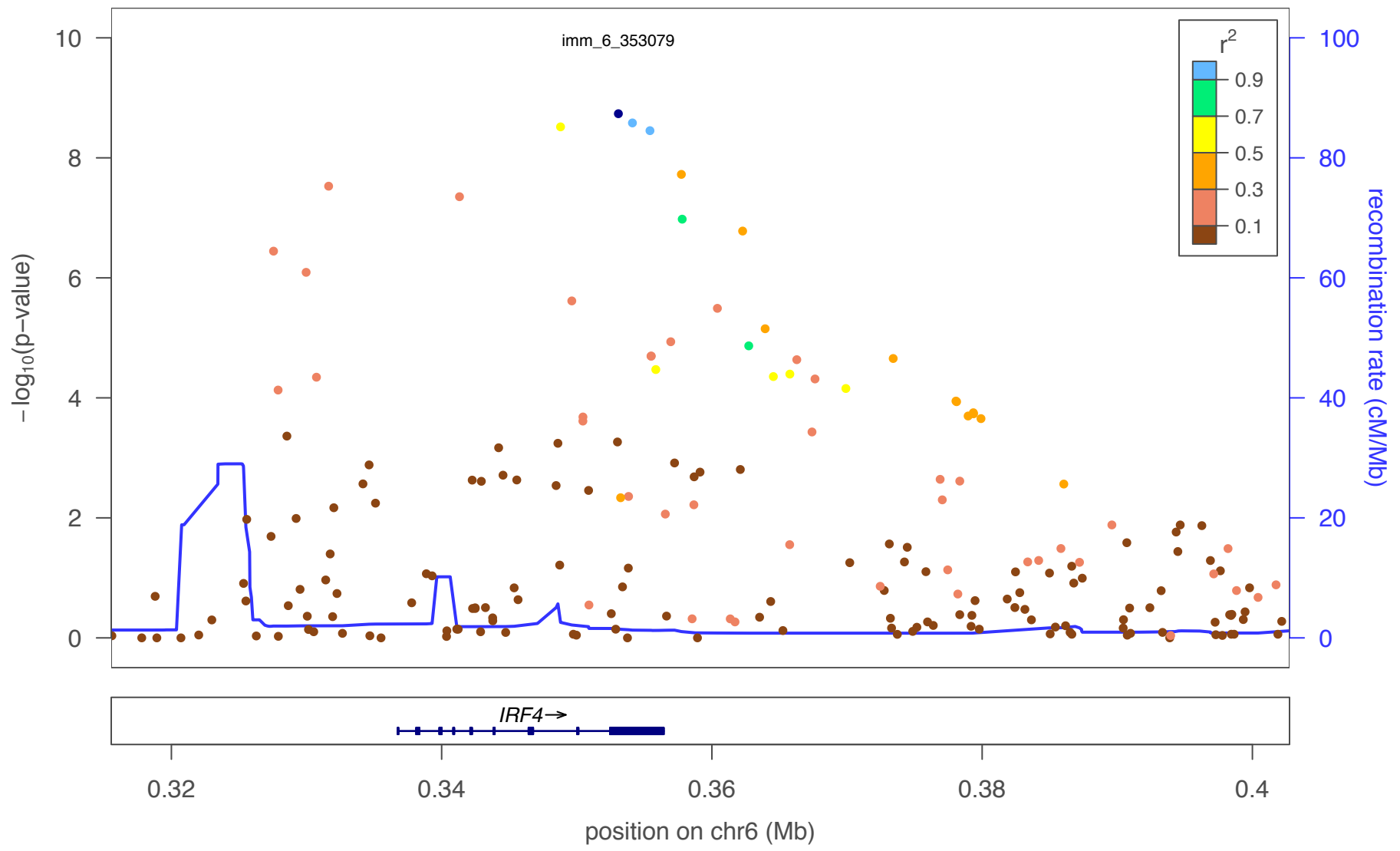
ImmunoChip 



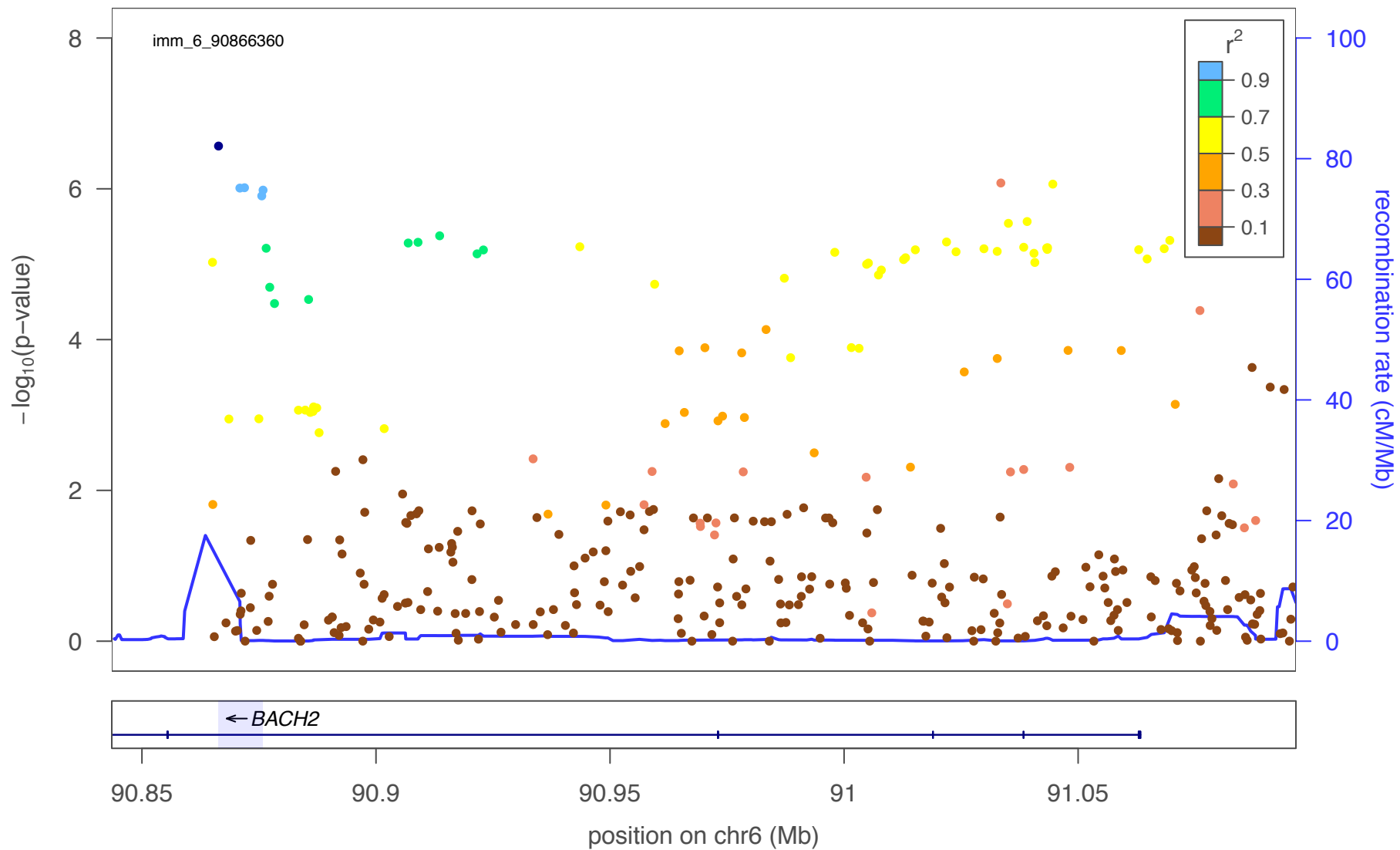
ImmunoChip



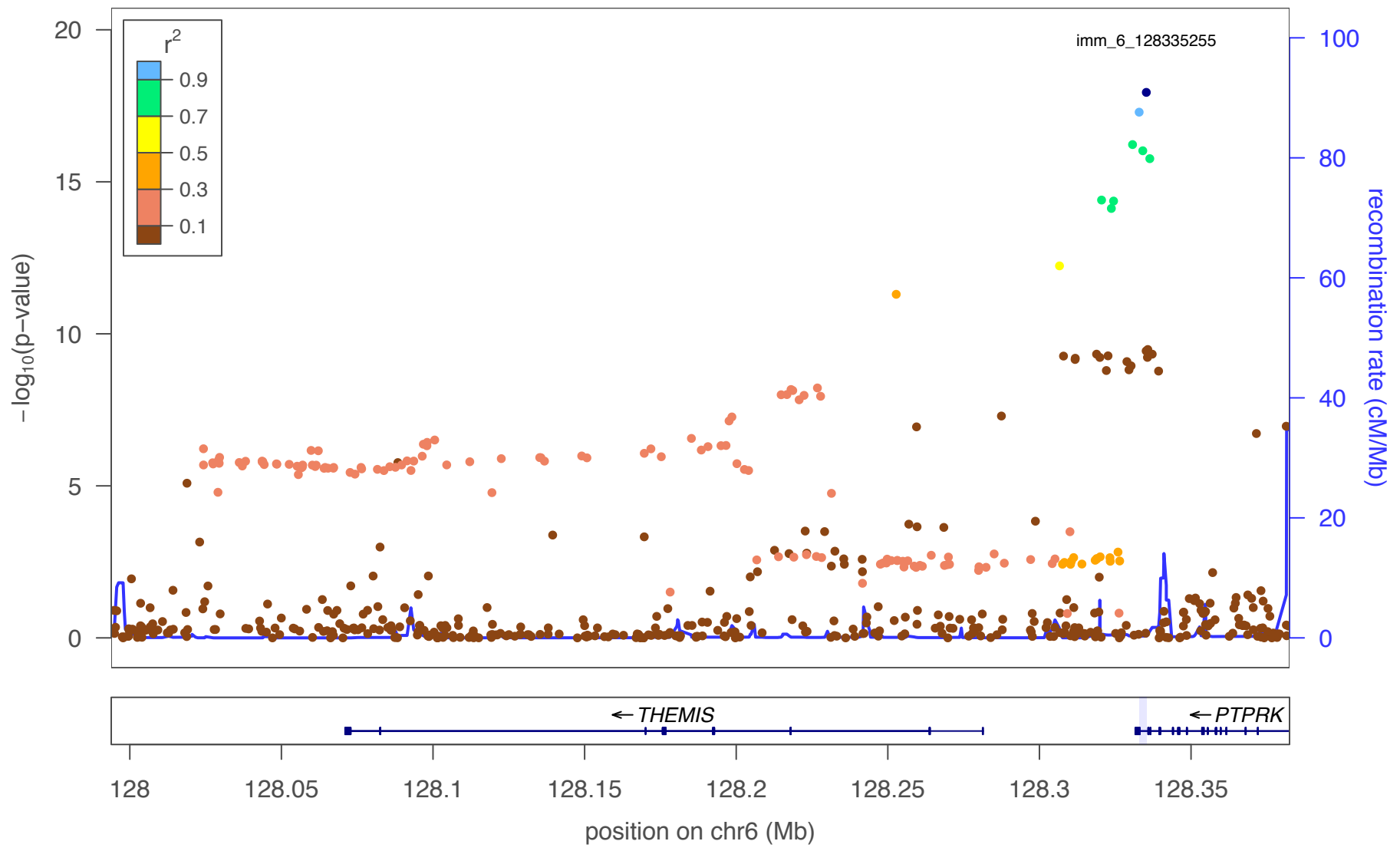
ImmunoChip



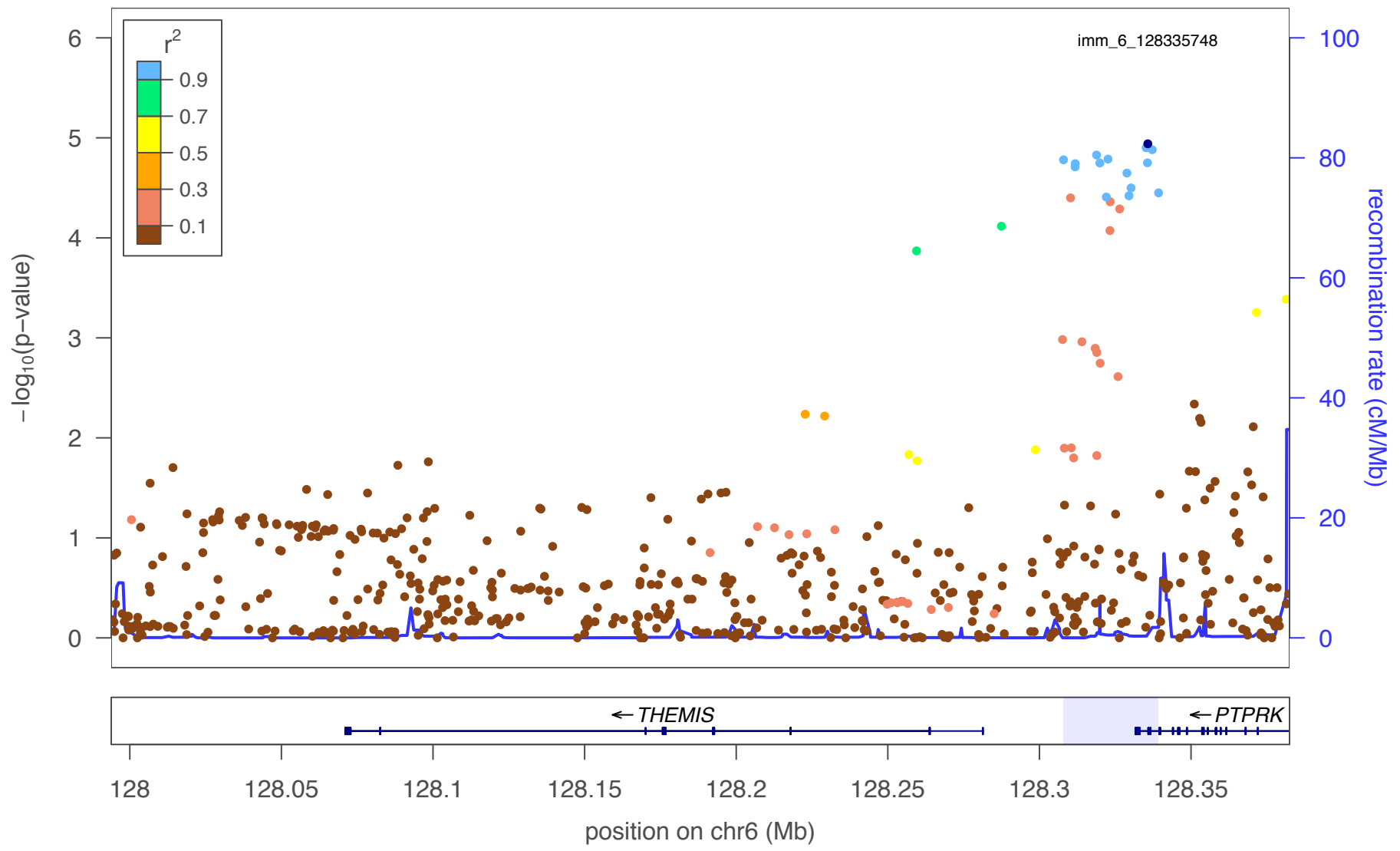
ImmunoChip



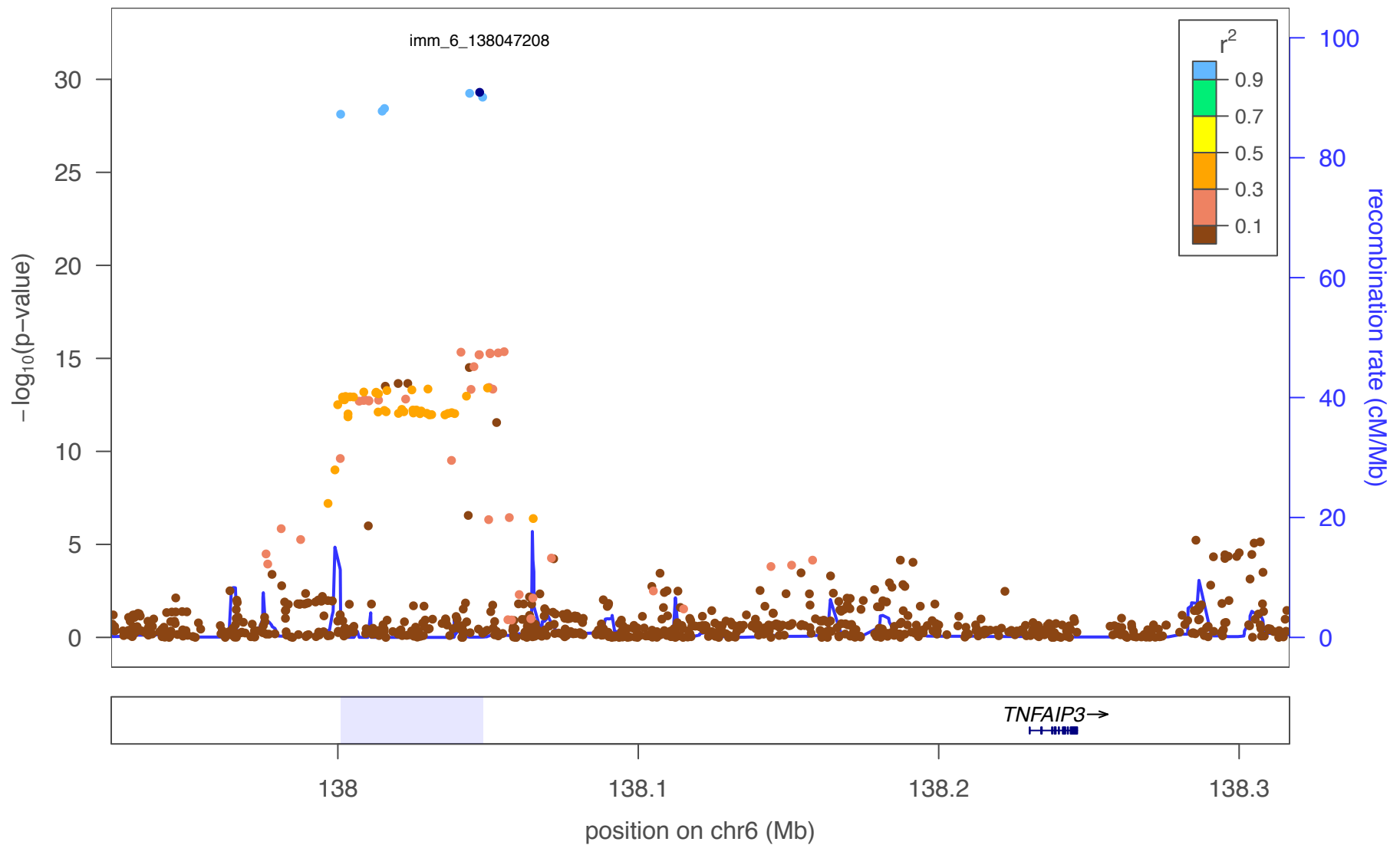
ImmunoChip 

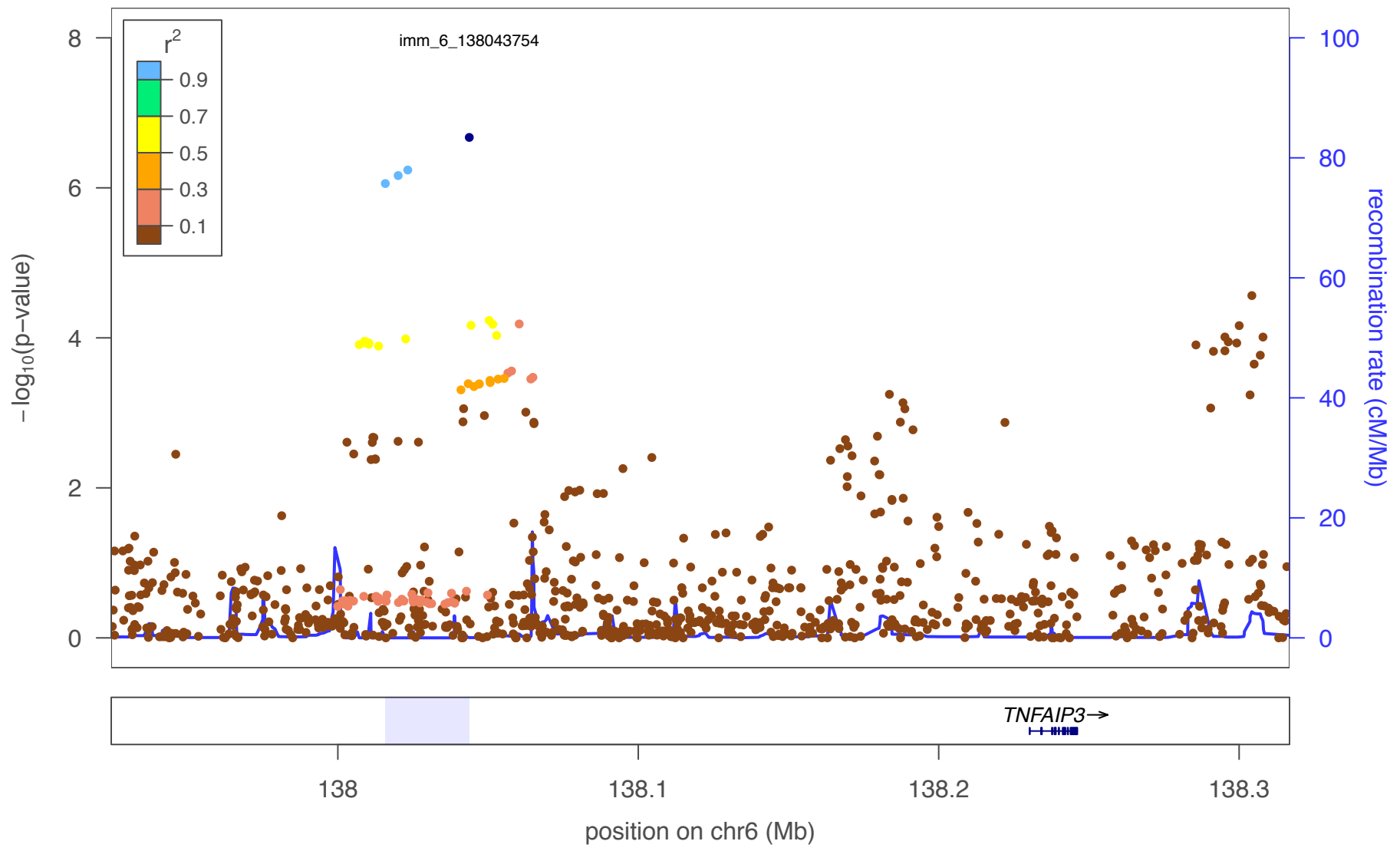


ImmunoChip 

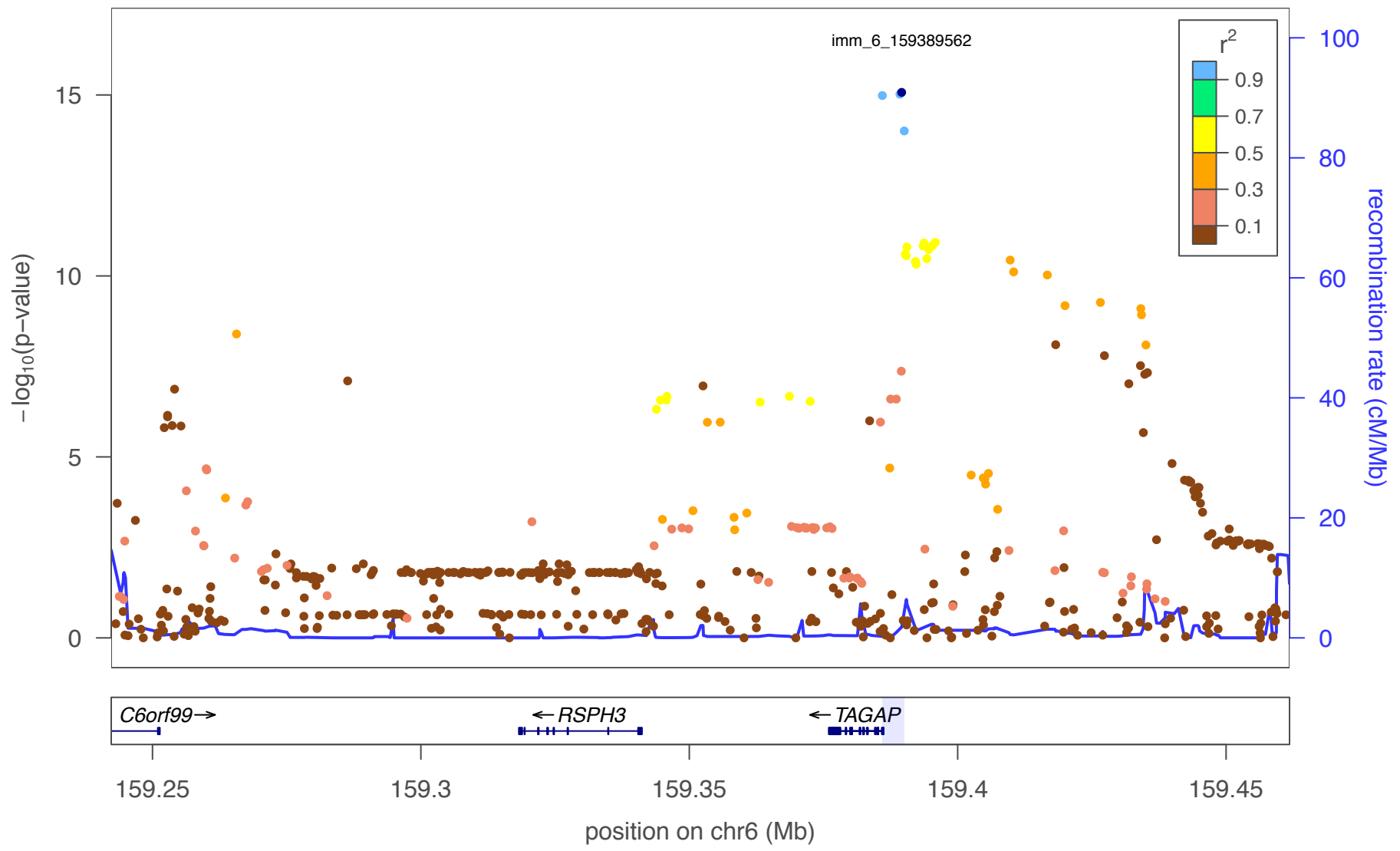


ImmunoChip

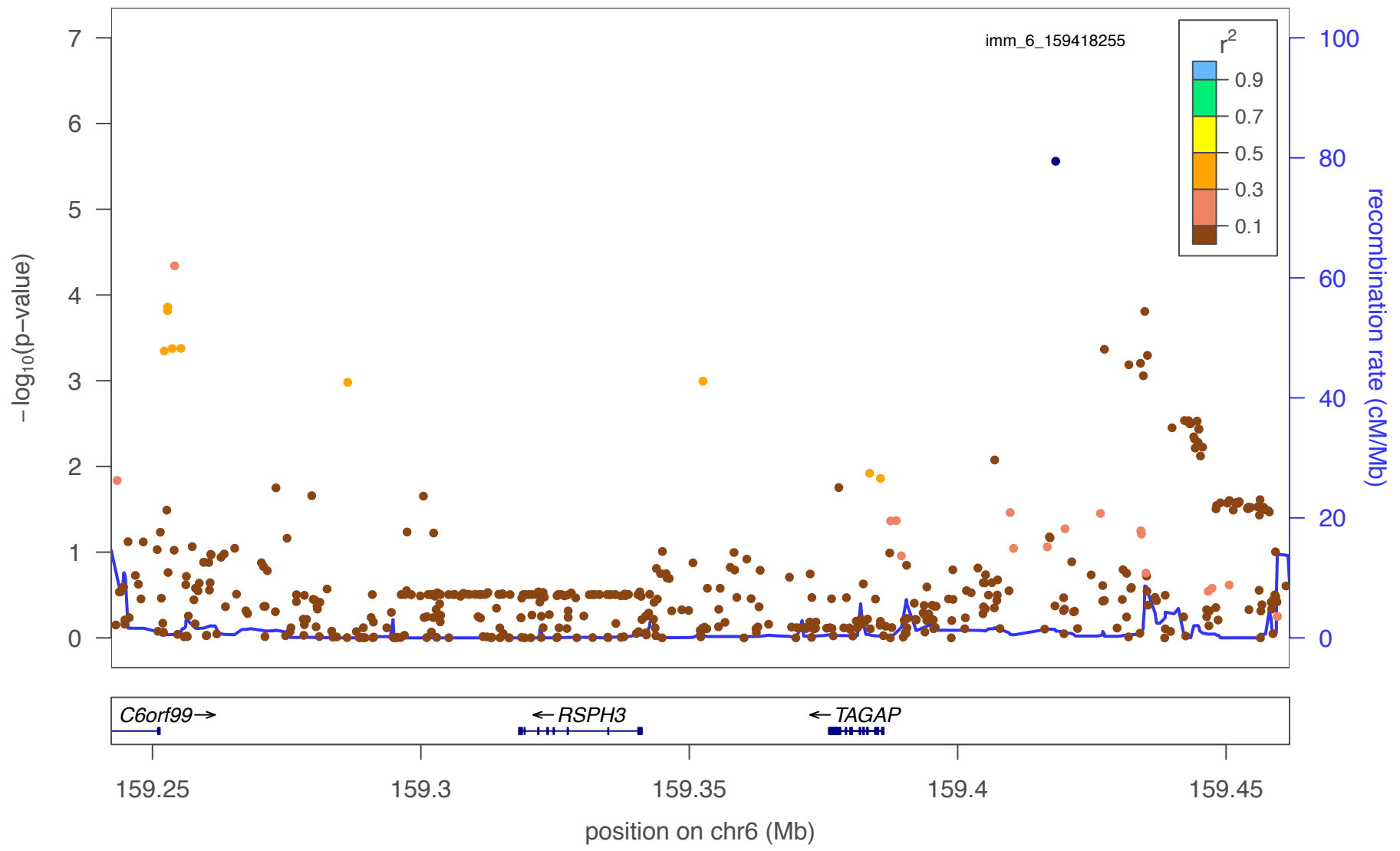




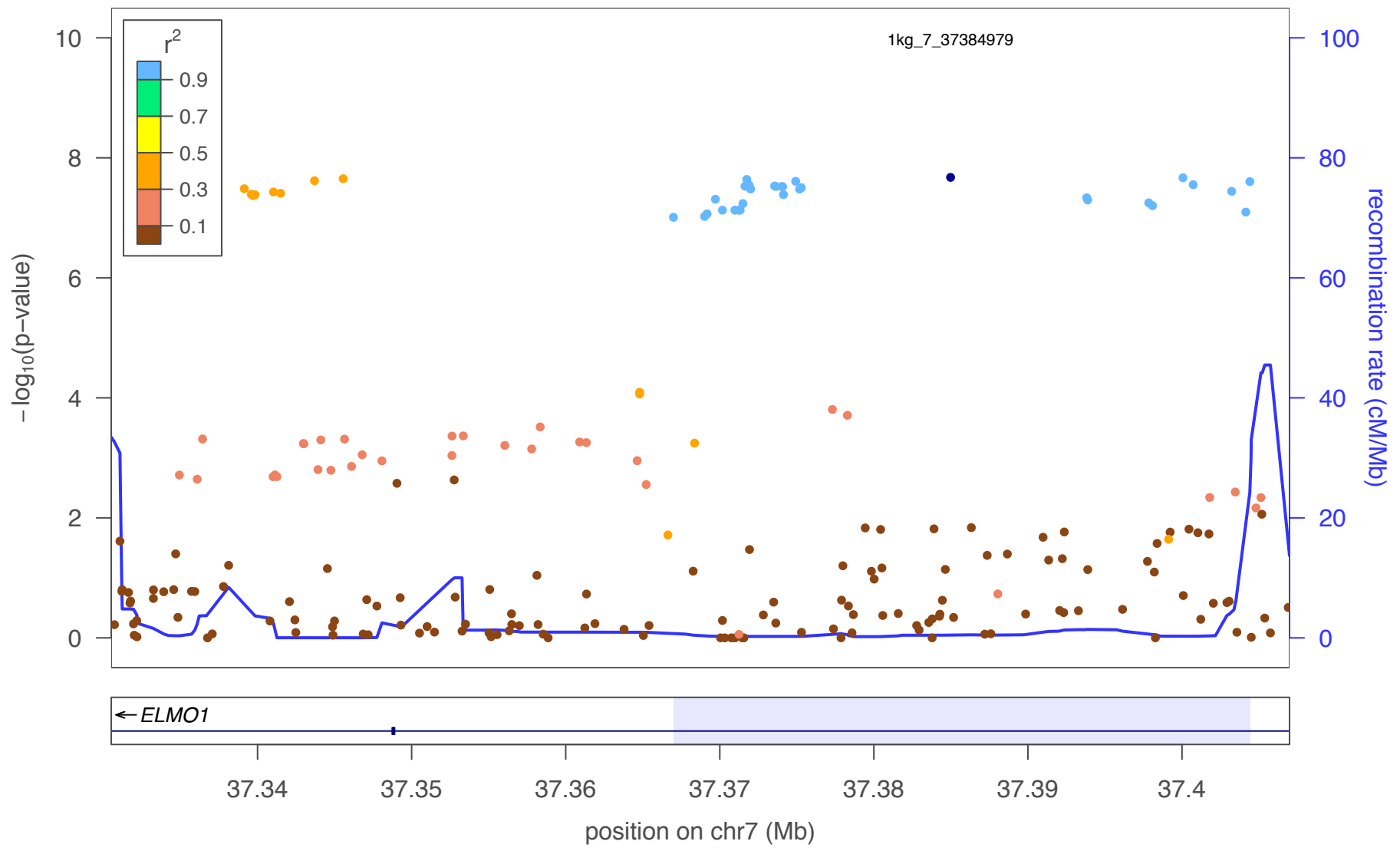
ImmunoChip



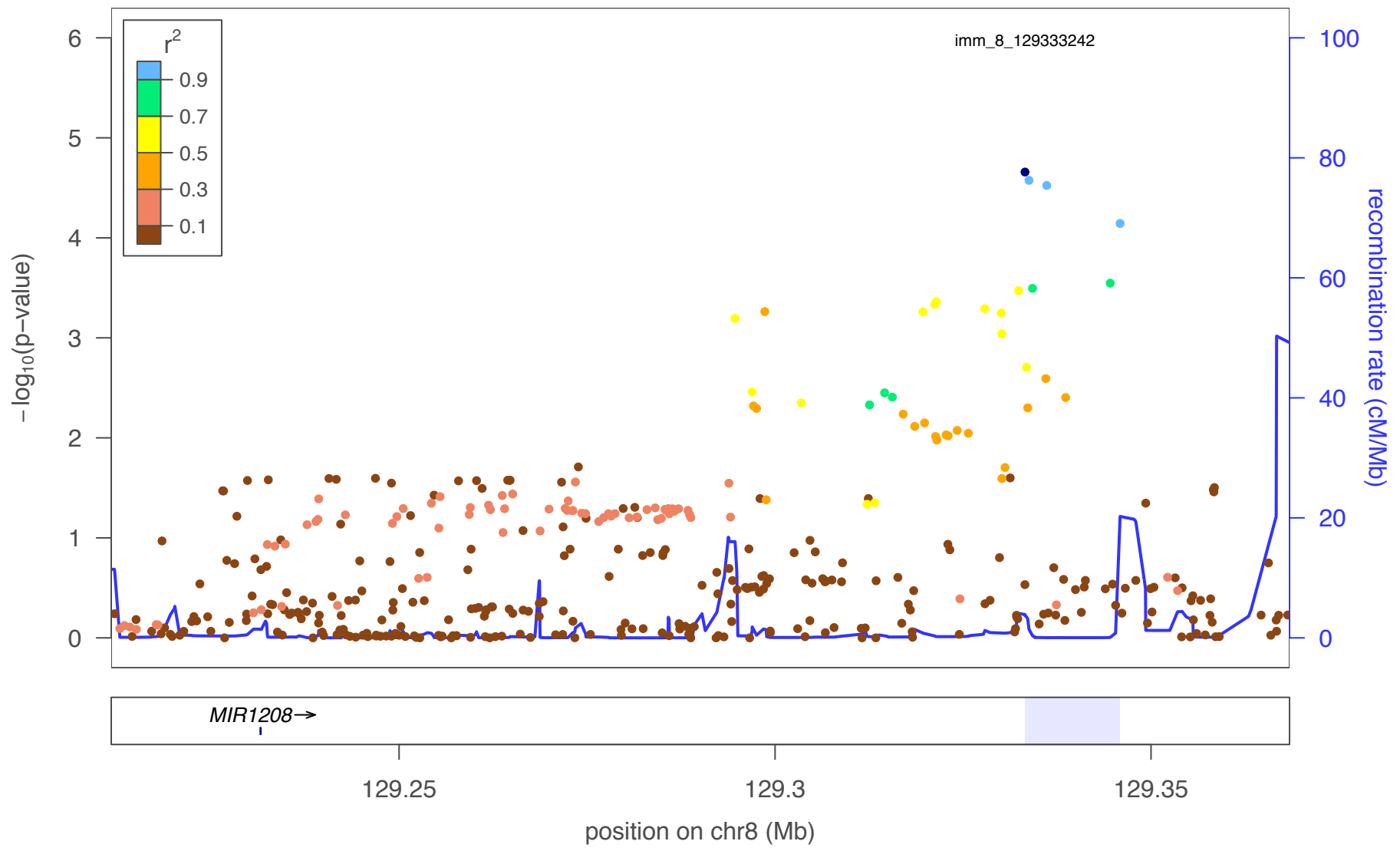
ImmunoChip



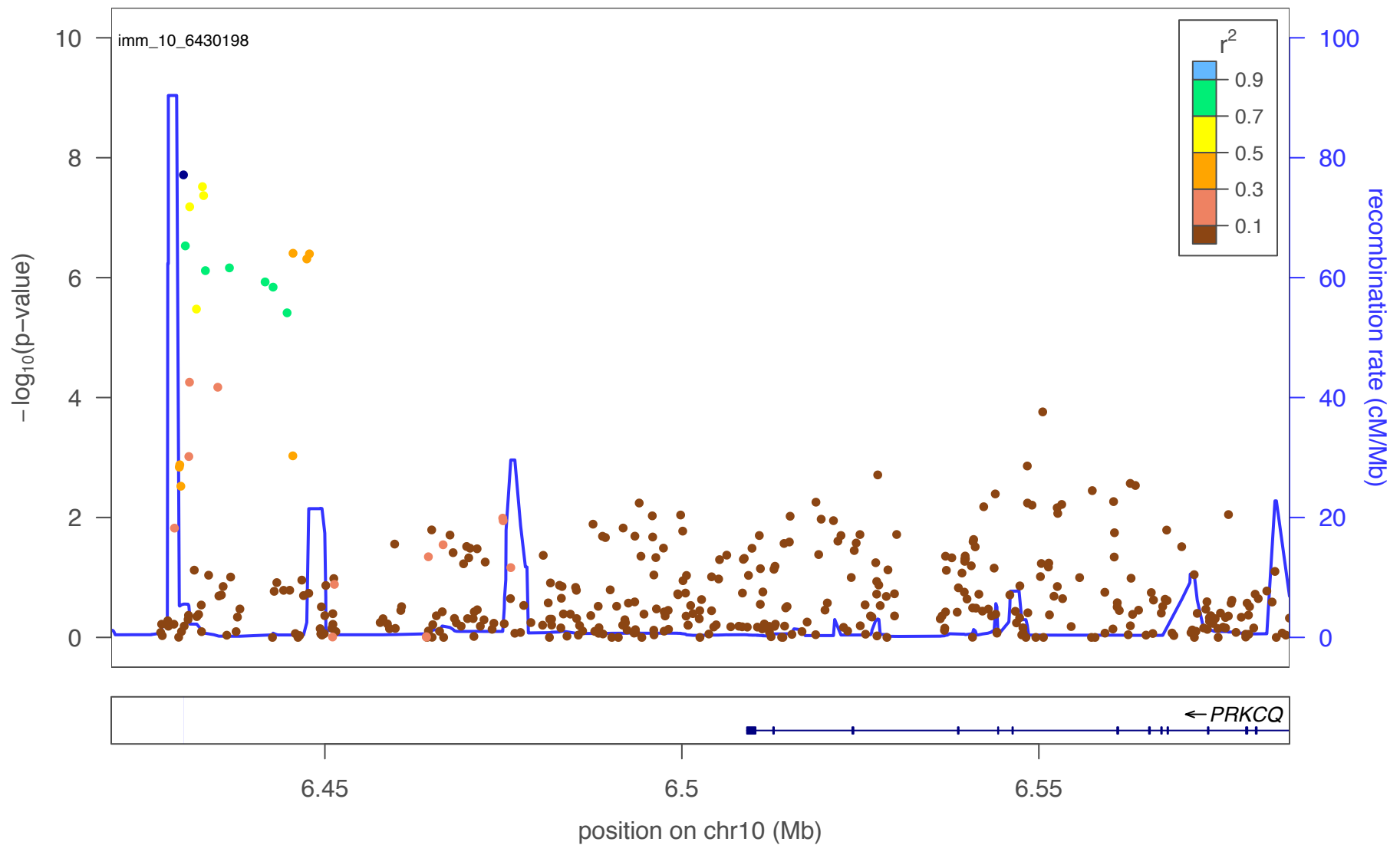
ImmunoChip



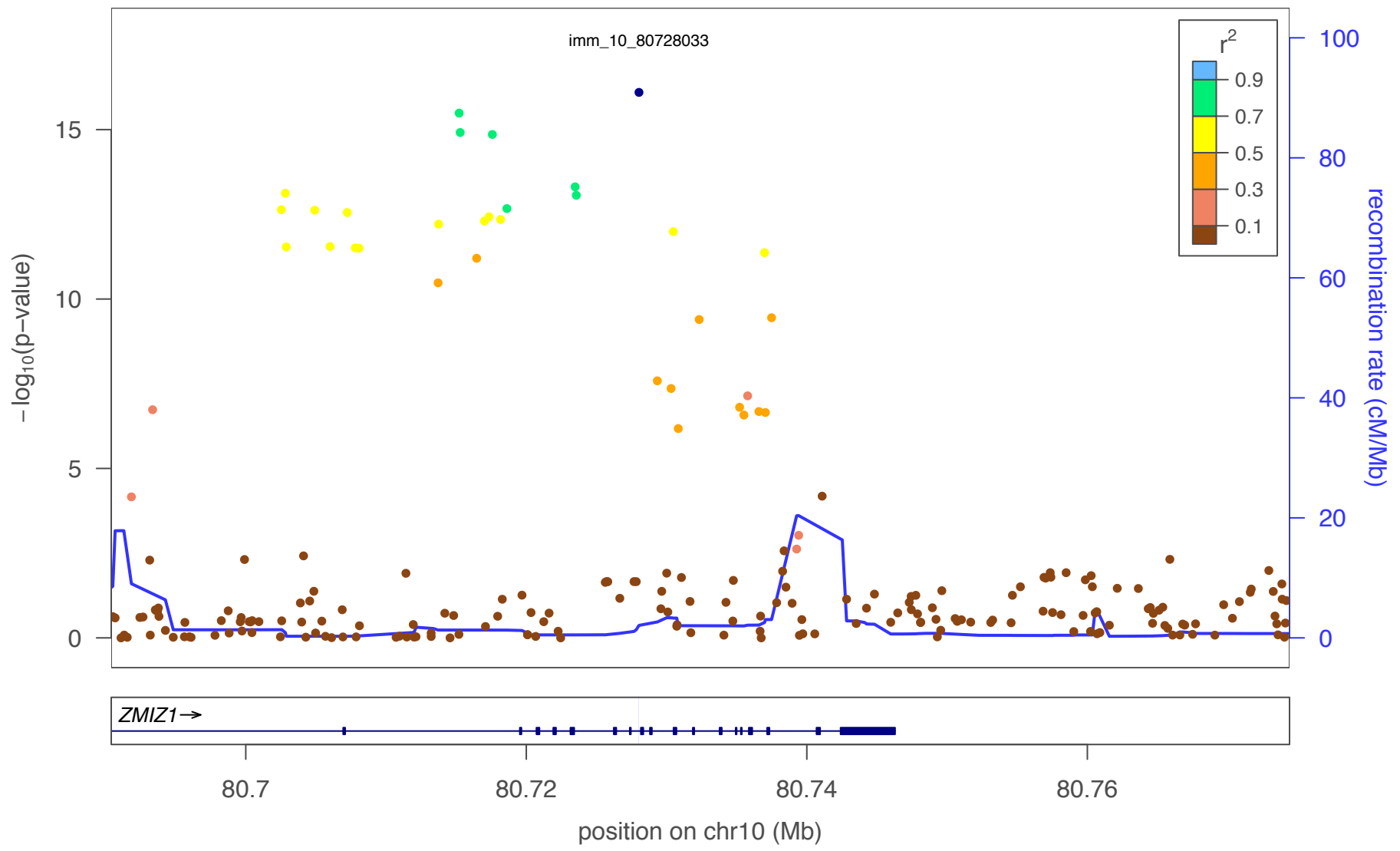
ImmunoChip 



ImmunoChip



ImmunoChip 

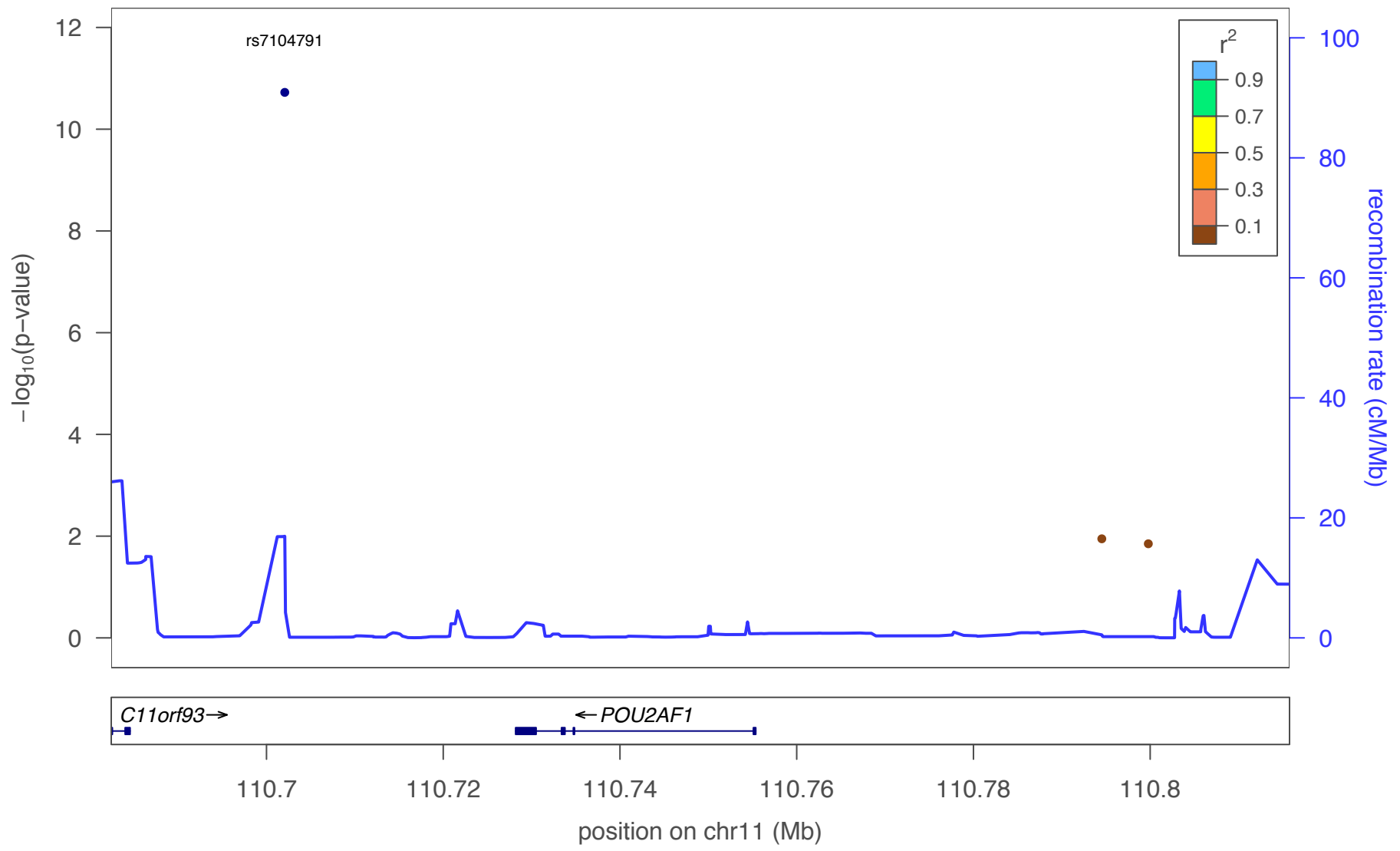


ImmunoChip

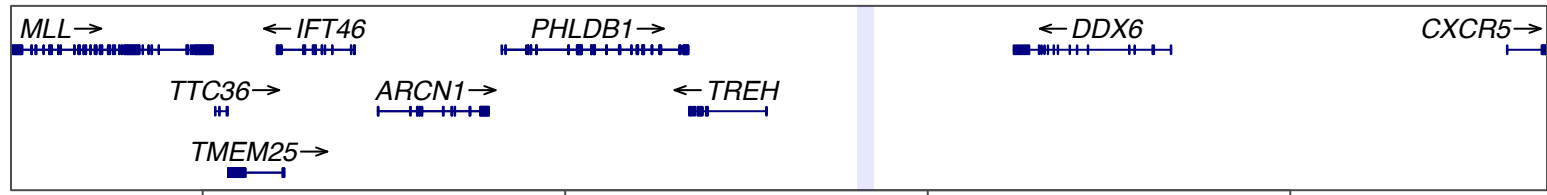
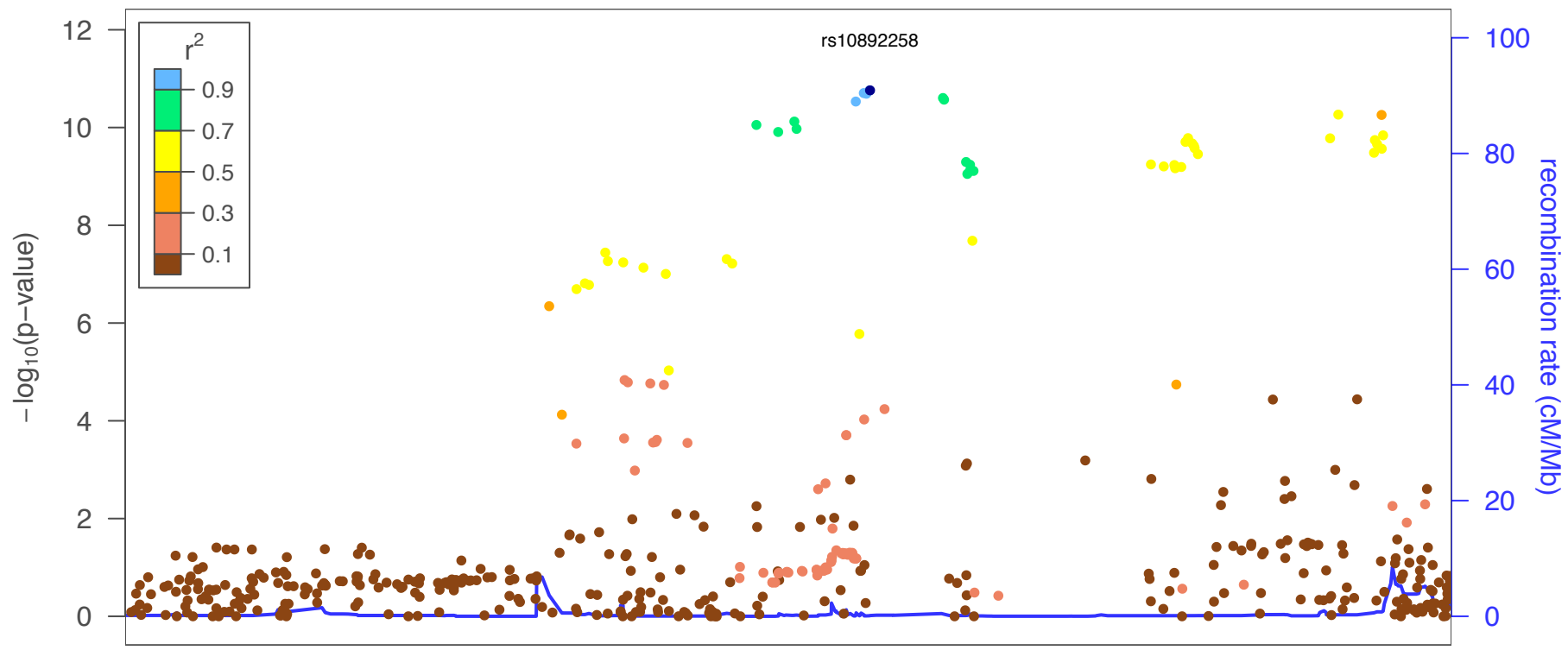
|

|

|



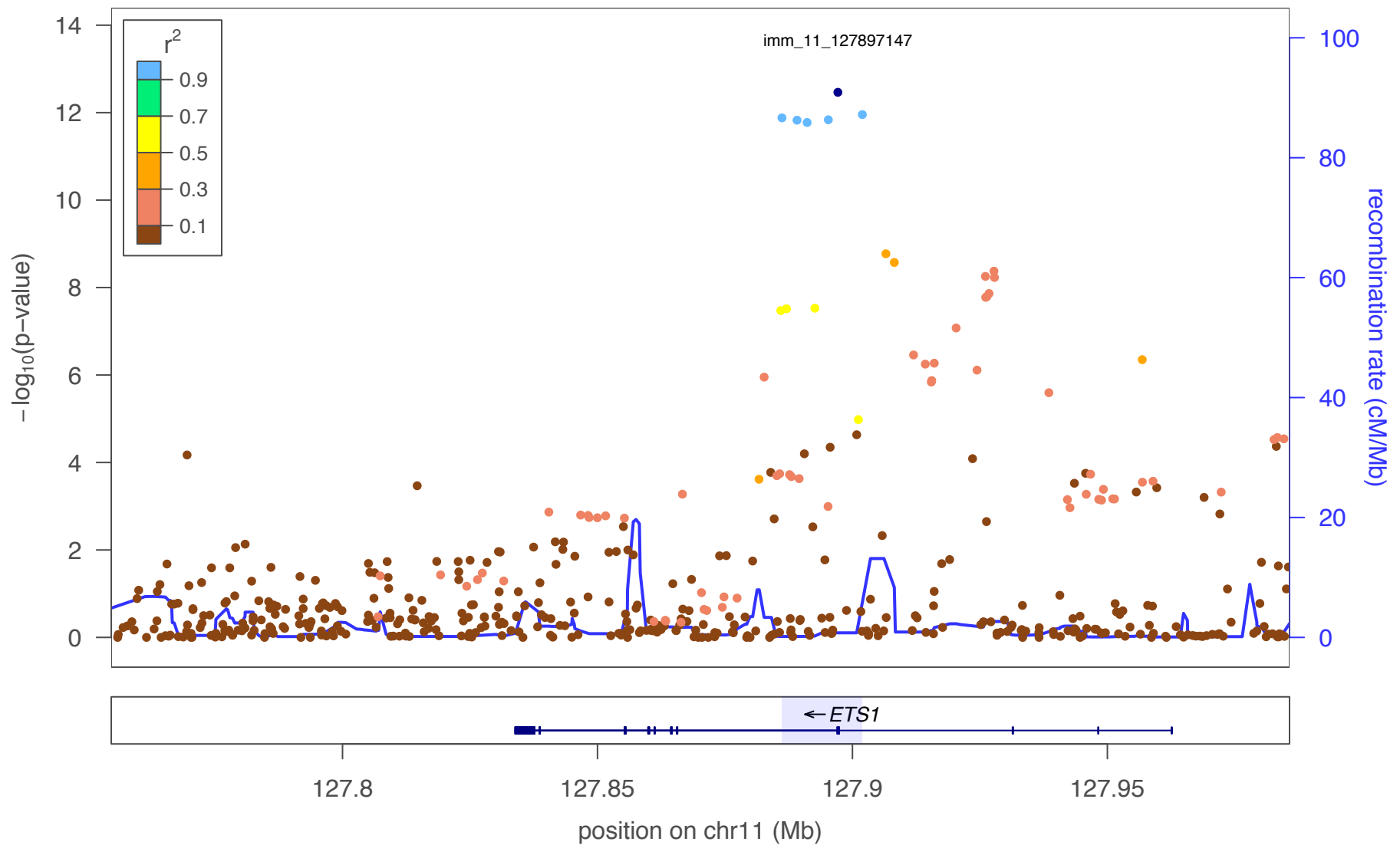
ImmunoChip 



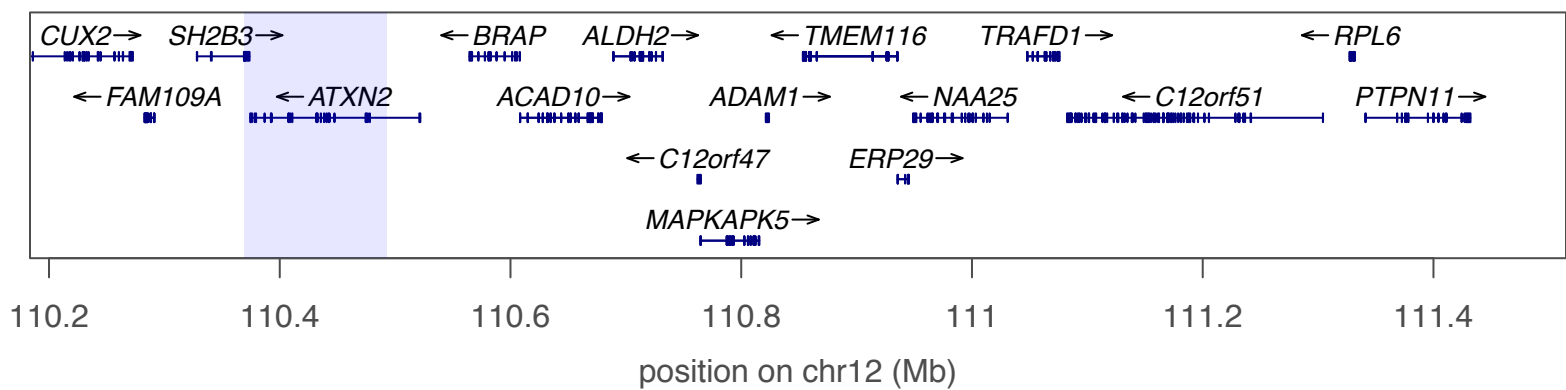
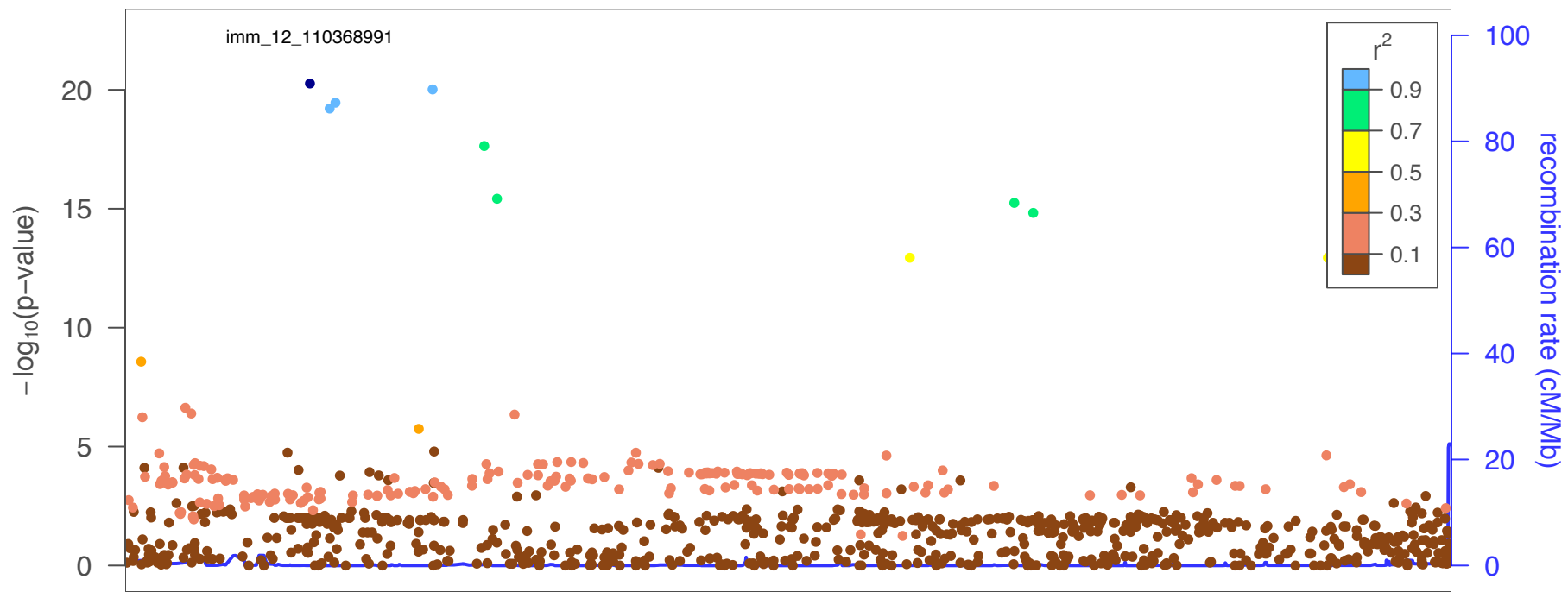
117.9 118 118.1 118.2

position on chr11 (Mb)

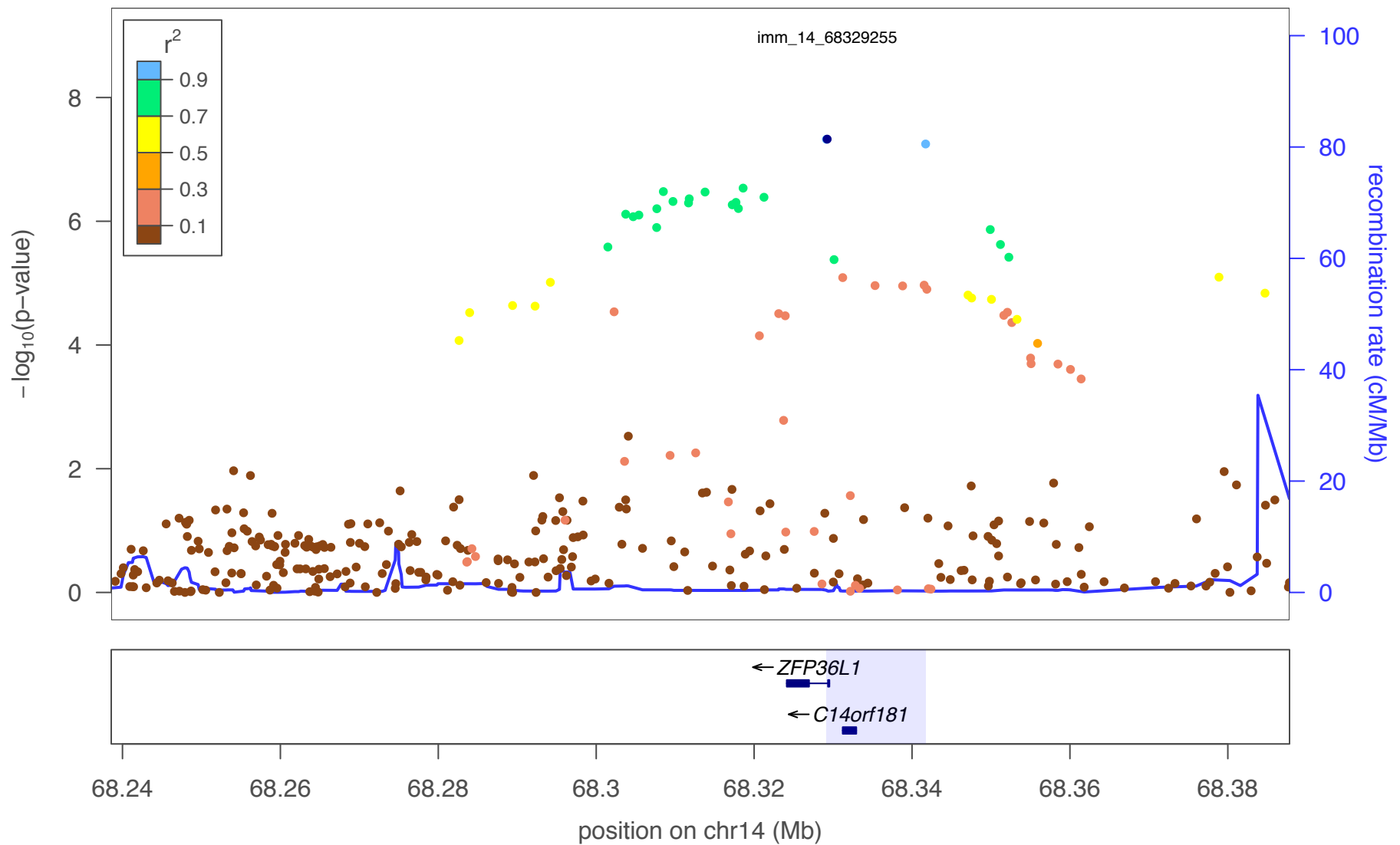
ImmunoChip



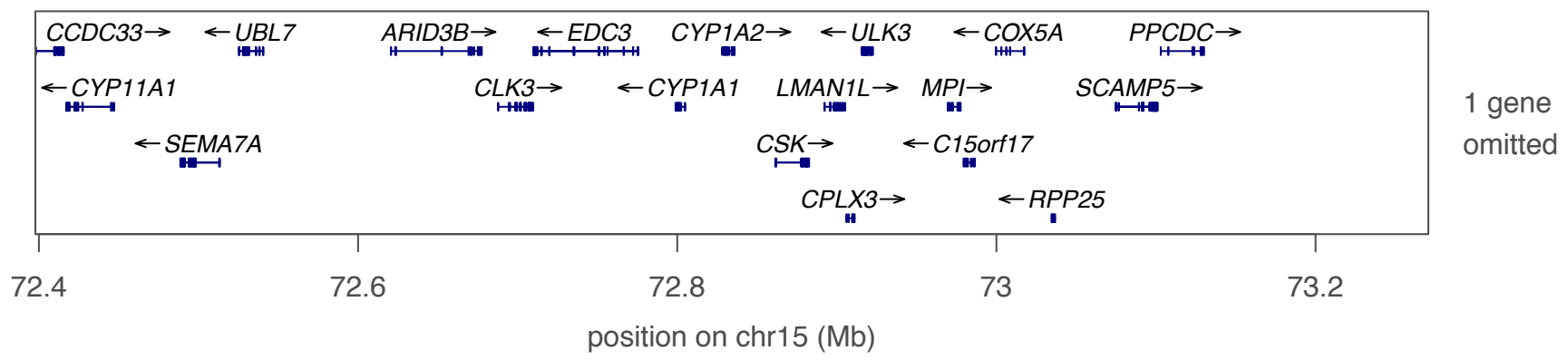
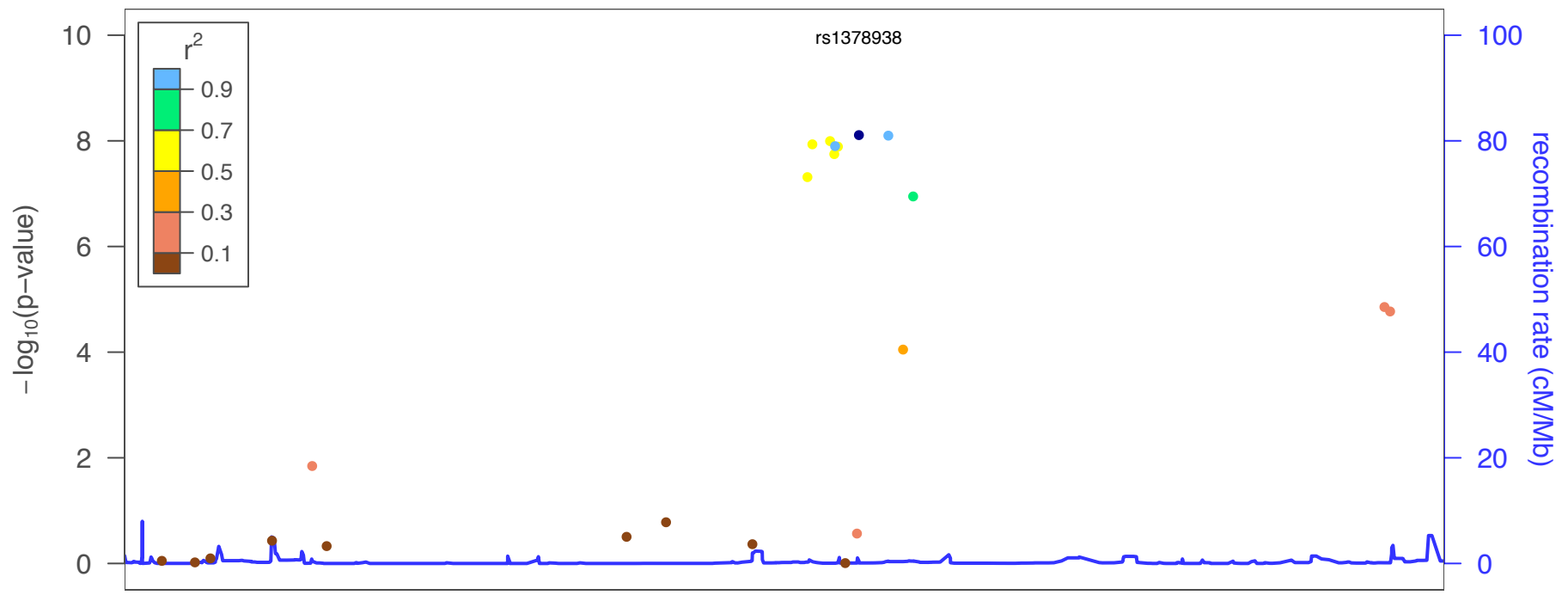
ImmunoChip



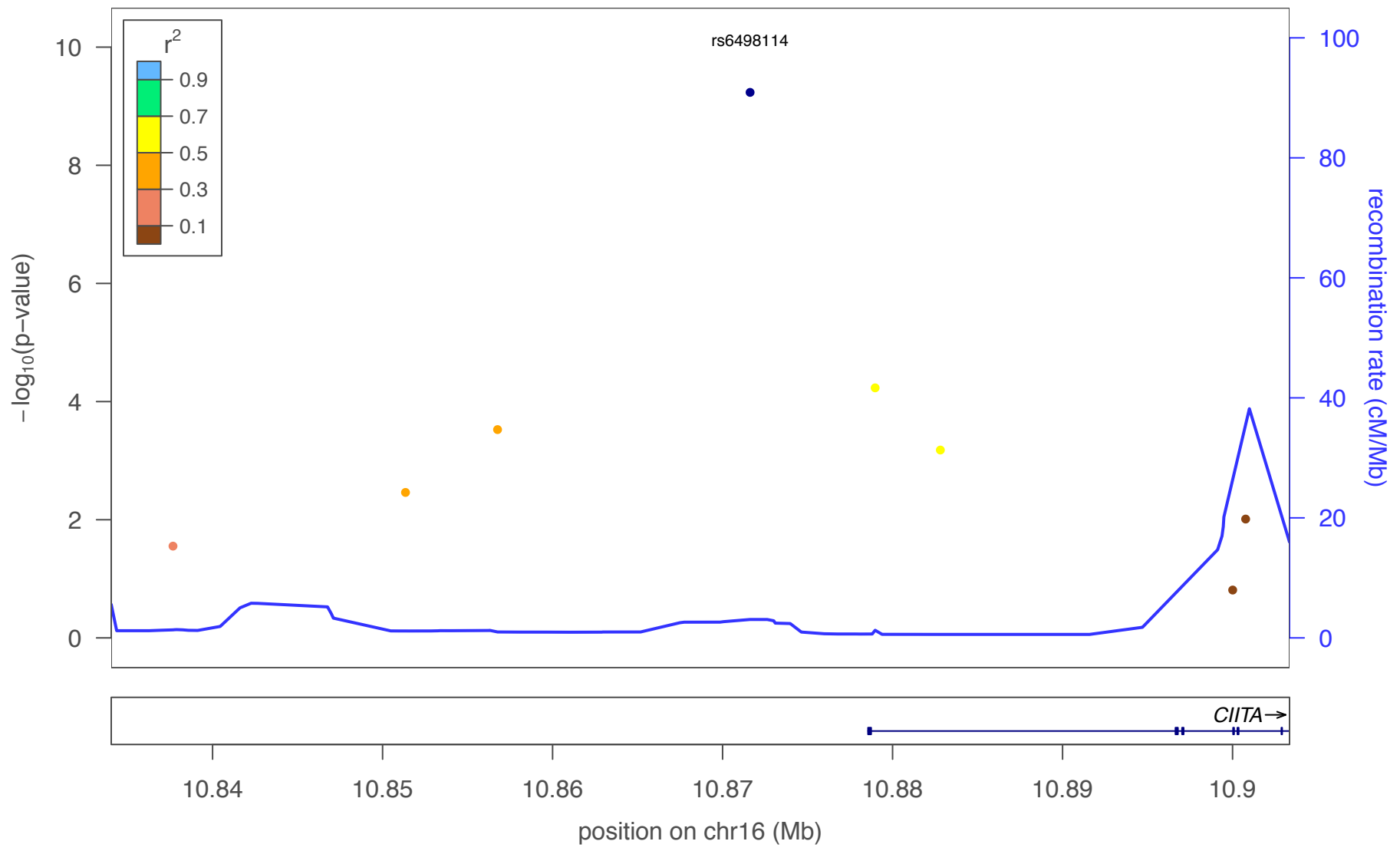
ImmunoChip 



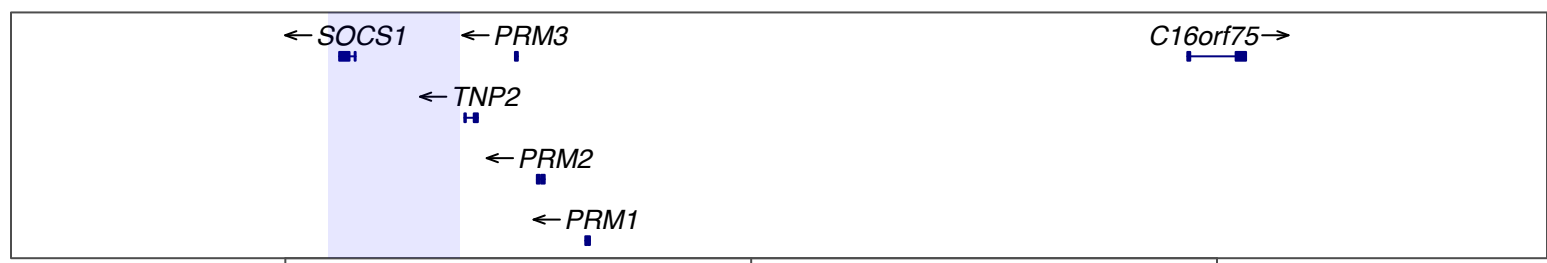
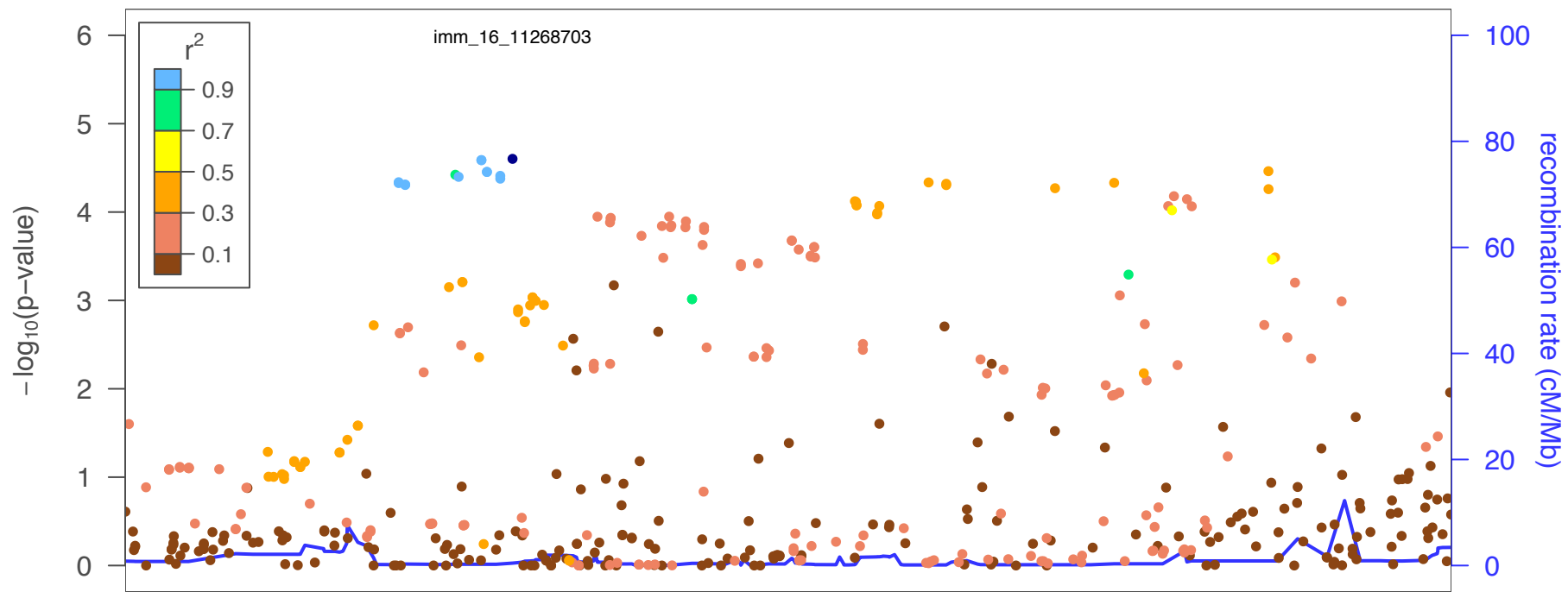
ImmunoChip



ImmunoChip



ImmunoChip



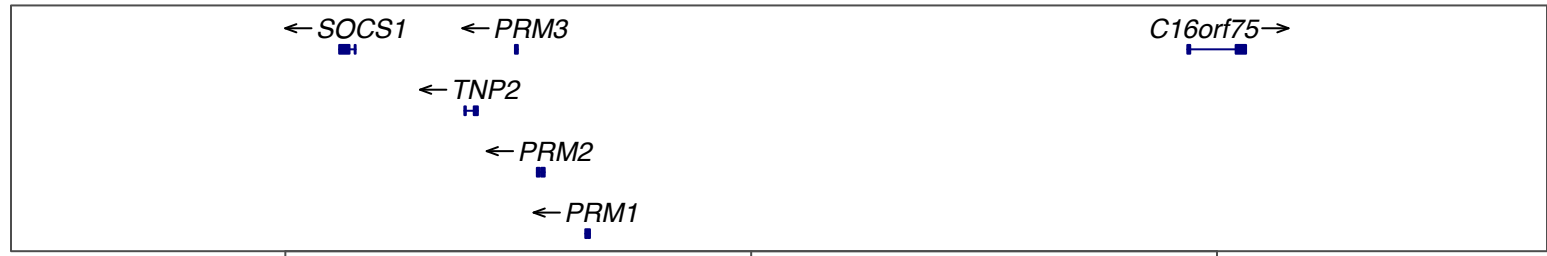
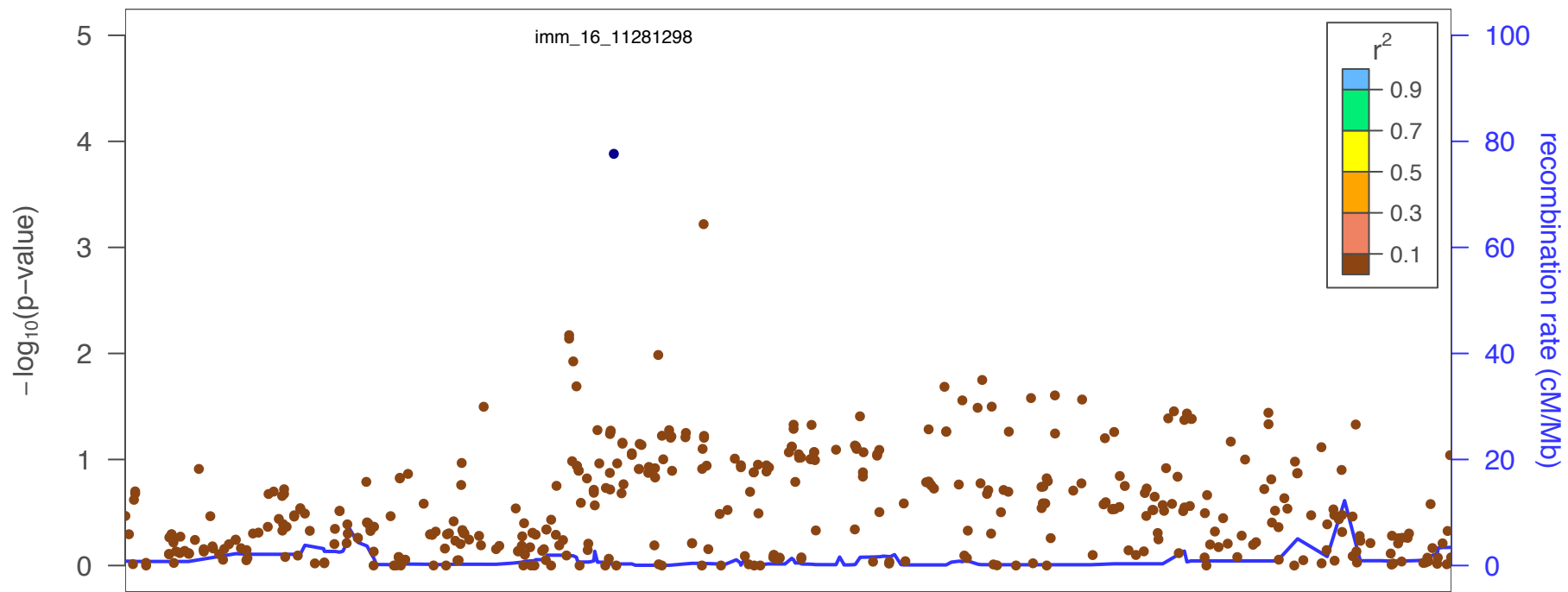
11.25

11.3

11.35

position on chr16 (Mb)

ImmunoChip



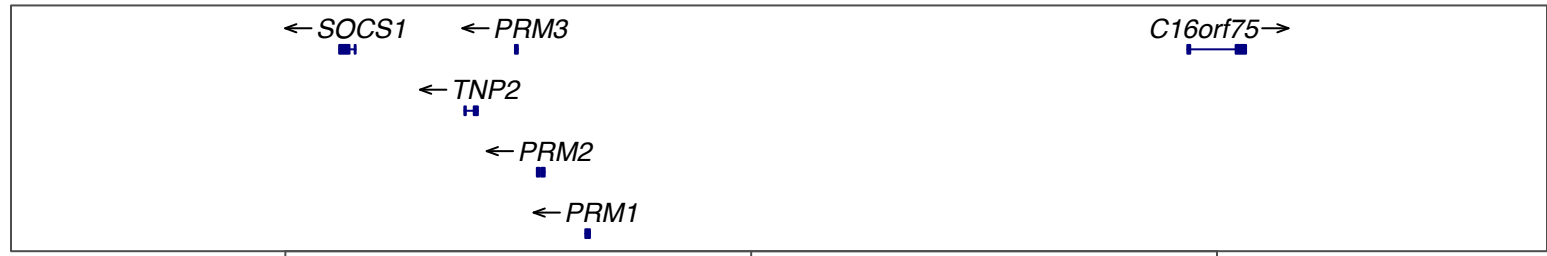
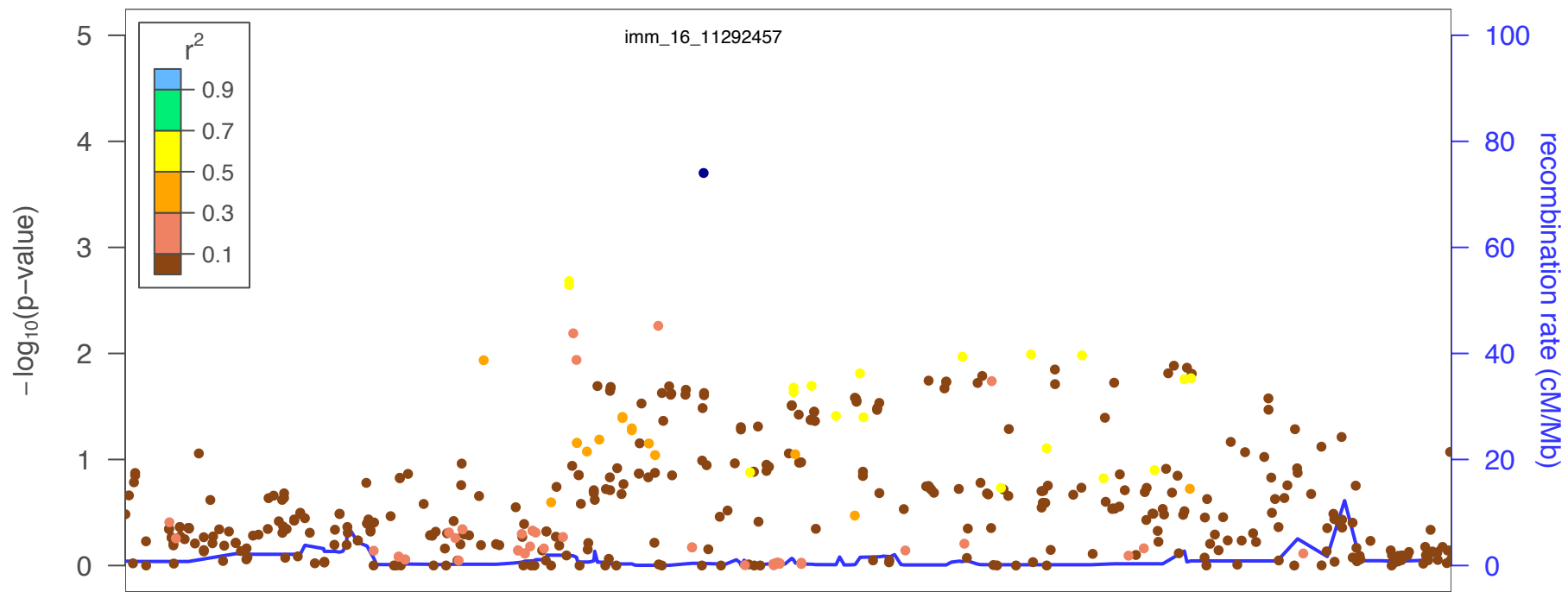
11.25

11.3

11.35

position on chr16 (Mb)

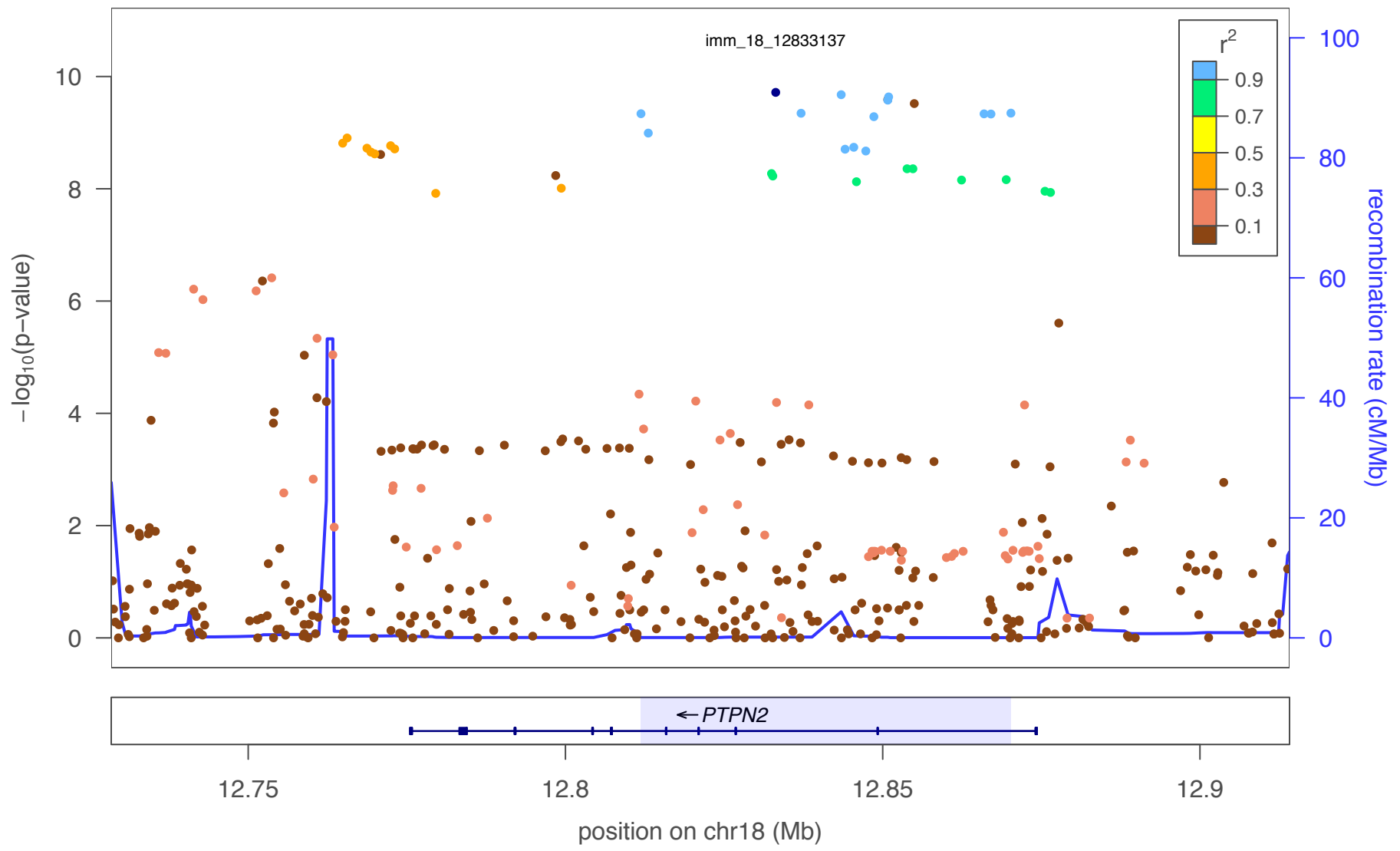
ImmunoChip



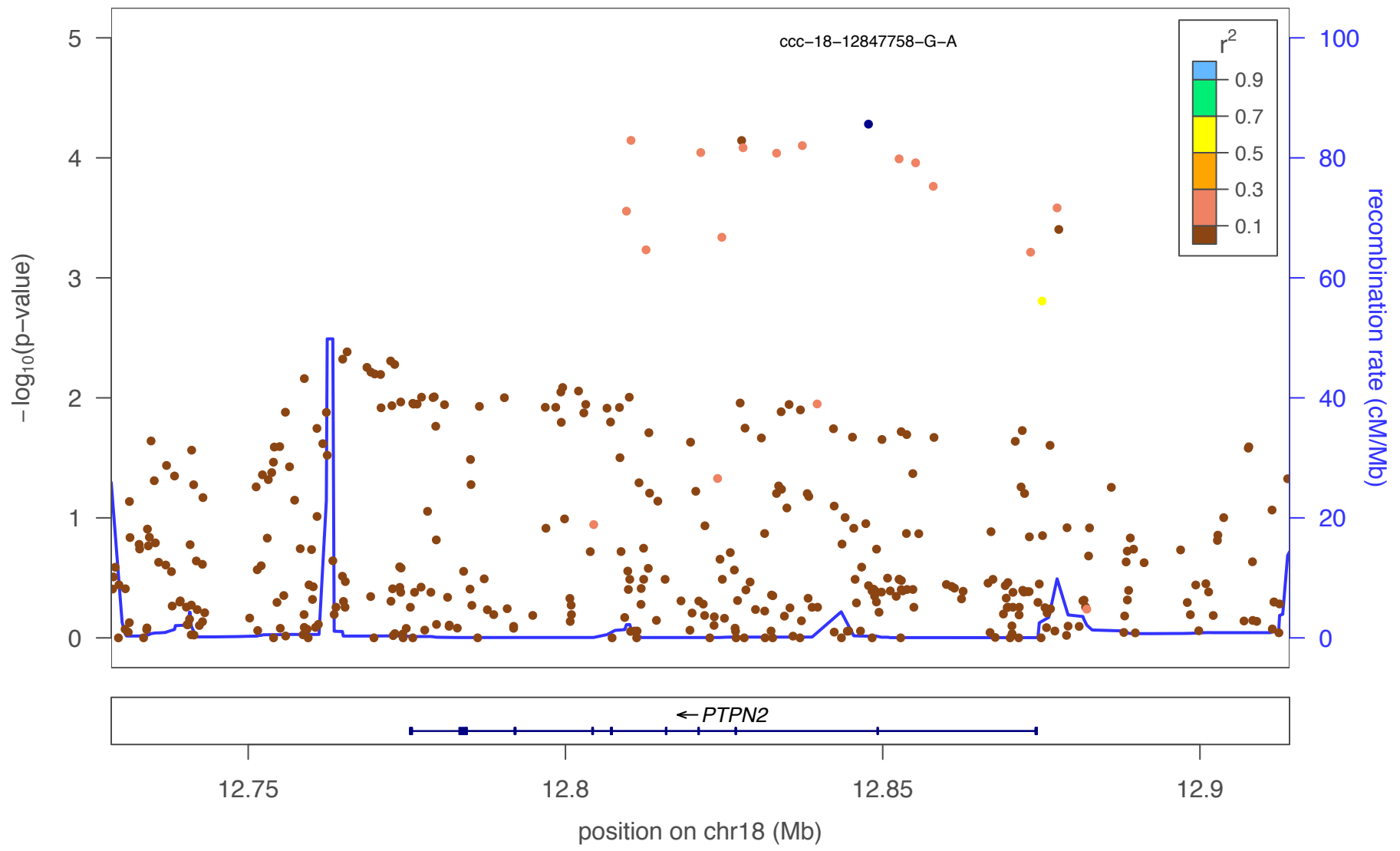
11.25 11.3 11.35

position on chr16 (Mb)

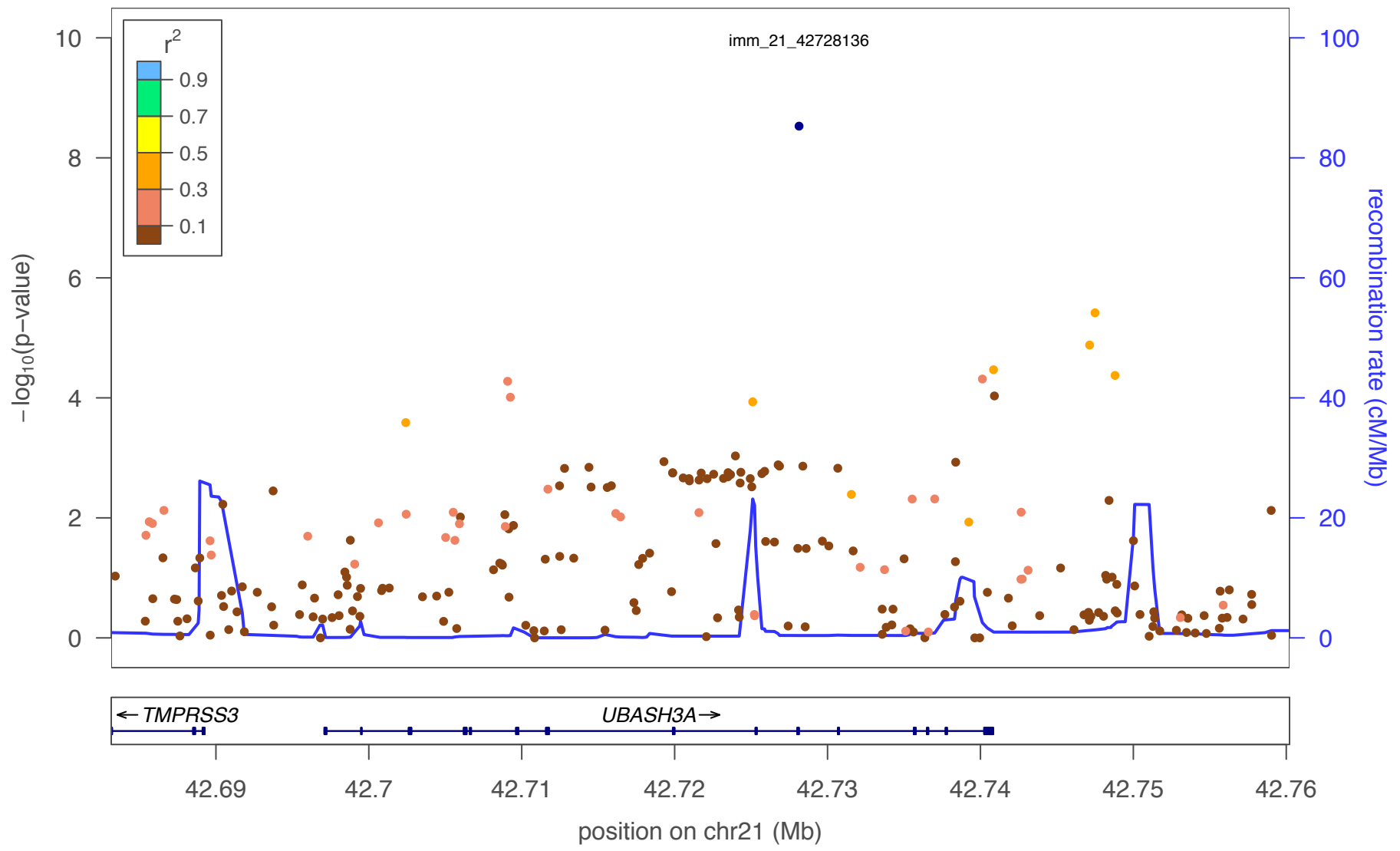
ImmunoChip



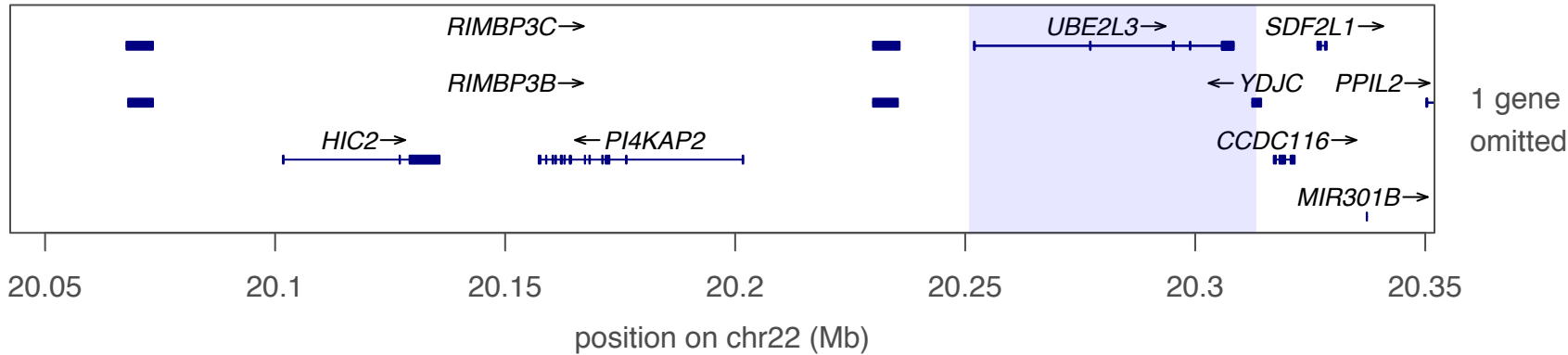
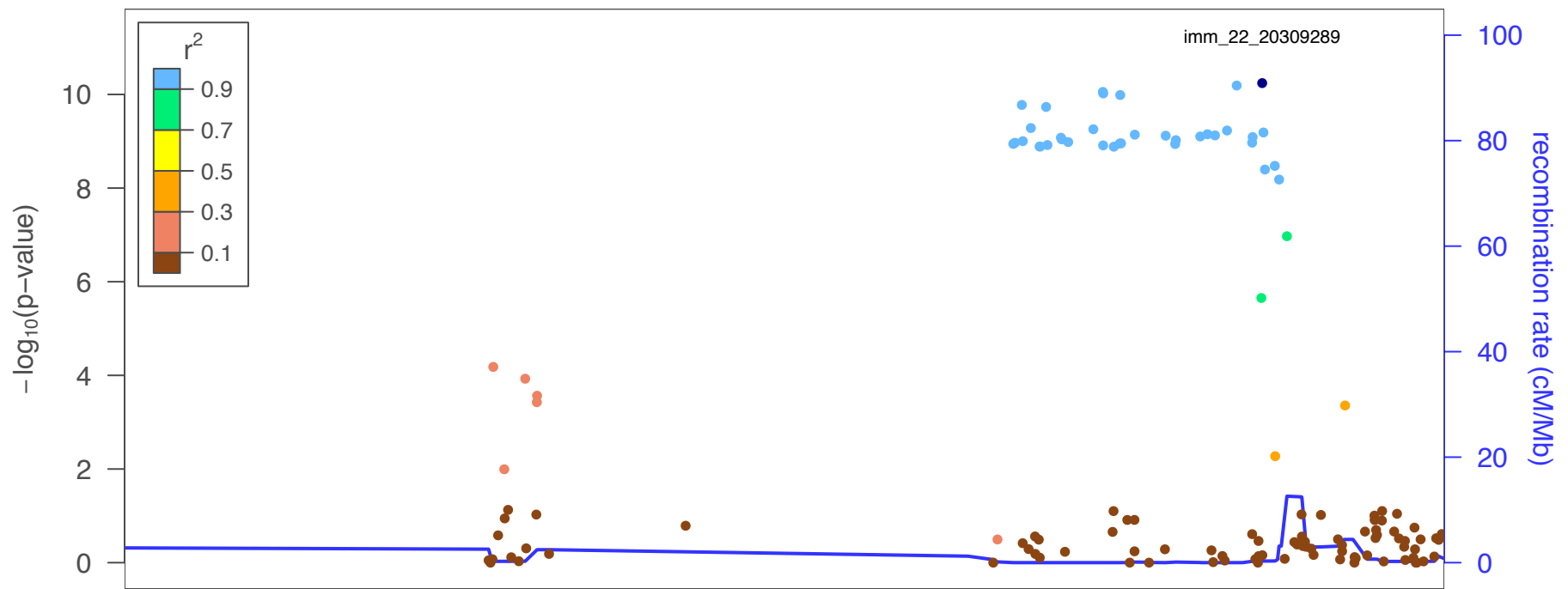
ImmunoChip



ImmunoChip | 



ImmunoChip



Supplementary Table 3

Significance thresholds for multiple logistic regression.

Chr	Top variant signal	Position (build36/hg18)	SubjectsWithGenotypes	Permutation based per-locus 5% significance threshold,		n ALL SNPs (Table 2)		n INDEPENDENT SNPs (pairwise r ² <0.05)	
				logistic regression test (P value)		Bonferroni corrected P value	Bonferroni corrected P value		
1	imm_1_2529260	2529260	24268	0.0004540	358	0.000139665	50	0.001	
1	imm_1_25162321	25162321	24265	0.001154	125	0.0004	23	0.002173913	
1	imm_1_170947654	170947654	24268	0.0003938	355	0.000140845	36	0.001388889	
1	imm_1_190808095	190808095	24260	0.0003215	181	0.000276243	39	0.001282051	
1	imm_1_199148015	199148015	24265	0.0005068	331	0.000151057	40	0.00125	
2	imm_2_61040333	61040333	24269	0.0001910	1047	4.77555E-05	80	0.000625	
2	imm_2_68499064	68499064	24266	0.0005327	357	0.000140056	36	0.001388889	
2	imm_2_102453202	102453202	24268	0.0002867	894	5.59284E-05	72	0.000694444	
2	imm_2_181716045	181716045	24269	0.0003658	898	5.56793E-05	49	0.001020408	
2	imm_2_191621279	191621279	24269	0.0008308	203	0.000246305	32	0.0015625	
2	imm_2_204318641	204318641	24269	0.0003573	642	7.78816E-05	48	0.001041667	
3	imm_3_33012725	33012725	24269	0.0005337	260	0.000192308	26	0.001923077	
3	imm_3_46353029	46353029	24267	0.0001662	1343	3.72301E-05	54	0.000925926	
3	imm_3_120605968	120605968	24262	0.0005499	372	0.000134409	33	0.001515152	
3	imm_3_161120372	161120372	24269	0.0004084	423	0.000118203	62	0.000806452	
3	imm_3_189602595	189602595	24262	0.0009655	142	0.000352113	24	0.002083333	
4	imm_4_123770564	123770564	24269	0.0001466	1294	3.86399E-05	165	0.00030303	
6	imm_6_353079	353079	24269	0.0006746	199	0.000251256	31	0.001612903	
6	imm_6_90866360	90866360	24269	0.0005366	341	0.000146628	43	0.001162791	
6	imm_6_128335255	128335255	24269	0.0003598	572	8.74126E-05	47	0.00106383	
6	imm_6_138047208	138047208	24268	0.0002504	864	5.78704E-05	59	0.000847458	
6	imm_6_159389562	159389562	24266	0.0004611	514	9.72763E-05	34	0.001470588	
7	1kg_7_37384979	37384979	24269	0.0007000	213	0.000234742	18	0.002777778	
8	imm_8_129333242	129333242	24267	0.0005021	400	0.000125	35	0.001428571	
10	imm_10_6430198	6430198	24259	0.0003170	411	0.000121655	84	0.000595238	
10	imm_10_80728033	80728033	24257	0.0006753	223	0.000224215	21	0.002380952	
11	rs10892258	118085075	24263	0.0003924	466	0.000107296	57	0.000877193	
11	imm_11_127897147	127897147	24269	0.0004242	480	0.000104167	56	0.000892857	
12	imm_12_110368991	110368991	24269	0.0003621	938	5.33049E-05	62	0.000806452	
14	imm_14_68329255	68329255	24265	0.0005669	338	0.000147929	29	0.001724138	
16	imm_16_11268703	11268703	24263	0.0004695	446	0.000112108	27	0.001851852	
18	imm_18_12833137	12833137	24269	0.0004279	411	0.000121655	57	0.000877193	
21	imm_21_42728136	42728136	24269	0.0008135	226	0.000221239	21	0.002380952	
21	imm_21_44453549	44453549	24259	0.0004120	239	0.000209205	35	0.001428571	
22	imm_22_20309289	20309289	24266	0.001491	131	0.000381679	26	0.001923077	
23	1kg_X_152901442	152901442	24266	0.0007816	88	0.000568182	29	0.001724138	

total n=24269 subjects

this method was used in the manuscript

Supplementary Table 5

ImmunoChip marker names and dbSNP names.

<u>Illumina ImmunoChip Manifest Name</u>	<u>dbSNP130 id</u>
imm_1_2529260	rs4445406
imm_1_25162321	rs72657048
imm_1_170947654	rs12068671
imm_1_171131275	rs859637
imm_1_190808095	rs1359062
imm_1_190779182	rs72734930
imm_1_199148015	rs10800746
imm_2_61040333	rs13003464
imm_2_68499064	rs10167650
imm_2_102453202	rs990171
imm_2_181716045	rs1018326
imm_2_191621279	rs6715106
imm_2_191681808	rs6752770
imm_2_191656882	rs12998748
imm_2_204318641	rs1980422
imm_2_204478299	rs34037980
imm_2_204168206	rs10207814
imm_3_33012725	rs4678523
imm_3_46353029	rs2097282
imm_3_46180690	rs7616215
imm_3_46459287	rs60215663
imm_3_120605968	rs61579022
imm_3_161120372	none
imm_3_161106253	rs1353248
imm_3_161157622	rs2561288
imm_3_189602595	rs2030519
imm_4_123770564	rs13132308
imm_4_123257745	rs62323881
imm_6_353079	rs1050976
imm_6_341321	rs12203592
imm_6_90866360	rs7753008
imm_6_128335255	rs55743914
imm_6_128335748	rs72975916
imm_6_138047208	rs17264332
imm_6_138043754	none
imm_6_159389562	rs182429
imm_6_159418255	rs1107943
1kg_7_37384979	none
imm_8_129333242	rs10808568
imm_10_6430198	rs2387397
imm_10_80728033	rs1250552
rs7104791	rs7104791
rs10892258	rs10892258
imm_11_127897147	rs61907765
imm_12_110368991	rs3184504
imm_14_68329255	rs11851414
rs1378938	rs1378938
rs6498114	rs6498114
imm_16_11268703	rs243323
imm_16_11281298	none
imm_16_11292457	rs9673543
imm_18_12833137	rs11875687
ccc-18-12847758-G-A	rs62097857
imm_21_42728136	rs1893592
imm_21_44453549	rs58911644
imm_22_20309289	rs4821124
1kg_X_152901442	rs13397

(note, where chr_position is given in the Name, these are build36/hg18 coordinates)