bioRxiv preprint doi: https://doi.org/10.1101/214973; this version posted February 1, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Title: Genome-wide Analysis of Insomnia (N=1,331,010) Identifies Novel Loci and Functional Pathways

Philip R. Jansen[1,2], Kyoko Watanabe[1], Sven Stringer[1], Nathan Skene[3], Julien Bryois[4], Anke R. Hammerschlag[1], Christiaan A. de Leeuw[1], Jeroen Benjamins[5], Ana B. Muñoz-Manchado[3], Mats Nagel[1,6], Jeanne E. Savage[1], Henning Tiemeier[2,7], Tonya White[2], The 23andMe Research Team[8], Joyce Y. Tung[8], David A. Hinds[8], Vladimir Vacic[8], Patrick F. Sullivan[4,9,10], Sophie van der Sluis[1,6], Tinca J.C. Polderman[1], August B. Smit[11], Jens Hjerling-Leffler[3], Eus J.W. Van Someren[12,13*], Danielle Posthuma[1,6*†]

Affiliations:

[1] Department of Complex Trait Genetics, Center for Neurogenomics and Cognitive Research, Amsterdam Neuroscience, VU University Amsterdam, The Netherlands
[2] Department of Child and Adolescent Psychiatry, Erasmus University Medical Center, Rotterdam, the Netherlands
[3] Laboratory of Molecular Neurobiology, Department of Medical Biochemistry and Biophysics, Karolinska Institutet, Stockholm, Sweden
[4] Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden
[5] Departments of Social, Health and Organizational Psychology, and of Experimental Psychology, Utrecht University, the Netherlands
[6] Department of Clinical Genetics, Section Complex Trait Genetics, Amsterdam Neuroscience, VU University Medical Center, Amsterdam, the Netherlands
[7] Department of Psychiatry, Erasmus University Medical Center, Rotterdam, the Netherlands
[8] 23andMe, Inc., Mountain View, CA, USA
[9] Department of Genetics, University of North Carolina, Chapel Hill, USA
[10] Department of Psychiatry, University of North Carolina, Chapel Hill, USA
[11] Department of Molecular and Cellular Neurobiology, Center for Neurogenomics and Cognitive Research, Amsterdam Neuroscience, VU University Amsterdam, Amsterdam, Netherlands
[12] Departments of Integrative Neurophysiology and Psychiatry InGeest, Amsterdam Neuroscience, VU University and Medical Center, Amsterdam, The Netherlands
[13] Department of Sleep and Cognition, Netherlands Institute for Neuroscience, an institute of the Royal Netherlands Academy of Arts and Sciences, Amsterdam, The Netherlands

[*] These authors contributed equally to this work

[†] Correspondence should be addressed to: Danielle Posthuma: Department of Complex Trait Genetics, VU University, De Boelelaan 1085, 1081 HV, Amsterdam, The Netherlands. Phone: +31 20 598 2823, Fax: +31 20 5986926, d.posthuma@vu.nl

Word count: abstract:196; main text: 3,747; Methods: 2,097
Display items: 5 (Tables 0, Figures 5)
Includes **Supplementary Information**, **Supplementary Figures** 1 and 2 in separate pdf file, and **Supplementary Tables** 1-28 in separate excel file

bioRxiv preprint doi: https://doi.org/10.1101/214973; this version posted February 1, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Abstract

Insomnia is the second-most prevalent mental disorder, with no sufficient treatment available. Despite a substantial role of genetic factors, only a handful of genes have been implicated and insight into the associated neurobiological pathways remains limited. Here, we use an unprecedented large genetic association sample (N=1,331,010) to allow detection of a substantial number of genetic variants and gain insight into biological functions, cell types and tissues involved in insomnia complaints. We identify 202 genome-wide significant loci implicating 956 genes through positional, eQTL and chromatin interaction mapping. We show involvement of the axonal part of neurons, of specific cortical and subcortical tissues, and of two specific cell-types in insomnia: striatal medium spiny neurons and hypothalamic neurons. These cell-types have been implicated previously in the regulation of reward processing, sleep and arousal in animal studies, but have never been genetically linked to insomnia in humans. We found weak genetic correlations with other sleep-related traits, but strong genetic correlations with psychiatric and metabolic traits. Mendelian randomization identified causal effects of insomnia on specific psychiatric and metabolic traits. Our findings reveal key brain areas and cells implicated in the neurobiology of insomnia and its related disorders, and provide novel targets for treatment.

bioRxiv preprint doi: https://doi.org/10.1101/214973; this version posted February 1, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Insomnia is the second-most prevalent mental disorder[1]. One third of the general population reports insomnia complaints. The diagnostic criteria for Insomnia Disorder[2] (i.e. difficulties with initiating or maintaining sleep with accompanying daytime complaints at least three times a week for at least three months, which cannot be attributed to inadequate circumstances for sleep[3]) are met by 10%, up to one third in samples of older age[4]. Insomnia contributes significantly to the risk and severity of cardiovascular, metabolic, mood, and neurodegenerative disorders[2]. Despite evidence of a considerable genetic component (heritability 38-59%[5]), only a small number of genetic loci moderating the risk of insomnia have thus far been identified[6,7]. Recent genome-wide association studies[6,7] (GWAS) for insomnia complaints (N=113,006) demonstrated its polygenic architecture and implicated three genome-wide significant (GWS) loci and seven genes. A prominent role was reported for *MEIS1*, which showed pleiotropic effects for insomnia complaints and restless legs syndrome (RLS)[7], yet the role of other genes was not unambiguously shown. We set out to substantially increase the sample size to allow the detection of more genetic risk variants for insomnia complaints, which may aid in understanding its neurobiological mechanisms. By combining data collected in the UK Biobank v2[8] (UKB; N=386,533) and 23andMe, Inc., a personal genetics company[9,10] (N=944,477), we obtained an unprecedented sample size of 1,331,010 individuals. Insomnia complaints were measured using questionnaire data, and the specific questions were validated to be good proxies of insomnia disorder, using an independent sample (The Netherlands Sleep Register, NSR)[11] in which we had access to similar question data, as well as clinical interviews assessing insomnia disorder (see **Supplementary Methods 1.1-1.3**). We find 202 risk loci for insomnia, and extensive functional in silico analyses reveal the involvement of specific tissue and cell types, whereas secondary statistical analyses reveal causal effects of insomnia on metabolic and psychiatric traits.

bioRxiv preprint doi: https://doi.org/10.1101/214973; this version posted February 1, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Meta-analysis yields 202 risk loci

UKB assessed insomnia complaints (hereafter referred to as 'insomnia') using a touchscreen device while 23andMe research participants completed online surveys. Assessment of insomnia in both samples shows high accuracy (sensitivity=84-98%; specificity=80-96%) for Insomnia Disorder (see **Supplementary Methods 1.3**). The prevalence of insomnia in the UKBv2 sample was 28.3%, 30.5% in the 23andMe sample, and 29.3% in the combined sample, in keeping with previous estimates for people with advanced age in the UK[4] and elsewhere[12,13]. Older people dominate the UKB sample (mean age=56.7, SD=8.0) and the 23andMe sample (two-thirds of the sample older than 45, one-third even older than 60 years of age). Prevalence was higher in females (34.6%) than males (24.5%), yielding an odds ratio (OR) of 1.6, close to the OR of 1.4 reported in a meta-analysis[14].

Quality control was conducted separately per sample, following standardized, stringent protocols (see **Methods**). GWAS was run separately per sample (UKB; N=386,533, 23andMe, Inc.; N=944,477) (**Extended Data Fig. 1**), and then meta-analyzed using METAL[15] by weighing SNP effects by sample size (see **Methods**). We first analyzed males and females separately (**Extended Data Fig. 2, 3**), and observed a high genetic correlation between the sexes (r_g=0.92, SE=0.02, **Extended Data Table 1**), indicating strong overlap of genetic effects. Owing to the large sample size, the r_g of 0.92 was significantly different from 1 (one-sided Wald test, $P=2.54\times10^{-6}$) suggesting a small role for sex-specific genetic risk factors, consistent with our previous report[7]. However, since sex-specific effects were relatively small, we here focus on identifying genetic effects important in both sexes and continued with the combined sample (**Supplementary Table 1**, **2** and **Supplementary Discussion 2.1** provide sex-specific results).

We observe significant polygenic signal in the GWAS (lambda inflation factor=1.808) which could not be ascribed to spurious association (LD Score intercept=1.075)[16] (**Extended Data**

bioRxiv preprint doi: https://doi.org/10.1101/214973; this version posted February 1, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Fig. 4a). Meta-analysis identified 11,990 genome-wide significant (GWS) SNPs ($P<5\times10^{-8}$), represented by 248 independent lead SNPs ($r^2<0.1$), located in 202 genomic risk loci (**Fig. 1a, Supplementary Fig. 1** and **Supplementary Table 3, 4**). All lead SNPs showed concordant signs of effect in both samples (**Extended Data Fig. 4b).** We confirm two (chr2:66,785,180 and chr5:135,393,752) out of six previously reported loci for insomnia [6,7] (**Supplementary Table 5**). Polygenic score (PGS) prediction in three randomly selected hold-out samples (N=3×3,000) estimated the current results to explain up to 2.6% of the variance in insomnia (**Fig. 1b, Extended Data Fig. 5** and **Supplementary Table 6**).

The SNP-based heritability (h^2_{SNP}) was estimated at 7.0% (SE=0.002). Partitioning the heritability by functional categories of SNPs (see **Methods**) showed the strongest enrichment of h^2_{SNP} in conserved regions (enrichment=15.8, $P=1.57\times10^{-14}$). In addition, h^2_{SNP} was enriched in common SNPs (MAF > 0.3) and depleted in more rare SNPs (MAF<0.01; **Fig. 1c** and **Supplementary Table 7**).

We used FUMA[17] to functionally annotate all 22,068 SNPs in the risk loci that were in LD ($r^2\geq 0.6$) with one of the independent significant SNPs (see **Methods**). The majority of these SNPs (76.8%) were in open chromatin regions[18] as indicated by a minimum chromatin state of 1-7 (**Fig. 1d** and **Supplementary Table 8**). In line with findings for other traits[7,19], about half of these SNPs were in intergenic (35.5%) or non-coding RNA (13.0%) regions (**Fig. 1e**), and of these, 0.72% were highly likely to have a regulatory function as indicated by a RegulomeDB Score < 2 (see **Methods**). However, of these 51.5% were located inside a protein coding gene and 0.81% were exonic. Of the 177 exonic SNPs, 71 were exonic non-synonymous (ExNS, **Supplementary Table 9**). *WDR90* included four ExNS (rs7190775, rs4984906, rs3752493, and rs3803697) all in high LD with the same independent significant SNP (rs3184470). There were two ExNS SNPs with extremely high Combined Annotation Dependent Depletion (CADD) scores[20] suggesting a strong deleterious effect on protein

bioRxiv preprint doi: https://doi.org/10.1101/214973; this version posted February 1, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

function: rs13107325 in *SLC39A8* (locus 56, $P=8.31\times10^{-16}$) with the derived allele T (MAF=0.03) associated with an increased risk of insomnia, and rs35713889 in *LAMB2 (*locus 42, $P=1.77\times10^{-7}$), where the derived allele T of rs35713889 (MAF=0.11) was also associated with an increased risk of insomnia complaints. **Supplementary Table 10** and **Supplementary Discussion 2.2** provide a detailed overview of the functional impact of all variants in the genomic risk loci.

Genes implicated in insomnia

To obtain insight into (functional) consequences of individual GWS SNPs we used FUMA[17] to apply three strategies to map associated variants to genes (see **Methods**). Positional gene-mapping aligned SNPs to 412 genes by location. Expression Quantitative Trait Loci (eQTL) gene-mapping matched cis-eQTL SNPs to 594 genes whose expression levels they influence. Chromatin interaction mapping annotated SNPs to 159 genes based on three-dimensional DNA-DNA interactions between genomic regions of the GWS SNPs and nearby or distant genes (**Supplementary Fig. 2, Supplementary Table 11** and **Supplementary Discussion 2.3**). 91 genes were mapped by all three strategies (**Supplementary Table 12**) and 336 genes were physically located outside of the risk loci but were implicated by eQTL associations (306 genes), chromatin interactions (16 genes) or both (14 genes). Several genes were implicated by GWS SNPs originating from two distinct risk loci on the same chromosome (**Fig. 2a** and **2b**): *MEIS1*, located on chromosome 2 in the strongest associated locus (locus 20), was positionally mapped by 51 SNPs and mapped by chromatin interactions in 10 tissue types including cross-loci interactions from locus 21, and is a known gene involved in insomnia[7]. *LRGUK*, located on chromosome 7 in locus 106, was positionally mapped by 22 SNPs and chromatin interactions in 3 tissue types including cross-loci interactions from locus 105. *LRGUK* was also implicated by eQTLs associations of 125 SNPs in 14 general tissue

bioRxiv preprint doi: https://doi.org/10.1101/214973; this version posted February 1, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

118 types. *LRGUK* was previously implicated in type 2 diabetes[21] and autism spectrum disorder[22]
119 - disorders with prominent insomnia - but not yet directly implicated in sleep-related
120 phenotypes, and is the most likely candidate to explain the observed association in loci 105
121 and 106.
122 Apart from linking individual associated genetic variants to genes, we conducted a genome-
123 wide gene-based association analysis (GWGAS) using MAGMA[23]. GWGAS provides
124 aggregate association *P*-values based on all variants located in a gene, and complements the
125 three FUMA mapping strategies (see **Methods**). GWGAS identified 517 associated genes
126 (**Fig. 2c** and **Supplementary Table 13**). The top gene *BTBD9* (P=8.51×10^{-23}) on
127 chromosome 6 in locus 81 was also mapped by positional and eQTL mapping (tissue type:
128 left ventricle of the heart), and is part of a pathway regulating circadian rhythms. *BTBD9* has
129 been associated with RLS, periodic limb movement disorder[24,25] and Tourette Syndrome[26].
130 Involvement in sleep regulation was shown in *Drosophila*[27], and mouse mutants show
131 fragmented sleep[28] and increased levels of dynamin 1[29], a protein that mediates the increased
132 sleep onset latency following pre-sleep arousal[30].
133 Of the 517 MAGMA-based associated genes, 222 were outside of the GWAS risk loci, and
134 309 were also mapped by FUMA. In total, 956 unique genes were mapped by at least one of
135 the three FUMA gene mapping strategies or by MAGMA (**Extended Data Fig. 6**). Of these,
136 *MEIS1*, *MED27*, *IPO7* and *ACBD4* confirmed previous results[6,7] (**Supplementary Table 14**).
137 Sixty-two genes were implicated by all four mapping strategies indicating that apart from a
138 GWS gene-based *P*-value, there were (i) GWS SNPs located inside these genes, (ii) GWS
139 SNPs associated with differential expression of these genes and (iii) GWS SNPs that were
140 involved in genomic regions interacting with these genes. We note that genes that were
141 indicated by positional mapping and GWS gene-based *P*-values, but not via eQTL or
142 chromatin interaction mapping (N=54 genes), may be of equal importance, yet are more

bioRxiv preprint doi: https://doi.org/10.1101/214973; this version posted February 1, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

likely to exert their influence on insomnia via structural changes in the gene products (i.e. at the protein level) and not via quantitative changes in the availability of the gene products.

Implicated pathways, tissues and cell-types

To test whether GWS genes converged in functional gene-sets and pathways, we conducted gene-set analyses using MAGMA (see **Methods**). We tested associations of 7,473 gene-sets: 7,246 sets derived from the MsigDB[31], gene expression values from 54 tissues from the GTEx database[32], and cell-specific gene expression in 173 types of brain cells (**Fig. 2d**, **Supplementary Table 15**). Competitive testing was used and a Bonferroni corrected threshold of $P<6.7\times10^{-6}$ (0.05/7,473) to correct for multiple testing. Of the MsigDB gene-sets, three Gene Ontology (GO) gene-sets survived multiple testing: GO:*locomotory behavior* ($P=8.95\times10^{-7}$), GO:*behavior* ($P=5.23\times10^{-6}$), and GO:*axon part* ($P=4.25\times10^{-6}$). Twelve genes (*LRRK2*, *CRH*, *DLG4*, *DNM1*, *DRD1*, *DRD2*, *DRD4*, *GRIN1*, *NTSR1*, *SNCA*, *CNTN2*, and *CALB1*) were included in all of these gene-sets and two of these (*SNCA*, *DNM1*) had a GWS gene-based *P*-value (**Supplementary Table 16**). *SNCA* encodes alpha-synuclein and has been implicated in REM sleep behavior disorder[33] and Parkinson's disease[34]. Altered expression in mice changes sleep and wake EEG spectra[35] along the same dimensions that have been implicated in insomnia disorder[36]. *DNM1* encodes the synaptic neuronal protein dynamin 1, which is increased in *BTBD9* mutant mice[29] and mediates the sleep-disruptive effect of pre-sleep arousal (see above; *BTBD9* is the top associated gene). Conditional gene-set analyses suggested that the association with the gene-set *behavior* is almost completely explained by the association of *locomotory behavior*, and that the effect of *axon part* is independent of this (**Supplementary Discussion 2.4**). GO:*locomotory behavior* includes 175 genes involved in stimulus-evoked movement[37]. This set included 16 GWS genes: *BTBD9*, *MEIS1*, *DAB1*, *SNCA*, *GNAO1 ATP2B2*, *NEGR1*, *SLC4A10*, *GIP*, *DNM1*, *GPRC5B*, *GRM5*,

bioRxiv preprint doi: https://doi.org/10.1101/214973; this version posted February 1, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

NRG1, PARK2, TAL1, and *OXR1*). GO*:axon part* reflects a very general cellular component representing 219 genes, of which 14 were GWS (*KIF3B, SNCA, GRIA1, CDH8, ROBO2, DNM1, RANGAP1, GABBR1, P2RX3, NRG1, POLG, DAG1, NFASC,* and *CALB2*).

Tissue specific gene-set analyses showed strong enrichment of genetic signal in genes expressed in the brain. Correcting for overall expression, four specific brain tissues reached the threshold for significance: overall cerebral cortex ($P=3.68\times10^{-6}$), Brodmann area 9 (BA9) of frontal cortex ($P=5.04\times10^{-7}$), BA24 of the anterior cingulate cortex ($P=3.25\times10^{-6}$), and cerebellar hemisphere ($P=5.93\times10^{-6}$)[1]. Several other brain tissues also showed strong enrichment just below threshold, including three striatal basal ganglia (BG) structures (nucleus accumbens, caudate nucleus, putamen). To test whether genes expressed in all three BG structures together would show significant enrichment of low *P*-values, we used the first principal component (BG_{PC}) of these BG structures and found significant enrichment ($P=8.33\times10^{-8}$). When conditioning the three top cortical structures on the BG_{PC}, they were no longer significantly associated after multiple testing correction (minimum $P=0.03$), which was expected given that the BG_{PC} correlated strongly with gene-expression in cortical (and other) areas ($r>0.96$). Similar results were obtained vice versa, i.e. using the first principal component of all cortical areas and conditioning the three BG structures on this resulted in no evidence of enrichment of low *P*-values for BG structures (minimum $P=0.53$). These results show that (i) genes expressed in brain are important in insomnia, (ii) genes expressed in cortical areas are more strongly associated than genes expressed in BG, (iii) there is a strong correlation between gene expression patterns across brain tissues, which suggests involvement of general cellular signatures more than specific brain tissue structures.

[1] We caution that only a limited set of brain tissues were included and thus we cannot rule out associations with many important areas such as pons, midbrain or thalamus based on this analysis.

bioRxiv preprint doi: https://doi.org/10.1101/214973; this version posted February 1, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Brain cell type-specific gene-set analyses was first carried out on 24 broad cell-type categories. Cell type-specific gene expression was quantified using single cell RNA-sequencing of disassociated cells from somatosensory cortex, hippocampus, hypothalamus, striatum and midbrain from mouse (see **Methods**), which closely resembles gene-expression in humans[38]. Results indicated that genes expressed specifically in the medium spiny neurons (MSN, $P=4.83\times10^{-7}$) were associated with insomnia, and no other broad cell-types specific gene-set survived our strict threshold of $P<6.7\times10^{-6}$. MSNs represent 95% of neurons within the human striatum, which is one of the four major nuclei of the subcortical BG. Specifically, the striatum consists of the ventral (nucleus accumbens and olfactory tubercle) and dorsal (caudate nucleus and putamen) subdivisions. The association with MSNs thus likely explains the observed association of the BG striatal structures (nucleus accumbens, caudate nucleus, putamen).

Using broad cell classes risks not detecting associations that are specific to distinctive yet rare cell types; to account for this we then tested 149 specific brain cell-type categories, and found significant associations with 7 specific cell types: medio-lateral neuroblasts type 3, 4 and 5 ($P=2.36\times10^{-6}$, $P=1.88\times10^{-6}$, and $P=1.87\times10^{-6}$), D2 type medium spiny neurons ($P=2.12\times10^{-6}$), claustrum pyramidal neurons ($P=3.09\times10^{-6}$), hypothalamic Vglut2 Morn4 Prrc2a neurons ($P=4.36\times10^{-6}$), and hypothalamic Vglut2 Hcn16430411 K18 Rik neurons ($P=4.98\times10^{-6}$), known to have the densest number of melatonin receptors. These results suggest a role of distinct mature and developing cell types in the midbrain and hypothalamus. The hypothalamus contains multiple nuclei that are key to the control of sleep and arousal, including the suprachiasmatic nucleus (SCN) that accommodates the biological clock of the brain [39].

bioRxiv preprint doi: https://doi.org/10.1101/214973; this version posted February 1, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Low genetic overlap with sleep traits

Other sleep-related traits may easily be confounded with specific symptoms of insomnia, like early morning awakening, difficulties maintaining sleep, and daytime sleepiness. The most recent genome-wide studies for other sleep-related traits included 59,128 to 128,266 individuals, and assessed genetic effects on morningness[6,40,41] (i.e. being a morning person), sleep duration [6,41], and daytime sleepiness/dozing [41]. Using increased sample sizes for each of these sleep-related traits (max N=434,835), we here investigated to what extent insomnia and other sleep-related traits are genetically distinct or overlapping. We performed GWAS analyses for the following six sleep-related traits: morningness, sleep duration, ease of getting up in the morning, naps during the day, daytime dozing, and snoring (**Supplementary Methods 1.1-1.2, Extended Data Fig. 7, 9**). Of the 202 risk loci for insomnia, 39 were also GWS in at least one of the other sleep-related traits (**Fig. 3, Supplementary Table 17)**. The strongest overlap in loci was found with sleep duration, with 14 out of 49 sleep duration loci overlapping with insomnia. Insomnia showed the highest genetic correlation with sleep duration (−0.47, SE=0.02; **Supplementary Table 18**) compared to other sleep-related traits, which was not surprising given that insomnia also shared the most risk loci with sleep duration (See further discussion sleep phenotypes in **Supplementary Discussion 2.5**).

Gene-mapping of SNP associations of sleep-related traits resulted in 973 unique genes (**Extended Data Fig. 9, Supplementary Table 19-23**). Gene-based analysis showed that of the 517 GWS genes for insomnia, 120 were GWS in at least one of the other sleep-related traits, and one gene (*RBFOX1*) was GWS in all except napping and dozing (**Supplementary Table 24**). The largest proportion of overlap in GWS genes for insomnia was again with sleep duration, with 37 of the 135 (27%) GWS genes for sleep duration also GWS for insomnia. There was overlap in tissue enrichment in cortical structures and basal ganglia between insomnia and both morningness and sleep duration. On the single cell level, the

bioRxiv preprint doi: https://doi.org/10.1101/214973; this version posted February 1, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

medium spiny neurons were also implicated for morningness and sleep duration, but not for the other sleep-related traits (**Supplementary Table 25**). Taken together, these results suggest that at a genetic level, insomnia shows partial overlap with sleep duration, but minimal overlap with other sleep-related traits. Consistent short sleep across nights occurs only in a minor part of insomnia patients, even in a clinical sample[42].

Strong overlap between insomnia and psychiatric and metabolic traits

We confirm previously reported genetic correlations between insomnia and neuropsychiatric and metabolic traits[6,7] (**Supplementary Table 26**)**,** and also identify several GWS SNPs for insomnia that have previously been associated with these traits (**Supplementary Table 27**). The strongest correlations were with depressive symptoms (r_g=0.64, SE=0.04 P=1.21×10^{-71}), followed by anxiety disorder (r_g=0.56, SE=0.11 P=1.40×10^{-7}), subjective well-being (r_g=−0.51, SE=0.03 P=4.93×10^{-52}), major depression (r_g=0.50, SE=0.07 P=8.08×10^{-12}) and neuroticism (r_g=0.48, SE=0.02 P=8.72×10^{-80}). Genetic correlations with metabolic traits ranged between 0.09-0.20. The genetic correlations between insomnia and psychiatric traits were also stronger than the correlations between insomnia and the other sleep-related traits. Since a similar high reliability has been reported for both sleep and psychiatric phenotypes, the findings suggest that genetically insomnia more closely resembles neuropsychiatric traits than it resembles other sleep-related traits (**Fig. 4**). To infer directional associations between insomnia and these correlated traits, we performed bidirectional Multi-SNP Mendelian Randomization (MR) analysis[43] (see **Methods**)**.** Results support a direct risk effect of insomnia on metabolic syndrome phenotypes including BMI (b_{xy}=0.36, SE=0.05, P=1.25×10-^{12}) type 2 diabetes (b_{xy}=0.62, SE=0.11, P=2.29×10^{-8}), and coronary artery disease (b_{xy}=0.61, SE=0.09, P=2.88×10^{-12}). In addition, insomnia was bidirectionally associated with educational attainment, with a stronger effect from insomnia on educational attainment

bioRxiv preprint doi: https://doi.org/10.1101/214973; this version posted February 1, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

(b_{xy}=−0.32, SE=0.02, P=1.68×10^{-77}) (i.e. a higher risk for insomnia leads to lower educational attainment) than vice versa (b_{xy}=−0.10, SE=0.01, P=2.27×10^{-23}), the same pattern was observed for intelligence. We also found risk effects of insomnia on several psychiatric traits, including schizophrenia (b_{xy}=0.68, SE=0.10, P=5.12×10^{-11}), ADHD (b_{xy}=0.77, SE=0.06, P=2.50×10^{-45}), neuroticism (b_{xy}=0.46, SE=0.03, P=3.92×10^{-53}) and anxiety disorder (b_{xy}=0.47, SE=0.10, P=4.11×10^{-6}), with no evidence of large reverse effects, except for a small risk effect of neuroticism on insomnia (b_{xy}=0.09, SE=0.02, P=1.24×10^{-6}) and depressive symptoms (b_{xy}=0.09, SE=0.02, P=1.24×10^{-6})[2]. Overall, there was only a small proportion of SNPs showing pleiotropy between insomnia and other traits (**Supplementary Table 28** and **Supplementary Discussion 2.6**).

Discussion

In the largest GWAS study to date of 1,331,010 participants we identified 202 genomic risk loci for insomnia. Using extensive functional annotation of associated genetic variants, we demonstrated that the genetic component of insomnia points towards a role of genes involved in locomotory behavior, and genes expressed in specific cell types from the claustrum, hypothalamus and striatum, and specifically in MSNs (**Fig. 5**). MSNs are GABAergic inhibitory cells and represent 95% of neurons in the human striatum, one of the four major nuclei of the BG (for reviews, see [44-46]). MSNs receive massive excitatory glutamatergic input from the cerebral cortex and the thalamus, and are targets of dopamine neurons in substantia nigra and the ventral tegmental area. In addition, they receive inhibitory inputs from striatal GABAergic interneurons. MSNs themselves are GABAergic output neurons with exceptionally long projections to globus pallidus (GP), substantia nigra and ventral

[2] We do note that for major depression the reverse MR could not be carried out due to an insufficient number of SNPs with a low *P*-value

bioRxiv preprint doi: https://doi.org/10.1101/214973; this version posted February 1, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

pallidum, and control the activity of thalamocortical neurons. Previous studies during the natural sleep-wake cycle, *in vitro*, and from anesthetized *in vivo* preparations have shown that MSNs show fast, synchronized cyclic firing, i.e. the so-called Up and Down states, during slow-wave sleep and irregular pattern of action potentials during wakefulness. In fact, MSNs were the first neurons in which the Up and Down states characteristic of slow wave sleep were described[47]. Cell body-specific striatal lesions of the rostral striatum induce a profound sleep fragmentation, which is most characteristic of insomnia. A role for BG in sleep regulation is also suggested by the high prevalence of insomnia in neurodegenerative disorders, such as Parkinson's Disease and Huntington's disease in which the BG are affected. Vetrivelan et al.[44] proposes a cortex-striatum-$GP_{external}$-cortex network involved in the control of sleep–wake behavior and cortical activation, in which midbrain dopamine disinhibits the $GP_{external}$ and promotes sleep through activation of D2 receptors in this network. Furthermore, brain imaging studies have suggested the caudate nucleus of the striatum as a key node in the neuronal network imbalance of insomnia[48], and also reported abnormal function in the cortical areas we found to be most enriched (BA9[49], BA24[50]). Our results support the involvement of the striato-cortical network in insomnia, by showing enrichment of risk genes for insomnia in cortical areas as well as the striatum, and specifically in MSNs. We recently showed that, along with several other cell types, MSNs also mediate the risk for mood disorders[51] and schizophrenia[38]. MSNs are strongly implicated in reward processing and future work could address whether the genetic overlap between insomnia and mood disorders is mediated by gene function in MSNs.

Our results also showed enrichment of insomnia genes in pyramidal neurons of the claustrum. This subcortical brain region is structurally closely associated with the amygdala and has been implicated in salience coding of incoming stimuli and binding of multisensory information into conscious percepts[52]. These functions are highly relevant to insomnia,

bioRxiv preprint doi: https://doi.org/10.1101/214973; this version posted February 1, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

because the disorder is characterized by increased processing of incoming stimuli[53] and by ongoing consciousness even during sleep, a phenomenon known as sleep state misperception[54]. We also found enrichment of insomnia genes in mediolateral neuroblasts from the embryonic midbrain and in two hypothalamic cell types. The role of the mediolateral neuroblasts is less clear; although they were obtained from the embryonic midbrain, it is at present unknown what type of mature neurons they differentiate into. We note that the midbrain is similar on a bulk transcriptomic level to the pons[55], and lacking cells from that region we cannot conclusively say that midbrain cell-types are enriched.

The current findings provide novel insight into the causal mechanism of insomnia, implicating specific cell types, brain areas and biological functions. These findings are starting points for the development of new therapeutic targets for insomnia and may also provide valuable insights for other, genetically related disorders.

bioRxiv preprint doi: https://doi.org/10.1101/214973; this version posted February 1, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Methods:

Meta-analysis

A meta-analysis on the GWAS results of insomnia and morningness in UKB and 23andMe cohorts was performed using fixed-effects meta-analysis METAL[15], using SNP *P*-values weighted by sample size. To investigate sex-specific genetic effects, we ran the meta-analysis between UKB and 23andMe datasets for males and females separately.

Genomic risk loci definition

We used FUMA[17] (http://fuma.ctglab.nl/), an online platform for functional mapping and annotation of genetic variants, to define genomic risk loci and obtain functional information of relevant SNPs in these loci. FUMA provides comprehensive annotation information by combining several external data sources. We first identified *independent significant SNPs* that have a genome-wide significant *P*-value ($<5\times10^{-8}$) and are independent from each other at $r^2<0.6$. These SNPs were further represented by *lead SNPs*, which are a subset of the independent significant SNPs that are in approximate linkage equilibrium with each other at $r^2<0.1$. We then defined associated *genomic risk loci* by merging any physically overlapping lead SNPs (linkage disequilibrium [LD] blocks <250kb apart). Borders of the genomic risk loci were defined by identifying all SNPs in LD (r^2≥0.6) with one of the independent significant SNPs in the locus, and the region containing all these *candidate SNPs* was considered to be a single independent genomic risk locus. LD information was calculated using the UK Biobank genotype data as a reference. Risk loci were defined based on evidence from independent significant SNPs that were available in both 23andMe and UKB.

We note that SNPs that were GWS but only available in the 23andMe dataset were not included when defining genomic risk loci and were not included in any follow-up annotations or analyses, because there was no external replication in the UKB sample. If such SNPs were

bioRxiv preprint doi: https://doi.org/10.1101/214973; this version posted February 1, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

located in a risk locus, they are displayed in Locuszoom plots (grey, as there is no LD information in UKB). When risk loci contained GWS SNPs based solely on 23andMe, we did not count that risk locus, as there were no other SNPs available in both samples that supported these GWS SNPs.

Gene-based analysis

SNP-based *P*-values from the meta-analysis were used as input for the gene-based genome-wide association analysis (GWGAS). 18,182 to 18,185 protein-coding genes (each containing at least one SNP in the GWAS, the total number of tested genes can thus be slightly different across phenotypes) from the NCBI 37.3 gene definitions were used as basis for GWGAS in MAGMA [23]. Bonferroni correction was applied to correct for multiple testing ($P<2.73\times10^{-6}$).

Gene-set analysis

Results from the GWGAS analyses were used to test for association in three types of 7,473 predefined gene-sets:

1. 7,246 curated gene-sets representing known biological and metabolic pathways derived from 9 data resources, catalogued by and obtained from the MsigDB version 6.0[56] (http://software.broadinstitute.org/gsea/msigdb/collections.jsp)
2. Gene expression values from 54 (53 + 1 calculated 1st PC of three tissue subtypes) tissues obtained from GTEx[32], log2 transformed with pseudocount 1 after winsorization at 50 and averaged per tissue
3. Cell-type specific expression in 173 types of brain cells (24 broad categories of cell types, 'level 1' and 129 specific categories of cell types 'level 2'), which were calculated following the method described in [38]. Briefly, brain cell-type expression data was drawn from single-cell RNA sequencing data from mouse brains. For each

bioRxiv preprint doi: https://doi.org/10.1101/214973; this version posted February 1, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

gene, the value for each cell-type was calculated by dividing the mean Unique Molecular Identifier (UMI) counts for the given cell type by the summed mean UMI counts across all cell types. Single-cell gene-sets were derived by grouping genes into 40 equal bins based on specificity of expression. Mouse cell gene-expression was shown to closely approximate gene-expression in post-mortem human tissue[38].

These gene-sets were tested using MAGMA. We computed competitive *P*-values, which represent the test of association for a specific gene-set compared with genes not in the gene-set to correct for baseline level of genetic association in the data[57]. The Bonferroni-corrected significance threshold was 0.05/7,473 gene-sets=6.7×10^{-6}. Conditional analyses were performed as a follow-up using MAGMA to test whether each significant association observed was independent of all others. The association between each gene-set in each of the three categories was tested conditional on the most strongly associated set, and then, if any substantial (P<0.05/number of gene-sets) associations remained, by conditioning on the first and second most strongly associated set, and so on until no associations remained. Gene-sets that retained their association after correcting for other sets were considered to represent independent signals. We note that this is not a test of association per se, but rather a strategy to identify, among gene-sets with known significant associations and overlap in genes, which set (s) are responsible for driving the observed association.

SNP-based heritability and genetic correlation

LD Score regression[16] was used to estimate genomic inflation and SNP-based heritability of the phenotypes, and to estimate the cross-cohort genetic correlations. Pre-calculated LD scores from the 1000 Genomes European reference population were obtained from https://data.broadinstitute.org/alkesgroup/LDSCORE/.

bioRxiv preprint doi: https://doi.org/10.1101/214973; this version posted February 1, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

400 **Genetic correlations**

401 Genetic correlations between sleep-related traits, and between sleep-related traits and
402 previously published GWAS studies of sufficient sample size were calculated using LD Score
403 regression on HapMap3 SNPs only. Genetic correlations were corrected for multiple testing
404 based on the total number of correlations (between 6 sleep-related phenotypes and 27
405 previous GWAS studies) by applying a Bonferroni corrected threshold of
406 ($P<0.05/33=1.51\times10^{-3}$).

407

408 **Stratified heritability**

409 To test whether specific categories of SNP annotations were enriched for heritability, we
410 partitioned SNP heritability for binary annotations using stratified LD score regression[58].
411 Heritability enrichment was calculated as the proportion of heritability explained by a SNP
412 category divided by the proportion of SNPs that are in that category. Partitioned heritability
413 was computed by 28 functional annotation categories, by minor allele frequency (MAF) in
414 six percentile bins and by 22 chromosomes. Annotations for binary categories of functional
415 genomic characteristics (e.g. coding or regulatory regions) were obtained from the LD score
416 website (https://github.com/bulik/ldsc). The Bonferroni-corrected significance threshold for
417 56 annotations was set at: $P<0.05/56=8.93\times10^{-4}$.

418

419 **Functional annotation of SNPs**

420 Functional annotation of SNPs implicated in the meta-analysis was performed using
421 FUMA[17]. We selected all candidate SNPs in genomic risk loci having an r^2 0.6 with one of
422 the independent significant SNPs (see above), a *P*-value ($P<1\times10^{-5}$), a MAF>0.0001 for
423 annotations, and availability in both UKB and 23andMe datasets. Functional consequences

bioRxiv preprint doi: https://doi.org/10.1101/214973; this version posted February 1, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

for these SNPs were obtained by matching SNPs' chromosome, base-pair position, and reference and alternate alleles to databases containing known functional annotations, including ANNOVAR[59] categories, Combined Annotation Dependent Depletion (CADD) scores, RegulomeDB[20] (RDB) scores, and chromatin states[60]. ANNOVAR categories identify the SNP's genic position (e.g. intron, exon, intergenic) and associated function. CADD scores predict how deleterious the effect of a SNP is likely to be for a protein structure/function, with higher scores referring to higher deleteriousness. A CADD score above 12.37 is considered to be potentially pathogenic[20]. The RegulomeDB score is a categorical score based on information from expression quantitative trait loci (eQTLs) and chromatin marks, ranging from 1a to 7 with lower scores indicating an increased likelihood of having a regulatory function. Scores are as follows: 1a=eQTL + Transciption Factor (TF) binding + matched TF motif + matched DNase Footprint + DNase peak; 1b=eQTL + TF binding + any motif + DNase Footprint + DNase peak; 1c=eQTL + TF binding + matched TF motif + DNase peak; 1d=eQTL + TF binding + any motif + DNase peak; 1e=eQTL + TF binding + matched TF motif; 1f=eQTL + TF binding / DNase peak; 2a=TF binding + matched TF motif + matched DNase Footprint + DNase peak; 2b=TF binding + any motif + DNase Footprint + DNase peak; 2c=TF binding + matched TF motif + DNase peak; 3a=TF binding + any motif + DNase peak; 3b=TF binding + matched TF motif; 4=TF binding + DNase peak; 5=TF binding or DNase peak; 6=other;7=Not available. The chromatin state represents the accessibility of genomic regions (every 200bp) with 15 categorical states predicted by a hidden Markov model based on 5 chromatin marks for 127 epigenomes in the Roadmap Epigenomics Project[61]. A lower state indicates higher accessibility, with states 1-7 referring to open chromatin states. We annotated the minimum chromatin state across tissues to SNPs. The 15-core chromatin states as suggested by Roadmap are as follows: 1=Active Transcription Start Site (TSS); 2=Flanking Active TSS; 3=Transcription at gene 5' and 3';

bioRxiv preprint doi: https://doi.org/10.1101/214973; this version posted February 1, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

449 4=Strong transcription; 5= Weak Transcription; 6=Genic enhancers; 7=Enhancers; 8=Zinc
450 finger genes & repeats; 9=Heterochromatic; 10=Bivalent/Poised TSS; 11=Flanking
451 Bivalent/Poised TSS/Enh; 12=Bivalent Enhancer; 13=Repressed PolyComb; 14=Weak
452 Repressed PolyComb; 15=Quiescent/Low.
453
454 **Gene-mapping**
455 Genome-wide significant loci obtained by GWAS were mapped to genes in FUMA[17] using
456 three strategies:
457 1. Positional mapping maps SNPs to genes based on physical distance (within a 10kb
458 window) from known protein coding genes in the human reference assembly
459 (GRCh37/hg19).
460 2. eQTL mapping maps SNPs to genes with which they show a significant eQTL association
461 (i.e. allelic variation at the SNP is associated with the expression level of that gene). eQTL
462 mapping uses information from 45 tissue types in 3 data repositories (GTEx[32], Blood eQTL
463 browser[60], BIOS QTL browser[62]), and is based on cis-eQTLs which can map SNPs to genes
464 up to 1Mb apart. We used a false discovery rate (FDR) of 0.05 to define significant eQTL
465 associations.
466 3. Chromatin interaction mapping was performed to map SNPs to genes when there is a
467 three-dimensional DNA-DNA interaction between the SNP region and another gene region.
468 Chromatin interaction mapping can involve long-range interactions as it does not have a
469 distance boundary. FUMA currently contains Hi-C data of 14 tissue types from the study of
470 Schmitt et al [63]. Since chromatin interactions are often defined in a certain resolution, such as
471 40kb, an interacting region can span multiple genes. If a SNP is located in a region that
472 interacts with a region containing multiple genes, it will be mapped to each of those genes.
473 To further prioritize candidate genes, we selected only interaction-mapped genes in which

bioRxiv preprint doi: https://doi.org/10.1101/214973; this version posted February 1, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

one region involved in the interaction overlaps with a predicted enhancer region in any of the 111 tissue/cell types from the Roadmap Epigenomics Project[61], and the other region is located in a gene promoter region (250bp up and 500bp downstream of the transcription start site and also predicted by Roadmap to be a promoter region). This method reduces the number of genes mapped but increases the likelihood that those identified will indeed have a plausible biological function. We used a *P*-FDR < 1×10^{-5} to define significant interactions, based on previous recommendations[63], modified to account for the differences in cell lines used here.

GWAS catalog lookup

We used FUMA to identify SNPs with previously reported ($P<5\times10^{-5}$) phenotypic associations in published GWAS listed in the NHGRI-EBI catalog[64], which matched with SNPs in LD with one of the independent significant SNPs identified in the meta-analysis.

Polygenic risk scoring

To calculate the explained variance in insomnia by our GWAS results, we calculated polygenic scores (PGS) based on the SNP effect sizes in the meta-analysis. The PGS were calculated using two methods: LDpred[65] and PRSice[66], a script for calculating *P*-value thresholded PGS in PLINK. PGS were calculated using a leave-one-out method, where summary statistics were recalculated each time with one sample of N=3,000 from UKB excluded from the analysis. This sample was then used as a target sample for estimating the explained variance in insomnia by the PGS.

Mendelian Randomization

bioRxiv preprint doi: https://doi.org/10.1101/214973; this version posted February 1, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

To investigate causal associations between insomnia and genetically correlated traits, we analyzed direction of effects using Generalized Summary-data based Mendelian Randomization (GSMR[43]; http://cnsgenomics.com/software/gsmr/). This method uses effect sizes from GWAS summary statistics (standardized betas or log-transformed odds ratios) to infer causality of effects between two traits based on genome-wide significant SNPs. Built-in HEIDI outlier detection was applied to remove SNPs with pleiotropic effects on both traits, as these may bias the results. We tested for causal associations between insomnia and traits that were significantly genetically correlated with insomnia (b_{zx}). In addition, we tested for bi-directional associations by using other traits as the determinant and insomnia as the outcome (b_{zy}). We selected independent (r^2<0.1) lead SNPs with a GWS *P*-value ($<5\times10^{-8}$) as instrumental variables in the analyses. For traits with less than 10 lead SNPs (i.e. the minimum number of SNPs on which GSMR can perform a reliable analysis) we selected independent SNPs (r^2<0.1), with a *P*-value $<1\times10^{-5}$. If the outcome trait is binary, the estimated b_{zx} and b_{zy} are approximately equal to the natural log of the odds ratio (OR). An OR of 2 can be interpreted as a doubled risk compared to the population prevalence of a binary trait for every SD increase in the exposure trait. For quantitative traits, the b_{zx} and b_{zy} can be interpreted as a one standard deviation increase explained in the outcome trait for every SD increase in the exposure trait.

bioRxiv preprint doi: https://doi.org/10.1101/214973; this version posted February 1, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

References:

1. Wittchen, H.-U. *et al.* The size and burden of mental disorders and other disorders of the brain in europe 2010. *Eur. Neuropsychopharmacol.* **21**, 655-679 (2011).
2. Morin, C.M. *et al.* Insomnia disorder. *Nat. Rev. Dis. Primers* **1**, 15026-15026 (2015).
3. American Psychiatric Association. *Diagnostic and statistical manual of mental disorders (DSM-5®)*, (American Psychiatric Pub, 2013).
4. Morphy, H., Dunn, K.M., Lewis, M., Boardman, H.F. & Croft, P.R. Epidemiology of insomnia: A longitudinal study in a uk population. *Sleep* **30**, 274-280 (2007).
5. Lind, M.J., Aggen, S.H., Kirkpatrick, R.M., Kendler, K.S. & Amstadter, A.B. A longitudinal twin study of insomnia symptoms in adults. *Sleep* **38**, 1423-1430 (2015).
6. Lane, J.M. *et al.* Genome-wide association analysis identifies novel loci for chronotype in 100,420 individuals from the uk biobank. *Nat. Comm.* **7**, 10889 (2016).
7. Hammerschlag, A. *et al.* Genome-wide association analysis of insomnia identifies novel risk genes and genetic overlap with psychiatric and metabolic traits. *Nat. Genet.* **49**, 1584 (2017).
8. Sudlow, C. *et al.* Uk biobank: An open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Medicine* **12**, e1001779 (2015).
9. Eriksson, N. *et al.* Web-based, participant-driven studies yield novel genetic associations for common traits. *PLoS Genetics* **6**, e1000993 (2010).
10. Tung, J.Y. *et al.* Efficient replication of over 180 genetic associations with self-reported medical data. *PLoS One* **6**, e23473 (2011).
11. Benjamins, J.S. *et al.* Insomnia heterogeneity: Characteristics to consider for data-driven multivariate subtyping. *Sleep Med. Rev.* **36**, 71-81 (2017).
12. Paparrigopoulos, T. *et al.* Insomnia and its correlates in a representative sample of the greek population. *BMC Public Health* **10**, 531 (2010).
13. Cho, Y.W. *et al.* Epidemiology of insomnia in korean adults: Prevalence and associated factors. *Jour. Clin. Neurology* **5**, 20-23 (2009).
14. Zhang, B. & Wing, Y.-K. Sex differences in insomnia: A meta-analysis. *Sleep* **29**, 85-93 (2006).
15. Willer, C.J., Li, Y. & Abecasis, G.R. Metal: Fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **26**, 2190-2191 (2010).
16. Bulik-Sullivan, B.K. *et al.* Ld score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat. Genet.* **47**, 291-295 (2015).
17. Watanabe, K., Taskesen, E., Bochoven, A. & Posthuma, D. Functional mapping and annotation of genetic associations with fuma. *Nat. Comm.* **8**, 1826 (2017).
18. Ernst, J. & Kellis, M. Chromhmm: Automating chromatin-state discovery and characterization. *Nat. Methods* **9**, 215-216 (2012).
19. Sniekers, S. *et al.* Genome-wide association meta-analysis of 78,308 individuals identifies new loci and genes influencing human intelligence. *Nat. Genet.* **49**, 1107-1112 (2017).
20. Kircher, M. *et al.* A general framework for estimating the relative pathogenicity of human genetic variants. *Nat. Genet.* **46**, 310-315 (2014).
21. Laramie, J.M. *et al.* Polymorphisms near exoc4 and lrguk on chromosome 7q32 are associated with type 2 diabetes and fasting glucose; the nhlbi family heart study. *BMC Med. Genet.* **9**, 46 (2008).
22. Butler, M.G., Rafi, S.K. & Manzardo, A.M. High-resolution chromosome ideogram representation of currently recognized genes for autism spectrum disorders. *Int. J. Mol. Sci.* **16**, 6464-6495 (2015).

bioRxiv preprint doi: https://doi.org/10.1101/214973; this version posted February 1, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

23. de Leeuw, C.A., Mooij, J.M., Heskes, T. & Posthuma, D. Magma: Generalized gene-set analysis of gwas data. *PLoS Computational Biology* **11**, e1004219 (2015).
24. Kripke, D.F. *et al.* Genetic variants associated with sleep disorders. *Sleep Med.* **16**, 217-224 (2015).
25. Stefansson, H. *et al.* A genetic risk factor for periodic limb movements in sleep. *N. Engl. J. Med.* **357**, 639-647 (2007).
26. Janik, P., Berdyński, M., Safranow, K. & Żekanowski, C. The btbd9 gene polymorphisms in polish patients with gilles de la tourette syndrome. *Acta Neurobiol. Exp. (Wars.)* **74**, 218-226 (2014).
27. Freeman, A. *et al.* Sleep fragmentation and motor restlessness in a drosophila model of restless legs syndrome. *Curr. Biol.* **22**, 1142-1148 (2012).
28. DeAndrade, M.P. *et al.* Motor restlessness, sleep disturbances, thermal sensory alterations and elevated serum iron levels in btbd9 mutant mice. *Hum. Mol. Genet.* **21**, 3984-3992 (2012).
29. DeAndrade, M.P. *et al.* Enhanced hippocampal long-term potentiation and fear memory in btbd9 mutant mice. *PLoS One* **7**, e35518 (2012).
30. Suzuki, A., Sinton, C.M., Greene, R.W. & Yanagisawa, M. Behavioral and biochemical dissociation of arousal and homeostatic sleep need influenced by prior wakeful experience in mice. *Proceedings of the National Academy of Sciences* **110**, 10288-10293 (2013).
31. Liberzon, A. *et al.* The molecular signatures database hallmark gene set collection. *Cell systems* **1**, 417-425 (2015).
32. Consortium, G.T. The genotype-tissue expression (gtex) pilot analysis: Multitissue gene regulation in humans. *Science* **348**, 648-660 (2015).
33. Toffoli, M. *et al.* Snca 3′ utr genetic variants in patients with parkinson's disease and rem sleep behavior disorder. *Neurol. Sci.* **38**, 1233-1240 (2017).
34. Edwards, T.L. *et al.* Genome-wide association study confirms snps in snca and the mapt region as common risk factors for parkinson disease. *Ann. Hum. Genet.* **74**, 97-109 (2010).
35. McDowell, K.A., Shin, D., Roos, K.P. & Chesselet, M.-F. Sleep dysfunction and eeg alterations in mice overexpressing alpha-synuclein. *J. Parkinsons Dis.* **4**, 531-539 (2014).
36. Colombo, M.A. *et al.* Wake high-density electroencephalographic spatiospectral signatures of insomnia. *Sleep* **39**, 1015-1027 (2016).
37. Gene Ontology. The gene ontology (go) database and informatics resource. *Nucleic Acids Res.* **32**, D258-D261 (2004).
38. Skene, N.G. *et al.* Genetic identification of brain cell types underlying schizophrenia. *bioRxiv*, 145466 (2017).
39. Saper, C.B., Scammell, T.E. & Lu, J. Hypothalamic regulation of sleep and circadian rhythms. *Nature* **437**, 1257-1263 (2005).
40. Hu, Y. *et al.* Gwas of 89,283 individuals identifies genetic variants associated with self-reporting of being a morning person. *Nat. Comm.* **7**, 10448 (2016).
41. Jones, S.E. *et al.* Genome-wide association analyses in 128,266 individuals identifies new morningness and sleep duration loci. *PLoS Genetics* **12**, e1006125 (2016).
42. Johann, A.F. *et al.* Insomnia with objective short sleep duration is associated with longer duration of insomnia in the freiburg insomnia cohort compared to insomnia with normal sleep duration, but not with hypertension. *PLoS One* **12**, e0180339 (2017).
43. Zhu, Z. *et al.* Causal associations between risk factors and common diseases inferred from gwas summary data. *Nat. Comm.* **9**, 224 (2018).

bioRxiv preprint doi: https://doi.org/10.1101/214973; this version posted February 1, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

44. Vetrivelan, R., Qiu, M.-H., Chang, C. & Lu, J. Role of basal ganglia in sleep–wake regulation: Neural circuitry and clinical significance. *Front. Neuroanat.* **4**(2010).
45. Lazarus, M., Huang, Z.-L., Lu, J., Urade, Y. & Chen, J.-F. How do the basal ganglia regulate sleep–wake behavior? *Trends Neurosci.* **35**, 723-732 (2012).
46. Swardfager, W., Rosenblat, J.D., Benlamri, M. & McIntyre, R.S. Mapping inflammation onto mood: Inflammatory mediators of anhedonia. *Neurosci. Biobehav. Rev.* **64**, 148-166 (2016).
47. Wilson, C.J. & Groves, P.M. Spontaneous firing patterns of identified spiny neurons in the rat neostriatum. *Brain Res.* **220**, 67-80 (1981).
48. Stoffers, D. *et al.* The caudate: A key node in the neuronal network imbalance of insomnia? *Brain* **137**, 610-620 (2013).
49. Altena, E. *et al.* Prefrontal hypoactivation and recovery in insomnia. *Sleep* **31**, 1271-1276 (2008).
50. Dai, X.-J. *et al.* Altered intrinsic regional brain spontaneous activity and subjective sleep quality in patients with chronic primary insomnia: A resting-state fmri study. *Neuropsychiatr. Dis. Treat.* **10**, 2163 (2014).
51. Nagel, M. *et al.* Gwas meta-analysis of neuroticism (n= 449,484) identifies novel genetic loci and pathways. *bioRxiv*, 184820 (2017).
52. Mathur, B.N. The claustrum in review. *Front. Syst. Neurosci.* **8**, 48 (2014).
53. Wei, Y. *et al.* I keep a close watch on this heart of mine: Increased interoception in insomnia. *Sleep* **39**, 2113-2124 (2016).
54. Krystal, A.D., Edinger, J.D., Wohlgemuth, W.K. & Marsh, G.R. Nrem sleep eeg frequency spectral correlates of sleep complaints in primary insomnia subtypes. *Sleep* **25**, 630-40 (2002).
55. Hawrylycz, M. *et al.* Canonical genetic signatures of the adult human brain. *Nat. Neurosci.* **18**, 1832-1844 (2015).
56. Liberzon, A. *et al.* Molecular signatures database (msigdb) 3.0. *Bioinformatics* **27**, 1739-1740 (2011).
57. de Leeuw, C.A., Neale, B.M., Heskes, T. & Posthuma, D. The statistical properties of gene-set analysis. *Nature Reviews Genetics* **17**, 353-364 (2016).
58. Finucane, H.K. *et al.* Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nat. Genet.* **47**, 1228-1235 (2015).
59. Wang, K., Li, M. & Hakonarson, H. Annovar: Functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* **38**, e164-e164 (2010).
60. Westra, H.J. *et al.* Systematic identification of trans-eqtls as putative drivers of known disease associations. *Nat. Genet.* **45**, 1238 (2013).
61. Kundaje, A. *et al.* Integrative analysis of 111 reference human epigenomes. *Nature* **518**, 317 (2015).
62. Zhernakova, D.V. *et al.* Identification of context-dependent expression quantitative trait loci in whole blood. *Nat. Genet.* **49**, 139-145 (2017).
63. Schmitt, A.D. *et al.* A compendium of chromatin contact maps reveals spatially active regions in the human genome. *Cell Reports* **17**, 2042-2059 (2016).
64. MacArthur, J. *et al.* The new nhgri-ebi catalog of published genome-wide association studies (gwas catalog). *Nucleic Acids Res.* **45**, D896-D901 (2017).
65. Vilhjalmsson, B.J. *et al.* Modeling linkage disequilibrium increases accuracy of polygenic risk scores. *Am. J. Hum. Genet.* **97**, 576-592 (2015).
66. Euesden, J., Lewis, C.M. & O'Reilly, P.F. Prsice: Polygenic risk score software. *Bioinformatics* **31**, 1466-1468 (2015).

bioRxiv preprint doi: https://doi.org/10.1101/214973; this version posted February 1, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

667 **Acknowledgements:** This work was funded by The Netherlands Organization for Scientific
668 Research (NWO Brain & Cognition 433-09-228, NWO MagW VIDI 452-12-04, NWO VICI
669 435-14-005 and 453-07-001, 645-000-003). P.R.J. was funded by the Sophia Foundation for
670 Scientific Research (S14-27), E.J.W.V.S. was funded by the European Research Council
671 (ERC-ADG-2014-671084 INSOMNIA), J.B. was funded by the Swiss National Science
672 Foundation. Analyses were carried out on the Genetic Cluster Computer, which is financed
673 by the Netherlands Scientific Organization (NWO: 480-05-003), by the VU University,
674 Amsterdam, the Netherlands, and by the Dutch Brain Foundation, and is hosted by the Dutch
675 National Computing and Networking Services SurfSARA. This research has been conducted
676 using the UK Biobank Resource (application number 16406). We would like to thank the
677 participants and researchers who collected and contributed to the data. We thank the
678 23andMe research participants and employees for making this work possible, including the
679 following members of the 23andMe Research Team: Michelle Agee, Babak Alipanahi, Adam
680 Auton, Robert K. Bell, Katarzyna Bryc, Sarah L. Elson, Pierre Fontanillas, Nicholas A.
681 Furlotte, David A. Hinds, Karen E. Huber, Aaron Kleinman, Nadia K. Litterman, Jennifer C.
682 McCreight, Matthew H. McIntyre, Joanna L. Mountain, Elizabeth S. Noblin, Carrie A. M.
683 Northover, Steven J. Pitts, J. Fah Sathirapongsasuti, Olga V. Sazonova, Janie F. Shelton,
684 Suyash Shringarpure, Chao Tian, and Catherine H. Wilson.
685
686 **Author contributions:** D.P. and E.J.W.V.S. conceived the idea of the study. D.P. supervised
687 the pre- and post gwas analysis pipeline. P.R.J. and K.W. performed the analyses. S.St.
688 performed quality control on the UK Biobank data and wrote the analysis pipeline. K.W.
689 wrote the online platform (FUMA) that was used for follow-up analyses. C.d.L conducted
690 conditional gene-set analyses. J.B., N.S., A.M.M. and J.H.L contributed scRNA information.
691 J.T., D.H., V.V. and the 23andMe Research Team contributed and analyzed the 23andMe

bioRxiv preprint doi: https://doi.org/10.1101/214973; this version posted February 1, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

cohort data. D.P., E.J.W.V.S. and P.R.J. wrote the paper. All authors discussed the results, and approved the final version of the paper.

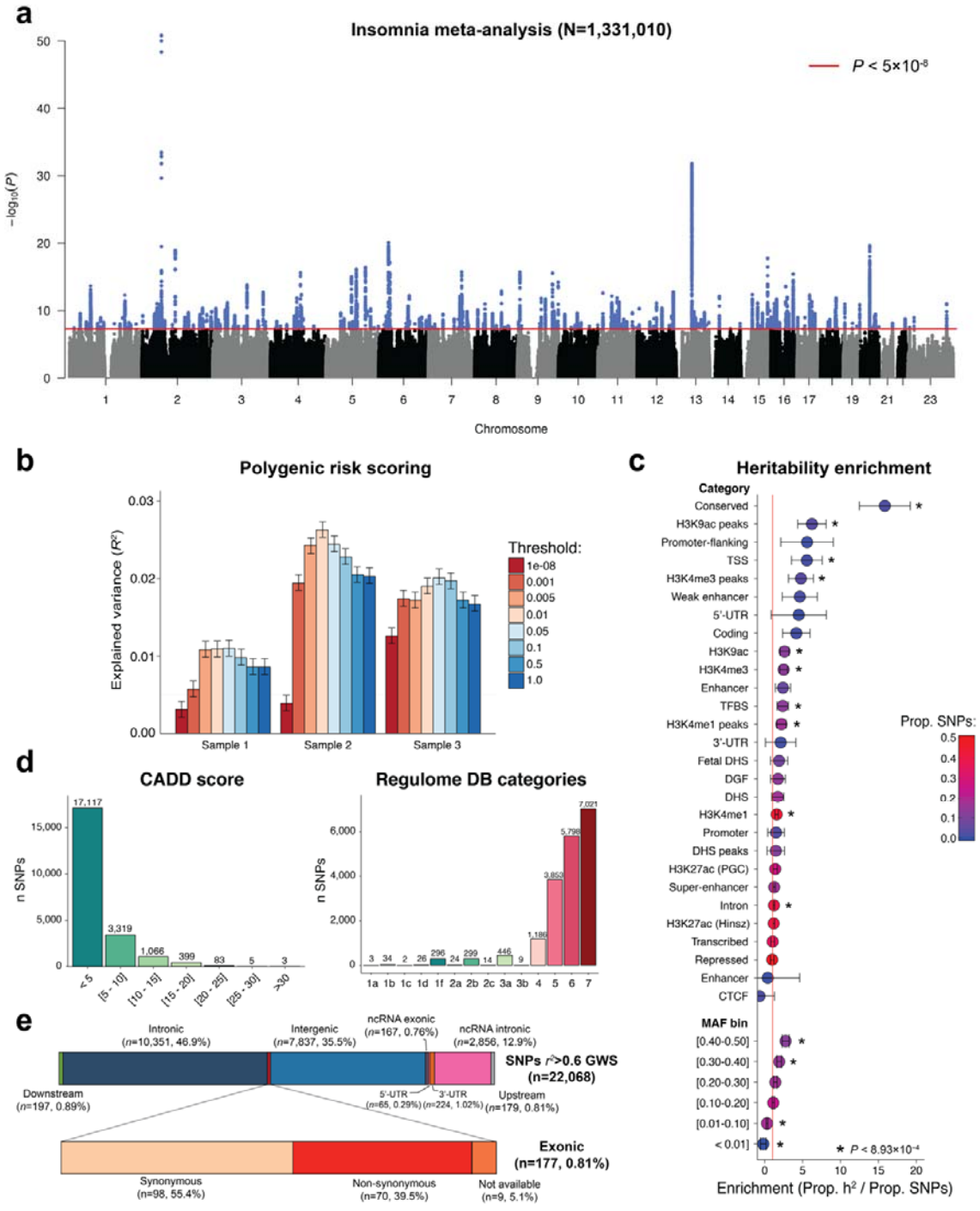
Materials & Correspondence: The data analyzed in the current study was partly provided by the UK Biobank Study (www.ukbiobank.ac.uk), received under the UK Biobank application number 16406. Our policy is to make genome-wide summary statistics (sumstats) publicly available. Sumstats from the GWAS's conducted are available for download at https://ctg.cncr.nl/. Note that our freely available meta-analytic sumstats (insomnia and morningness) concern results excluding the 23andMe sample. This is a non-negotiable clause in the 23andMe data transfer agreement, intended to protect the privacy of the 23andMe research participants. To fully recreate our meta-analytic results for insomnia and morningness: (a) obtain insomnia and morningness sumstats from 23andMe (see below); (b) conduct a meta-analysis of our sumstats with the 23andMe sumstats. 23andMe participant data are shared according to community standards that have been developed to protect against breaches of privacy. Currently, these standards allow for the sharing of summary statistics for at most 10,000 SNPs. The full set of summary statistics can be made available to qualified investigators who enter into an agreement with 23andMe that protects participant confidentiality. Interested investigators should email dataset-request@23andme.com for more information.

Author Information: V.V., D.H., and J.T. are employees of 23andMe. All other authors declare no competing financial interest. Correspondence and requests for materials should be addressed to d.posthuma@vu.nl.

bioRxiv preprint doi: https://doi.org/10.1101/214973; this version posted February 1, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

716 **FIGURES**

717



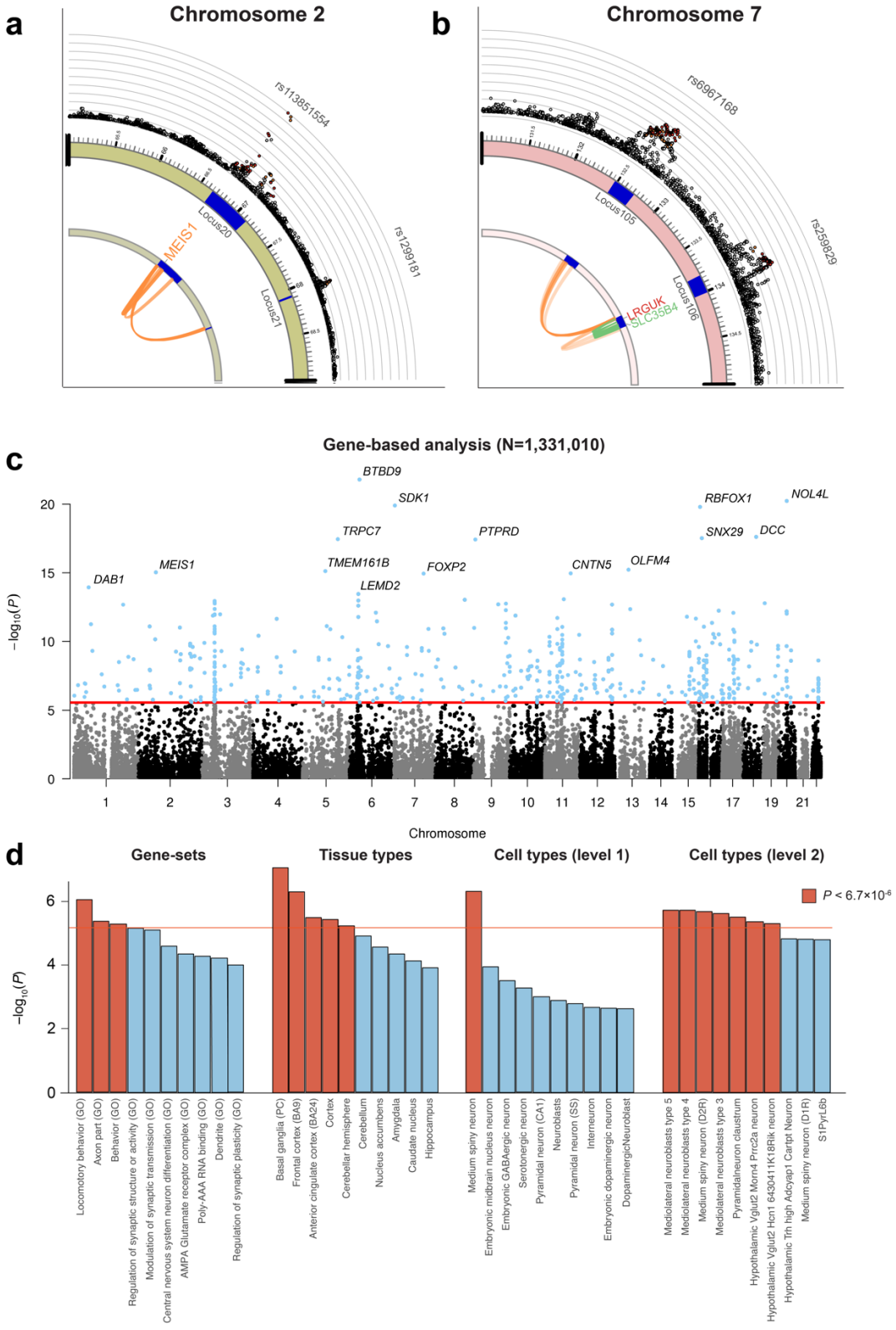
718

719

bioRxiv preprint doi: https://doi.org/10.1101/214973; this version posted February 1, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Fig. 1a-e. SNP-based results from the GWAS meta-analysis on insomnia (N=1,331,010). (a) Manhattan plot of the GWAS of insomnia, showing the –log10-transformed *P*-value for each SNP **(b)** Heritability enrichment for functional SNP categories and minor allele frequency bins (MAF). Enrichment was calculated by dividing the proportion of heritability for each category by the proportion of SNPs in that category, significant enrichments after Bonferroni correction (28 functional categories + 6 MAF bins + 22 chromosomes) are indicated by an asterisk ($P<0.05/56=8.93\times10^{-4}$) **(c)** Polygenic score (PGS) prediction in three hold-out samples (N=3,000), showing the increase in explained variance in insomnia (Nagelkerke's pseudo R^2) and 95% confidence interval for each *P*-value threshold. All *P*-value thresholds were statistically significant. **(d)** Distribution of CADD scores and RegulomeDB category of all annotated SNPs in LD ($r^2>=0.6$) with one of the GWS SNPs (n=22,068) and **(e)** functional consequences of these SNPs.

bioRxiv preprint doi: https://doi.org/10.1101/214973; this version posted February 1, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.



bioRxiv preprint doi: https://doi.org/10.1101/214973; this version posted February 1, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Fig. 2a-d. Gene-based and gene-set analyses of insomnia. Zoomed-in circos plots showing genes implicated by two genomic risk loci on chromosome 2 **(a)** and chromosome 7 **(b)**, genomic risk loci indicated as blue areas, eQTL associations in green, chromatin interactions in orange. Genes mapped by by both eQTL and chromatin interactions are red. The outer layer shows a Manhattan plot containing the negative log10-transformed *P*-value of each SNP in the GWAS meta-analysis of insomnia. Full circos plots of all autosomal chromosomes are provided in **Supplementary Fig. 2**. **(c)** Genome-wide gene-based analysis (GWAS) of 18,185 genes that were tested for association with insomnia in MAGMA. The y-axis shows the negative $\log_{10}$-transformed *P*-value of the gene-based test, the x-axis shows the starting position on the chromosome. The red line indicates the Bonferroni corrected threshold for genome-wide significance ($P=0.05/18{,}185=2.75\times10^{-6}$). The top 15 most significant genes are highlighted. **(d)** Gene-set analysis of top 20 for each of the MsigDB pathways, tissue expression of GTEx tissue types, and cell types from single-cell RNA sequencing. Gene-set analyses were performed using MAGMA. The red line shows the Bonferroni significance threshold ($P<0.05/7{,}473=6.7\times10^{-6}$), correcting for the total number of tested gene-sets. Red bars indicated significant gene-sets.

bioRxiv preprint doi: https://doi.org/10.1101/214973; this version posted February 1, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

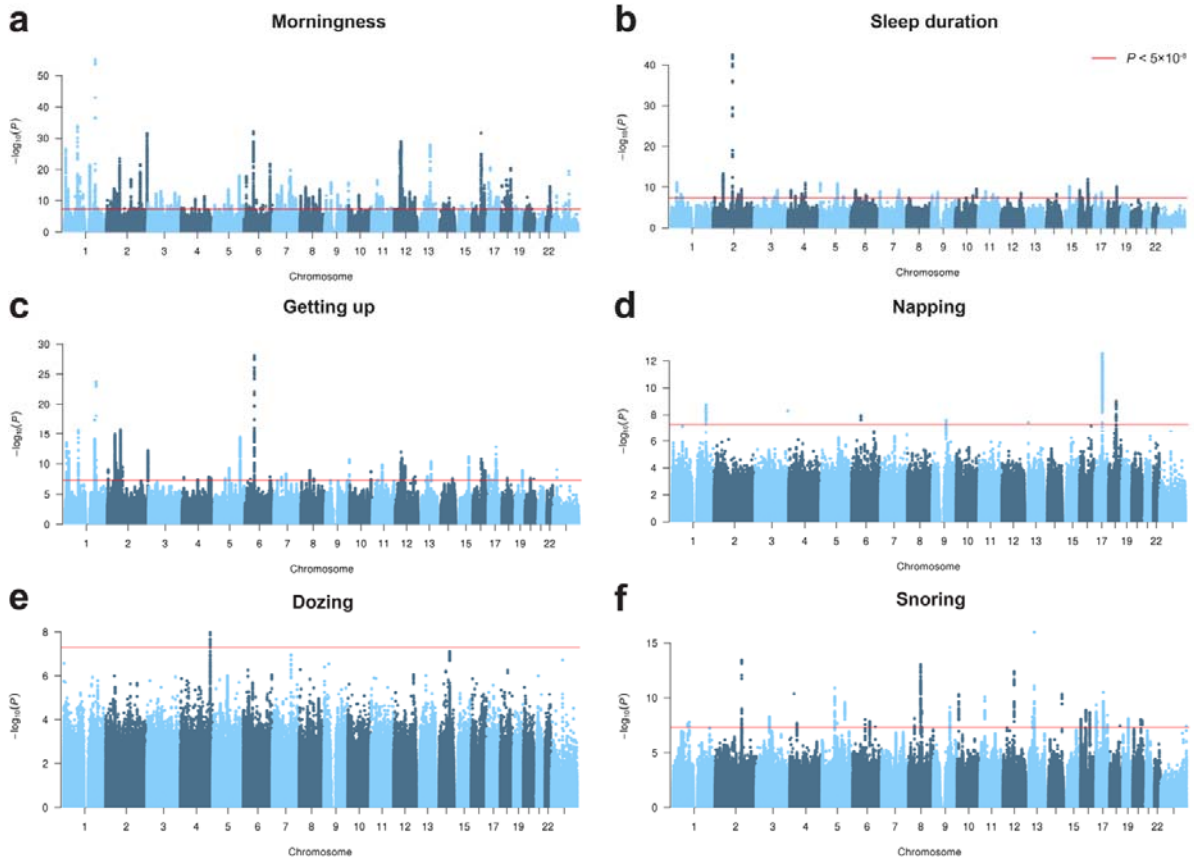


Fig. 3a-f. Genome-wide analyses of six sleep-related traits. Manhattan plots of the genome-wide association analyses of **(a)** Morningness (N=434,835). **(b)** Sleep duration (N=384,317) **(c)** Ease of getting up (N=385,949) **(d)** Napping (N=386,577) **(e)** Daytime dozing (N=385,333) and **(f)** Snoring (N=359,916). The y-axis shows the negative $\log_{10}$-transformed SNP *P*-value, the x-axis the base pair position of the SNPs on each chromosome. The red line indicates the Bonferroni corrected significance threshold ($P<5\times10^{-8}$).

bioRxiv preprint doi: https://doi.org/10.1101/214973; this version posted February 1, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

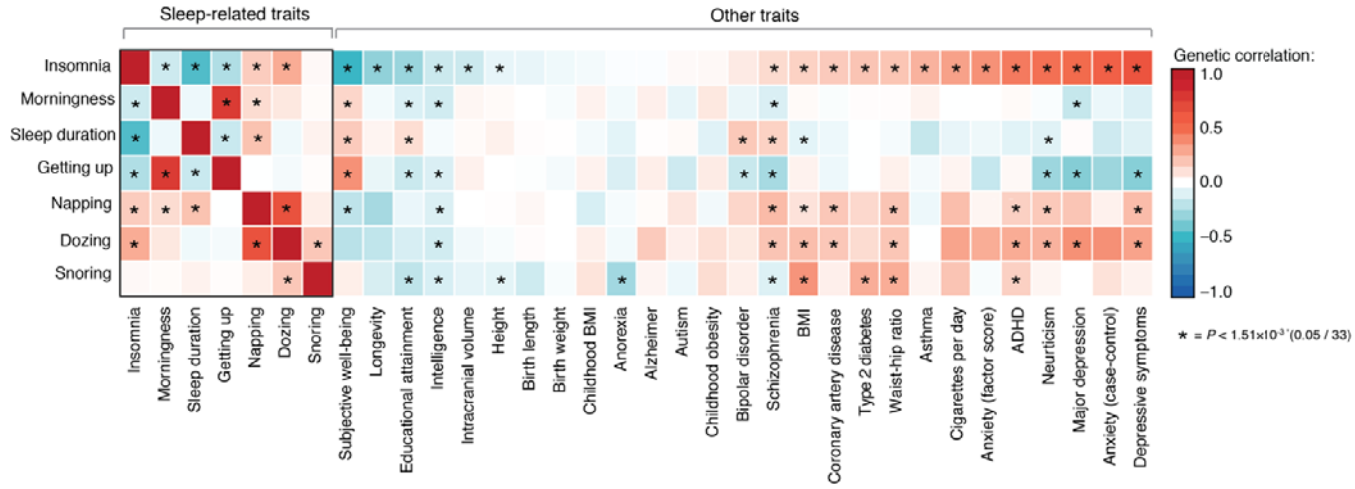


Fig. 4. Genetic overlap of insomnia with other sleep-related traits and psychiatric and metabolic traits. Heatmap of genetic correlations between insomnia, sleep-related phenotypes and neuropsychiatric and metabolic traits studies that were calculated using LD Score regression. Red color indicates a positive r_g while green indicates negative r_g. Correlations that were significant after Bonferroni correction ($P<0.05/33=1.5110^{-3}$) are indicated with an asterisk (see also **Supplementary Table 18, 26**).

bioRxiv preprint doi: https://doi.org/10.1101/214973; this version posted February 1, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

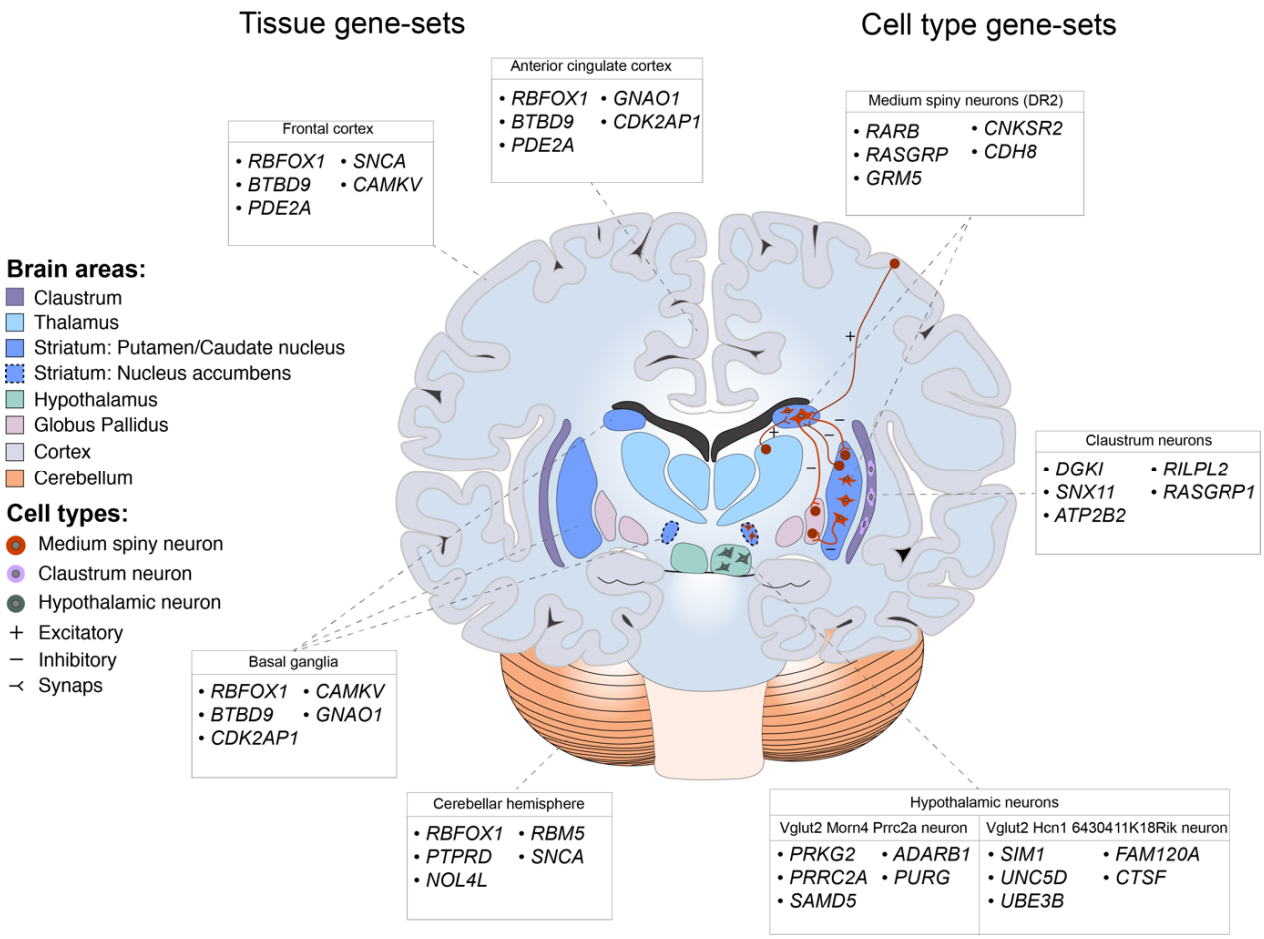


Fig. 5. Overview of brain tissues and cell types associated with insomnia based on GWAS results from 1,331,010 individuals. For each associated gene-set, the top 5 genes driving the association are reported for each brain area and cell-type. Results for GTEx brain tissue type gene-sets are shown on the left side of the figure, while results from the level 2 single-cell gene expression are shown on the right.

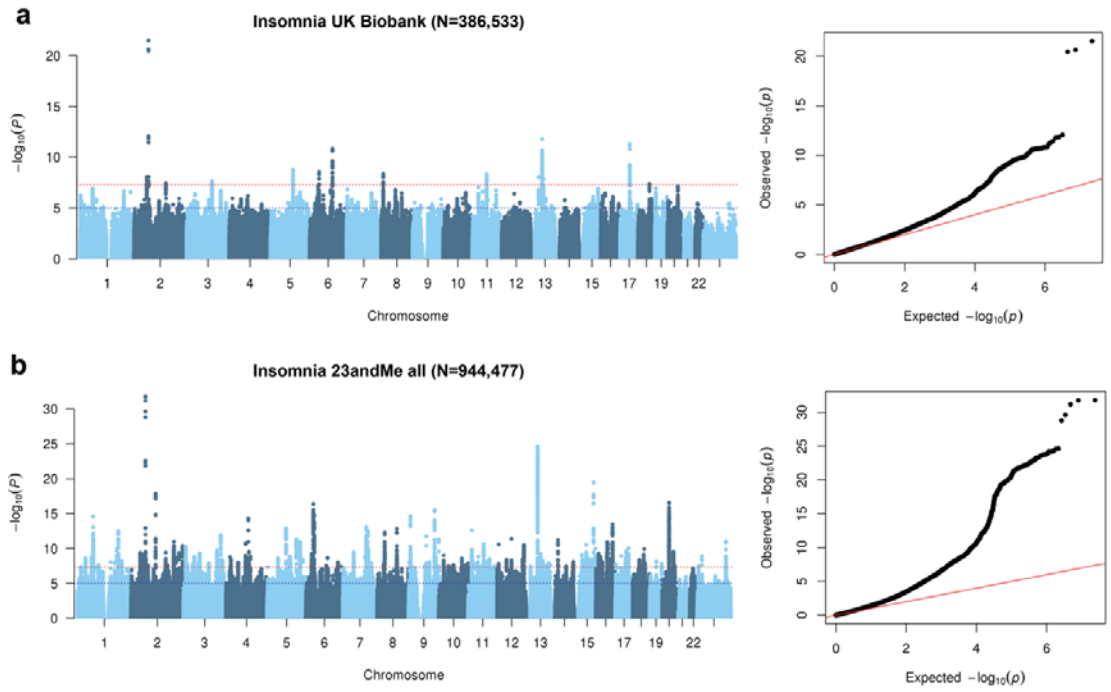
bioRxiv preprint doi: https://doi.org/10.1101/214973; this version posted February 1, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

EXTENDED DATA

Extended Data Table 1. LD Score regression estimates of the sex-specific GWAS of insomnia. Results are shown for UK Biobank, 23andMe and the sex-specific meta-analyzed sample. H2=estimated SNP-heritability, intercept=LD Score regression intercept, rg=genetic correlation in the same study sample.

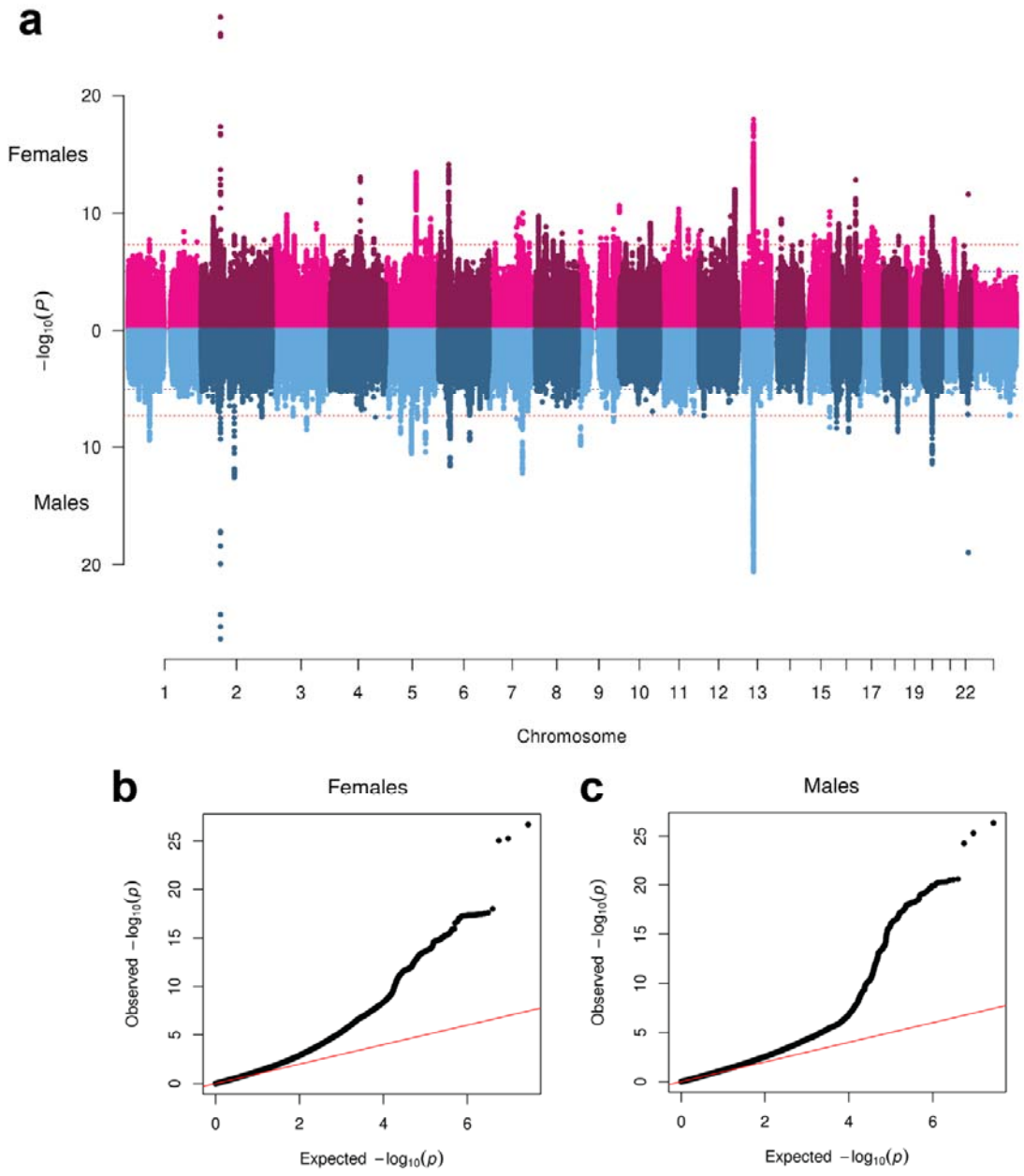
Sample	Sex	N	*h2(SE)*	Mean chi2	Lambda	Intercept (SE)	*rg male*	*rg female*
UK Biobank	male	177.817	0.083 (0.007)	1,157	1,143	1.001 (0.008)	1	0.857 (0.051)
	female	208.716	0.092 (0.005)	1,233	1,210	1.011 (0.008)	0.857 (0.051)	1
23andMe	male	443.207	0.080 (0.004)	1,385	1,317	1.016 (0.008)	1	0.925 (0.022)
	female	501.270	0.090 (0.003)	1,580	1,460	1.046 (0.009)	0.925 (0.022)	1
Meta	male	621.024	0.067 (0.003)	1,460	1,382	1.024 (0.009)	1	0.919 (0.018)
	female	709.986	0.078 (0.003)	1,700	1,547	1.042 (0.009)	0.919 (0.018)	1

bioRxiv preprint doi: https://doi.org/10.1101/214973; this version posted February 1, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.



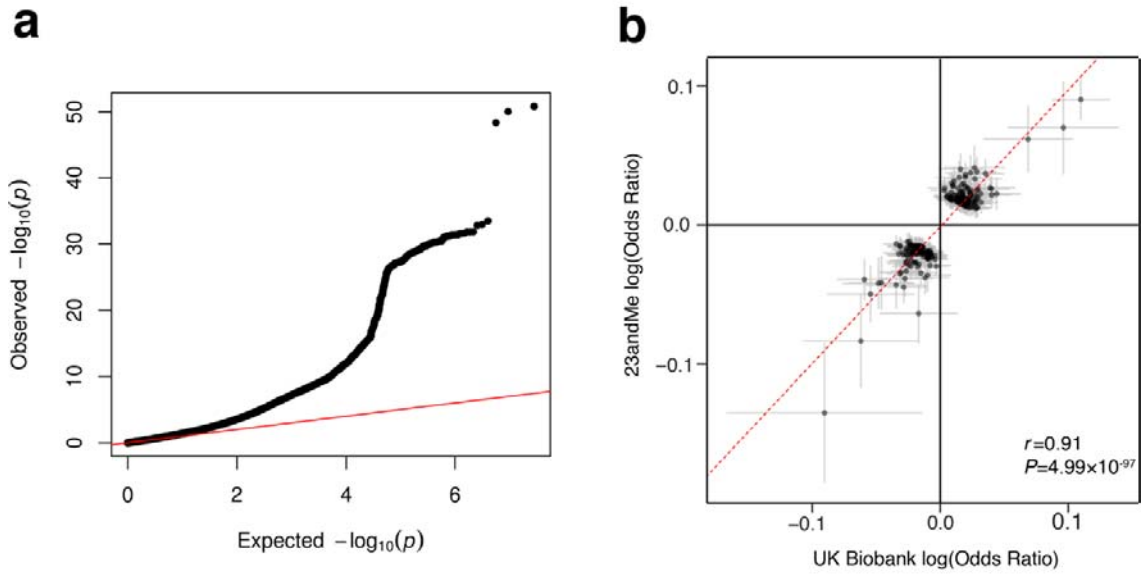
Extended Data Fig 1a-b. Manhattan and Q-Q plots of the genome-wide analysis of insomnia. Results are shown for the genome-wide analysis in **(a)** UK Biobank and **(b)** 23andMe.

bioRxiv preprint doi: https://doi.org/10.1101/214973; this version posted February 1, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.



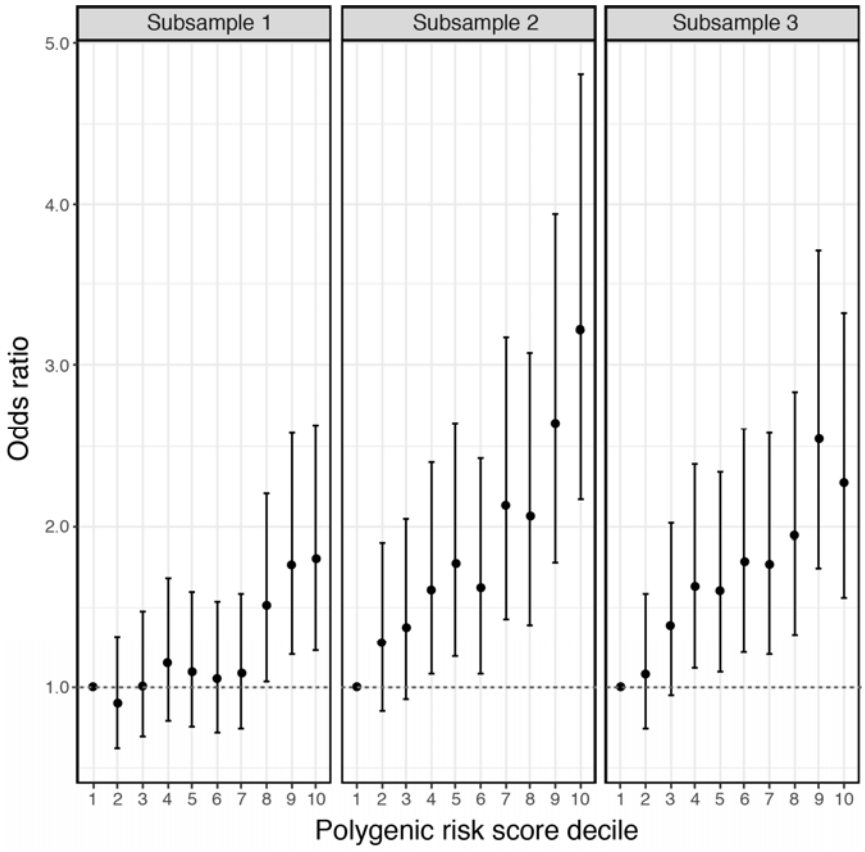
Extended Data Fig. 3a-c. Sex-specific Manhattan plot and Q-Q plot of the insomnia meta-analysis in males and females (UK Biobank + 23andMe). (**a**) Miami plot showing sex-specific SNP association P-values for females on the upper side and males on the lower side. (**b**) Q-Q plot in females, and (**c**) in males.

bioRxiv preprint doi: https://doi.org/10.1101/214973; this version posted February 1, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.



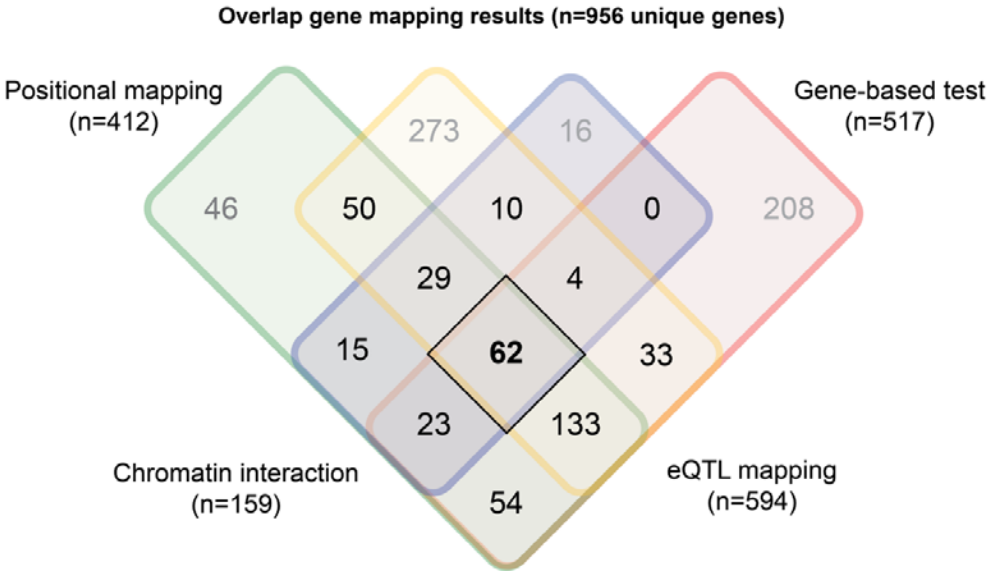
Extended Data Fig. 4a-b. Q-Q plot and lead SNPs of the GWAS meta-analysis for insomnia. (**a**) QQ-plot of the insomnia meta-analysis showing the expected negative log10-transformed *P*-value distribution on the x-axis, and observed negative log10-transformed *P*-value on the y-axis, (**b)** effect size plot of the 248 lead SNP of the insomnia meta-analysis (log-transformed odds ratio and 95% confidence interval) in UK Biobank and 23andMe.

bioRxiv preprint doi: https://doi.org/10.1101/214973; this version posted February 1, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.



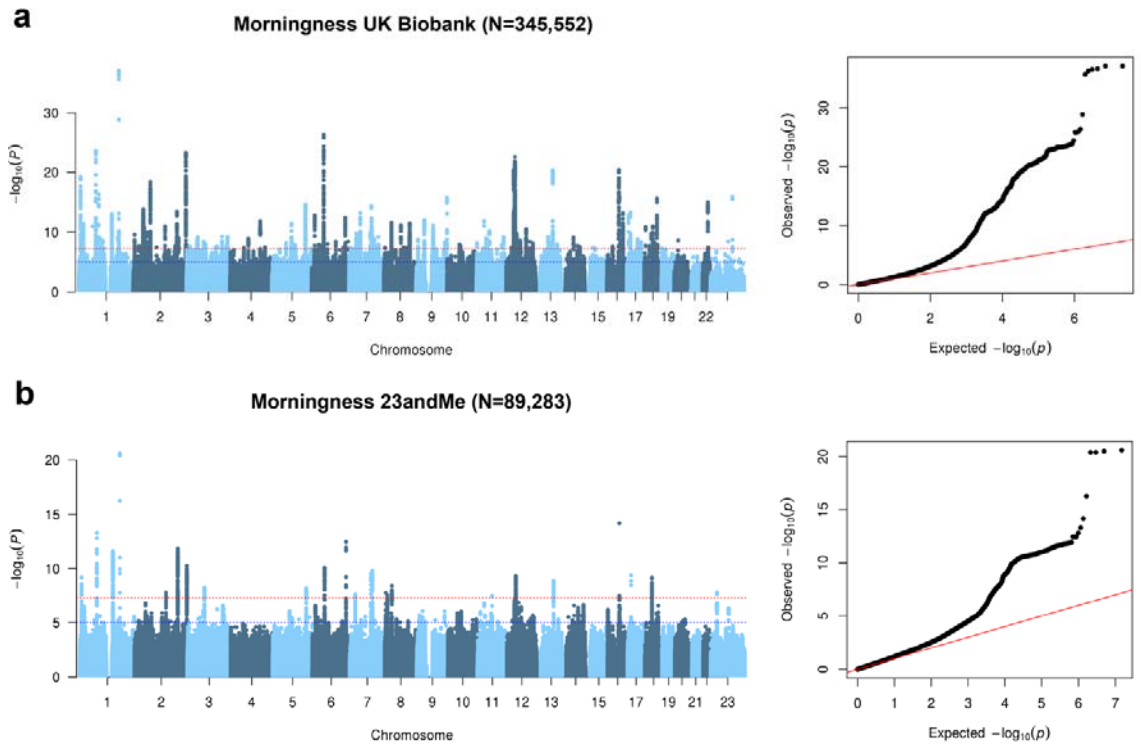
Extended Data Fig. 5. Risk of insomnia per polygenic risk score decile in three independent holdout samples (N=3x3000). Odds ratios and 95% confidence interval for deciles in polygenic risk score were calculated based on a logistic regression model, using the lowest polygenic risk score decile as the reference.

bioRxiv preprint doi: https://doi.org/10.1101/214973; this version posted February 1, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.



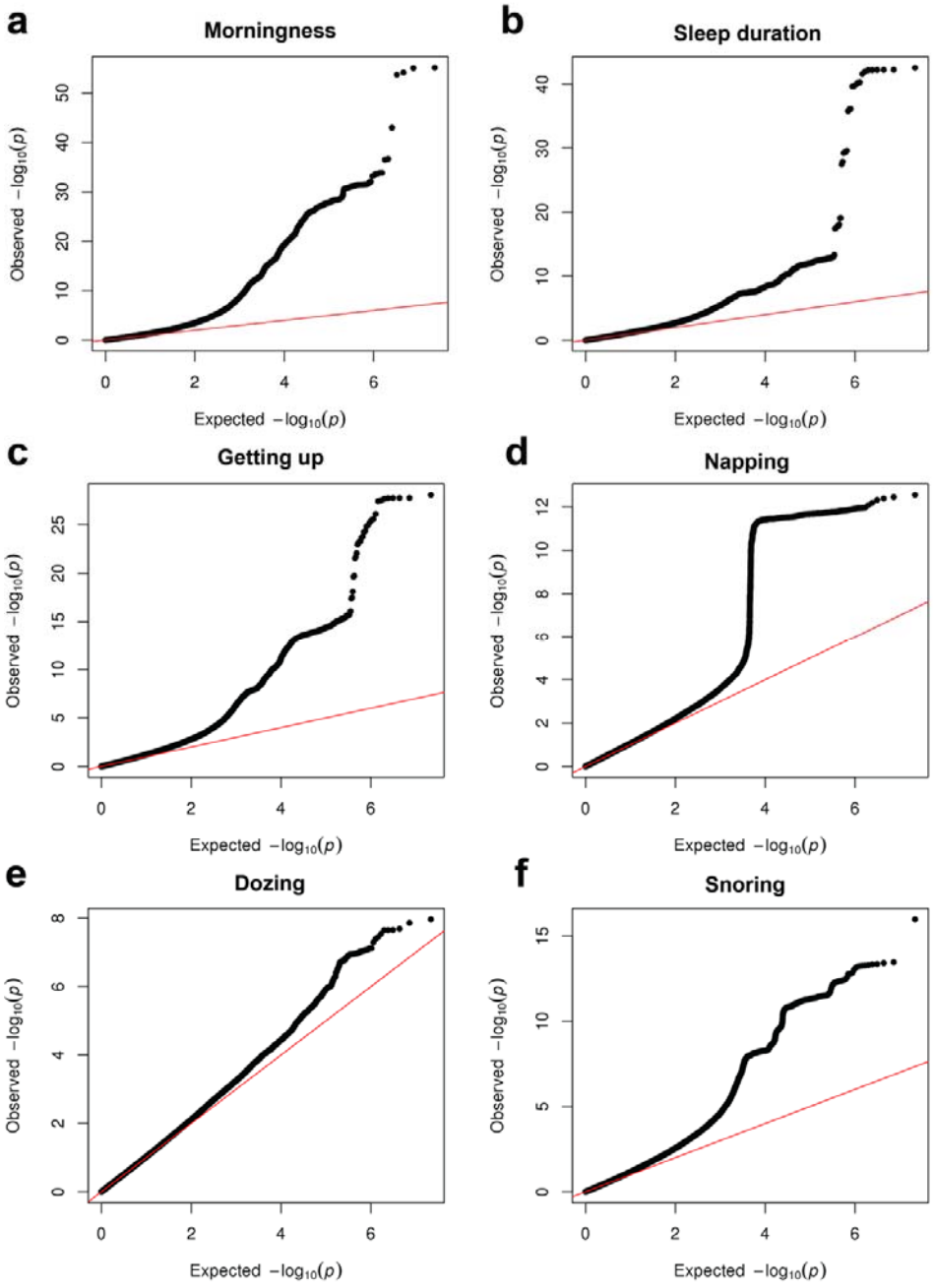
Extended Data Fig. 6. Venn diagram showing the number of genes that were mapped by four gene-mapping strategies. Each square shows the number of overlapping genes between three gene-mapping methods in FUMA (positional mapping, eQTL mapping and chromatin interaction mapping) and significant genes in gene-based tests in MAGMA. The number of genes in bold highlights the number of genes that were implicated by all four methods.

bioRxiv preprint doi: https://doi.org/10.1101/214973; this version posted February 1, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.



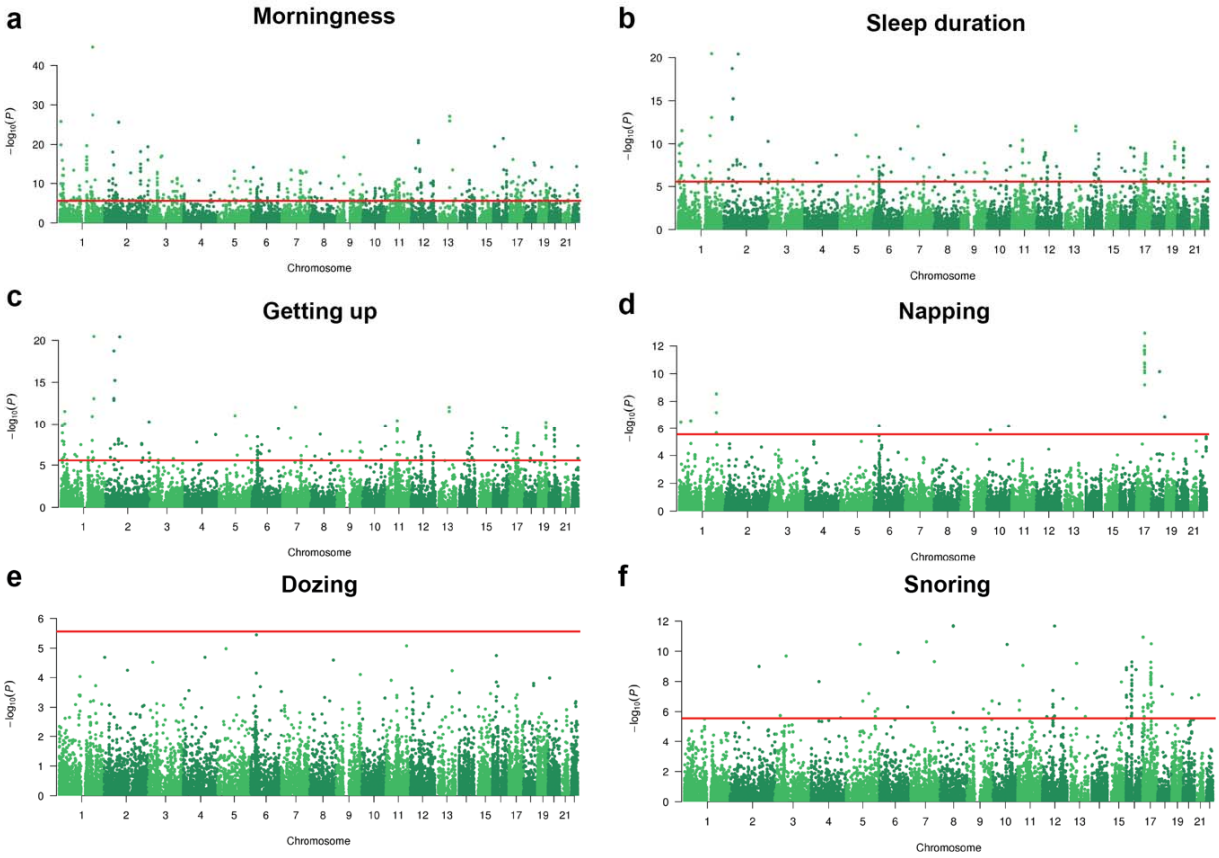
814 **Extended Data Fig. 7a-b. Manhattan plot and Q-Q plot of the genome-wide analysis of**
815 **morningness in UK Biobank and 23andMe.** Results are shown for (**a**) UK Biobank and (**b**)
816 23andMe.
817

bioRxiv preprint doi: https://doi.org/10.1101/214973; this version posted February 1, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.



Extended Data Fig. 8a-f. Q-Q plots of the genome-wide analysis of six sleep related traits. (a) morningness (including UKB and 23andMe), **(b)** sleep duration, **(c)** ease of getting up, **(d)** daytime napping, **(e)** daytime dozing, **(f)** snoring. Manhattan plots of the genome-wide analyses are shown in **Fig. 3**.

bioRxiv preprint doi: https://doi.org/10.1101/214973; this version posted February 1, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

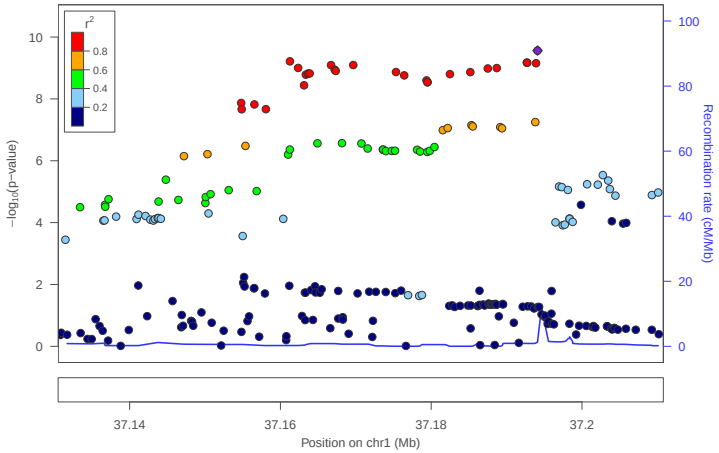


Extended Data Fig. 9a-f. Genome-wide gene-based association analysis of six sleep-related phenotypes. Manhattan plots genome-wide gene-based analysis (GWGAS) results for **(a)** morningness **(b)** sleep duration **(c)** ease of getting up **(d)** daytime napping **(e)** daytime dozing **(f)** snoring. GWGAS was performed in MAGMA. The analysis of morningness was based on GWAS meta-analysis of UKB and 23andMe, while other sleep-related phenotypes were analysed in UKB. The red line indicates Bonferroni corrected significance threshold depending on the number of genes tested.

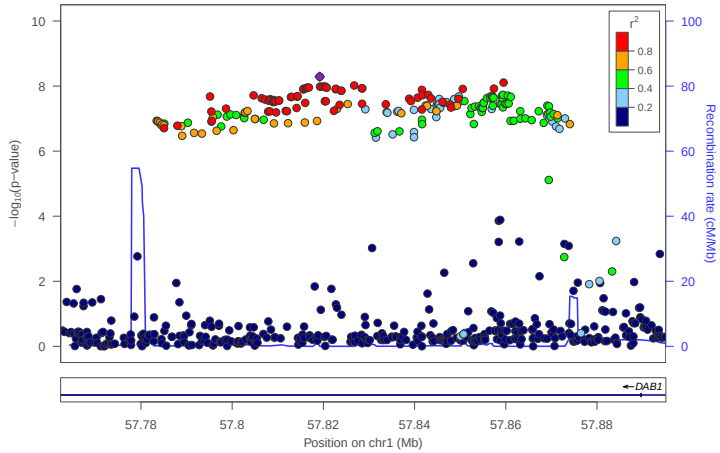
bioRxiv preprint doi: https://doi.org/10.1101/214973; this version posted February 1, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Supplementary Information includes:

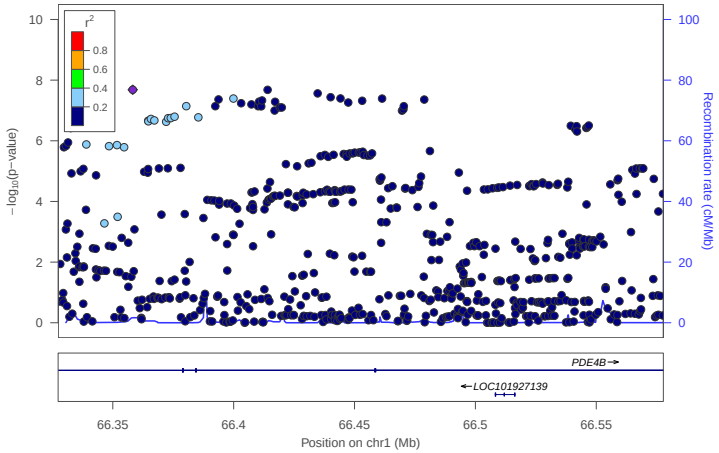
locus_1



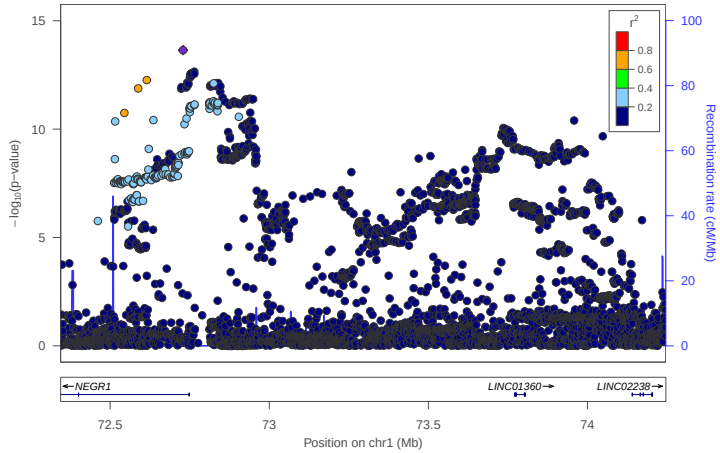
locus_2



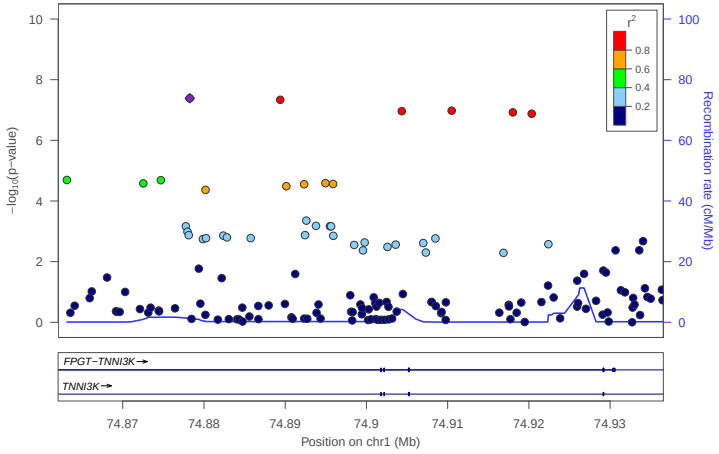
locus_3



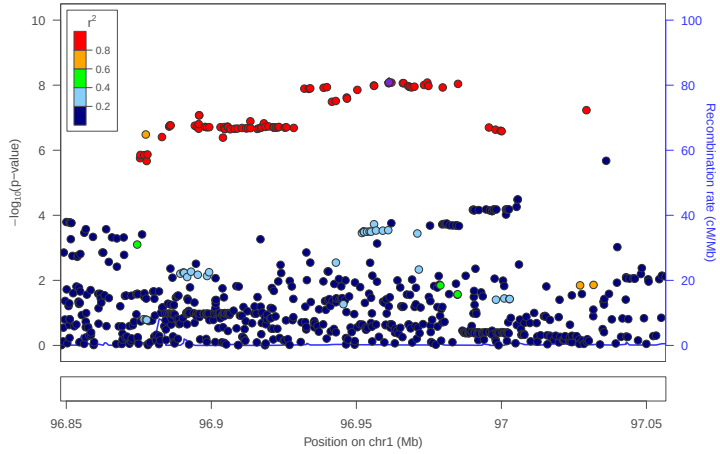
locus_4



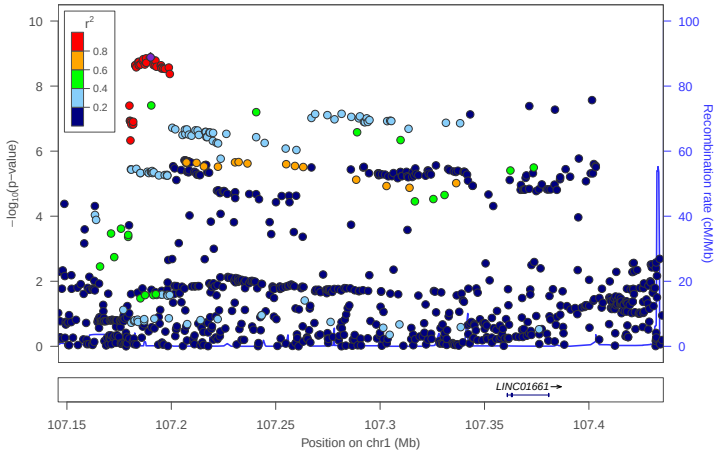
locus_5



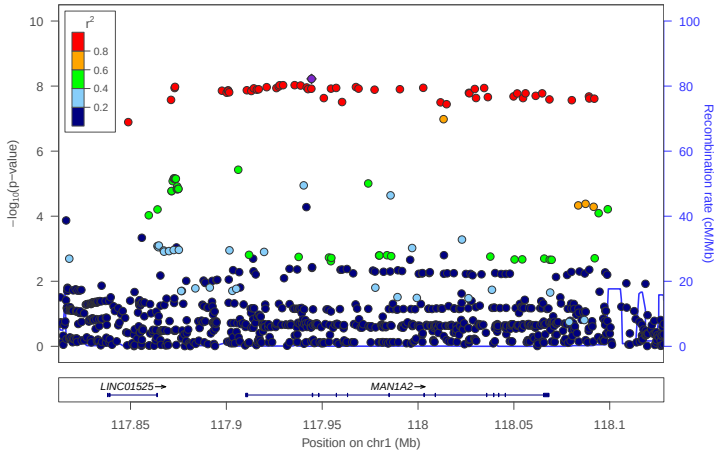
locus_6



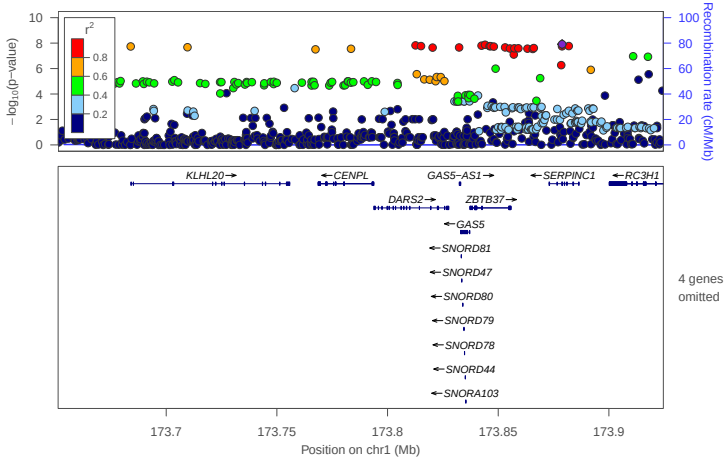
locus_7



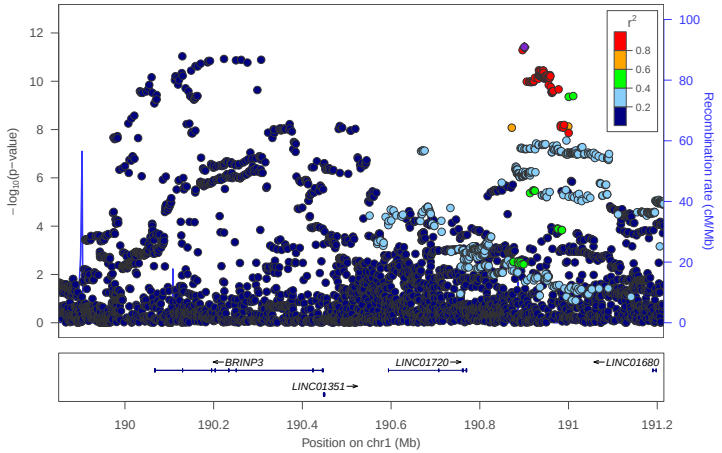
locus_8



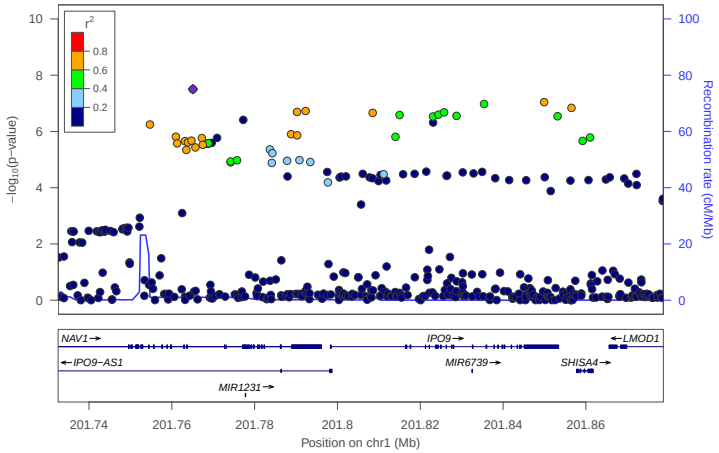
locus_9



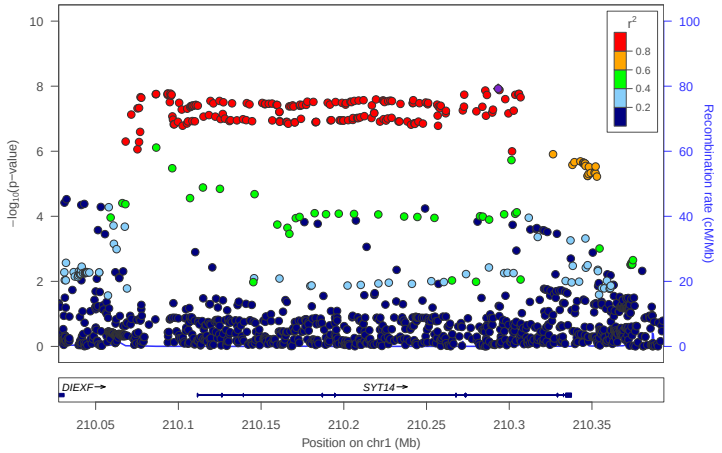
locus_10



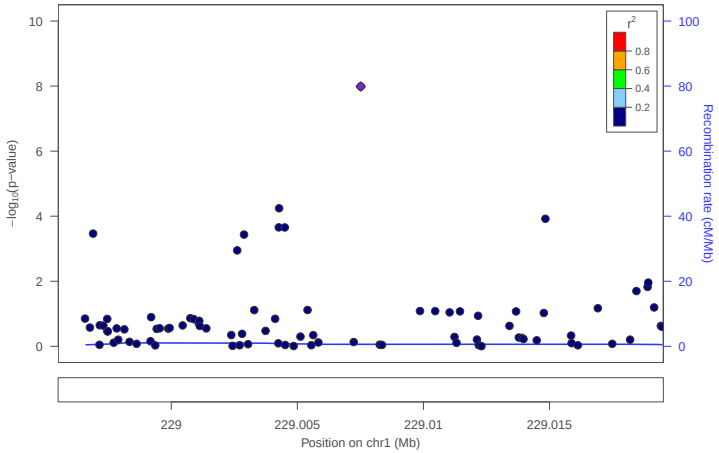
locus_11



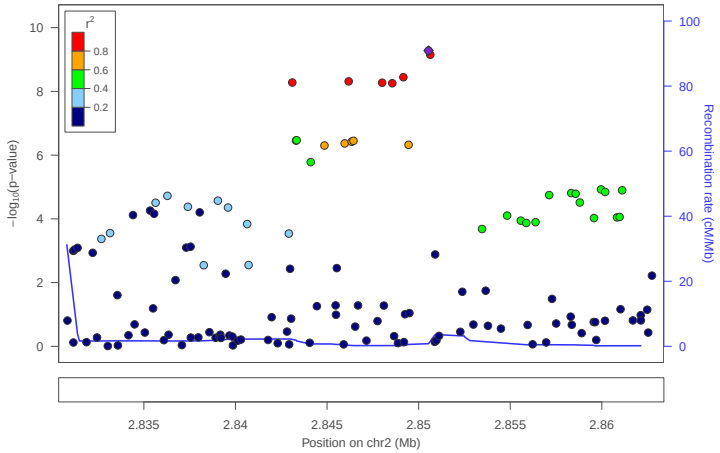
locus_12



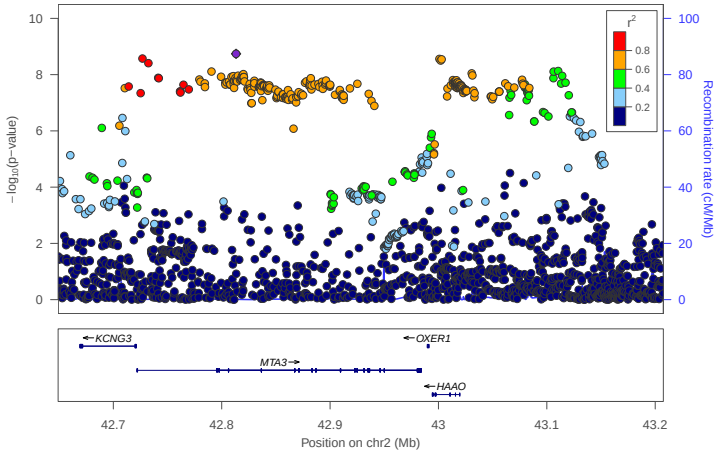
locus_13



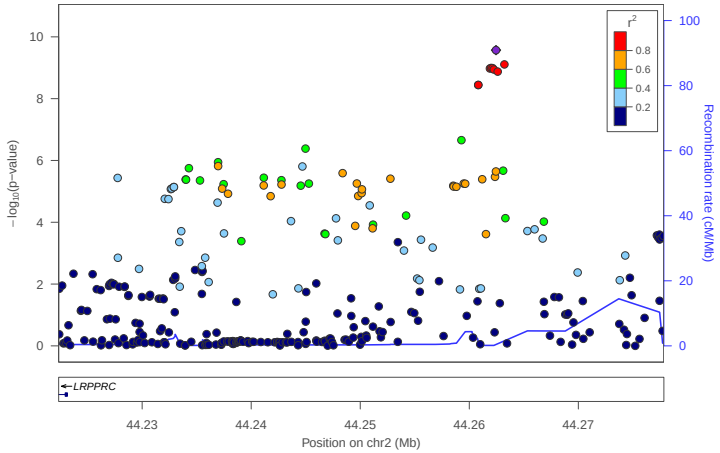
locus_14



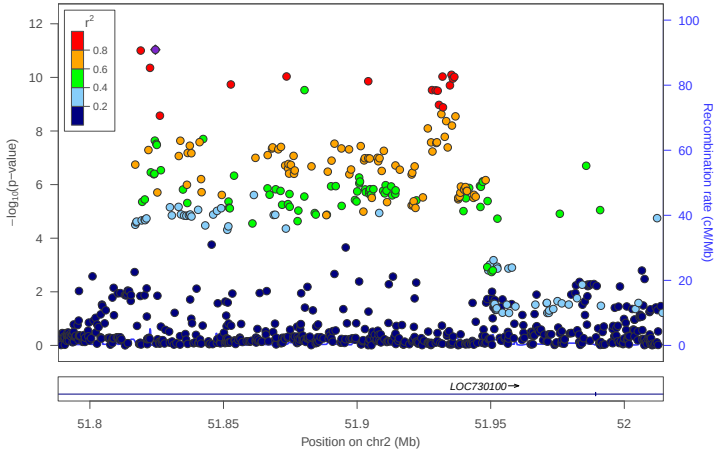
locus_15



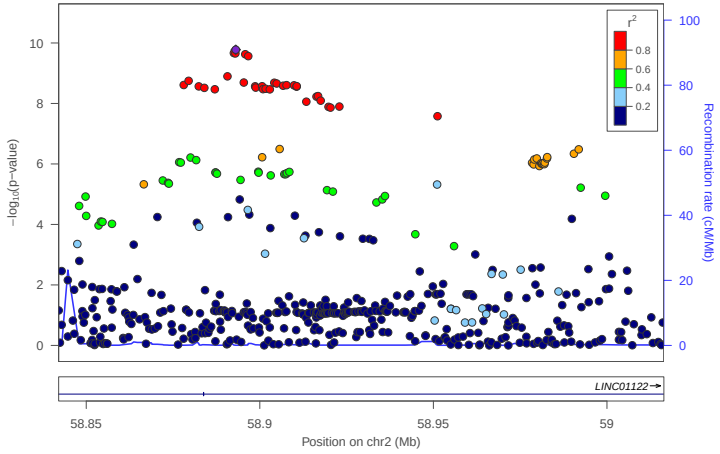
locus_16



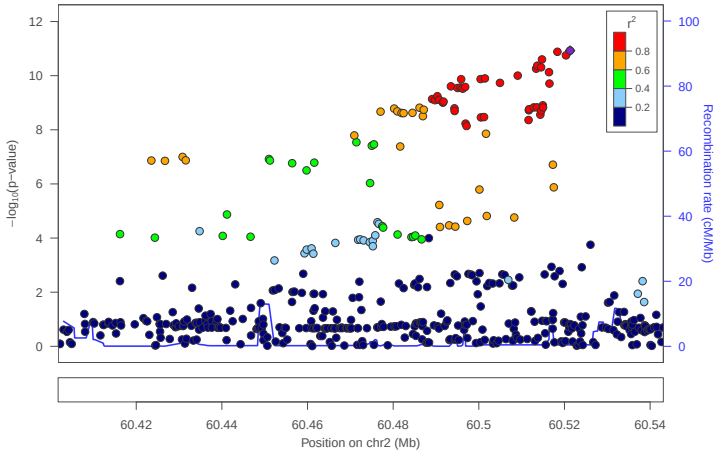
locus_17



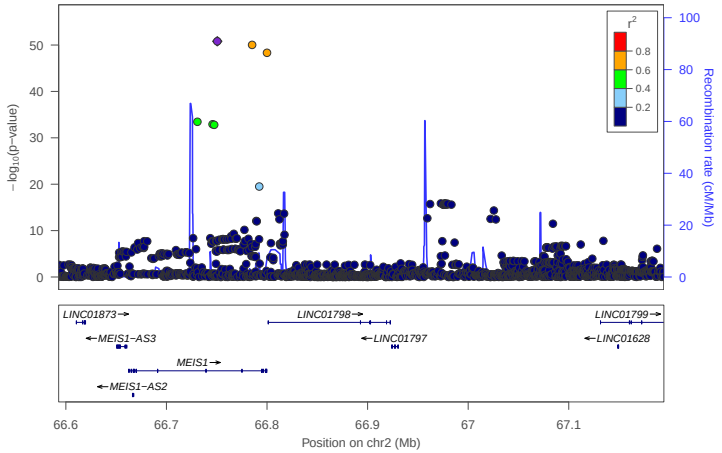
locus_18



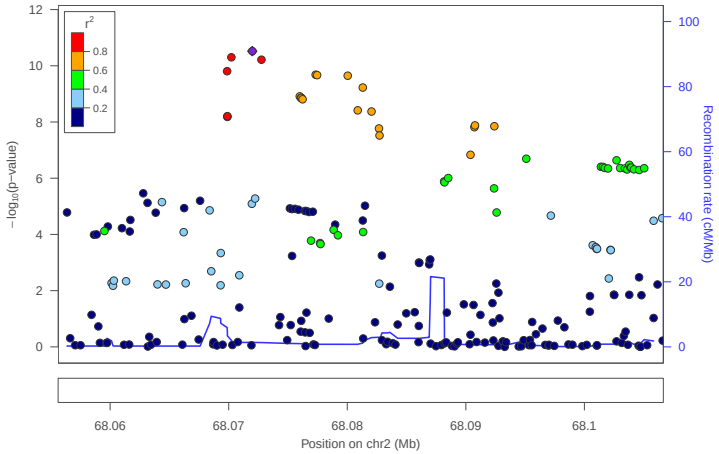
locus_19



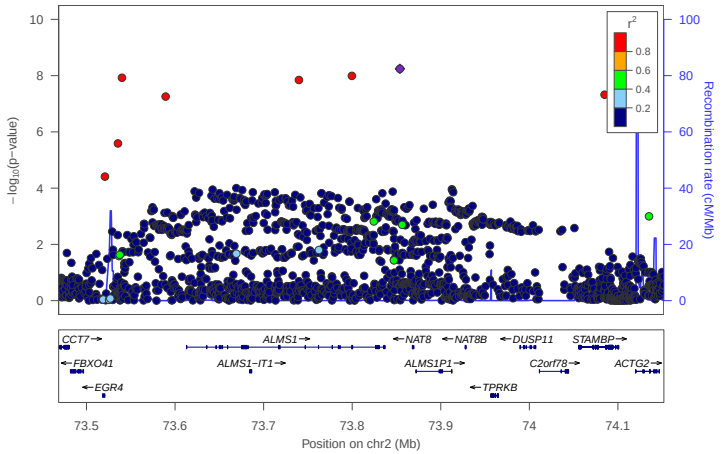
locus_20



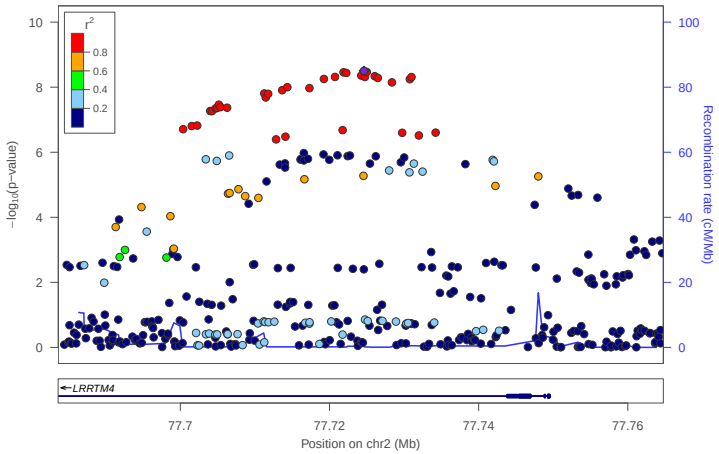
locus_21



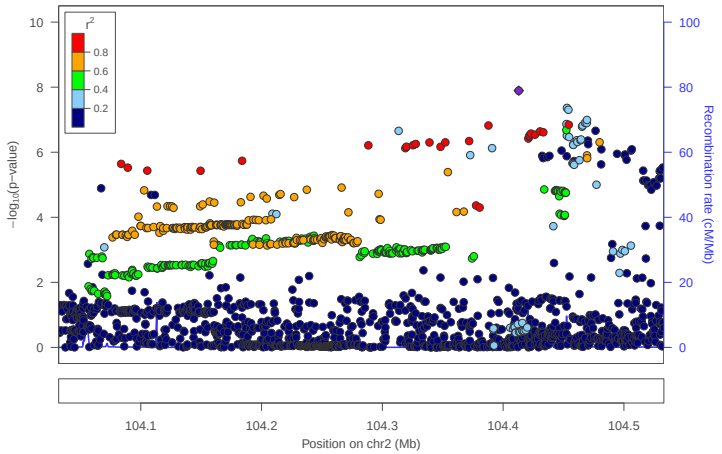
locus_22



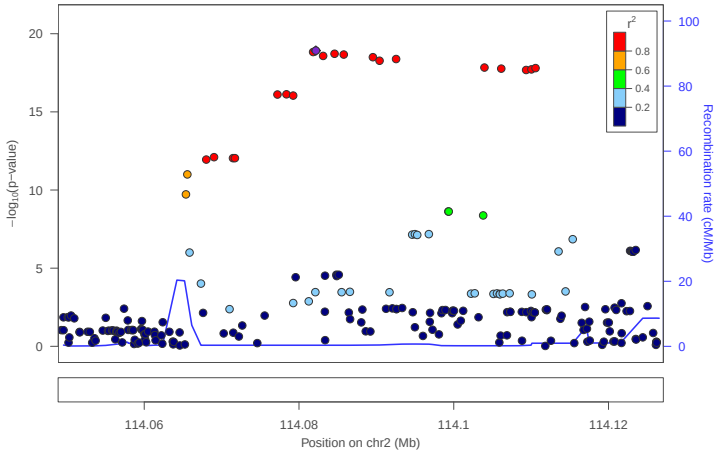
locus_23



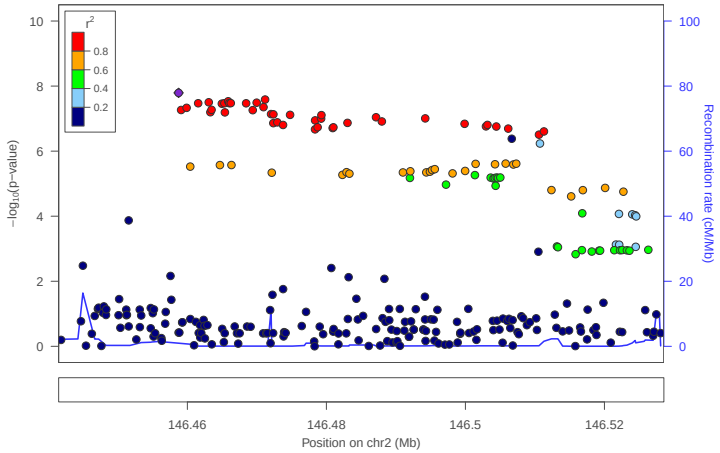
locus_24



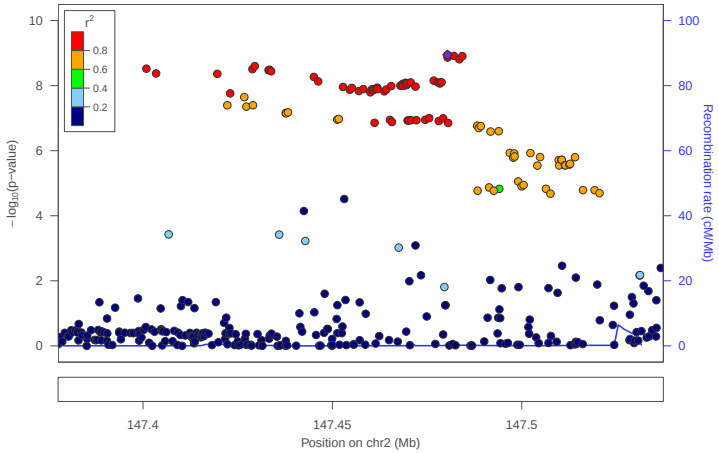
locus_25



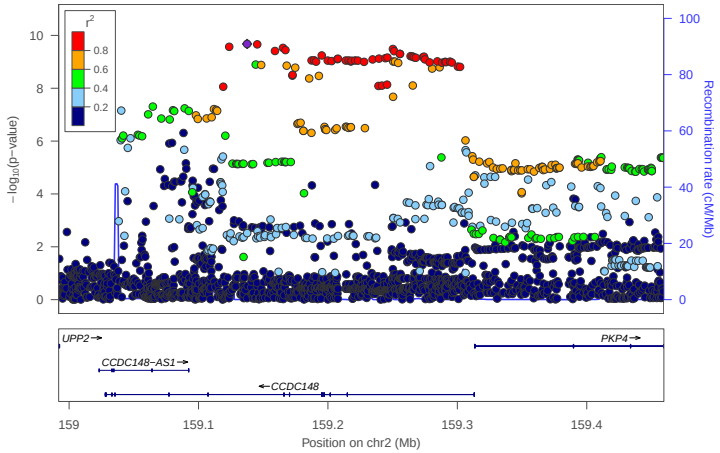
locus_26



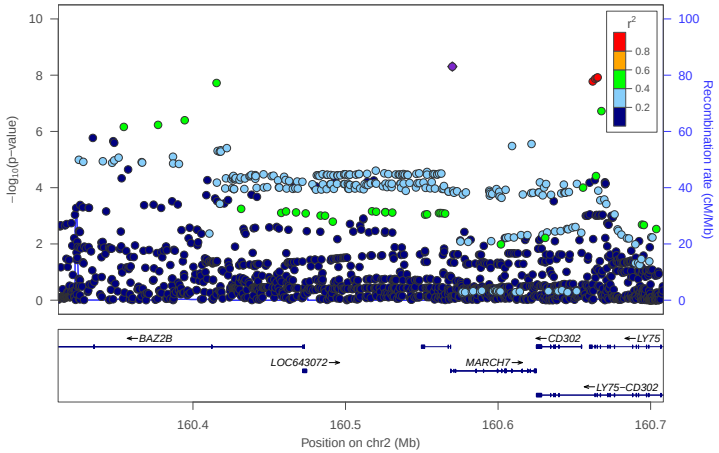
locus_27



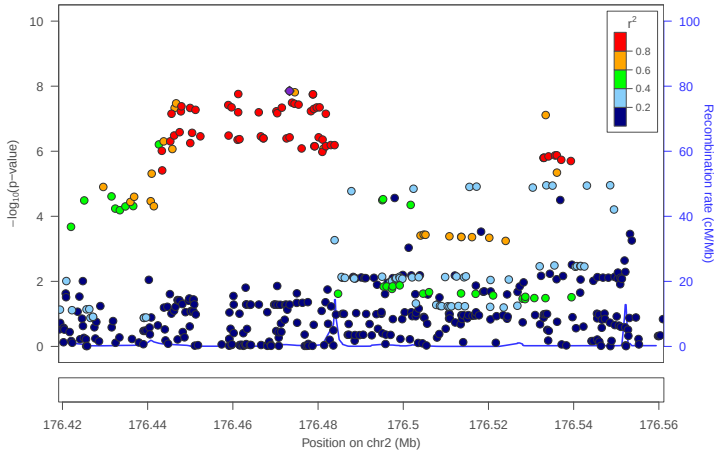
locus_28



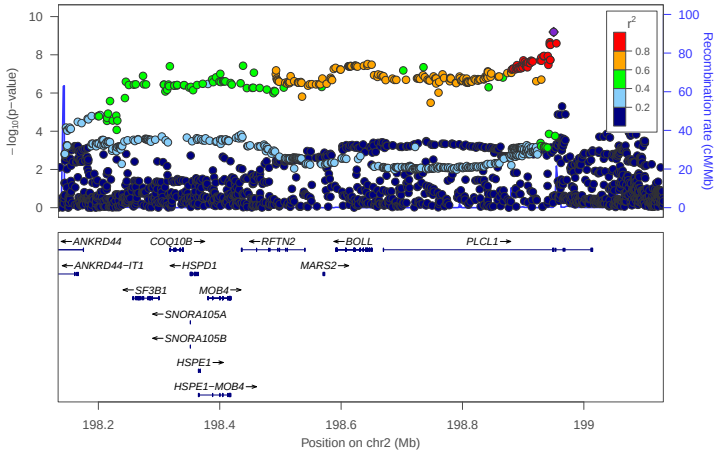
locus_29



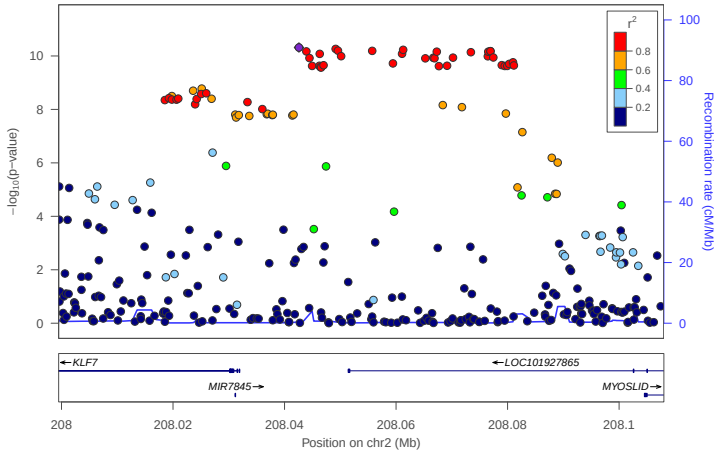
locus_30



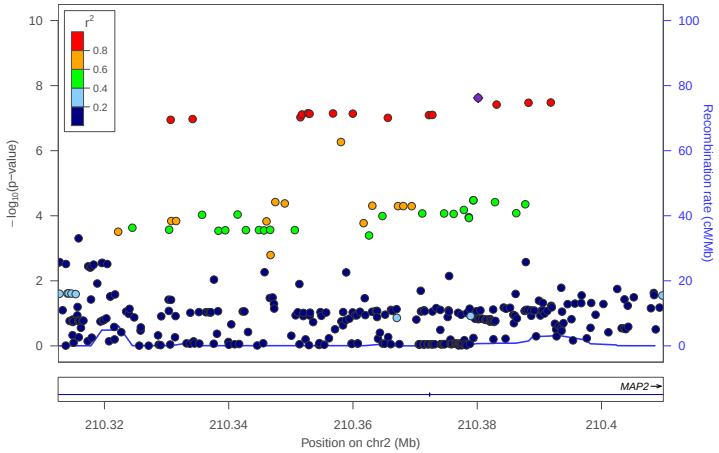
locus_31



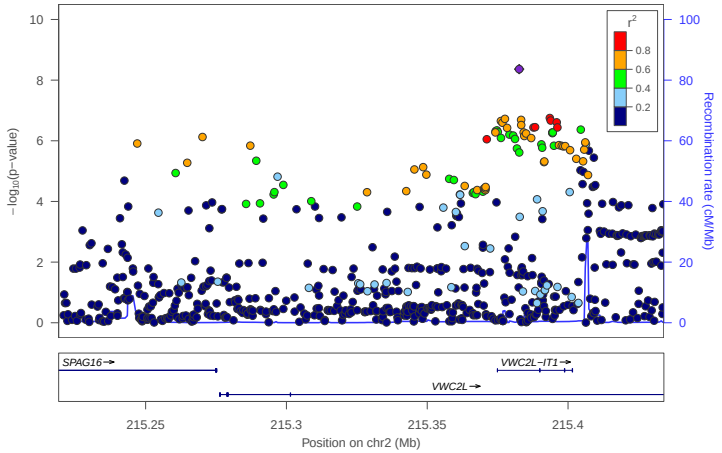
locus_32



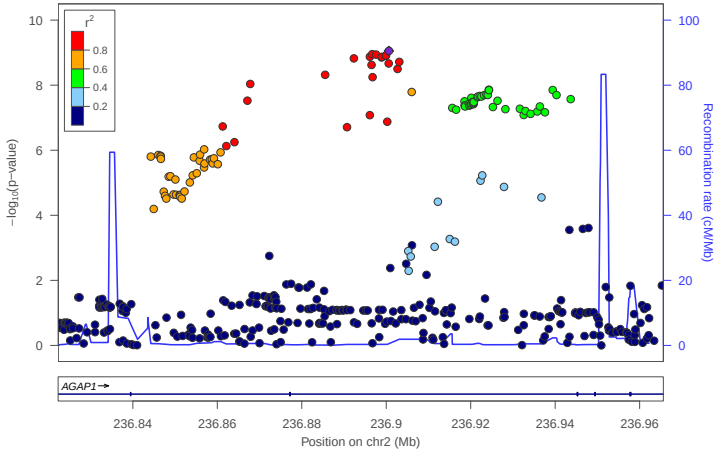
locus_33



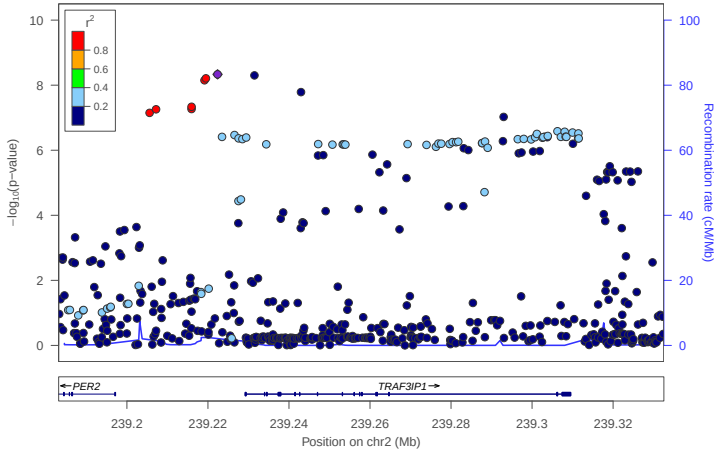
locus_34



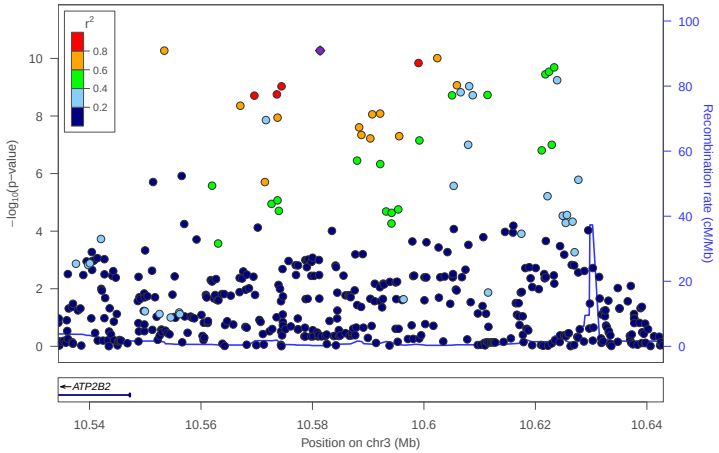
locus_35



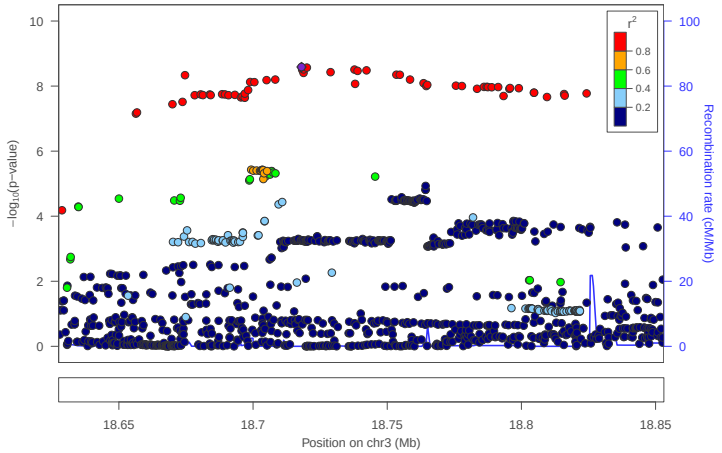
locus_36



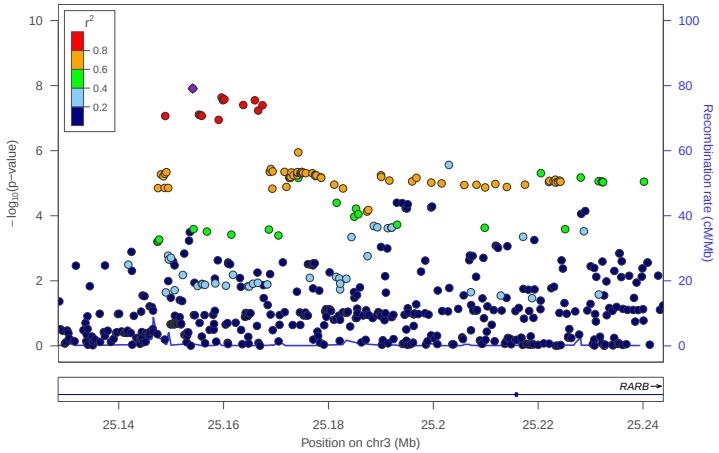
locus_37



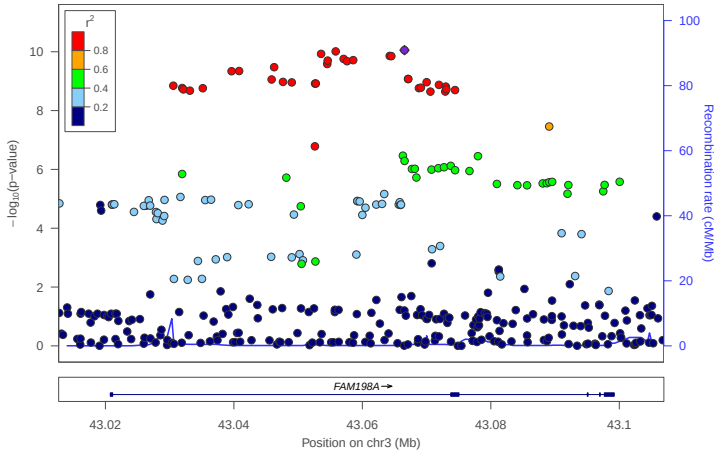
locus_38



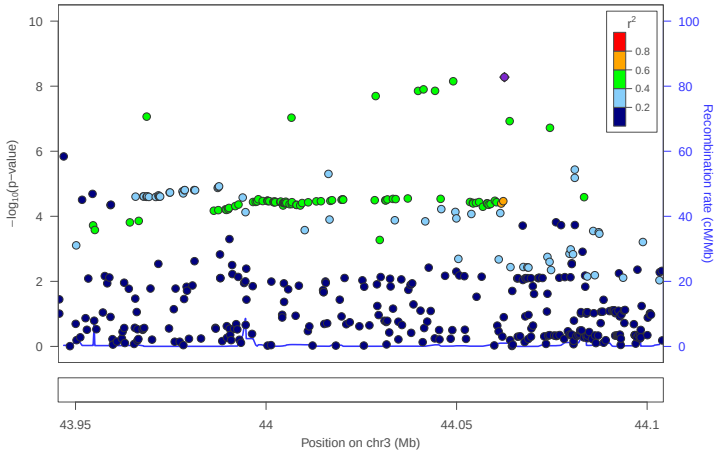
locus_39



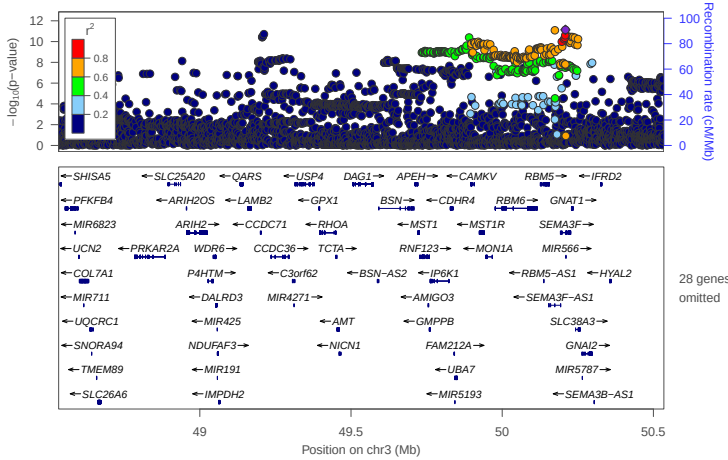
locus_40



locus_41

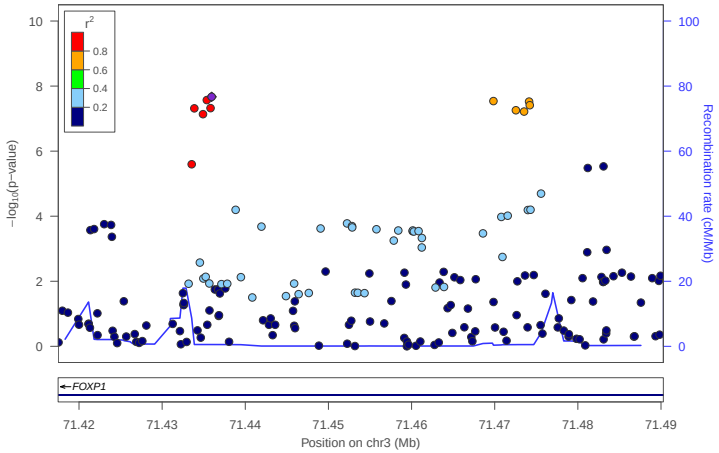


locus_42

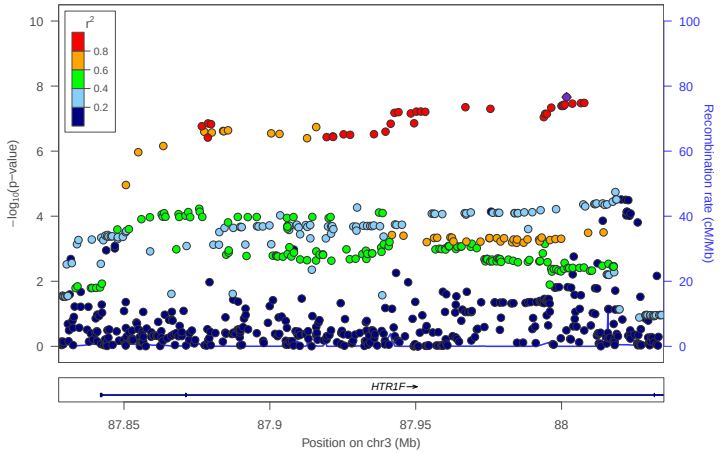


28 genes omitted

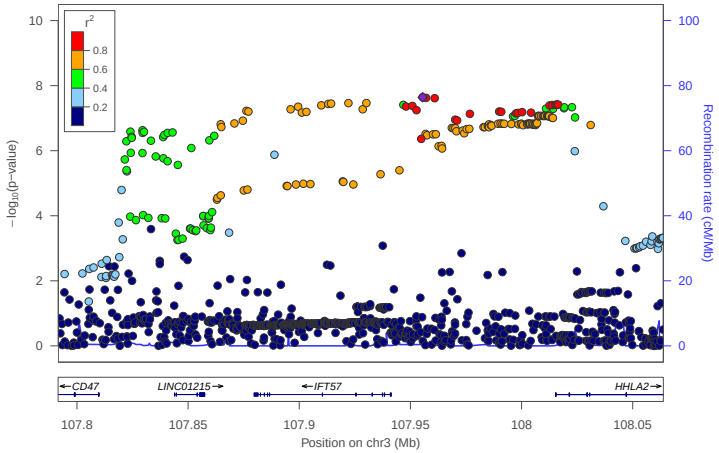
locus_43



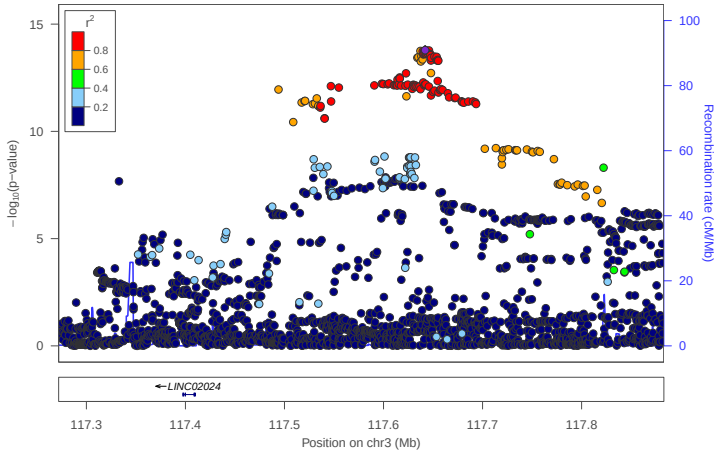
locus_44



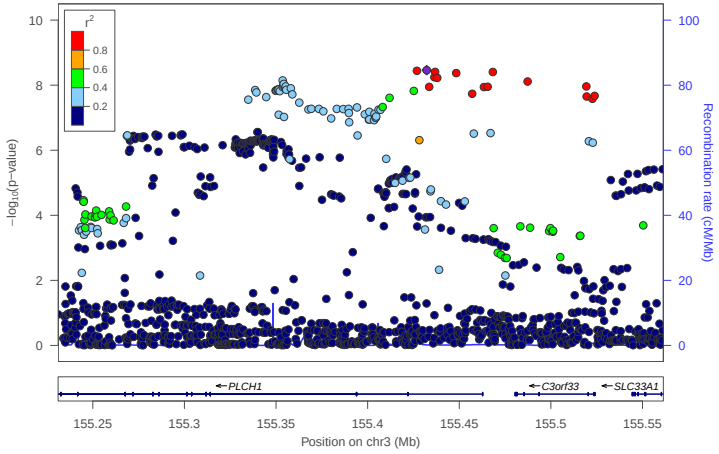
locus_45



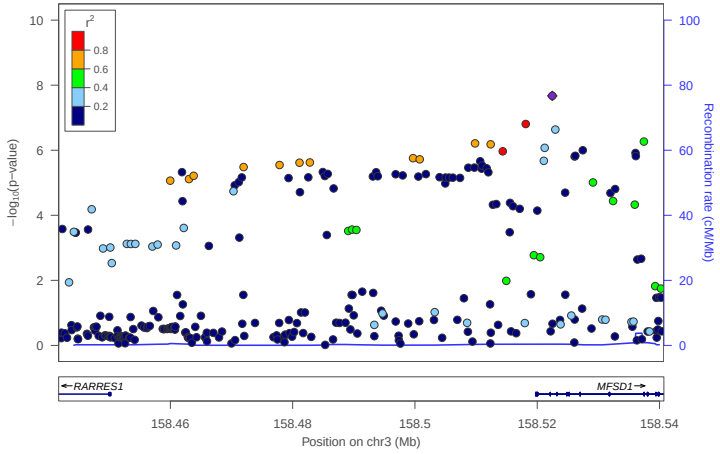
locus_46



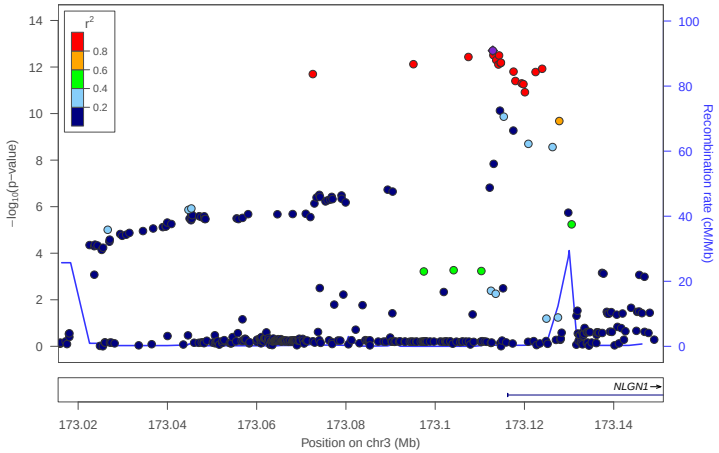
locus_47



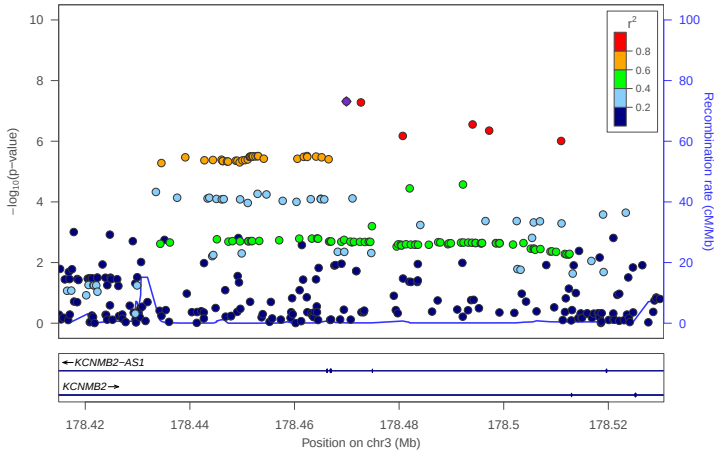
locus_48



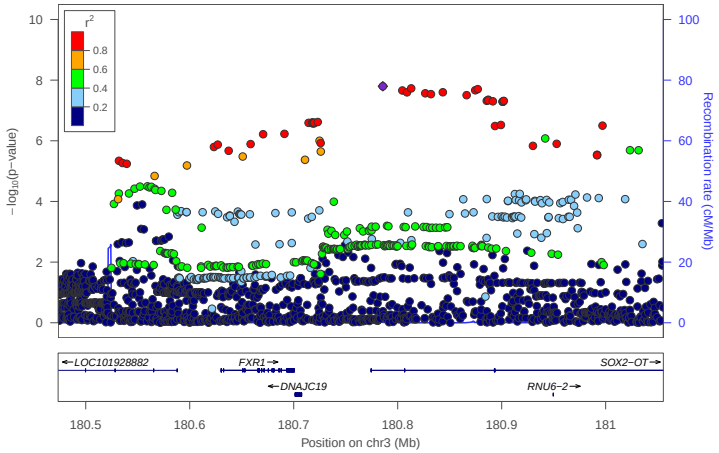
locus_49



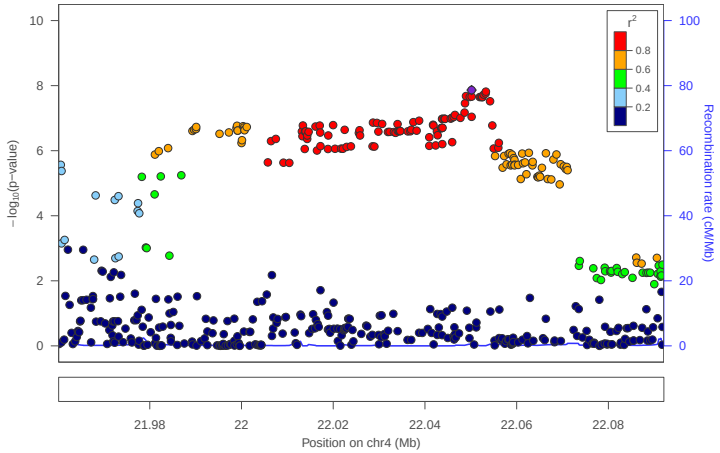
locus_50



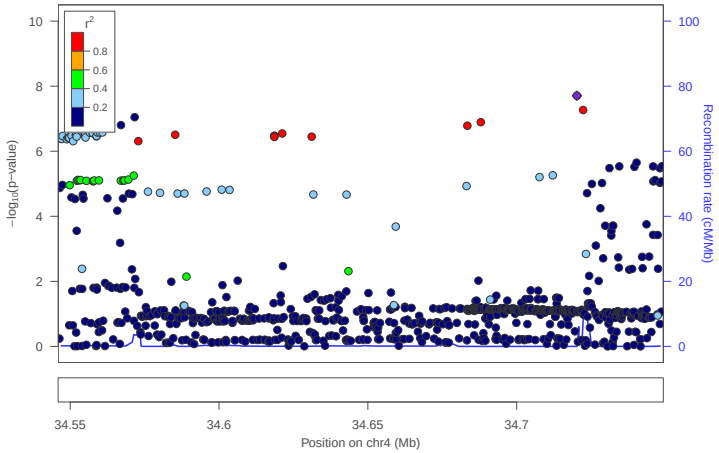
locus_51



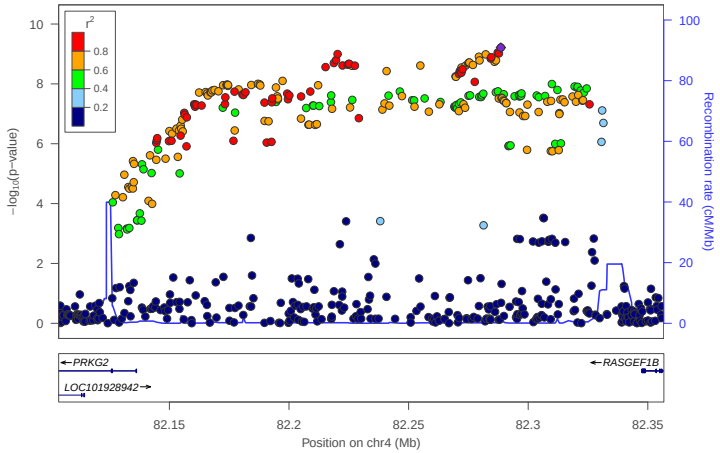
locus_52



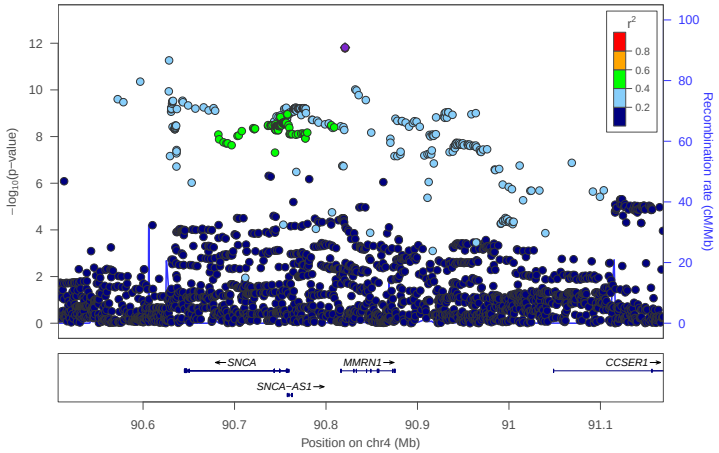
locus_53



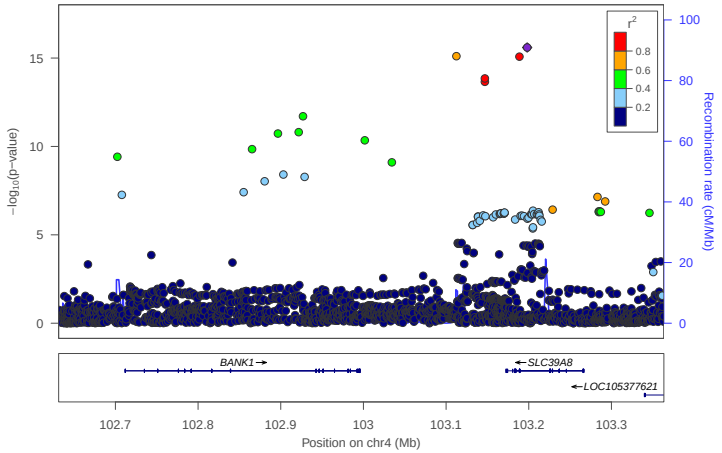
locus_54



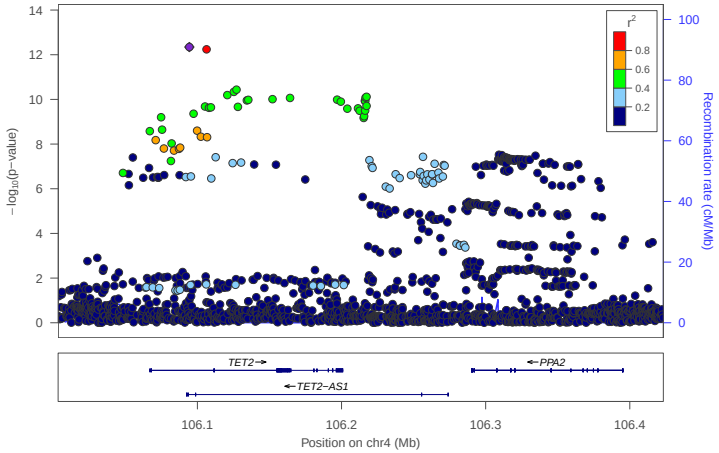
locus_55



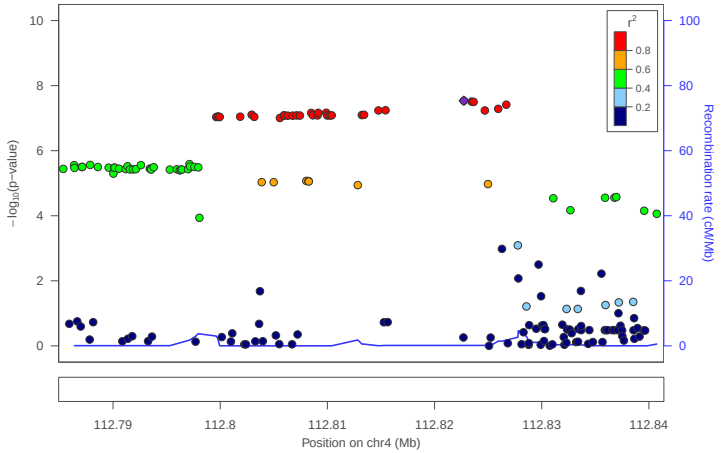
locus_56



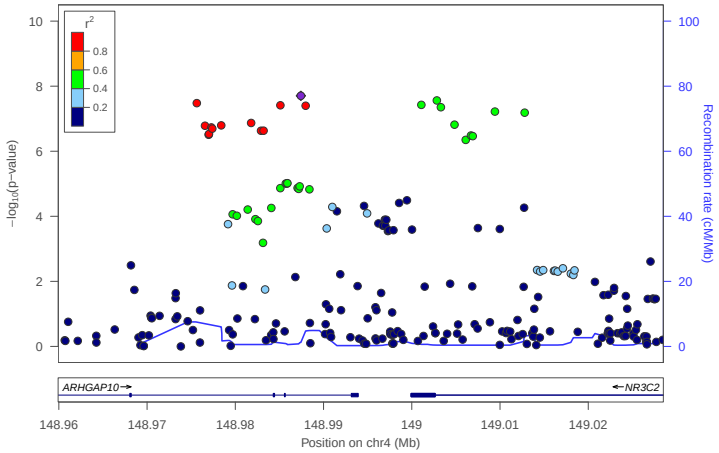
locus_57



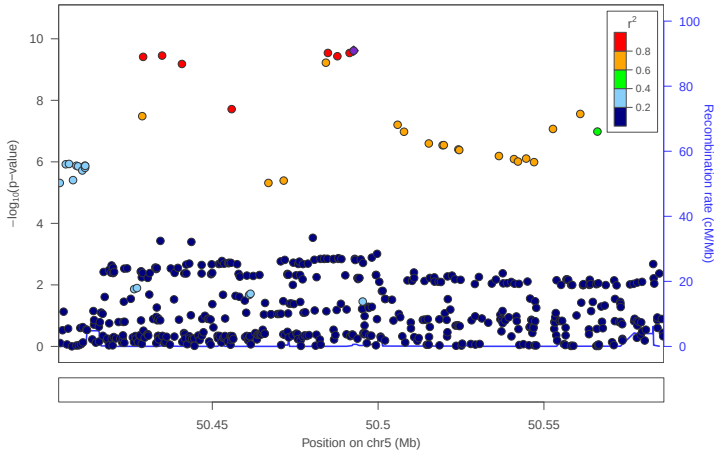
locus_58



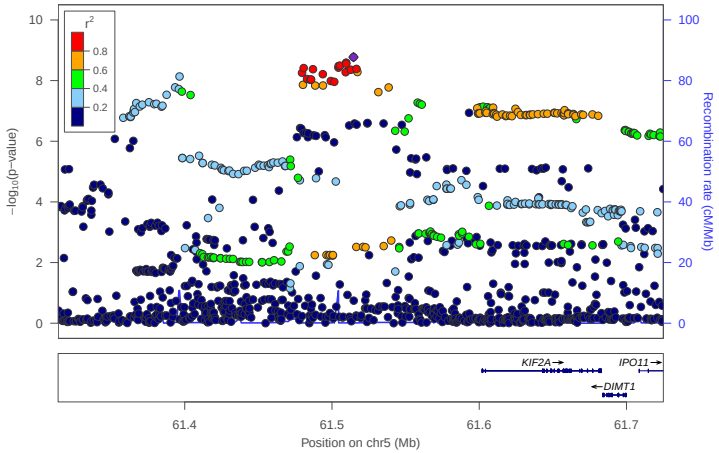
locus_59



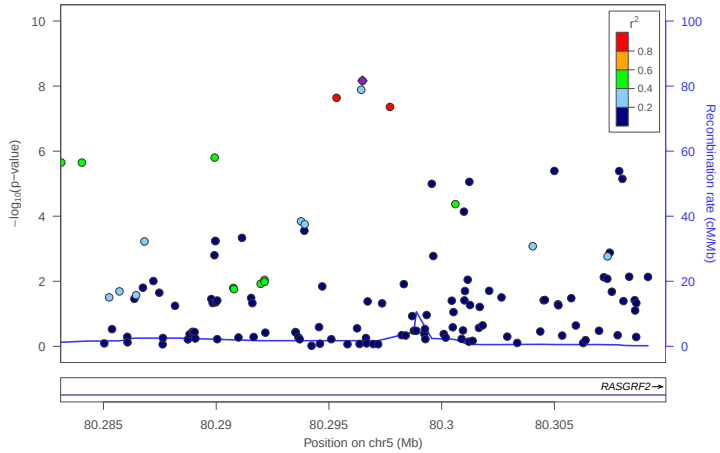
locus_60



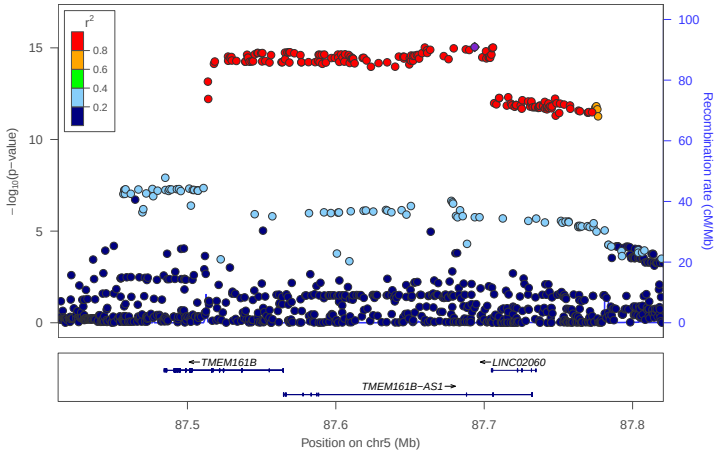
locus_61



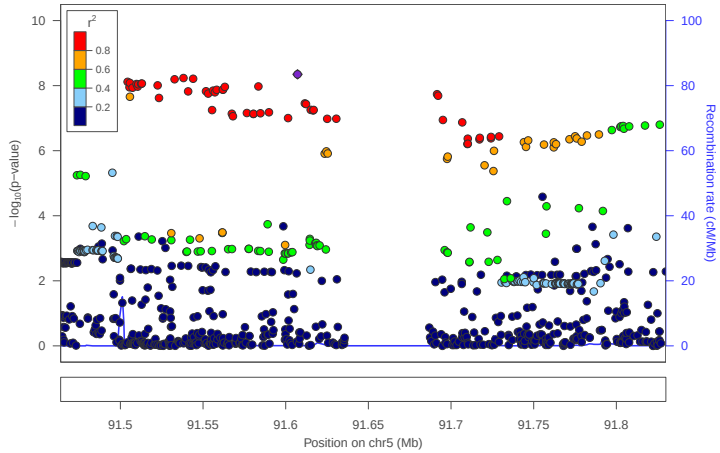
locus_62



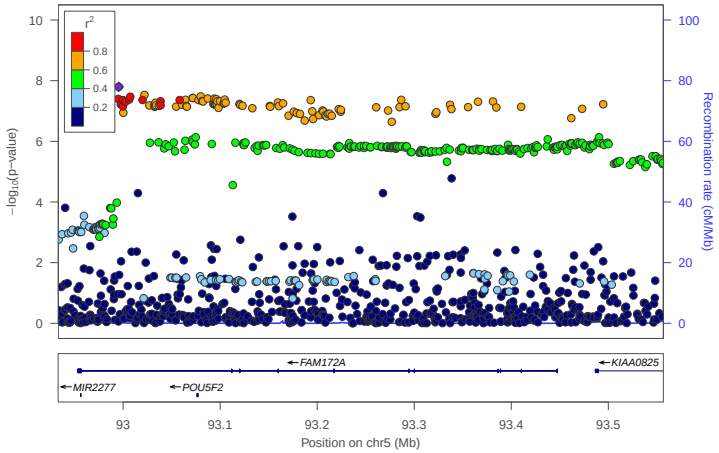
locus_63



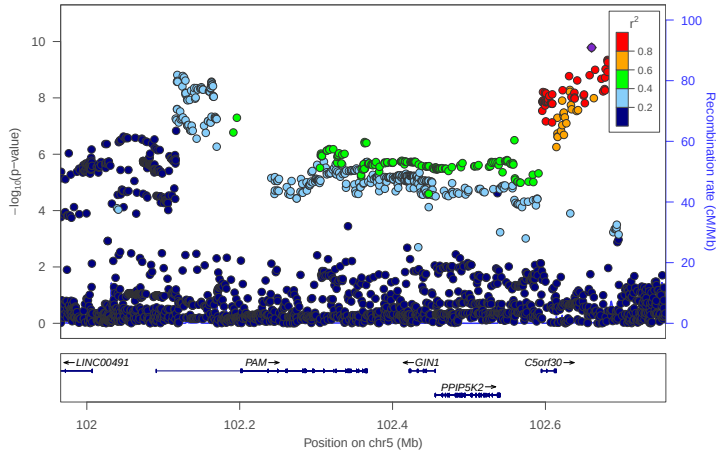
locus_64



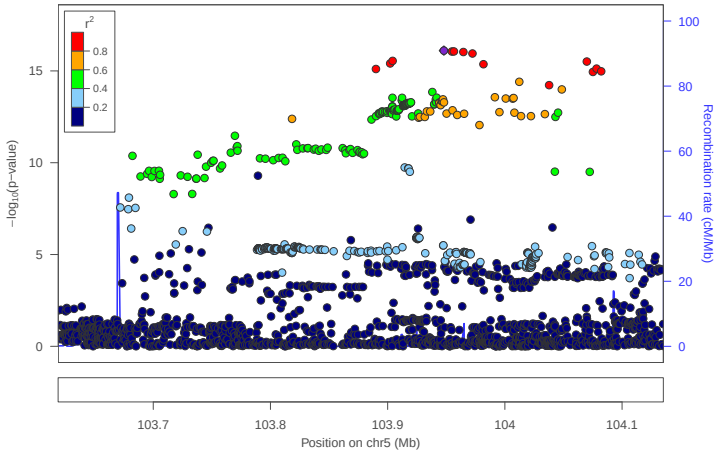
locus_65



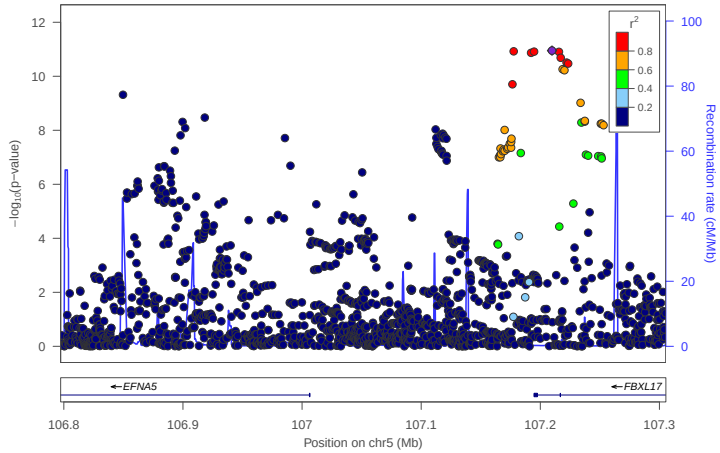
locus_66



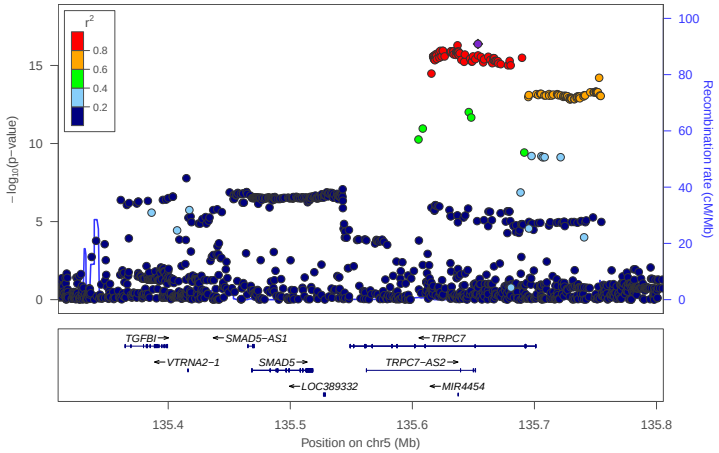
locus_67



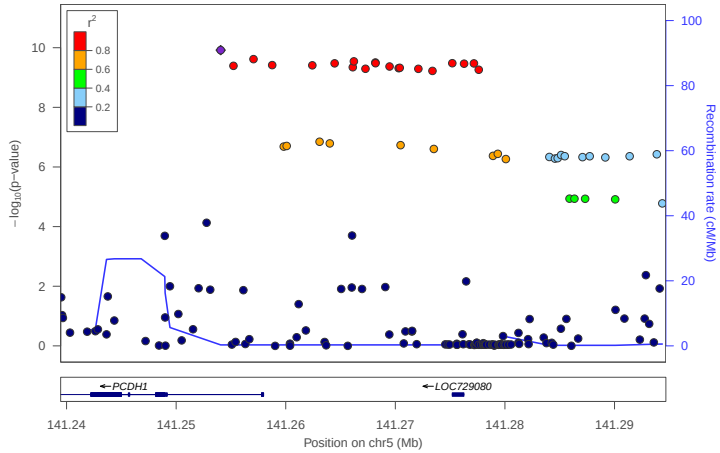
locus_68



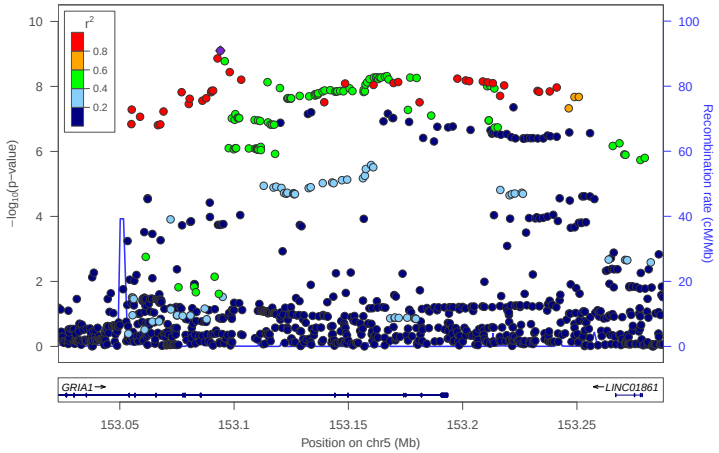
locus_69



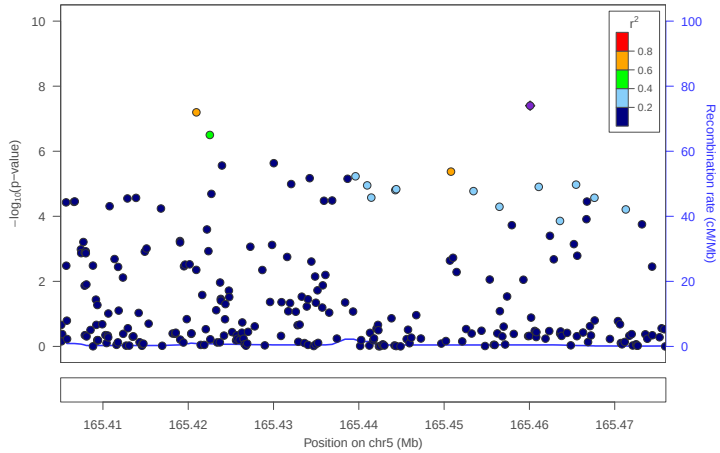
locus_70



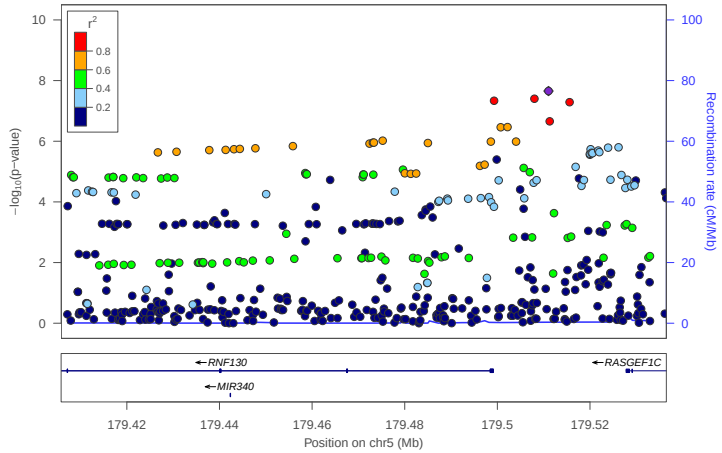
locus_71



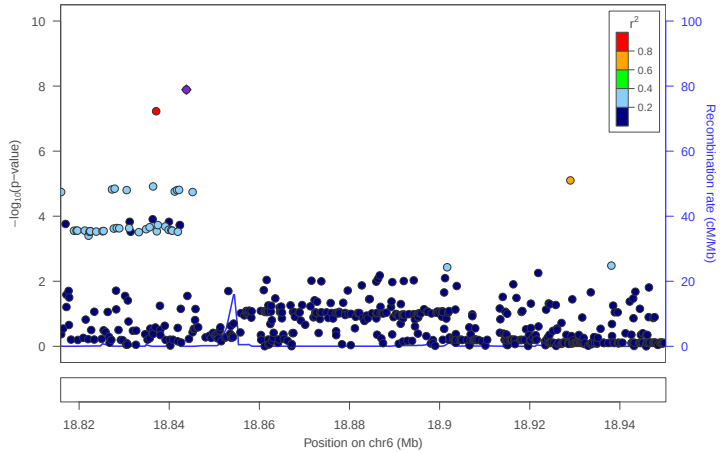
locus_72



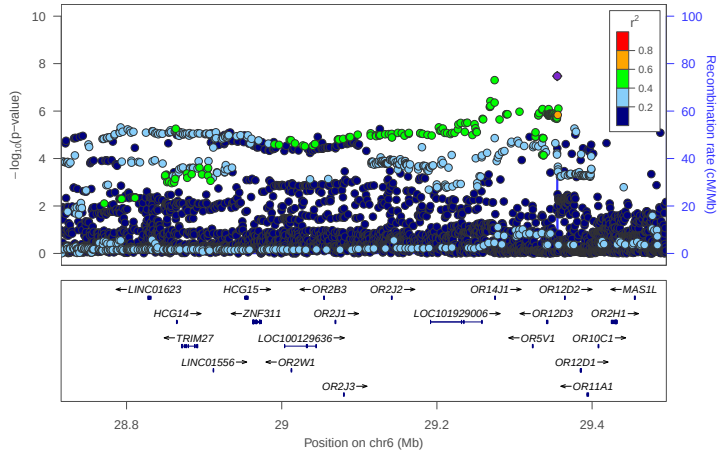
locus_73



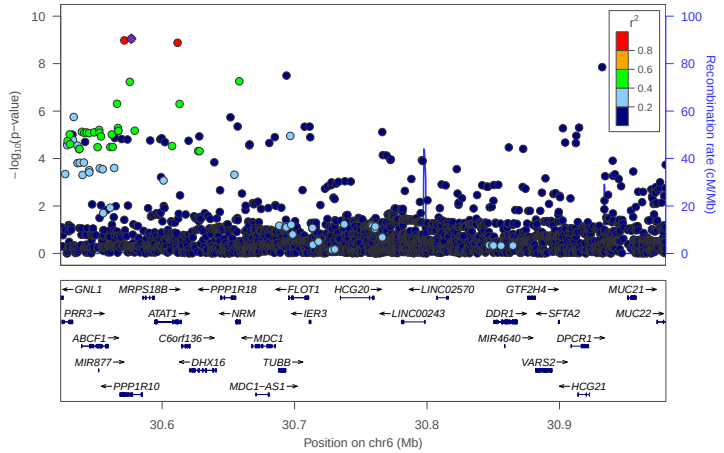
locus_74



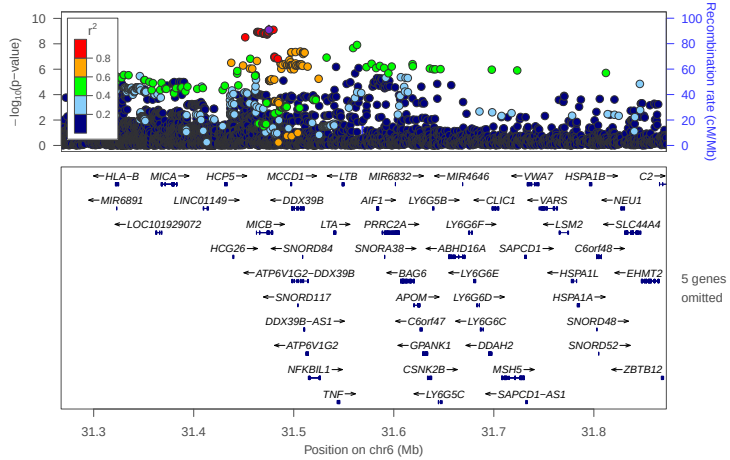
locus_75



locus_76

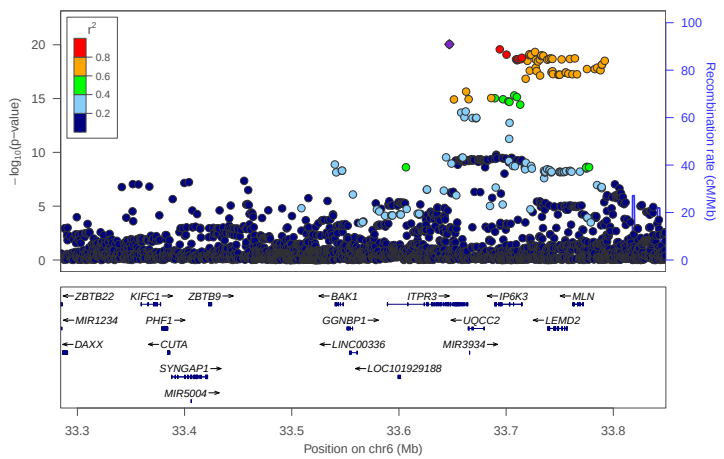


locus_77

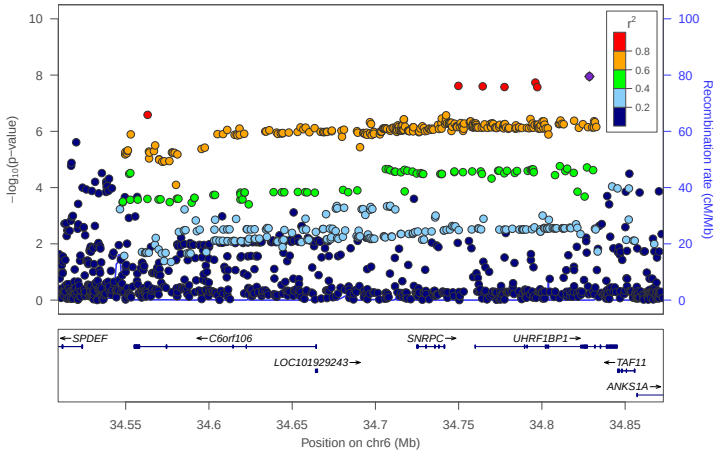


5 genes omitted

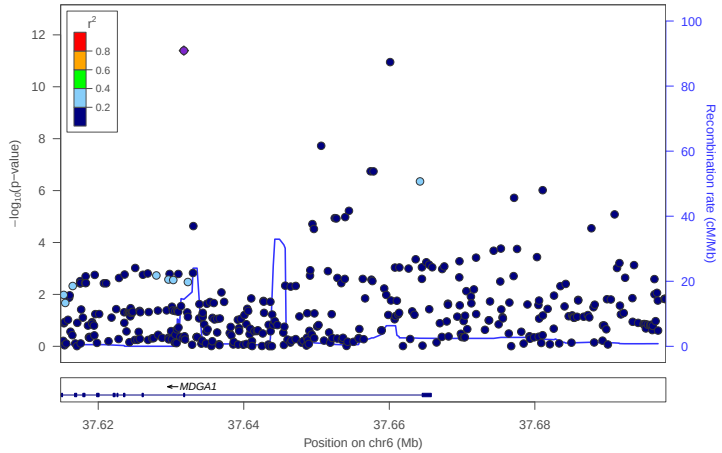
locus_78



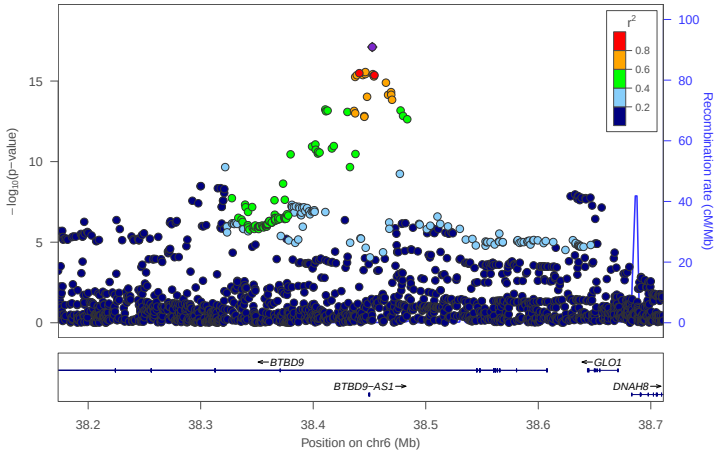
locus_79



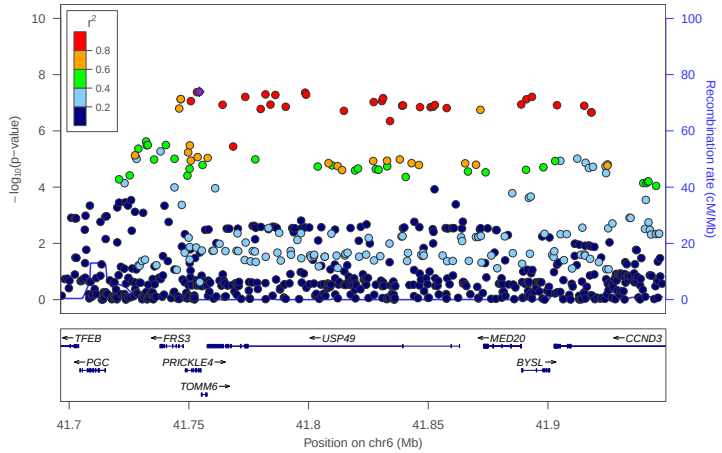
locus_80



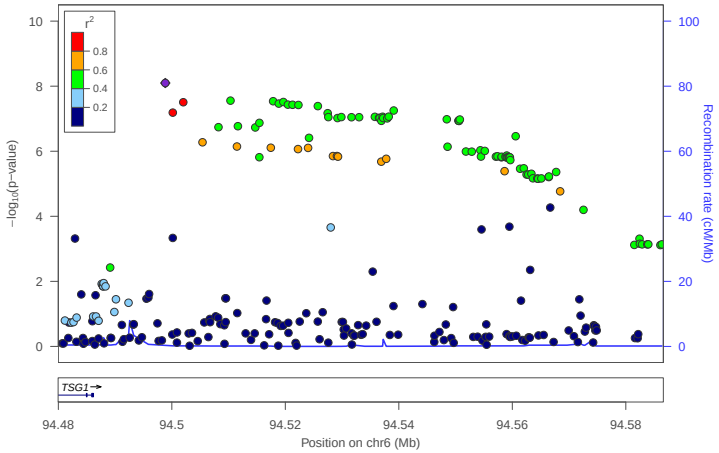
locus_81



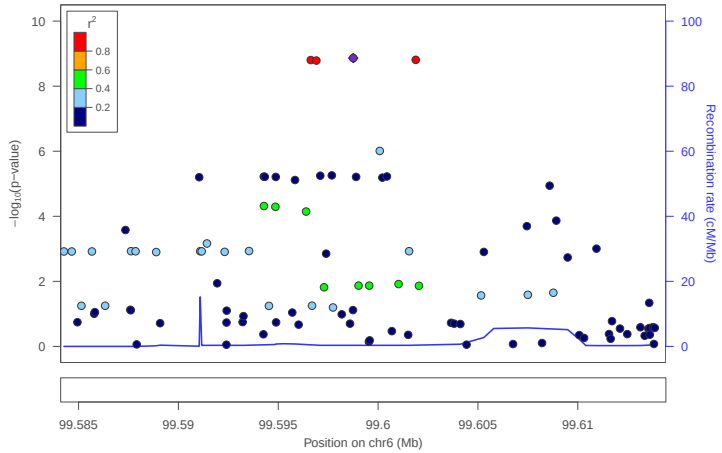
locus_82



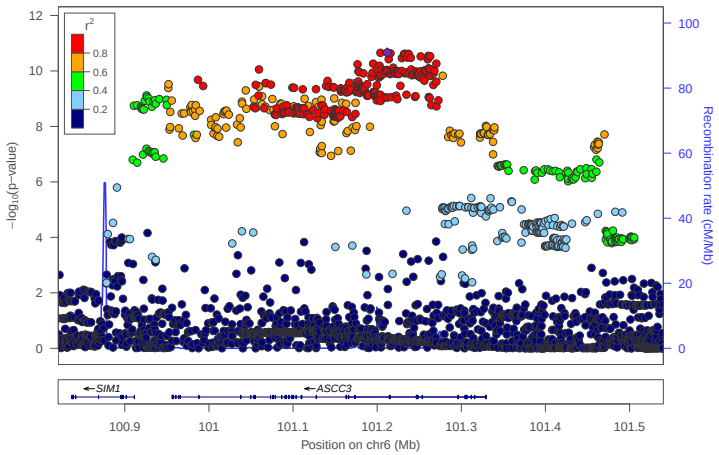
locus_83



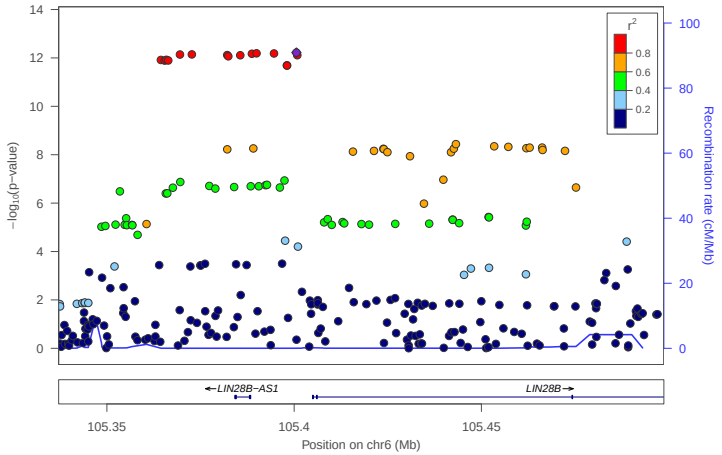
locus_84



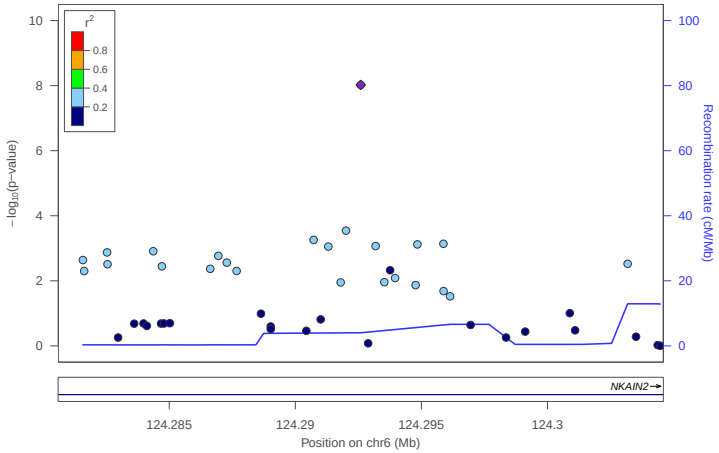
locus_85



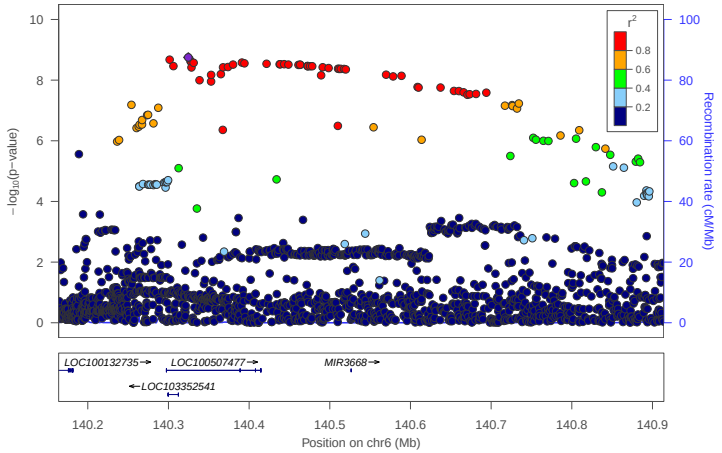
locus_86



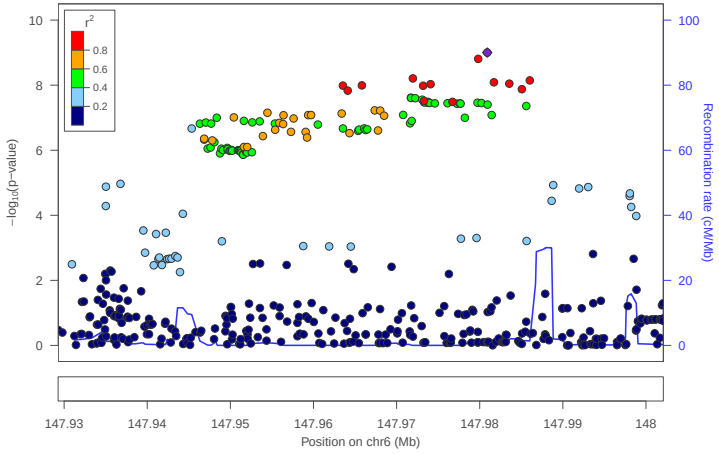
locus_87



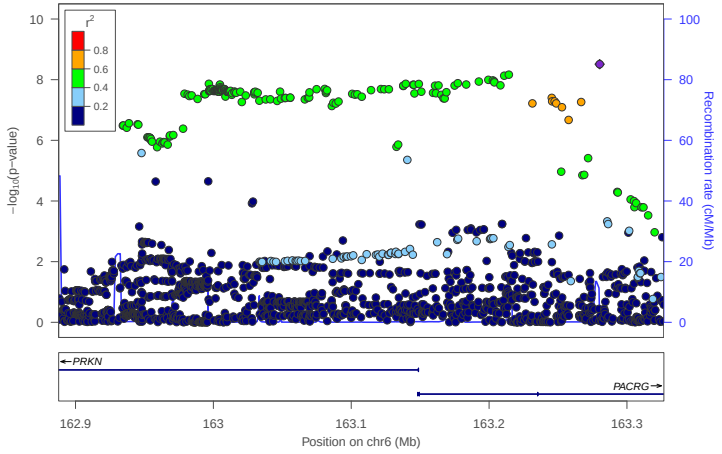
locus_88



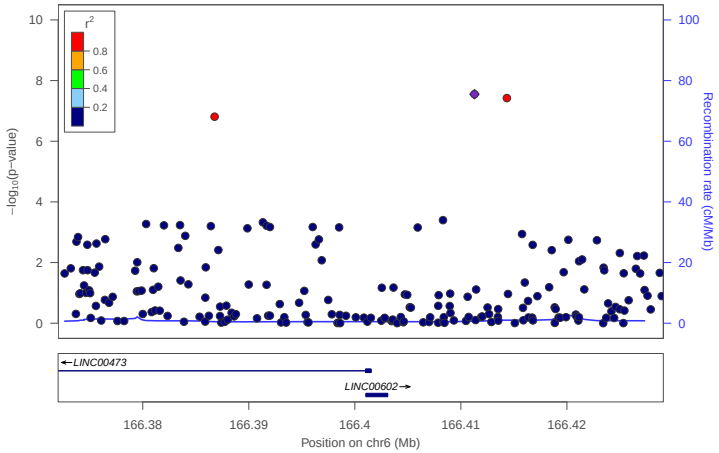
locus_89



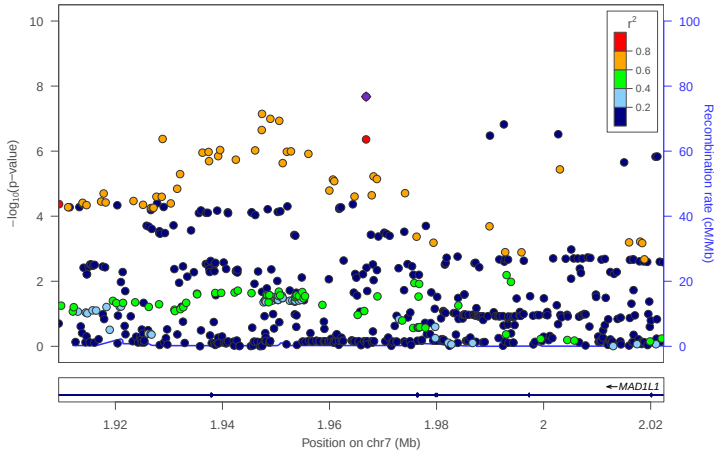
locus_90



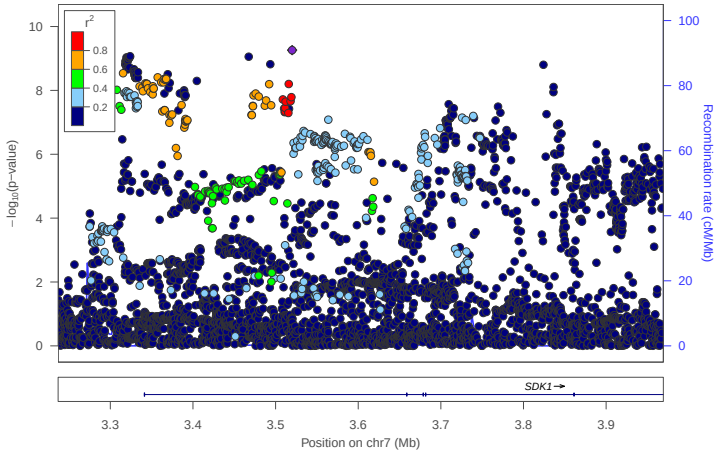
locus_91



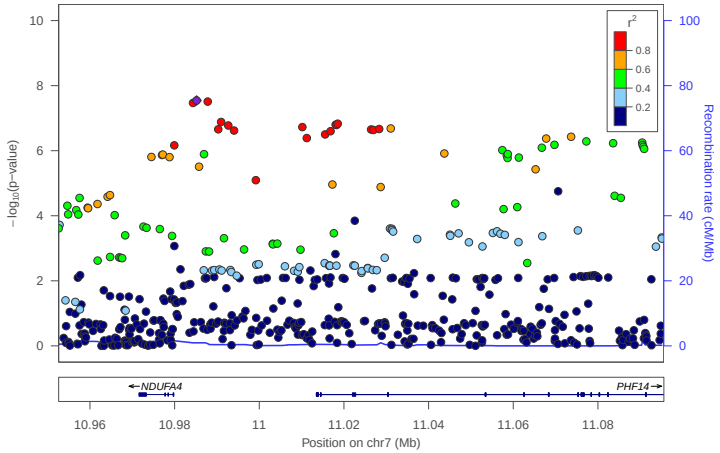
locus_92



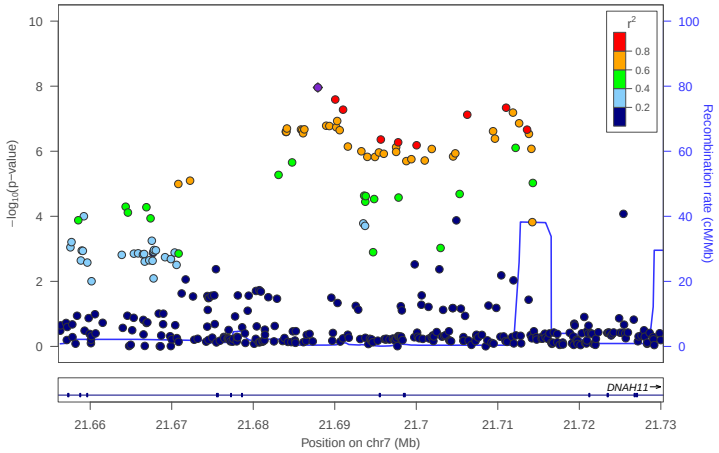
locus_93



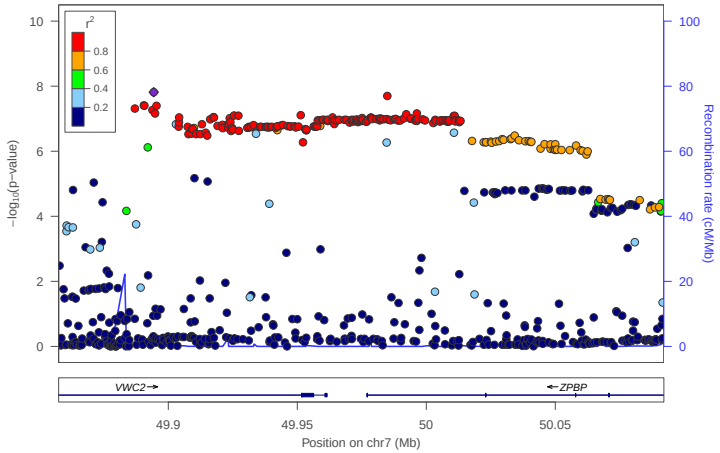
locus_94



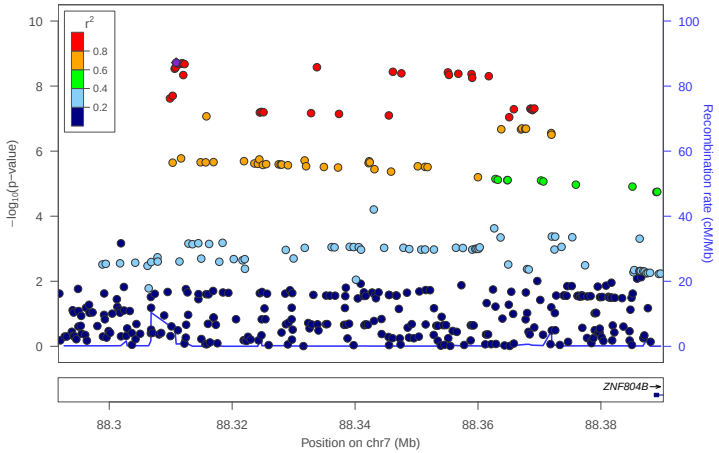
locus_95



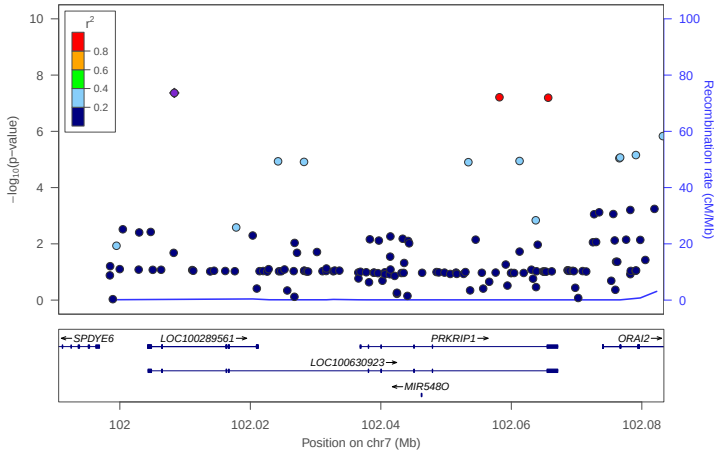
locus_96



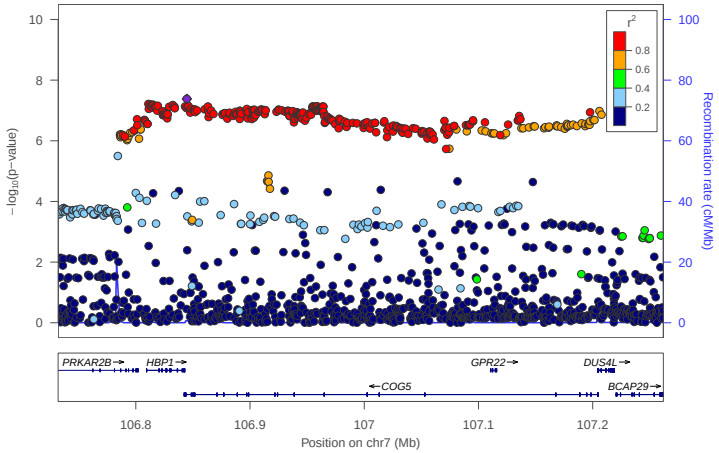
locus_97



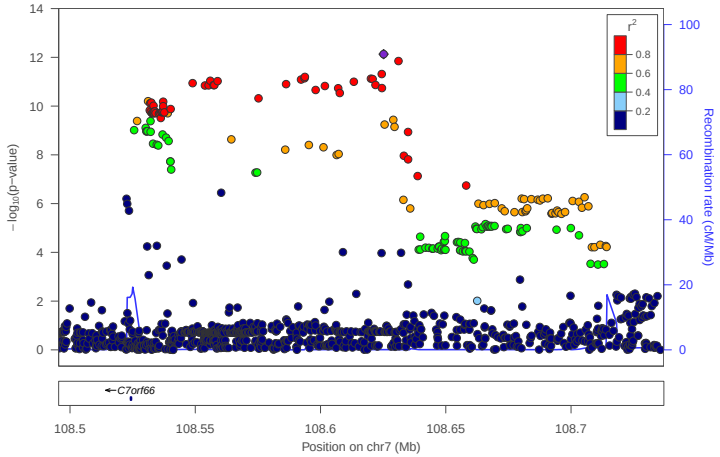
locus_98



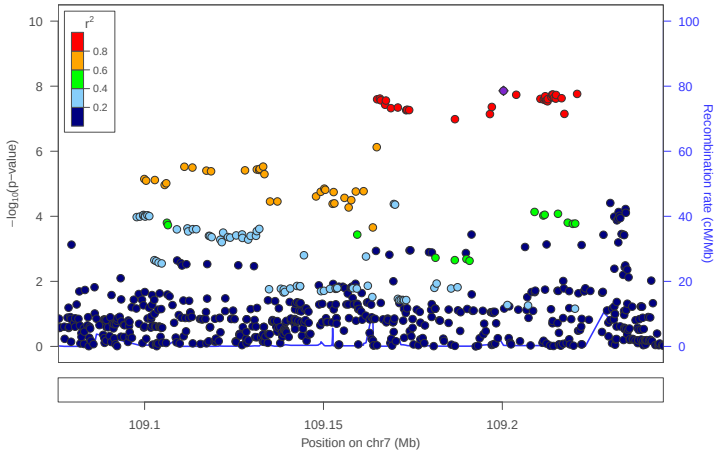
locus_99



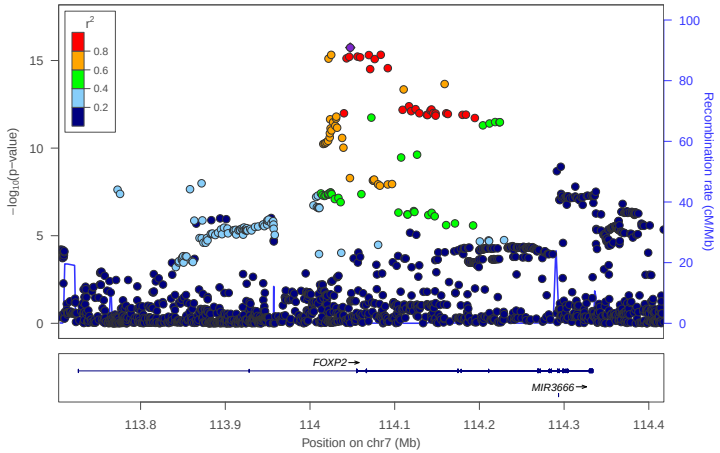
locus_100



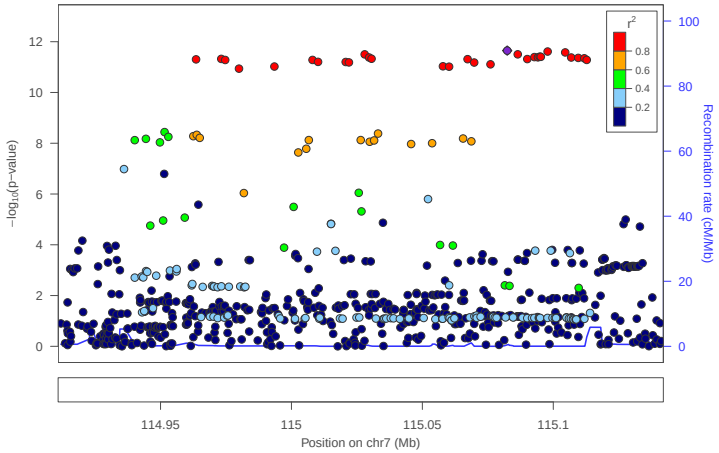
locus_101



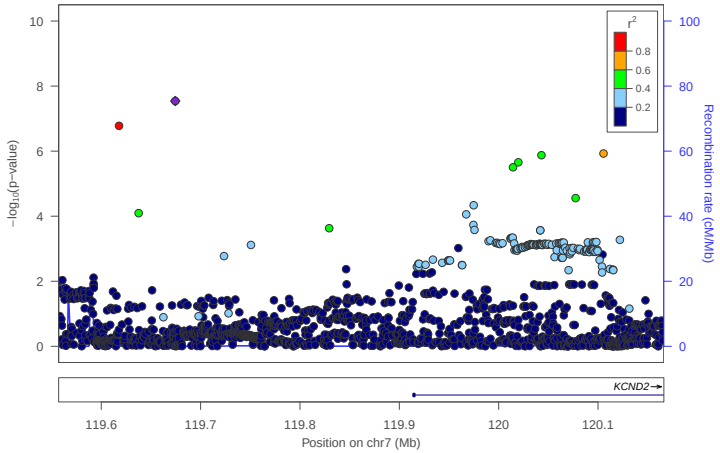
locus_102



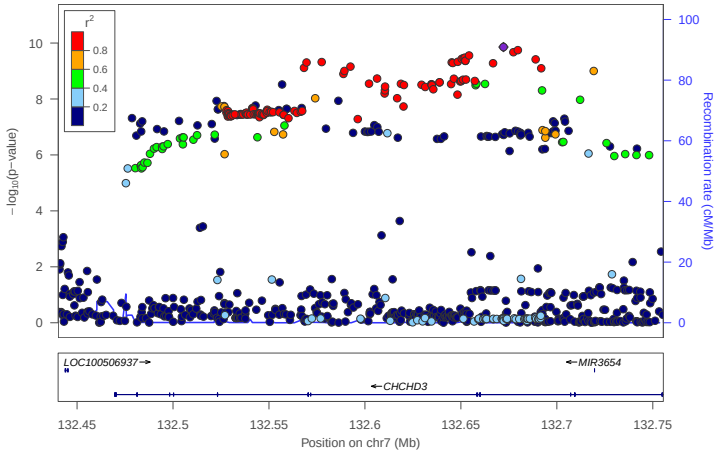
locus_103



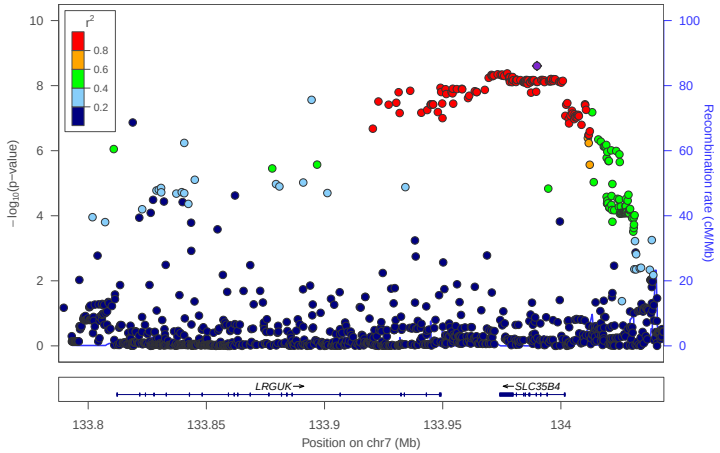
locus_104



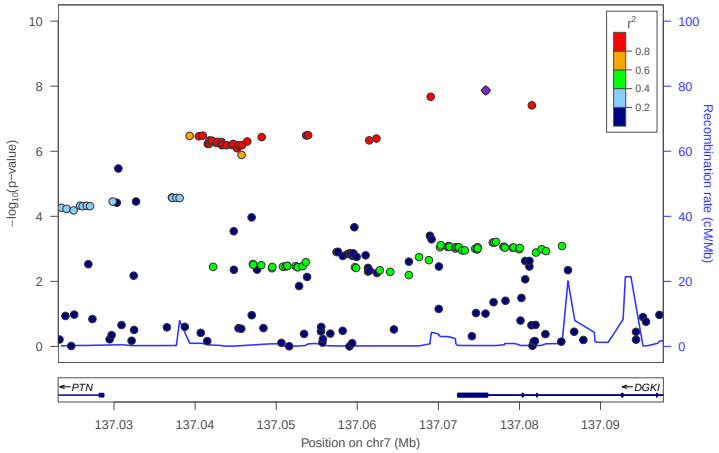
locus_105



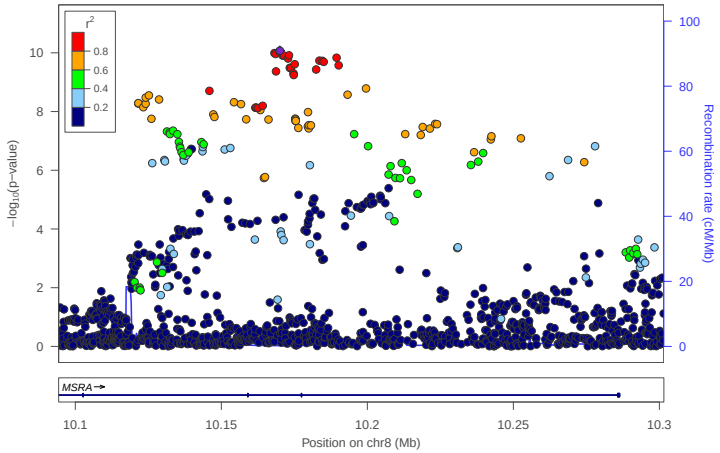
locus_106



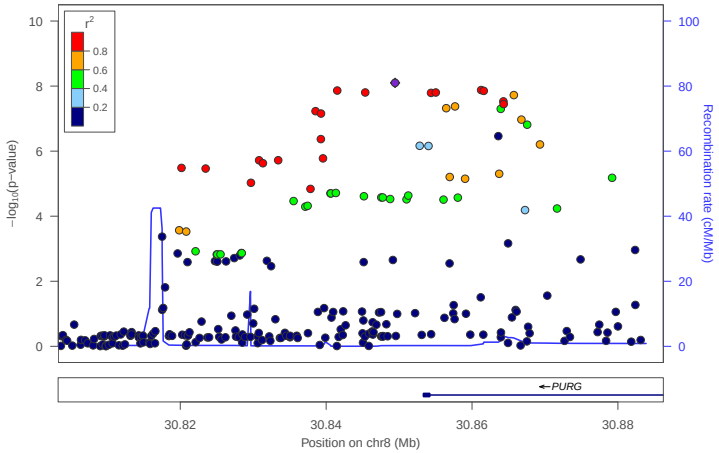
locus_107



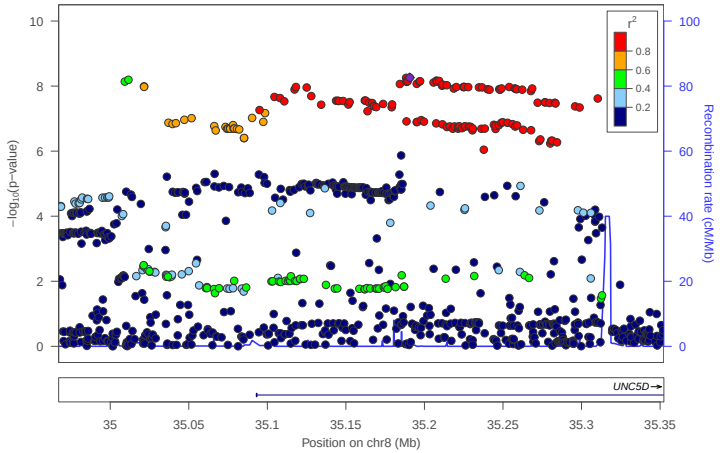
locus_108



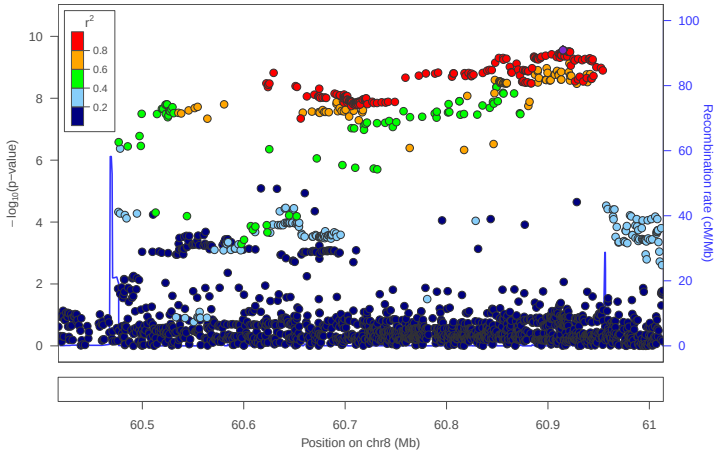
locus_109



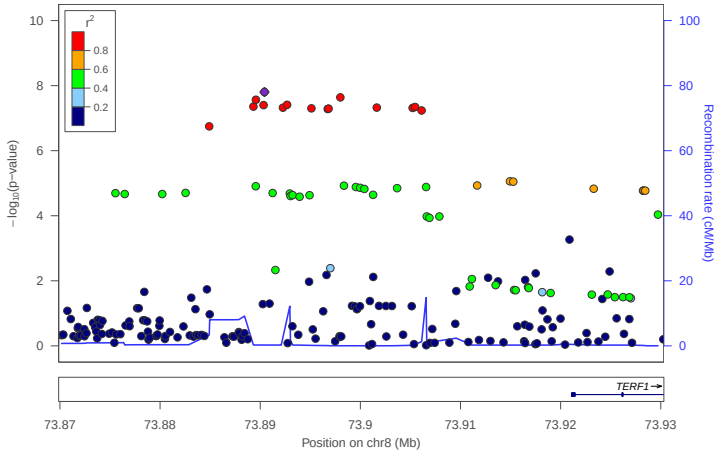
locus_110



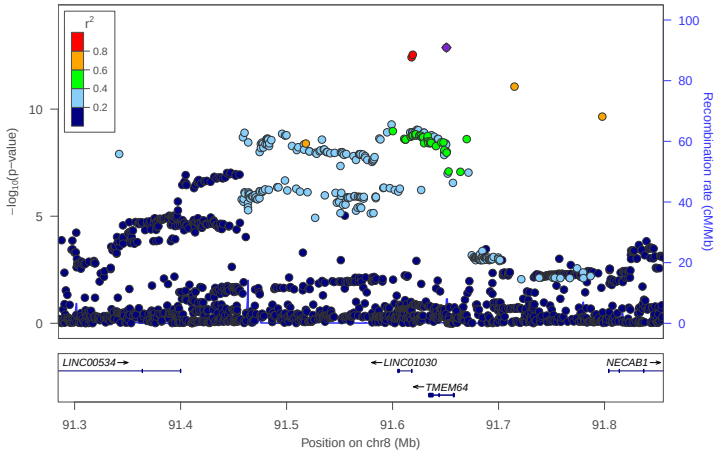
locus_111



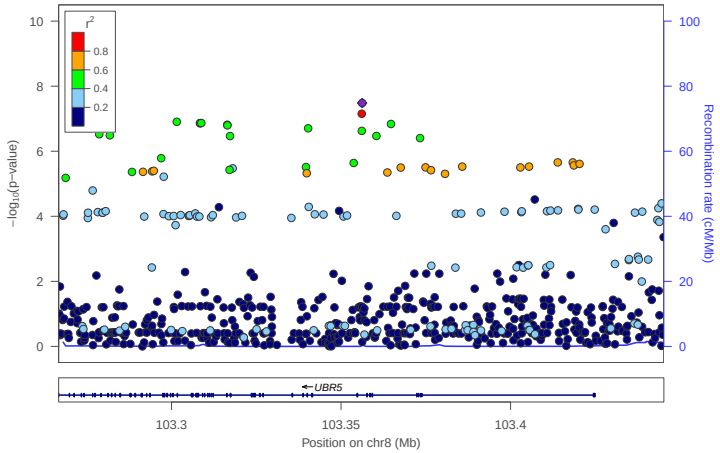
locus_112



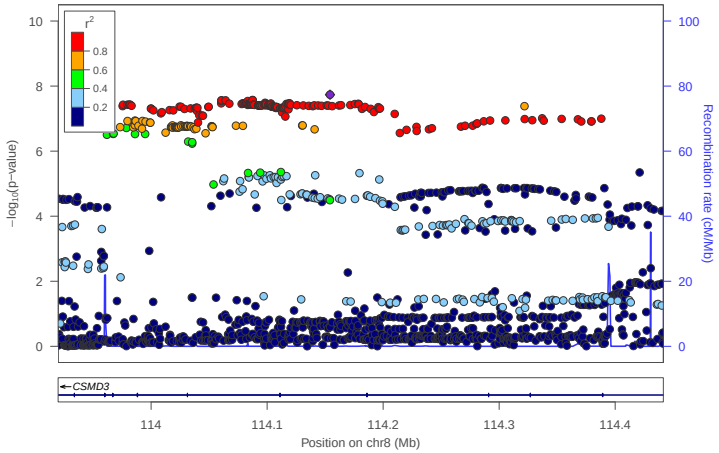
locus_113



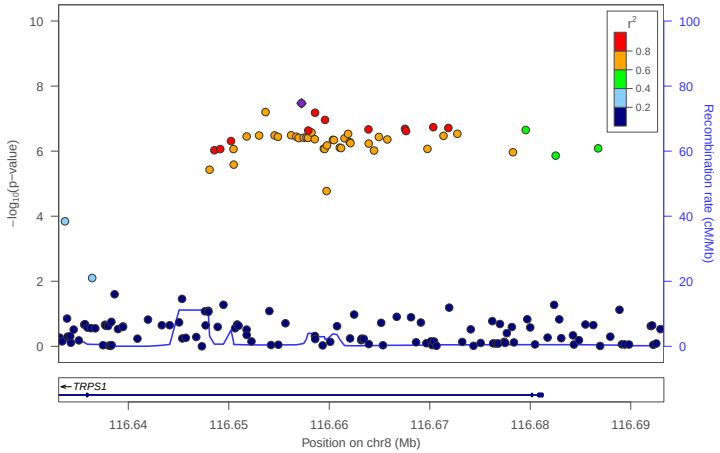
locus_114



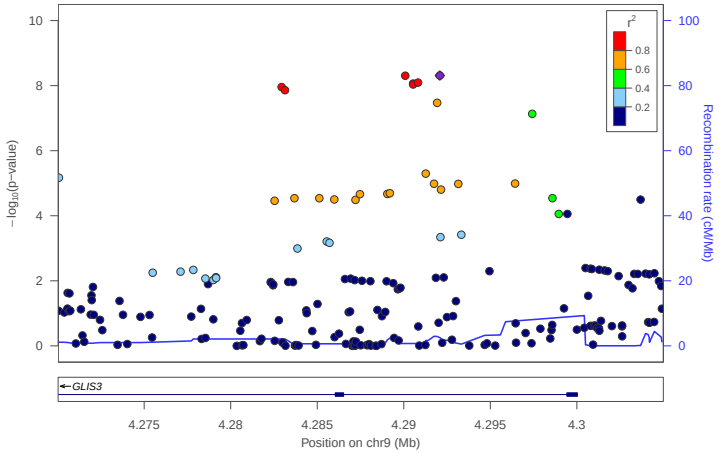
locus_115



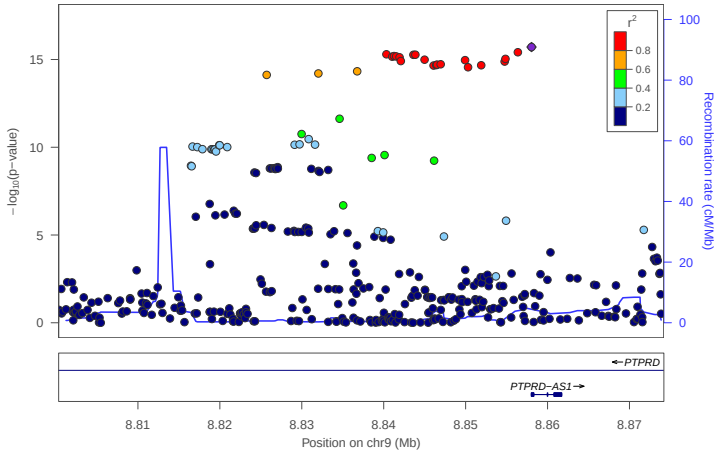
locus_116



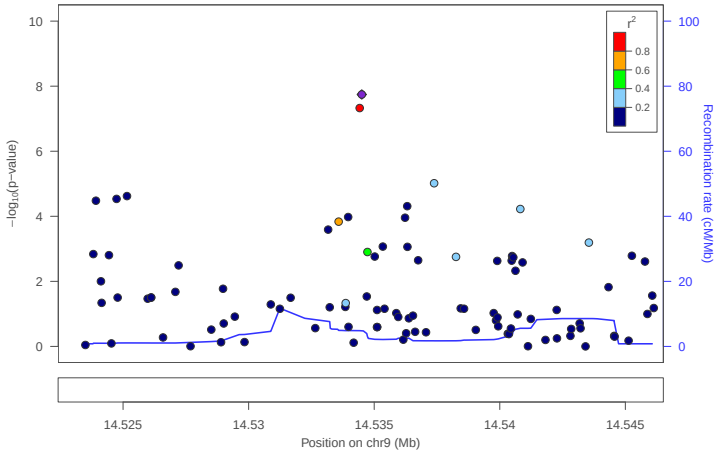
locus_117



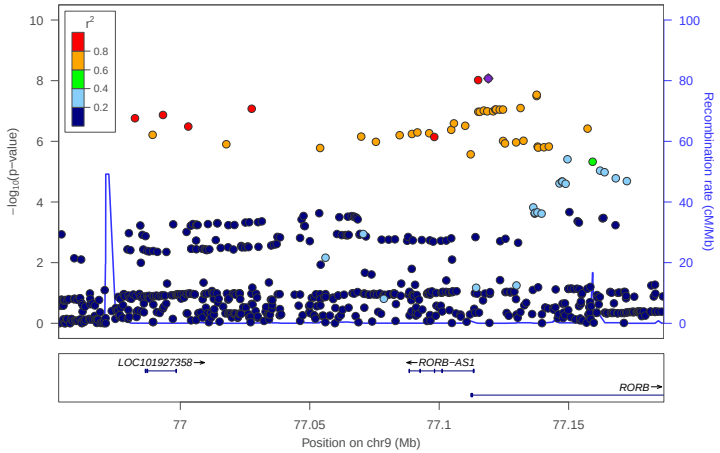
locus_118



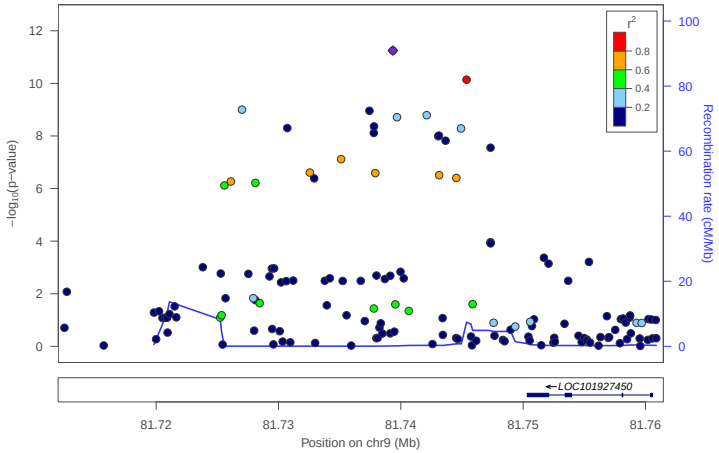
locus_119



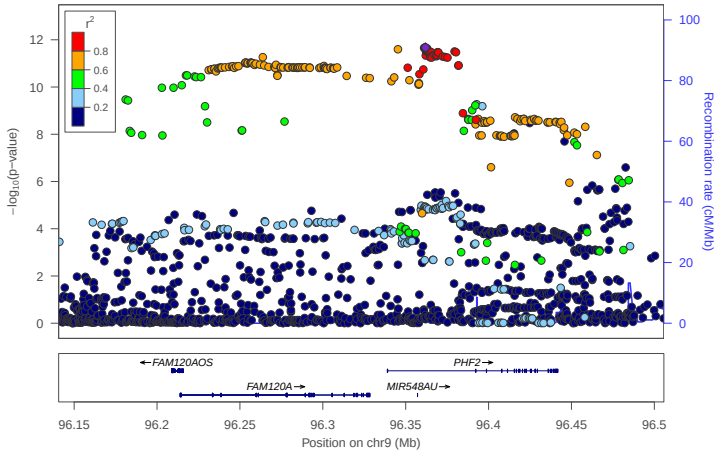
locus_120



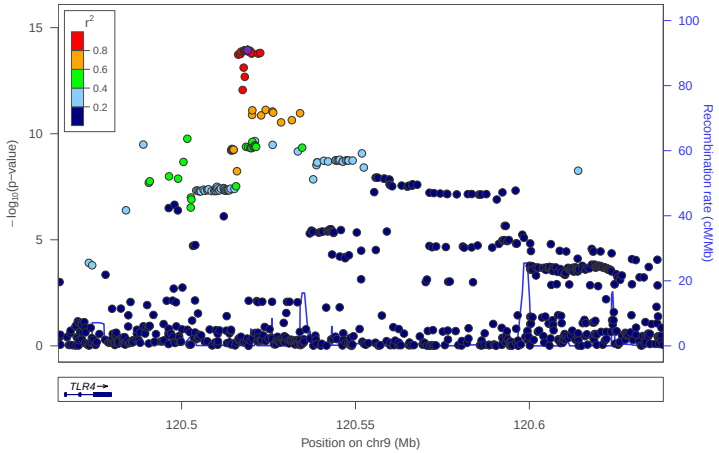
locus_121



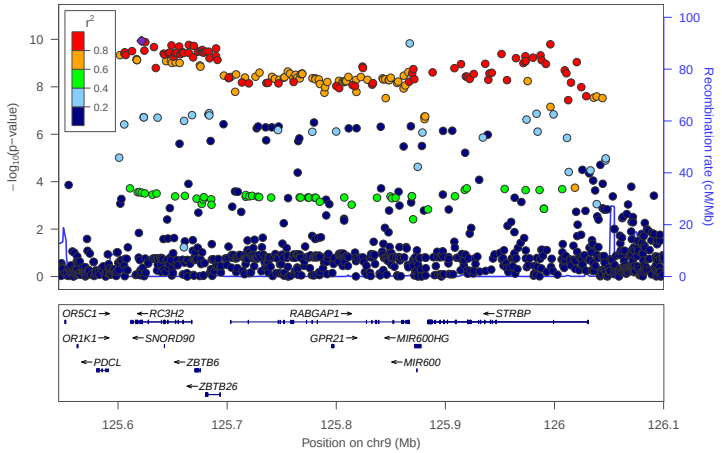
locus_122



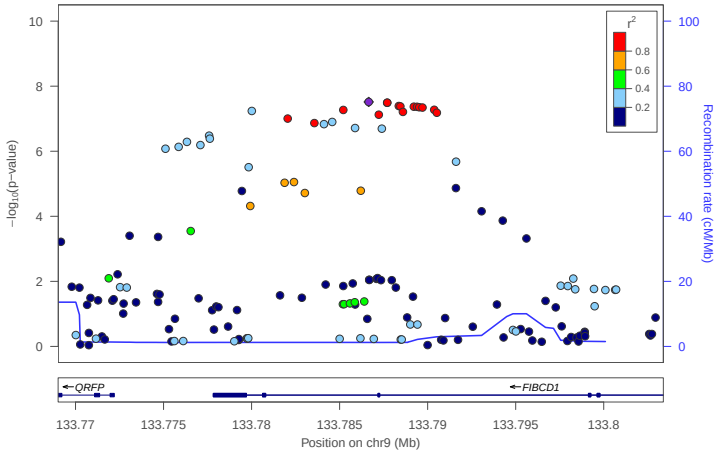
locus_123



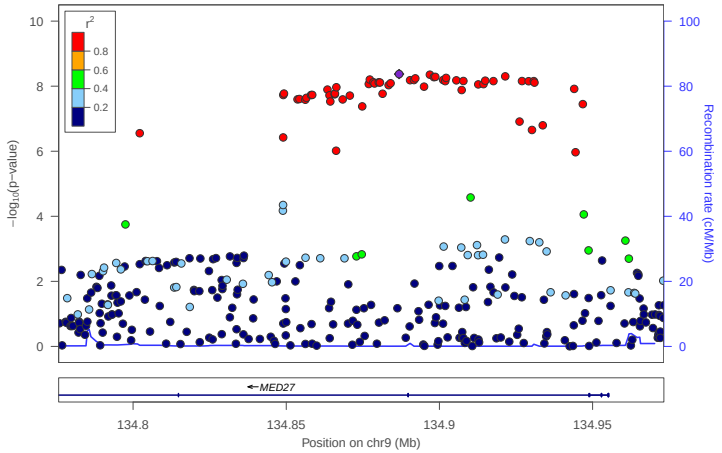
locus_124



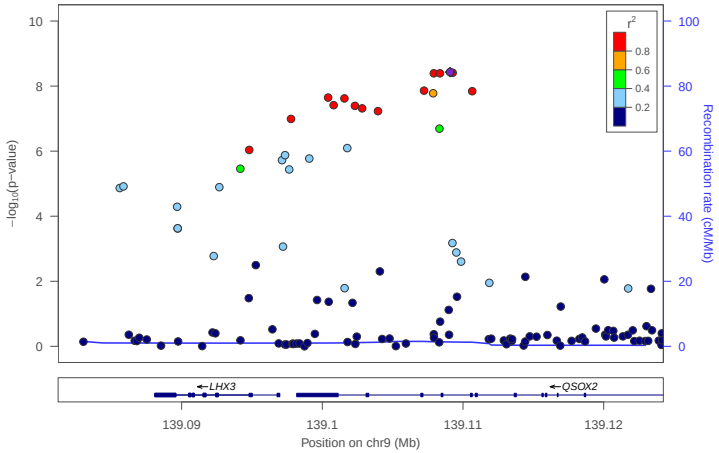
locus_125



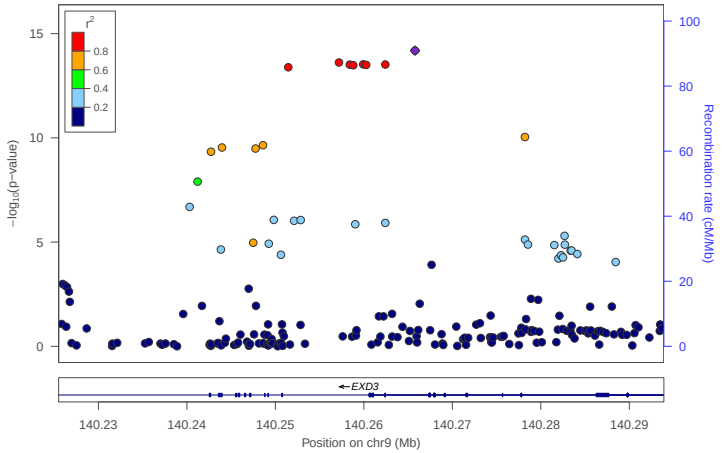
locus_126



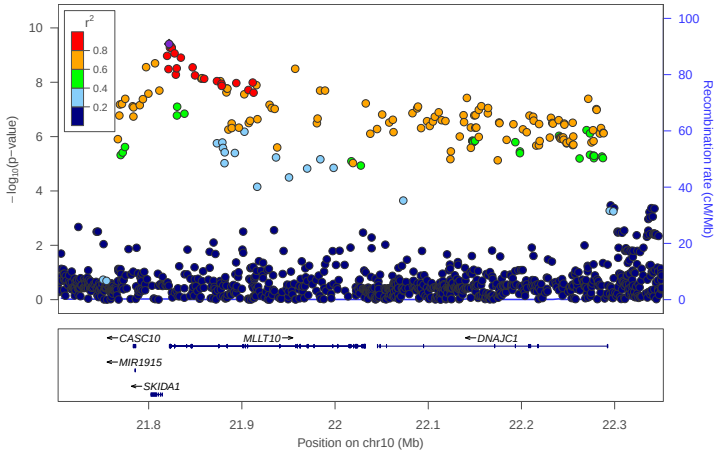
locus_127



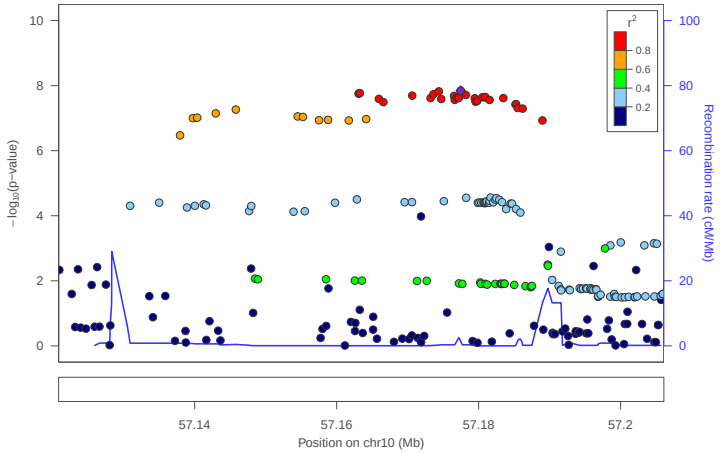
locus_128



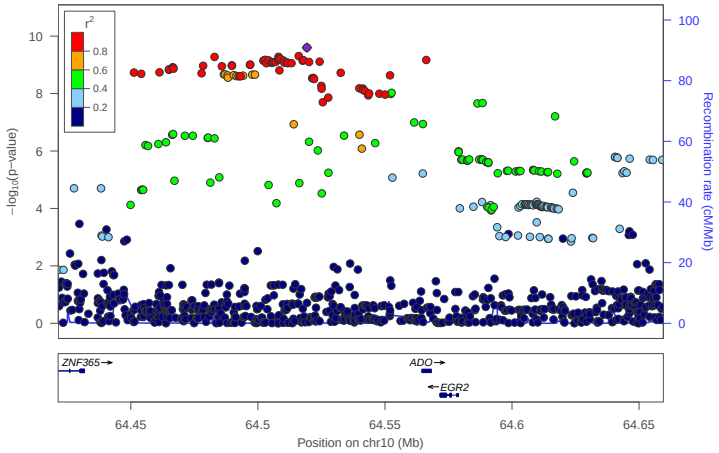
locus_129



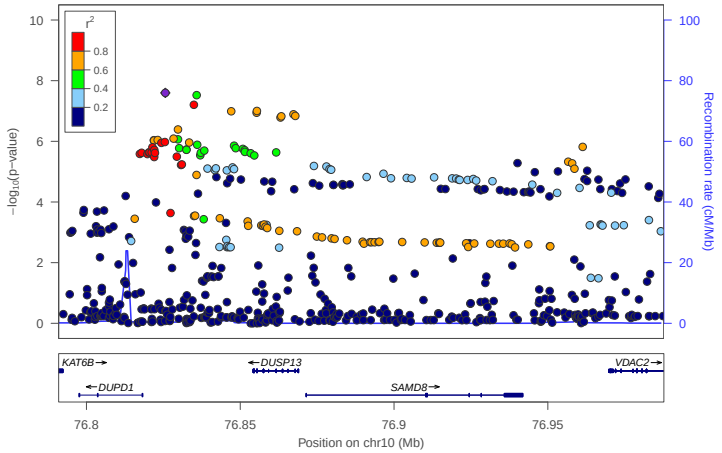
locus_130



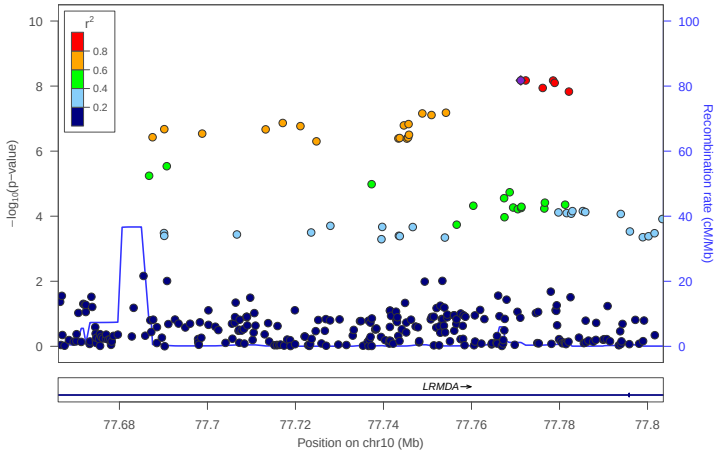
locus_131



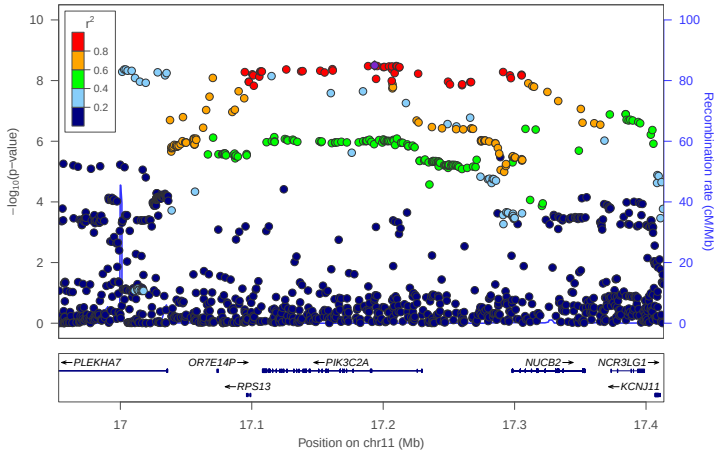
locus_132



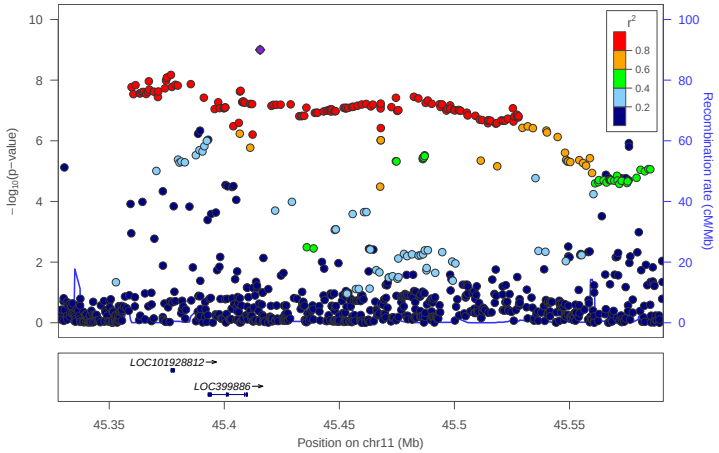
locus_133



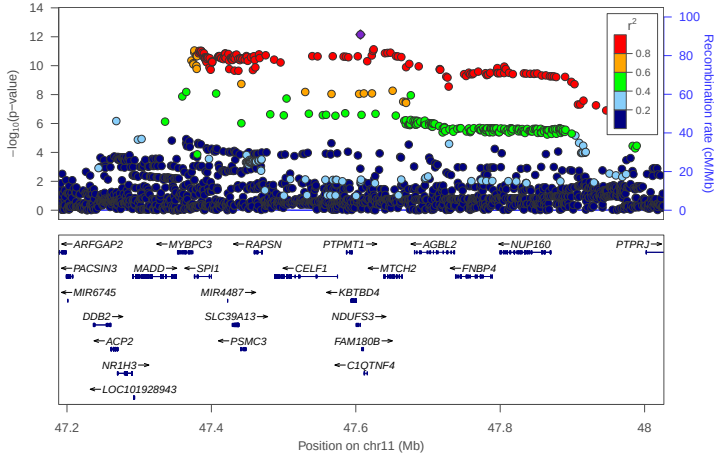
locus_134



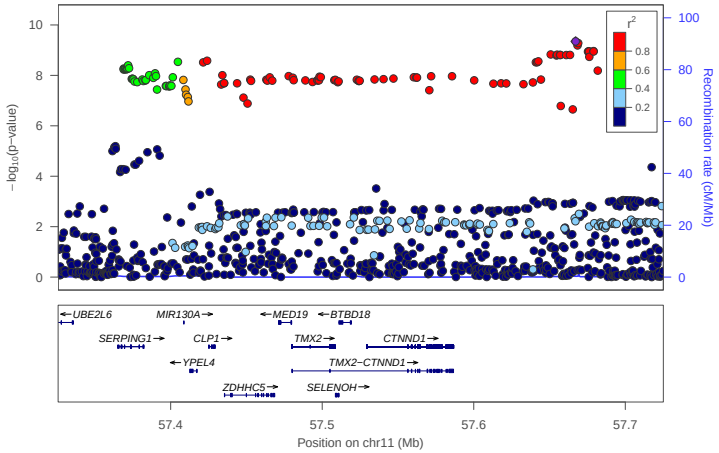
locus_135



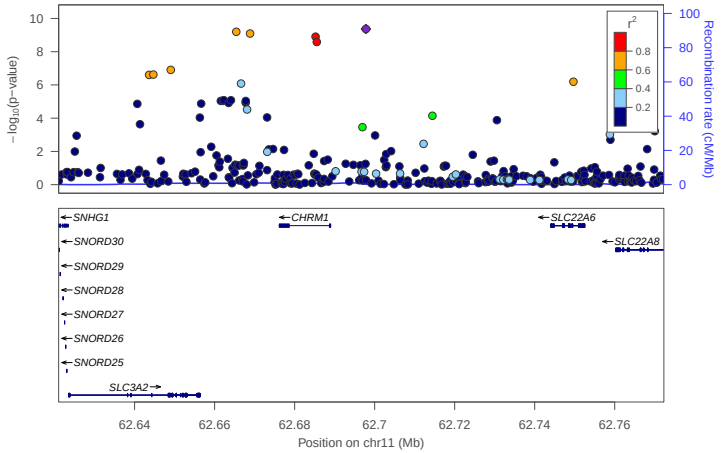
locus_136



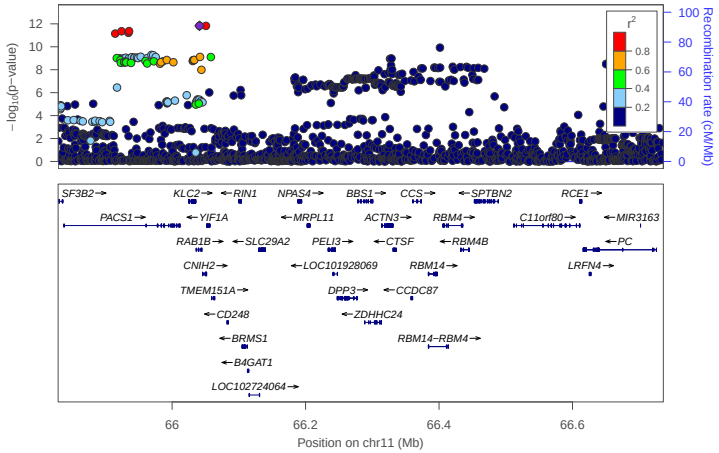
locus_137



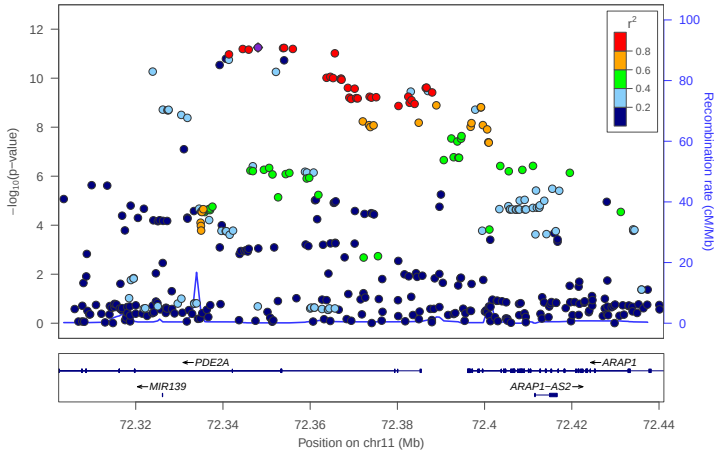
locus_138



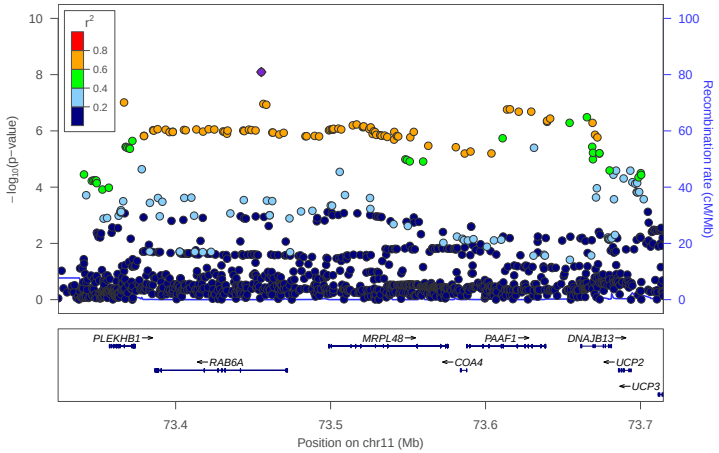
locus_139



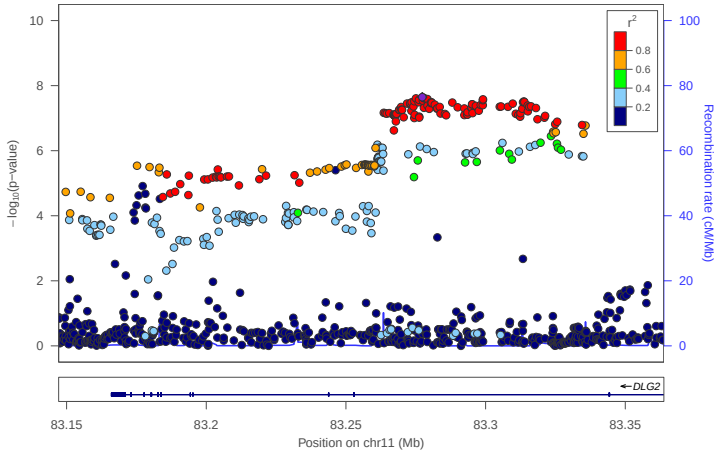
locus_140



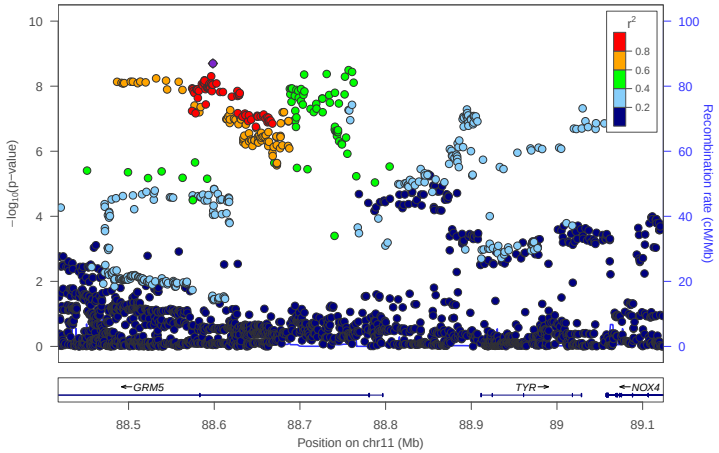
locus_141



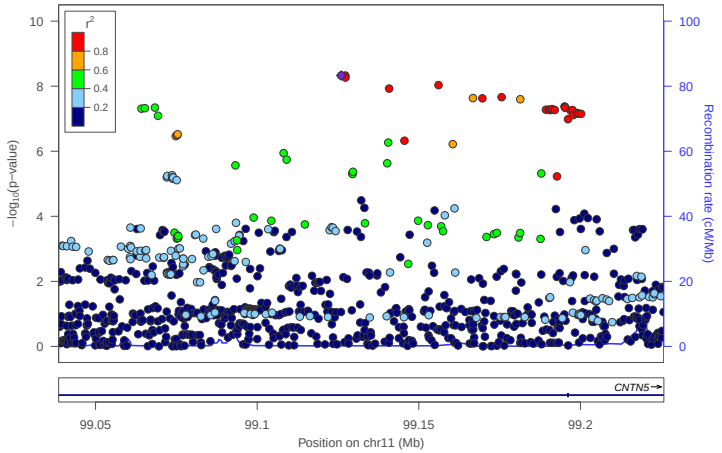
locus_142



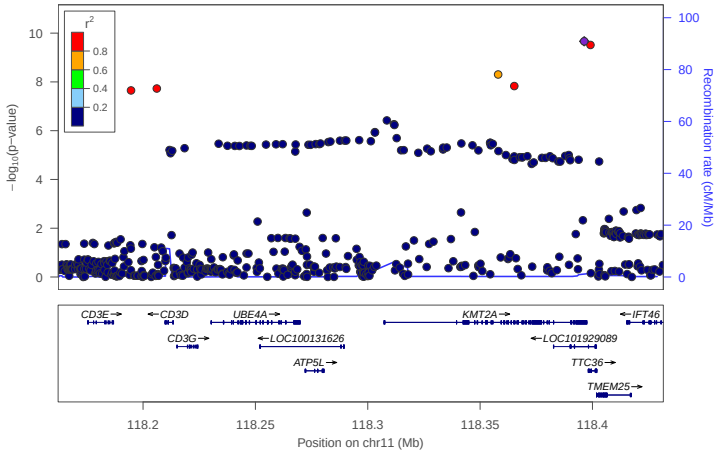
locus_143



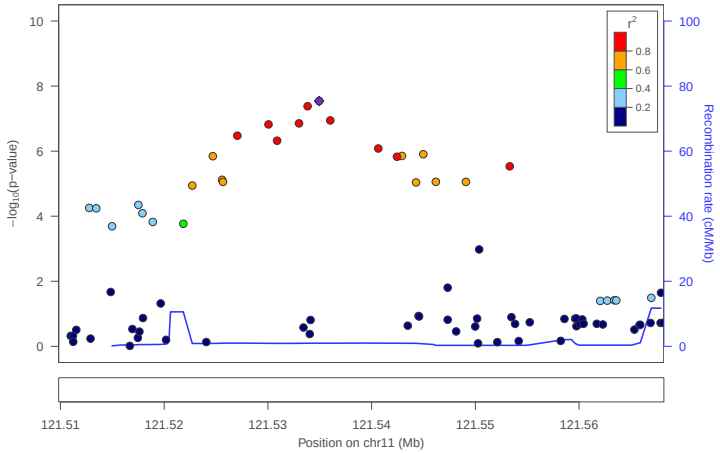
locus_144



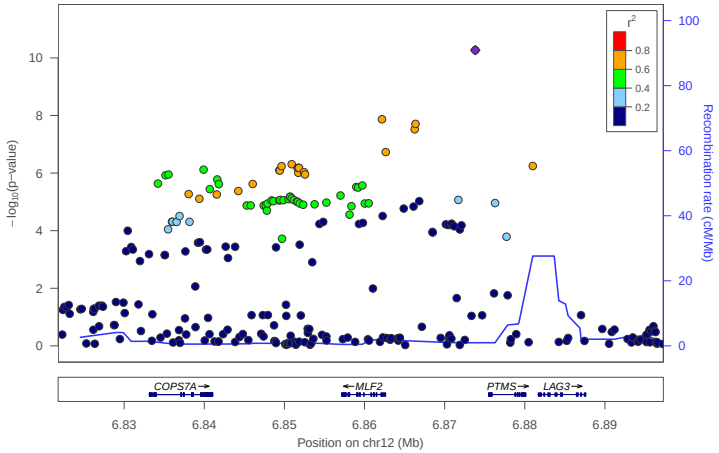
locus_145



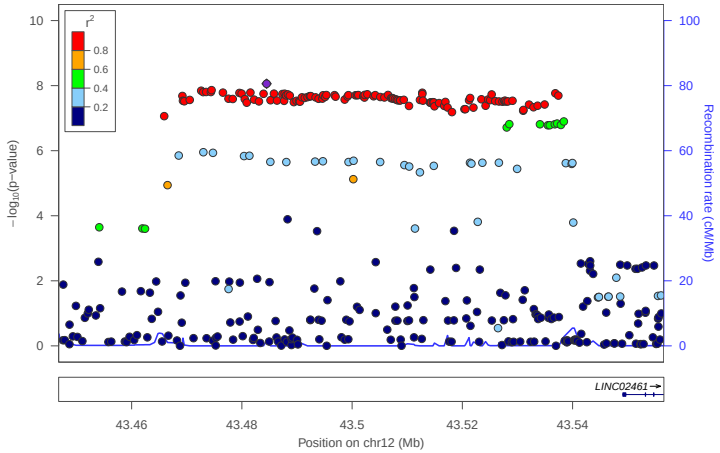
locus_146



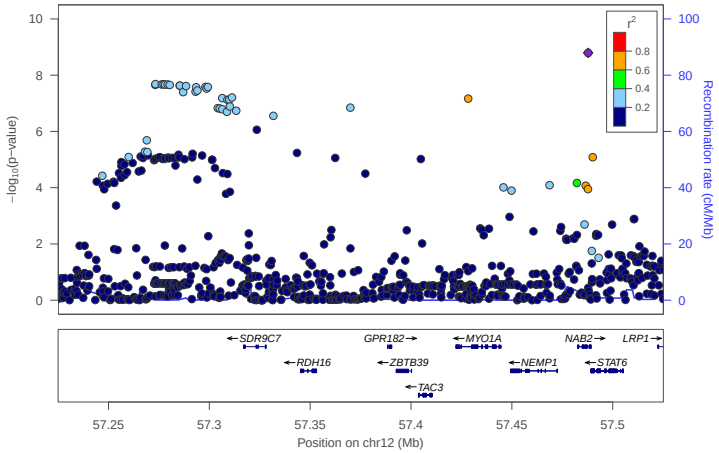
locus_147



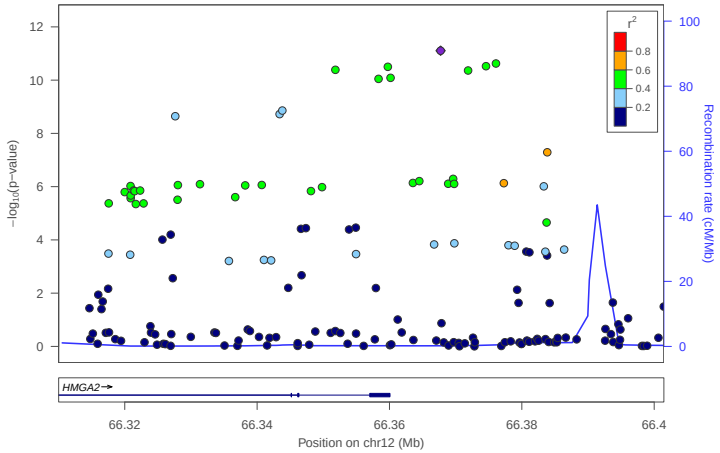
locus_148



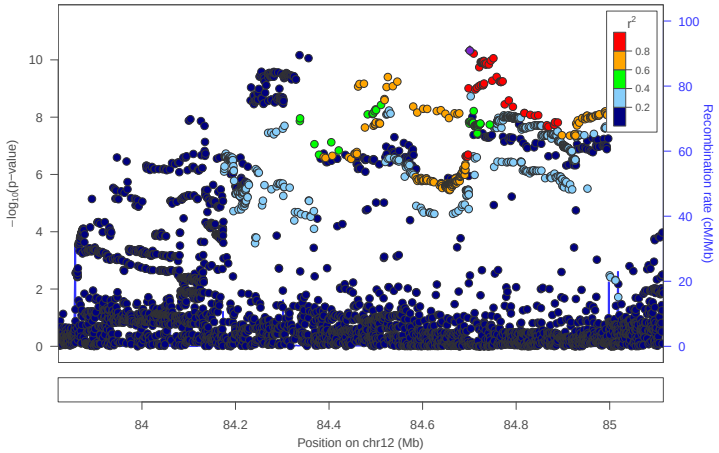
locus_149



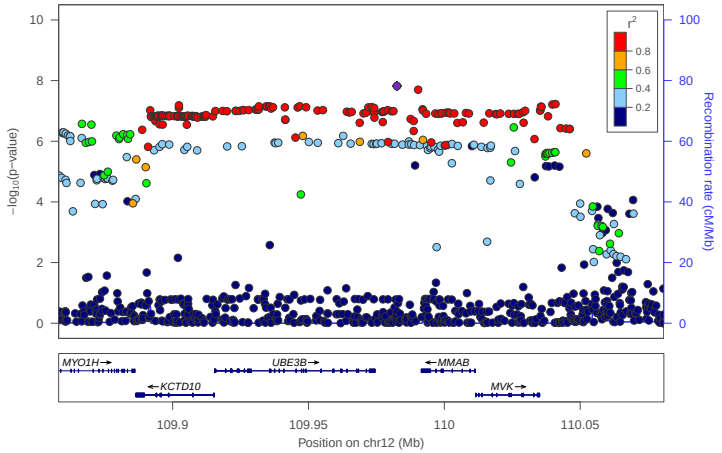
locus_150



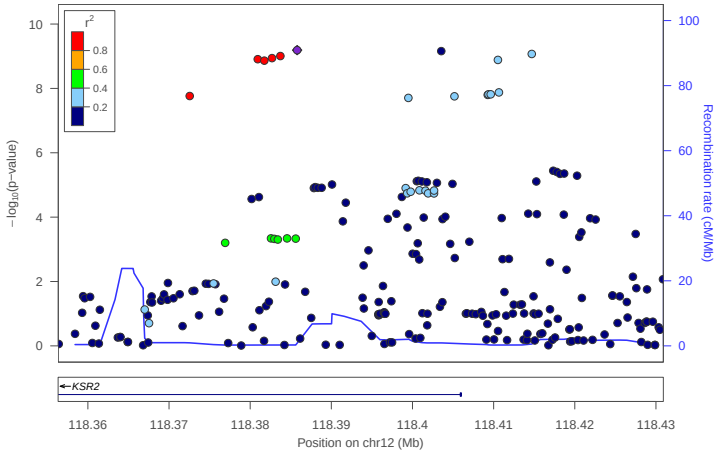
locus_151



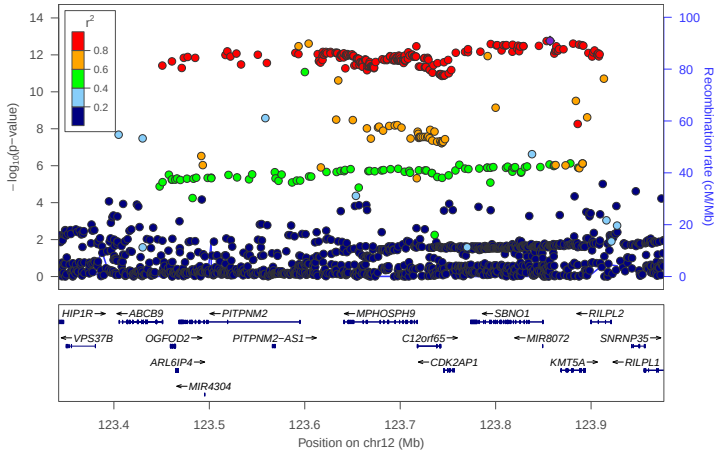
locus_152



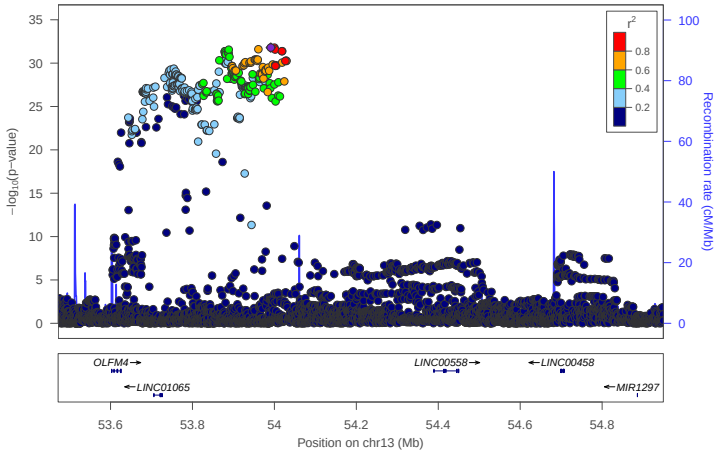
locus_153



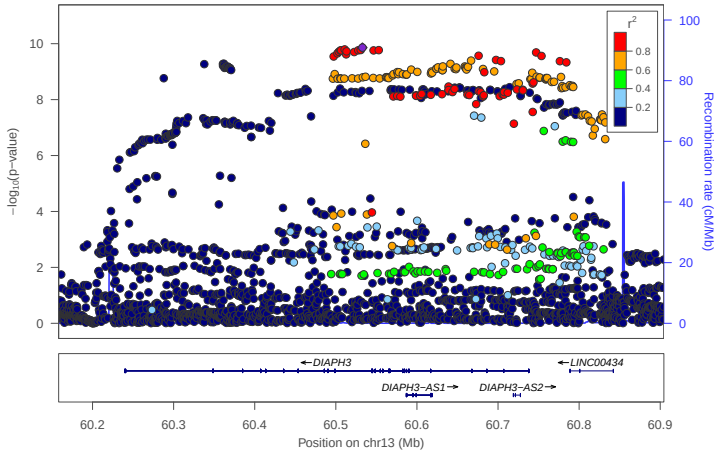
locus_154



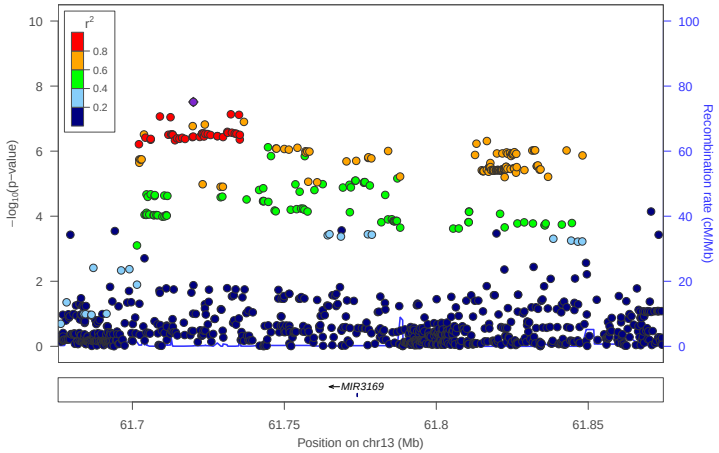
locus_155



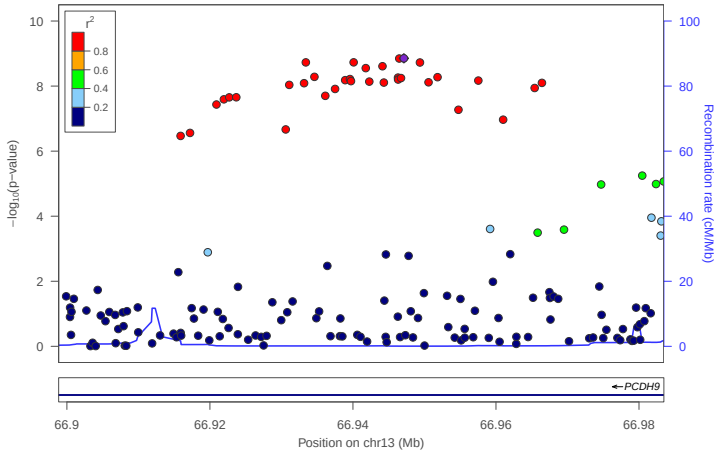
locus_156



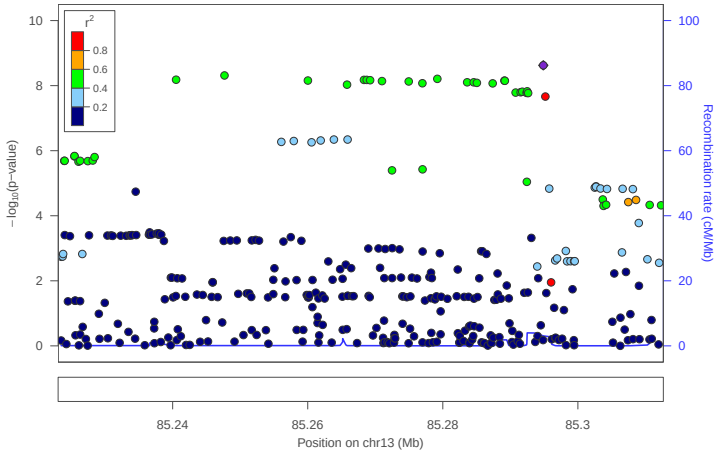
locus_157



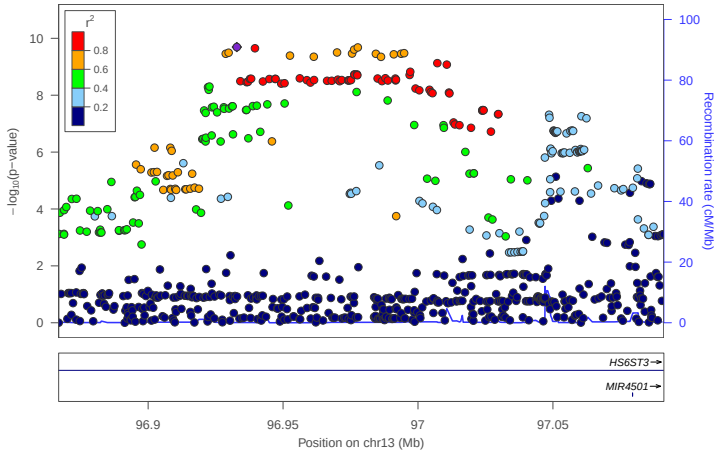
locus_158



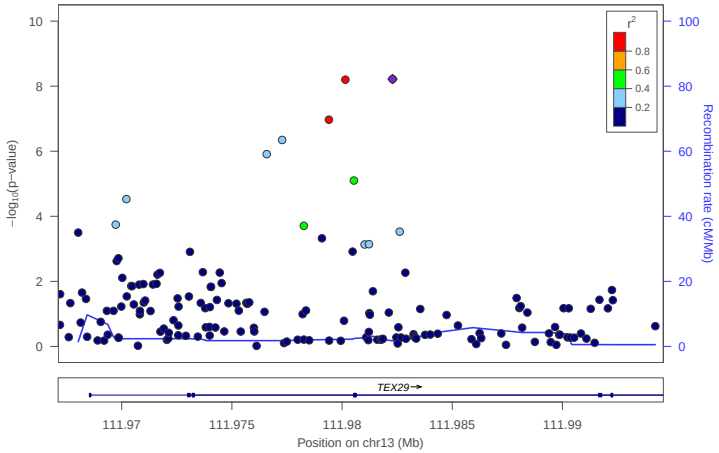
locus_159



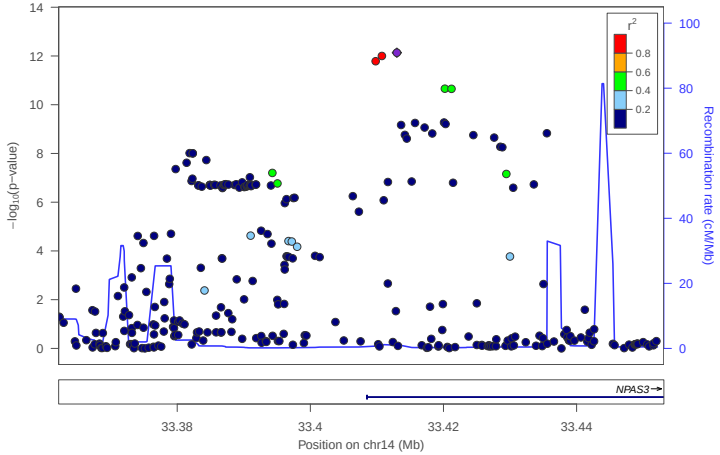
locus_160



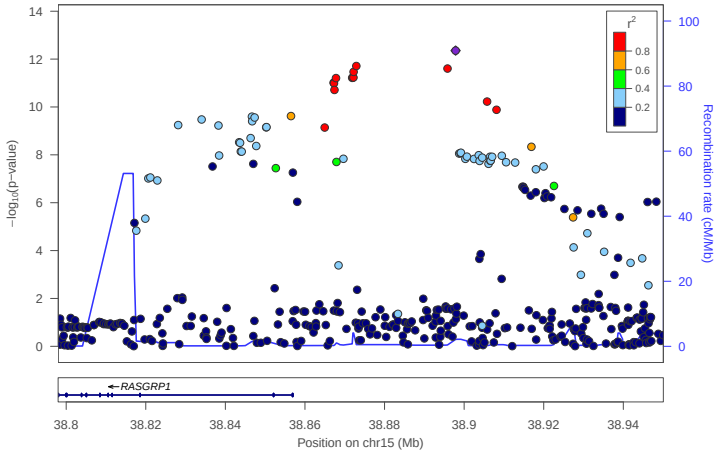
locus_161



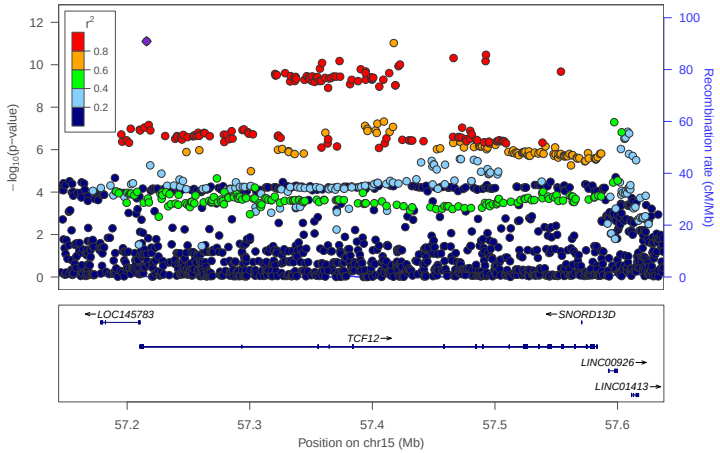
locus_162



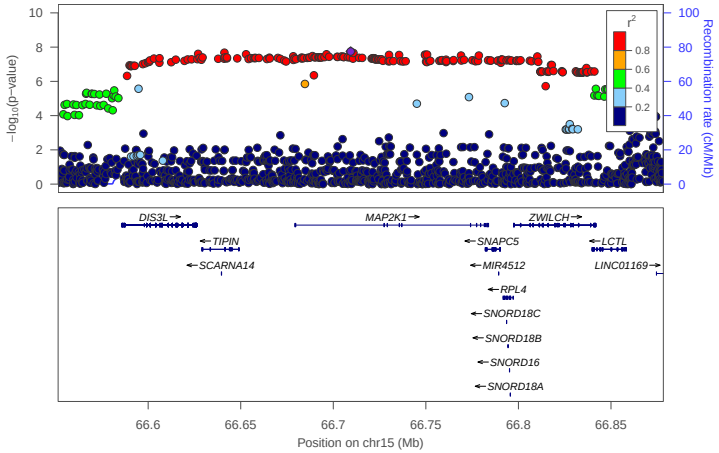
locus_163



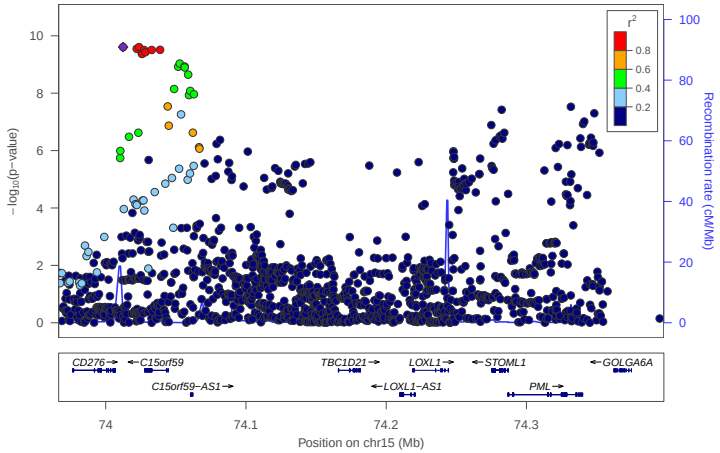
locus_164



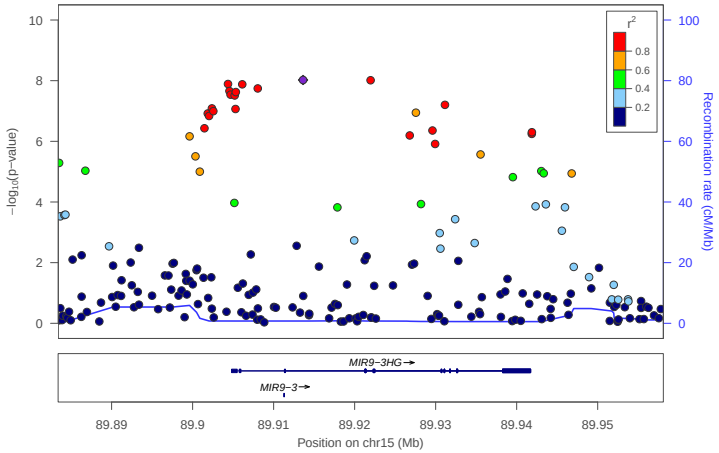
locus_165



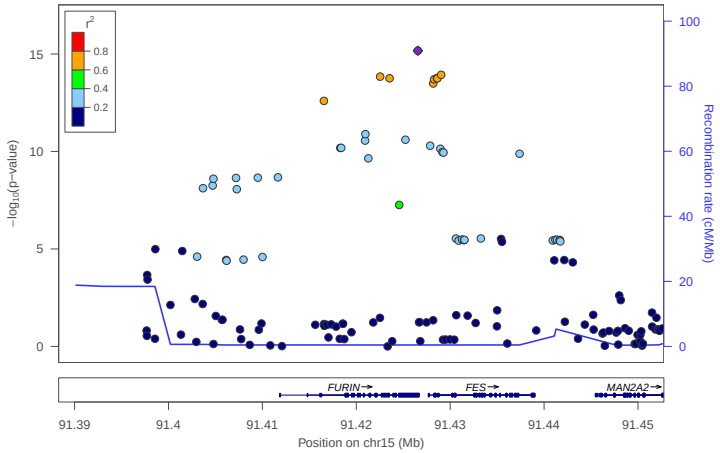
locus_166



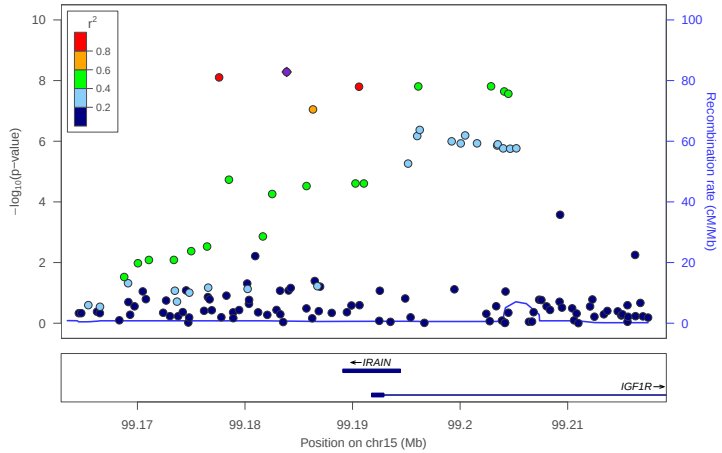
locus_167



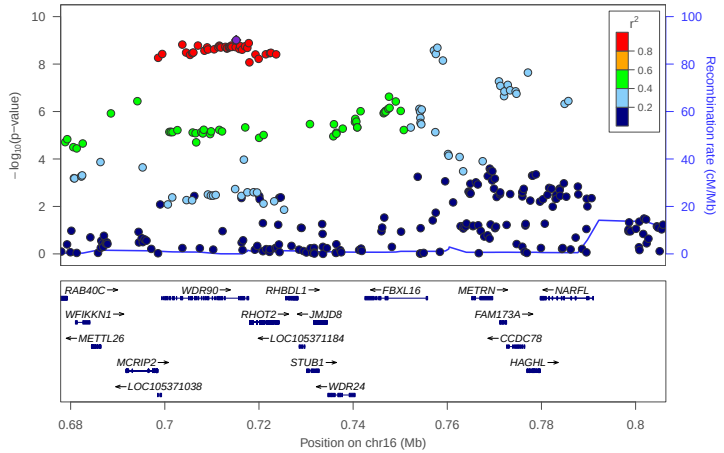
locus_168



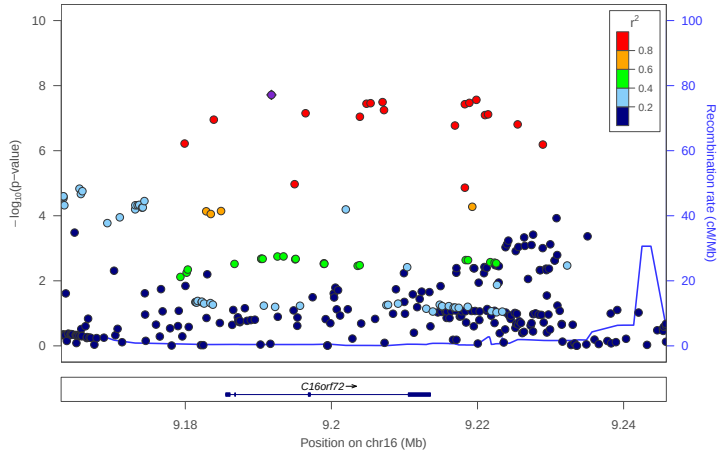
locus_169



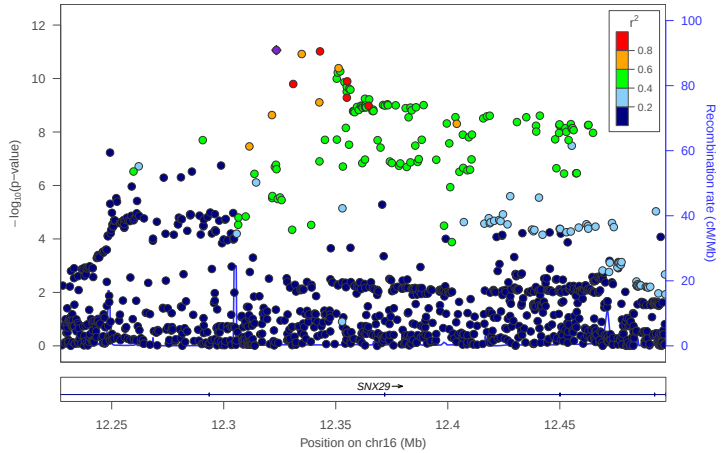
locus_170



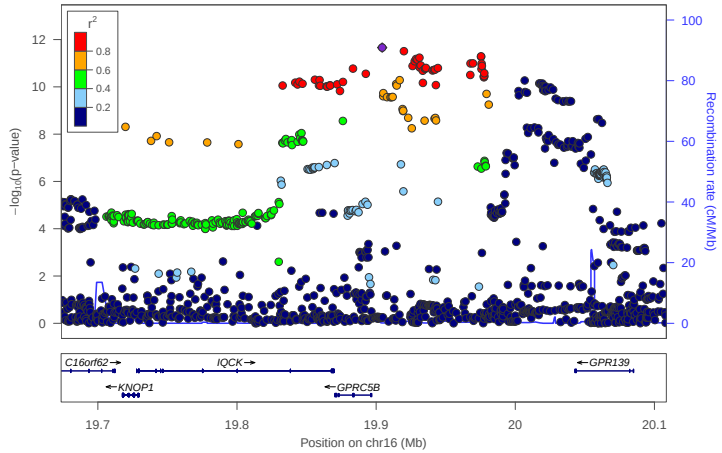
locus_171



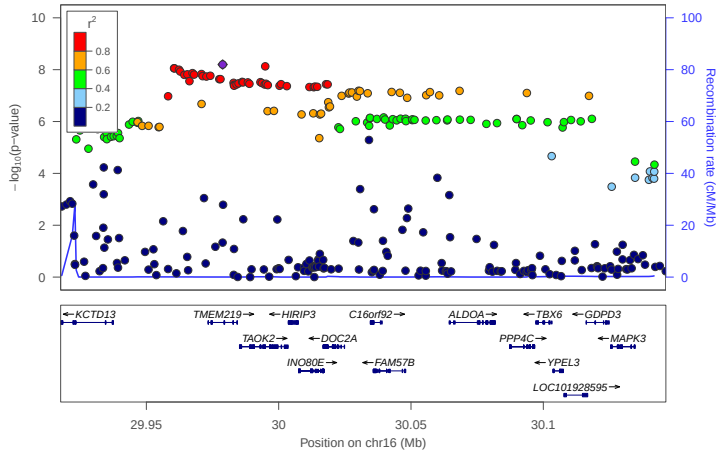
locus_172



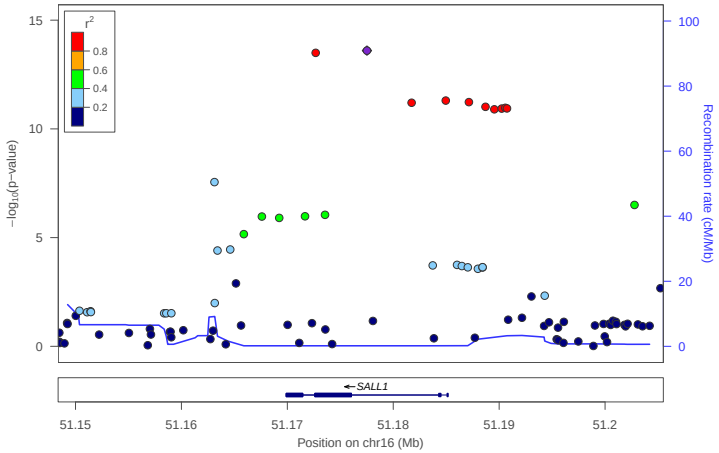
locus_173



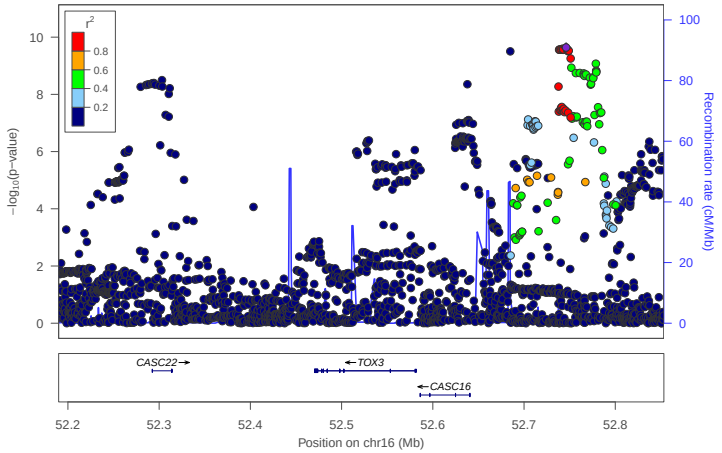
locus_174



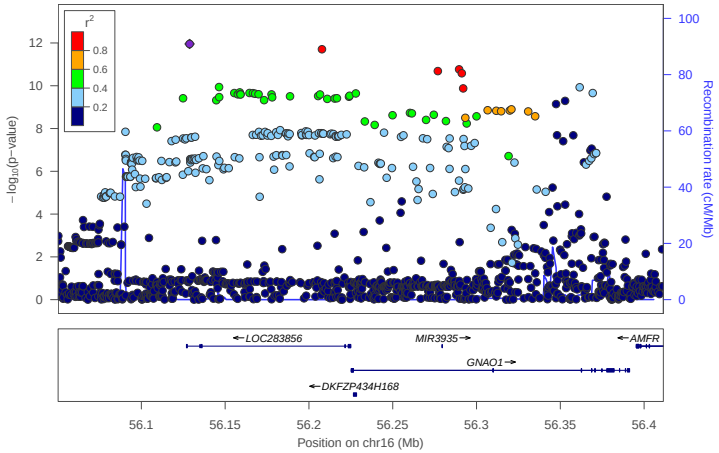
locus_175



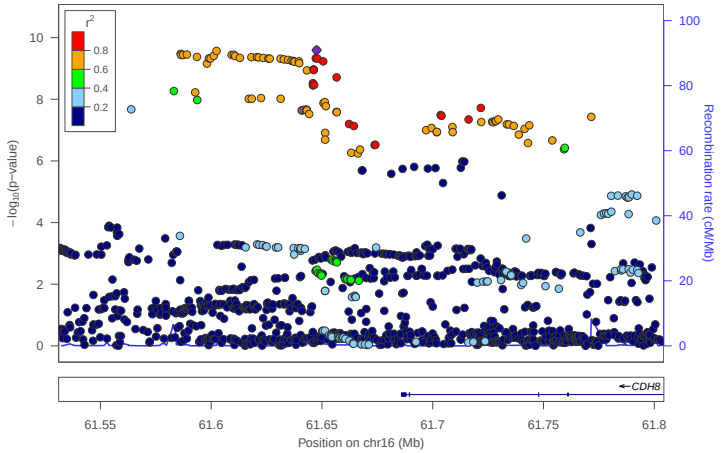
locus_176



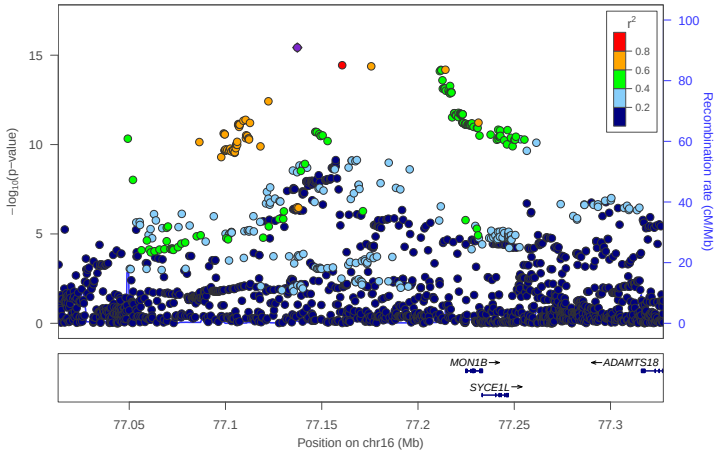
locus_177



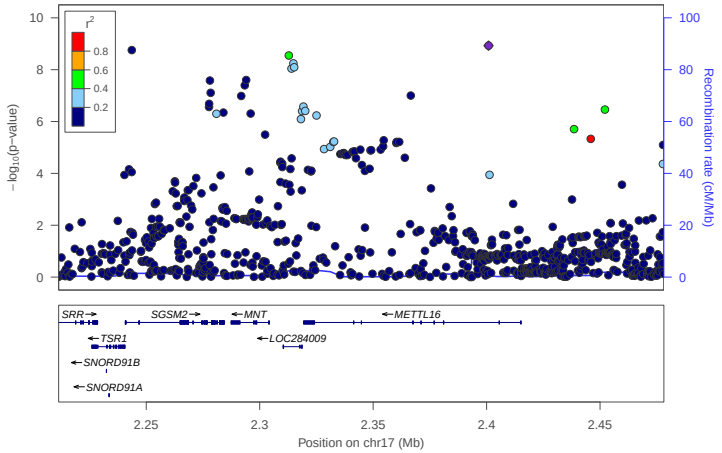
locus_178



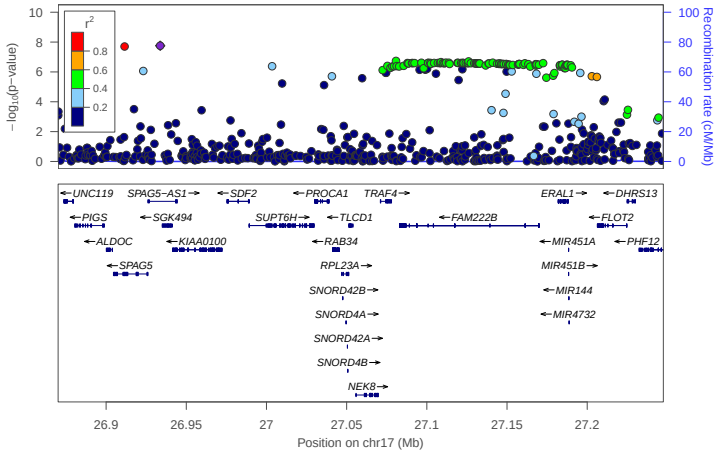
locus_179



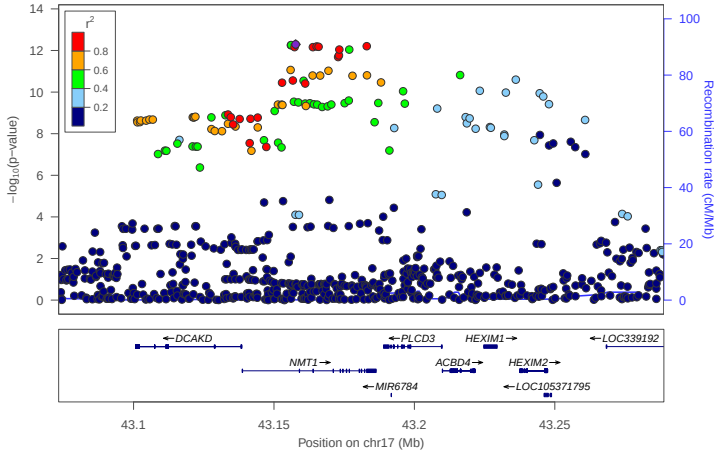
locus_180



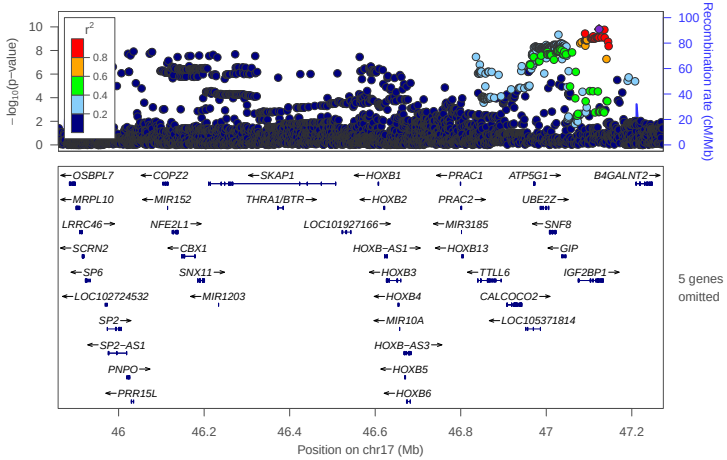
locus_181



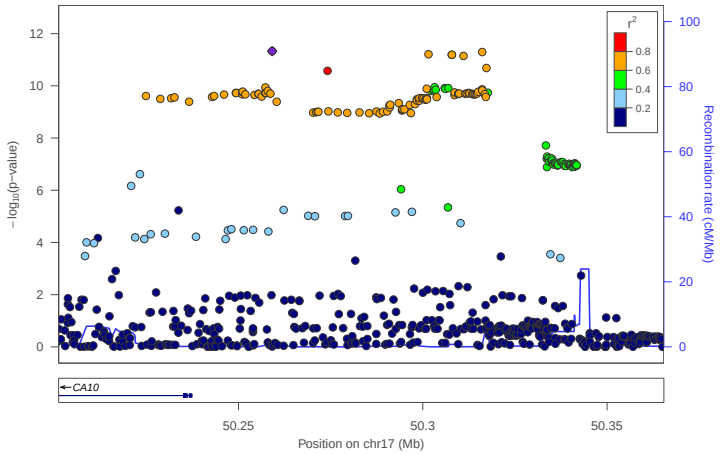
locus_182



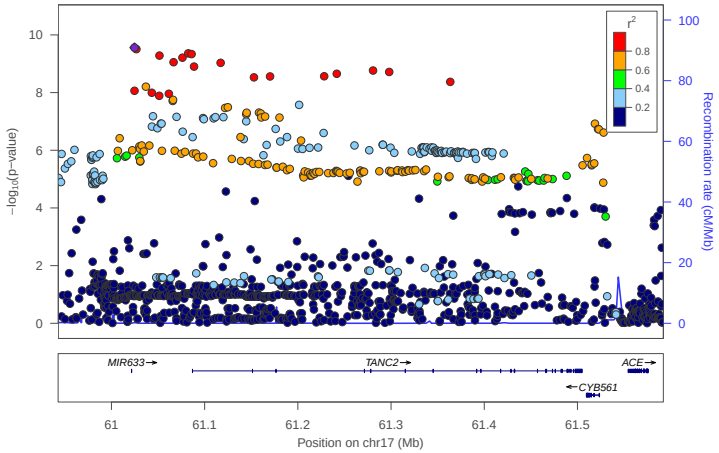
locus_183



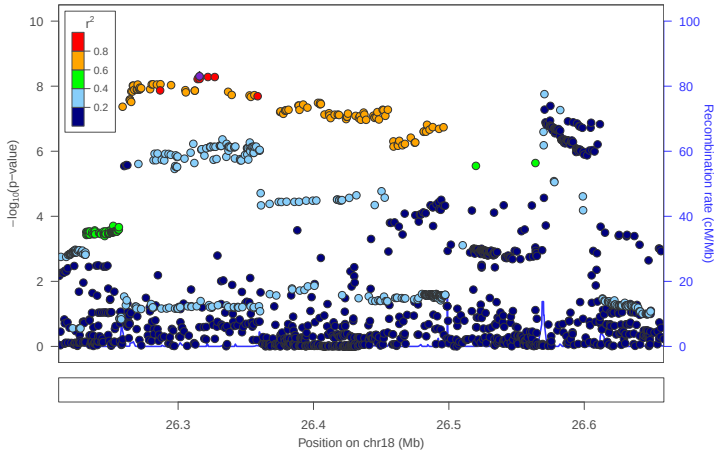
locus_184



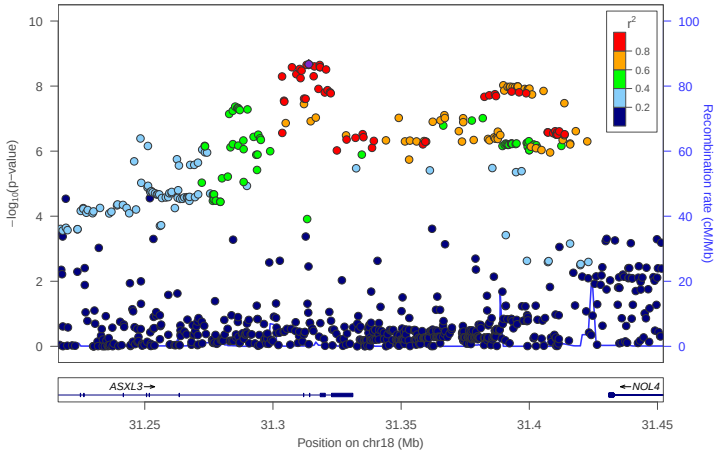
locus_185



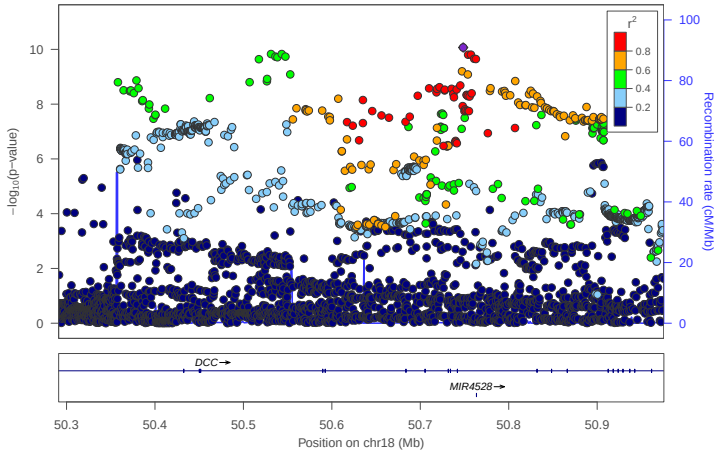
locus_186



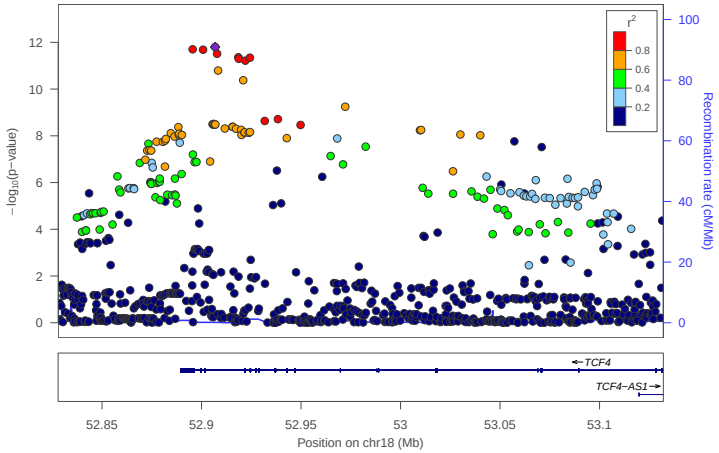
locus_187



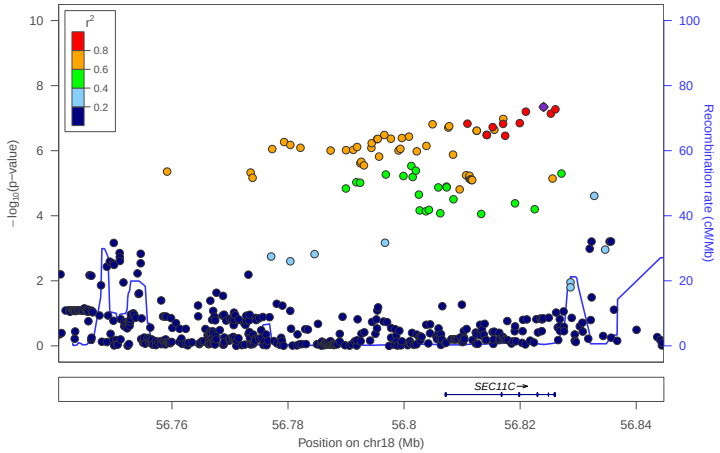
locus_188



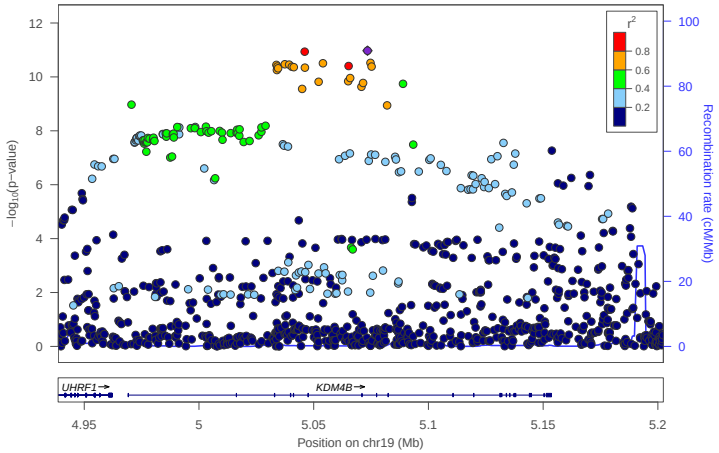
locus_189



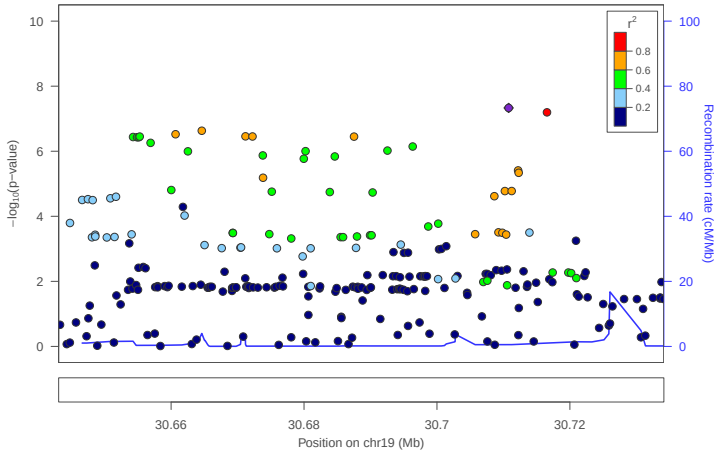
locus_190



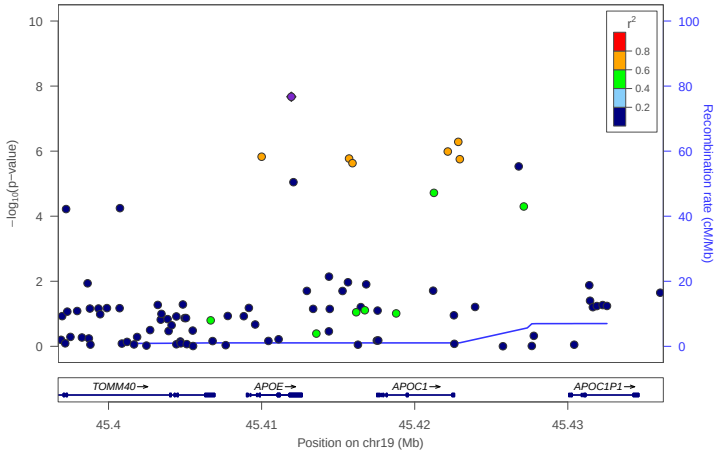
locus_191



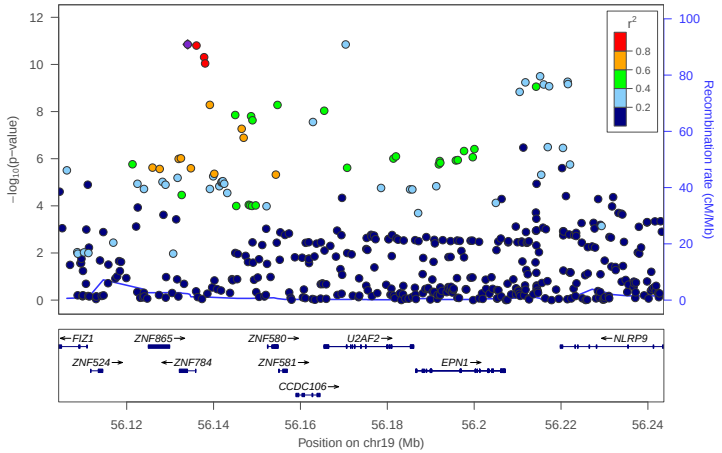
locus_192



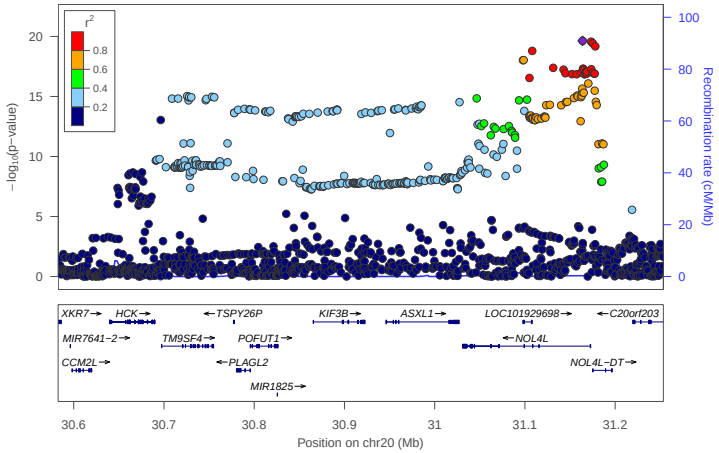
locus_193



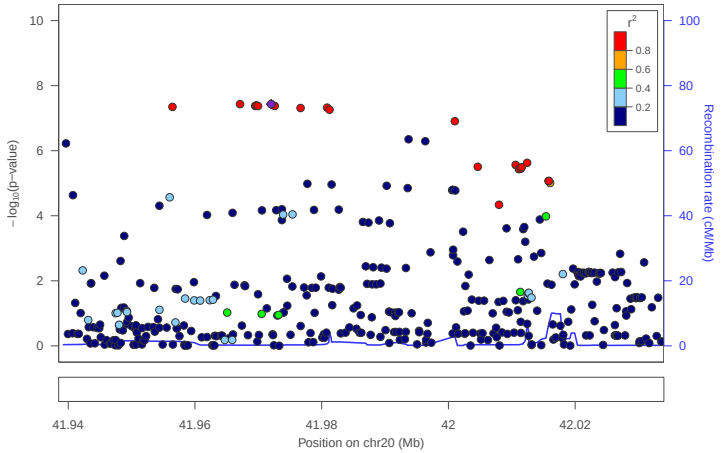
locus_194



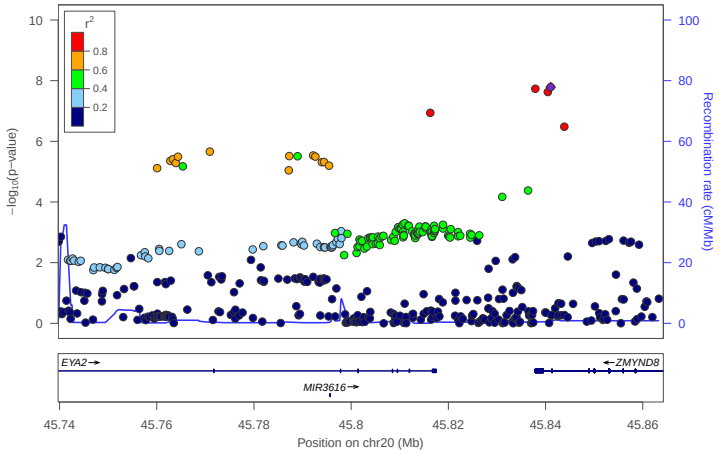
locus_195



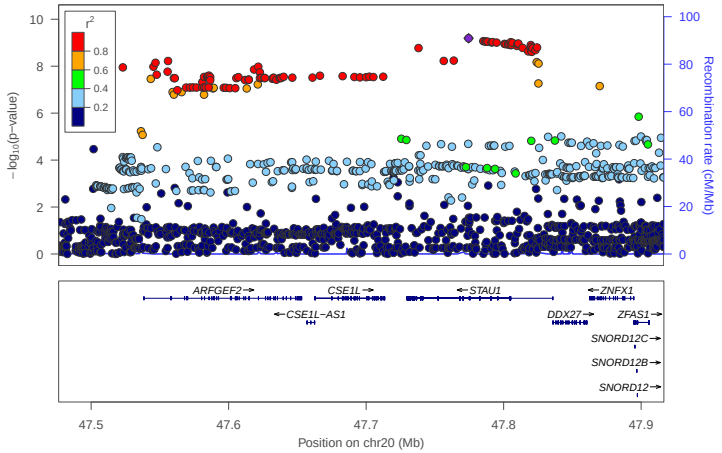
locus_196



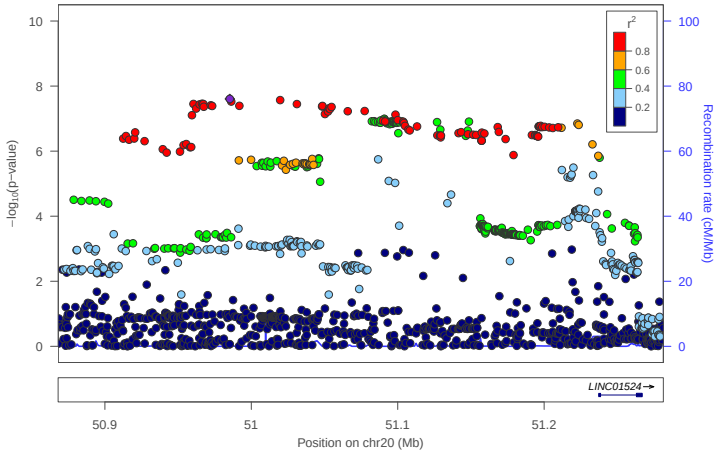
locus_197



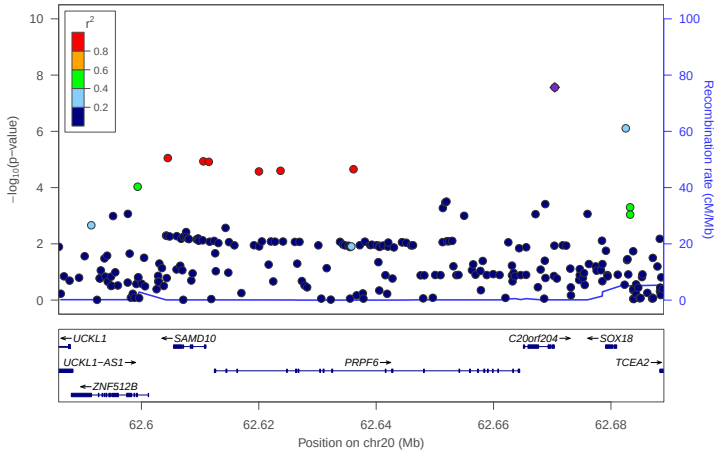
locus_198



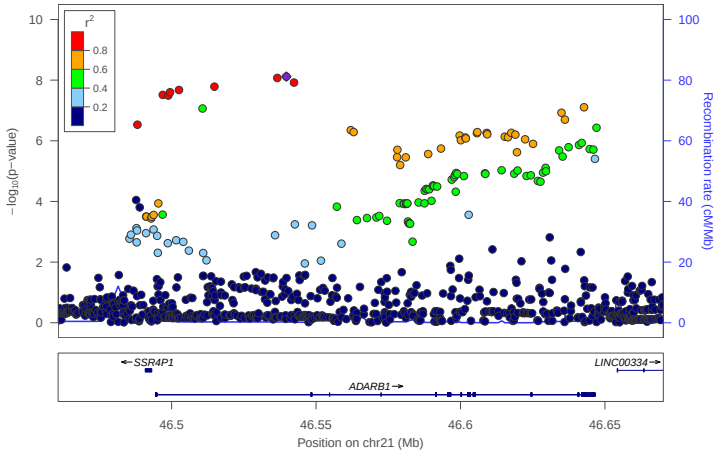
locus_199



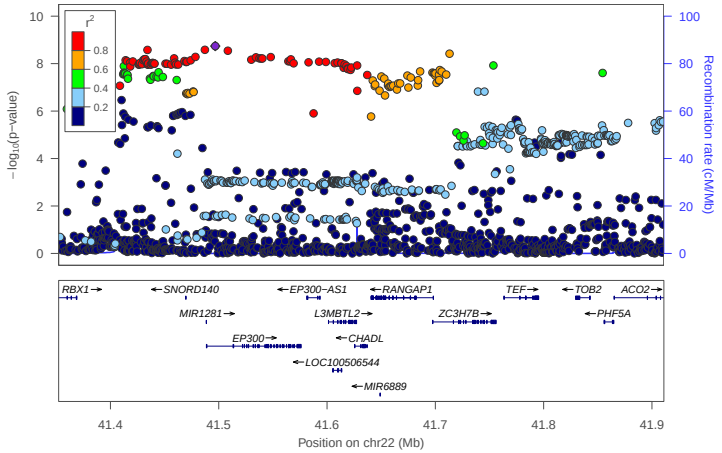
locus_200



locus_201

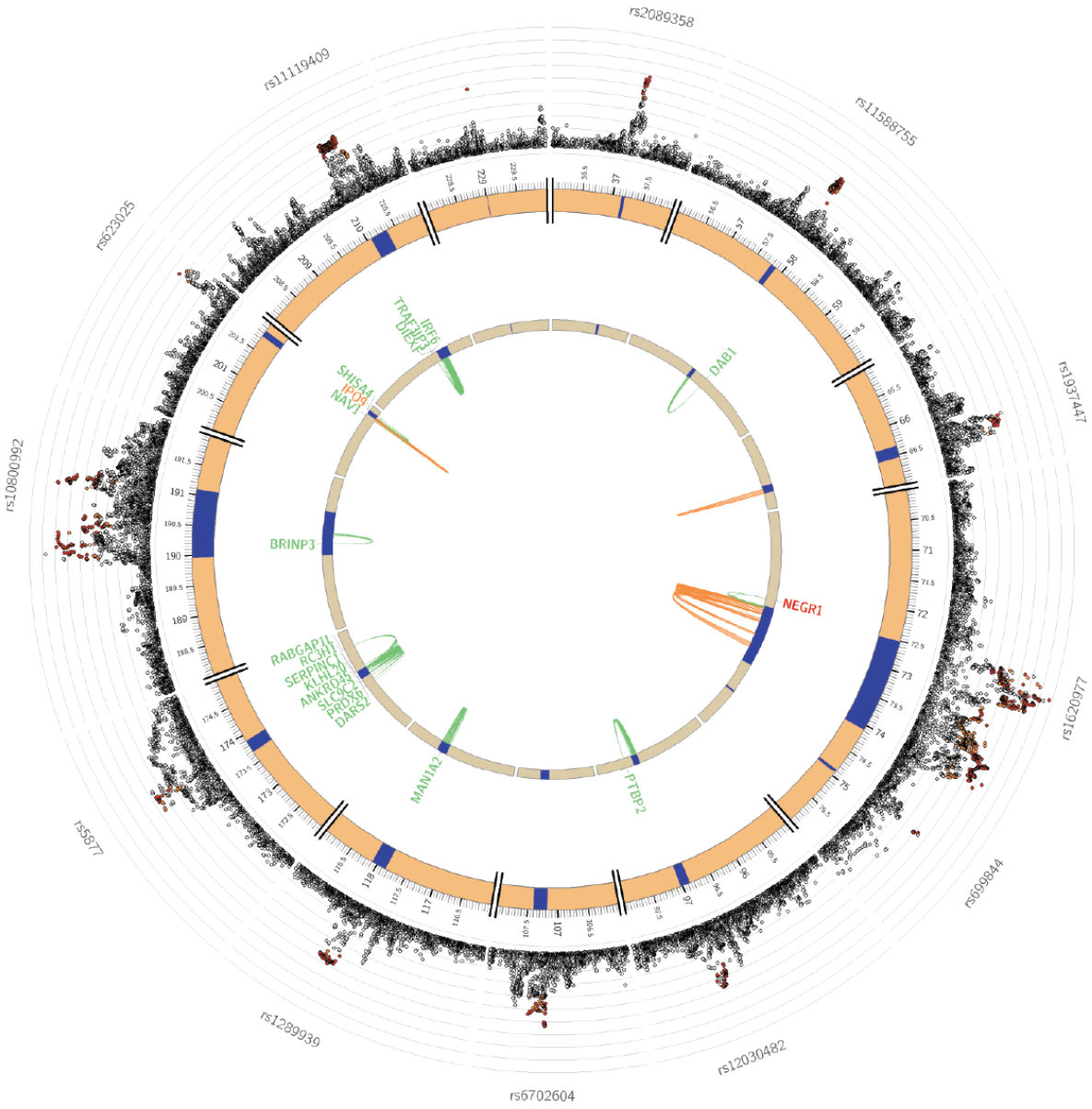


locus_202



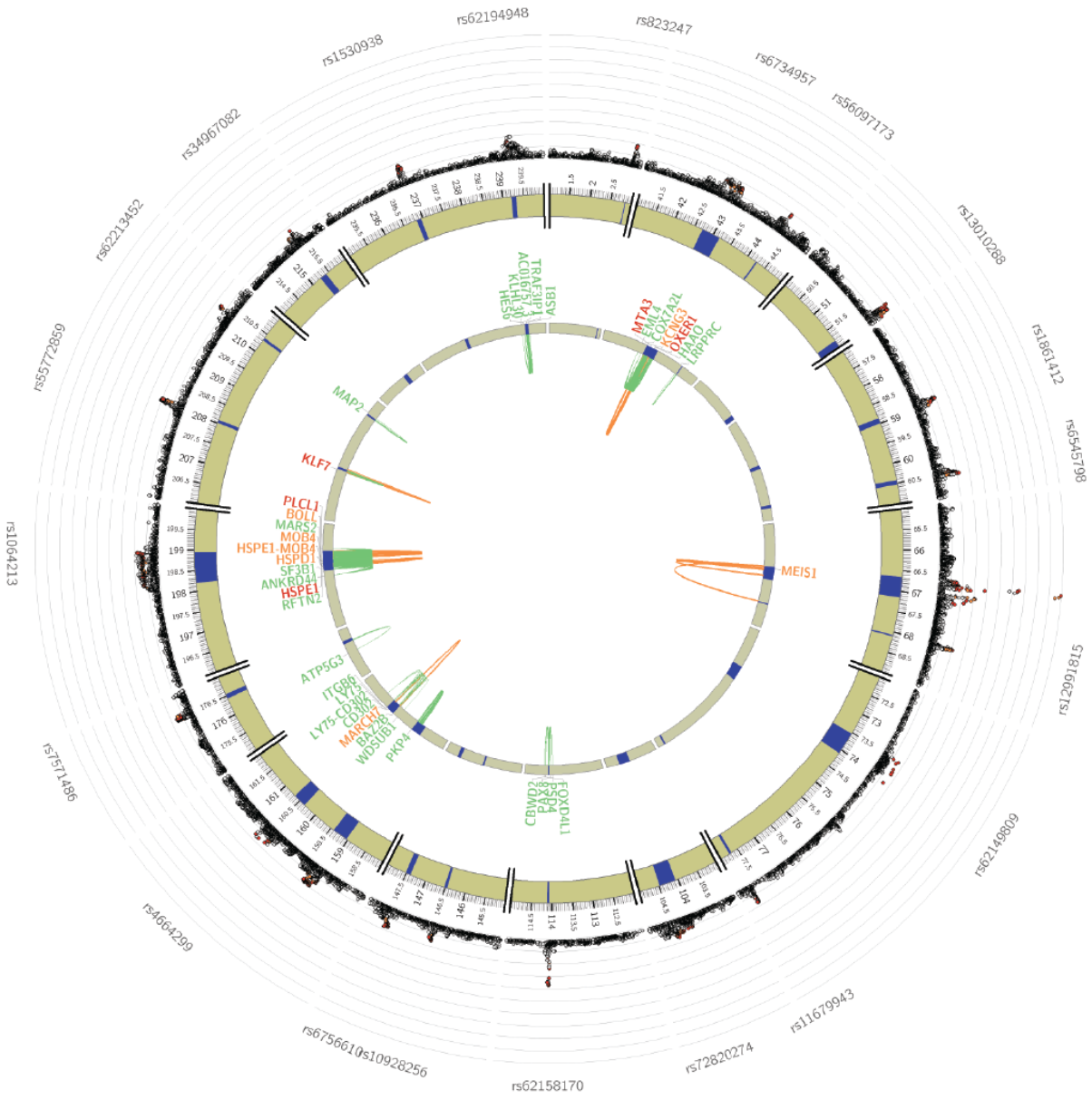
bioRxiv preprint doi: https://doi.org/10.1101/214973; this version posted February 1, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Chromosome 1



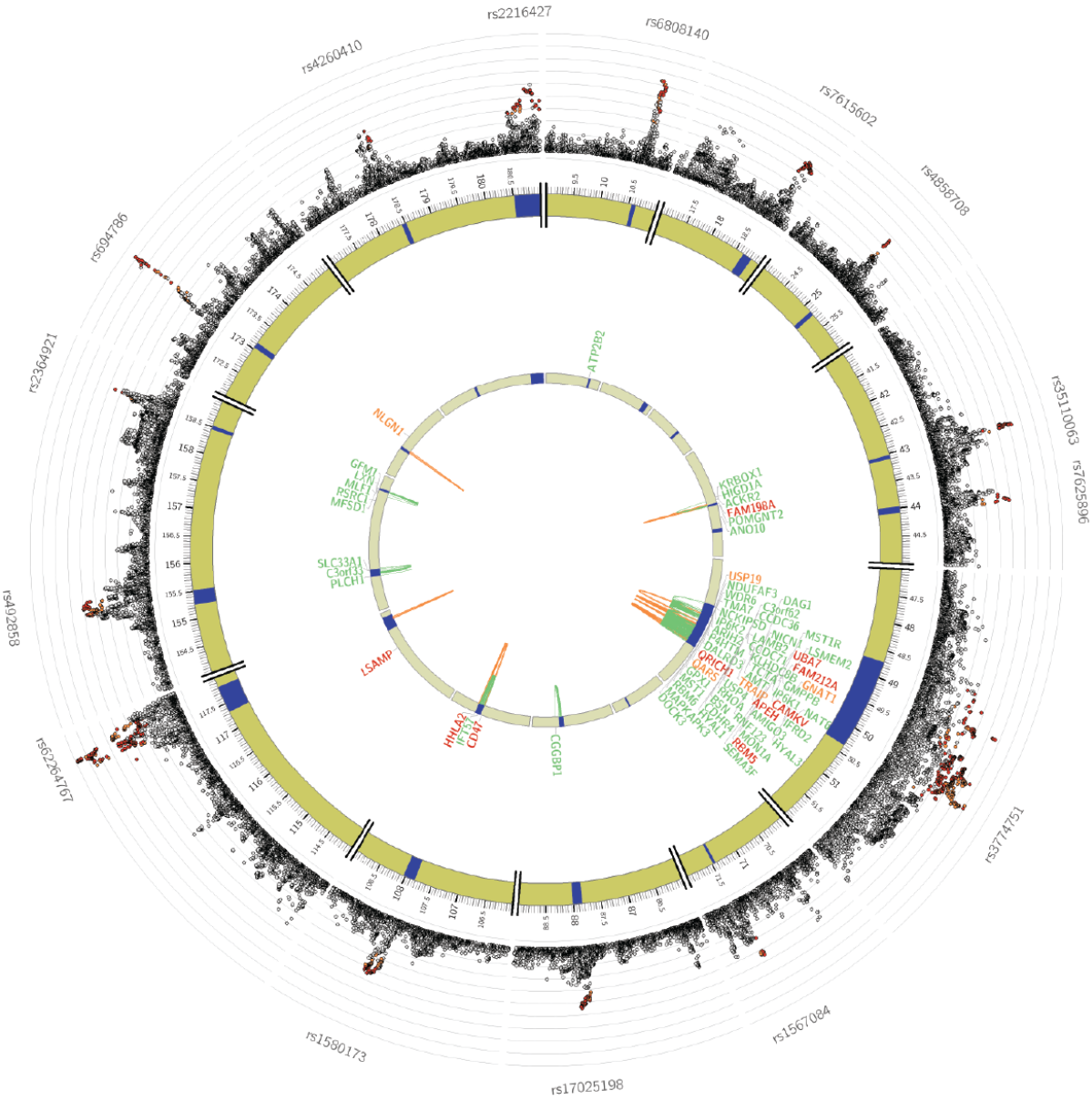
bioRxiv preprint doi: https://doi.org/10.1101/214973; this version posted February 1, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Chromosome 2



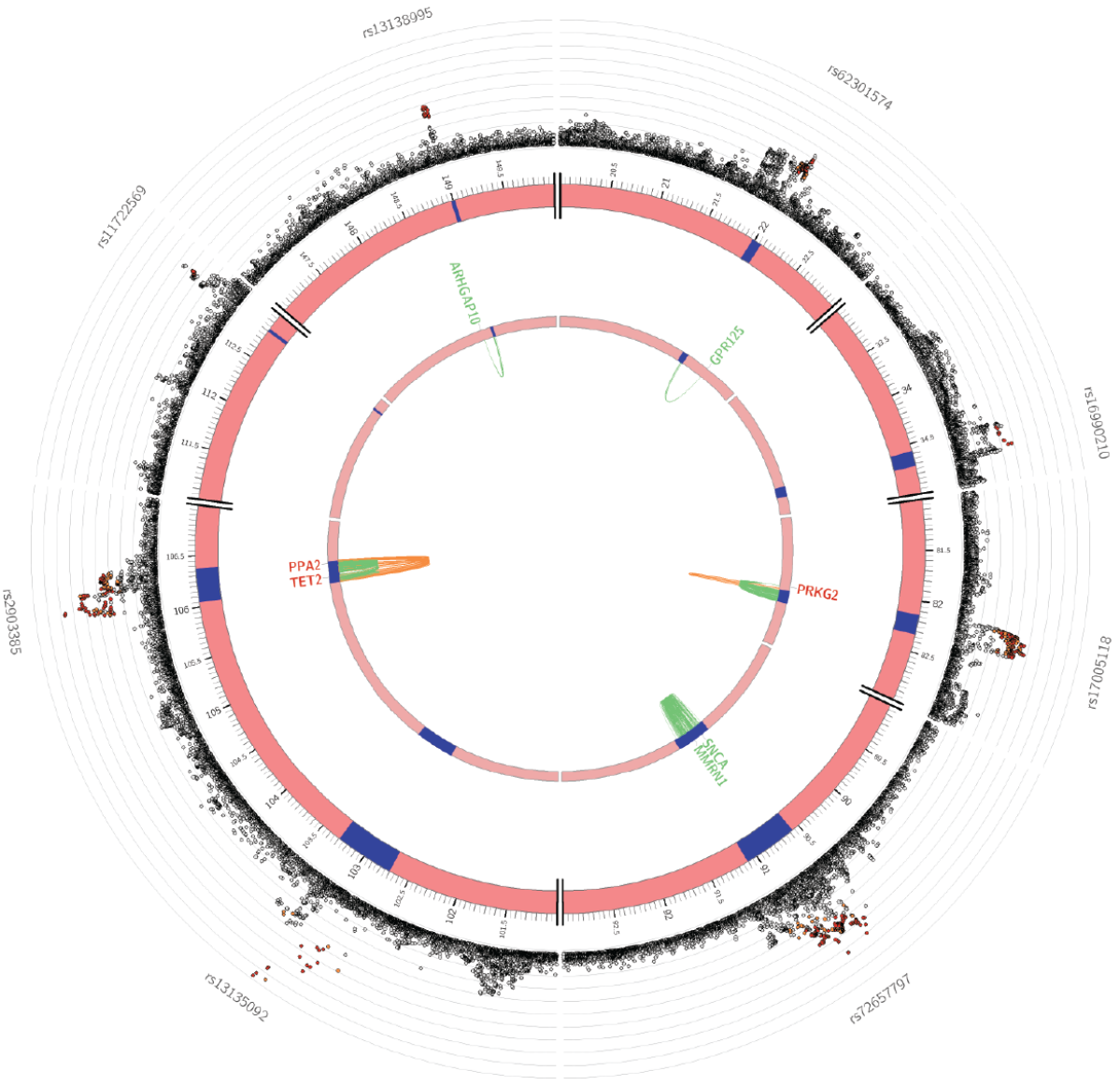
bioRxiv preprint doi: https://doi.org/10.1101/214973; this version posted February 1, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Chromosome 3



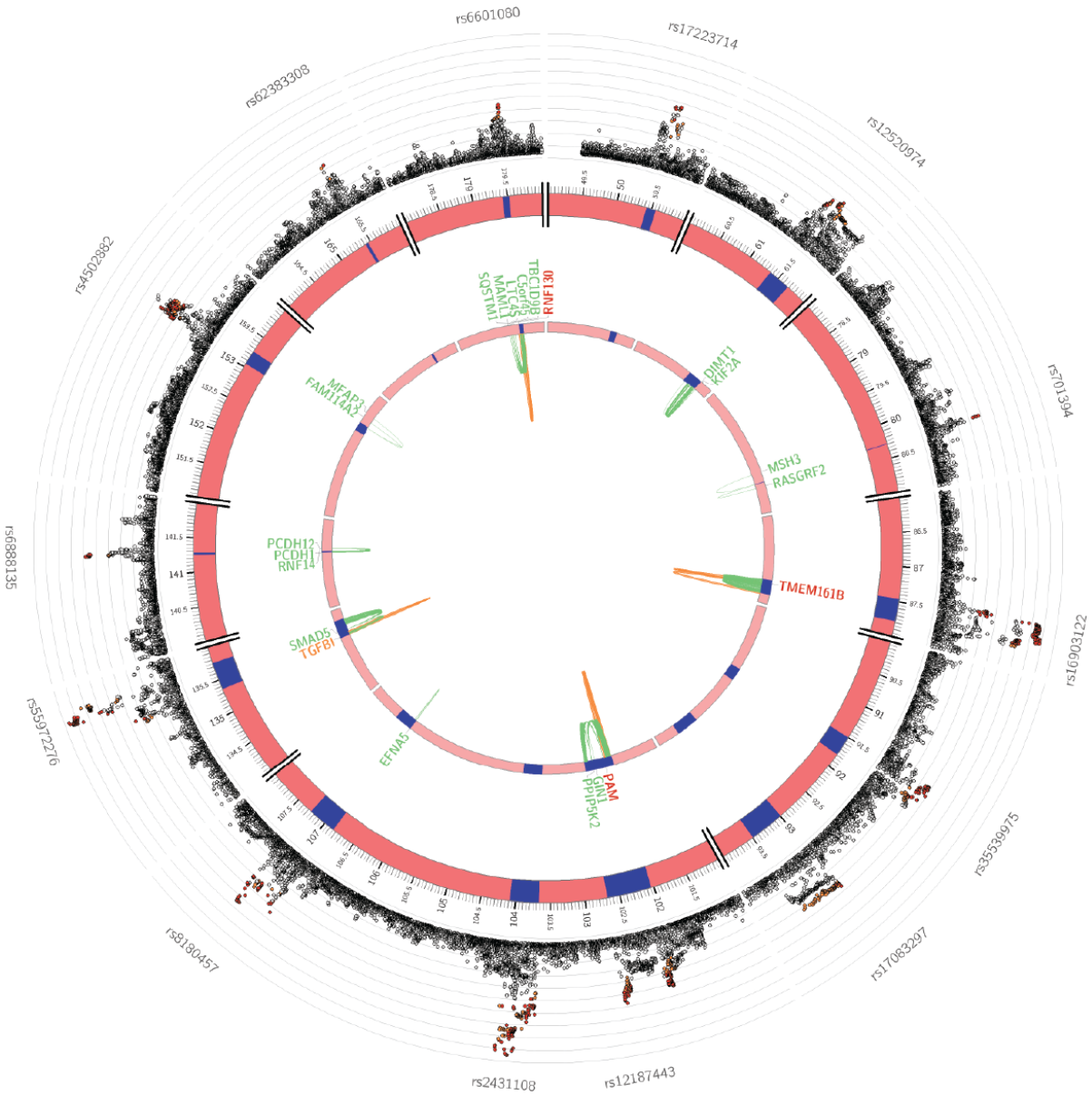
bioRxiv preprint doi: https://doi.org/10.1101/214973; this version posted February 1, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Chromosome 4



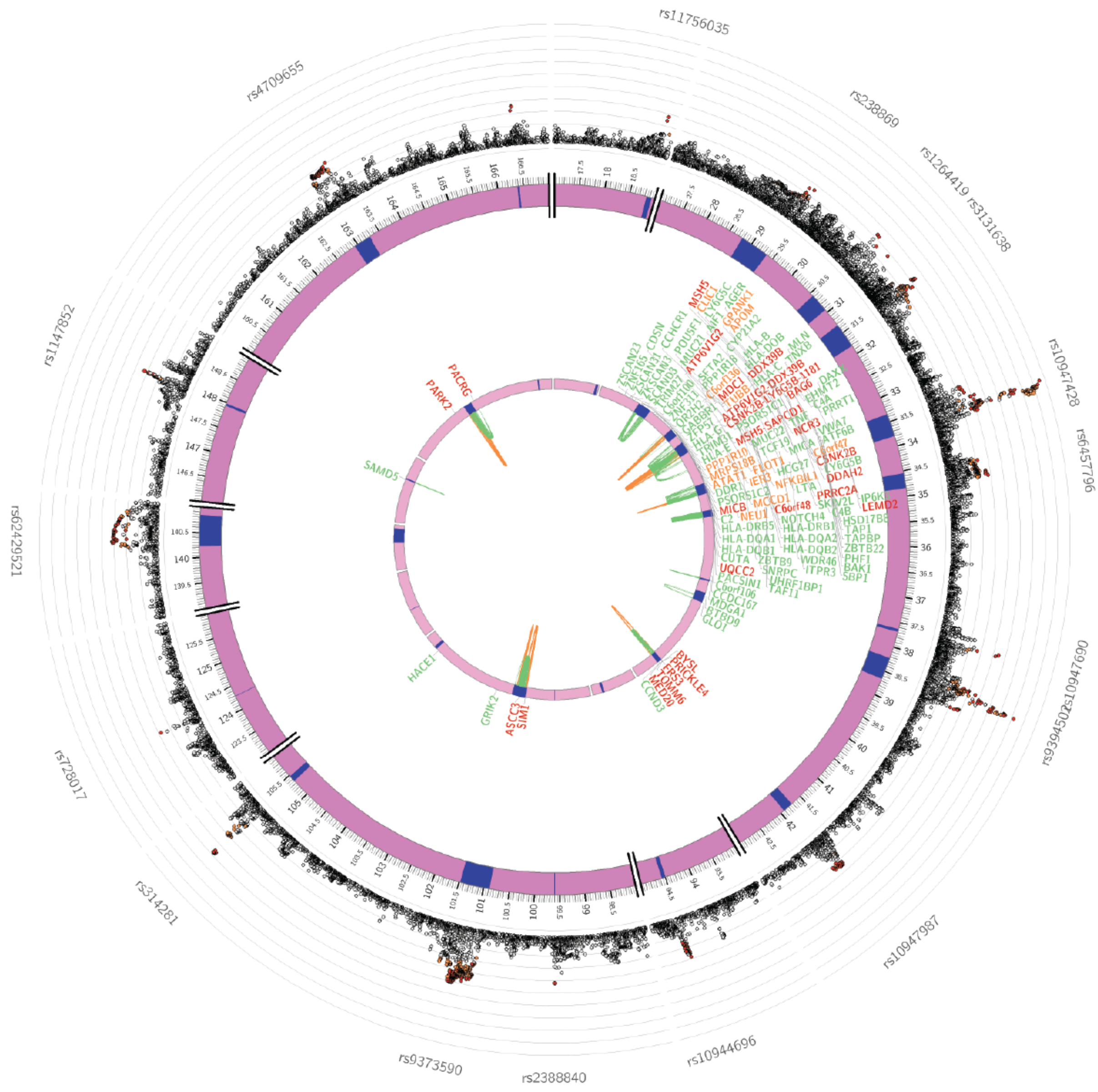
bioRxiv preprint doi: https://doi.org/10.1101/214973; this version posted February 1, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Chromosome 5



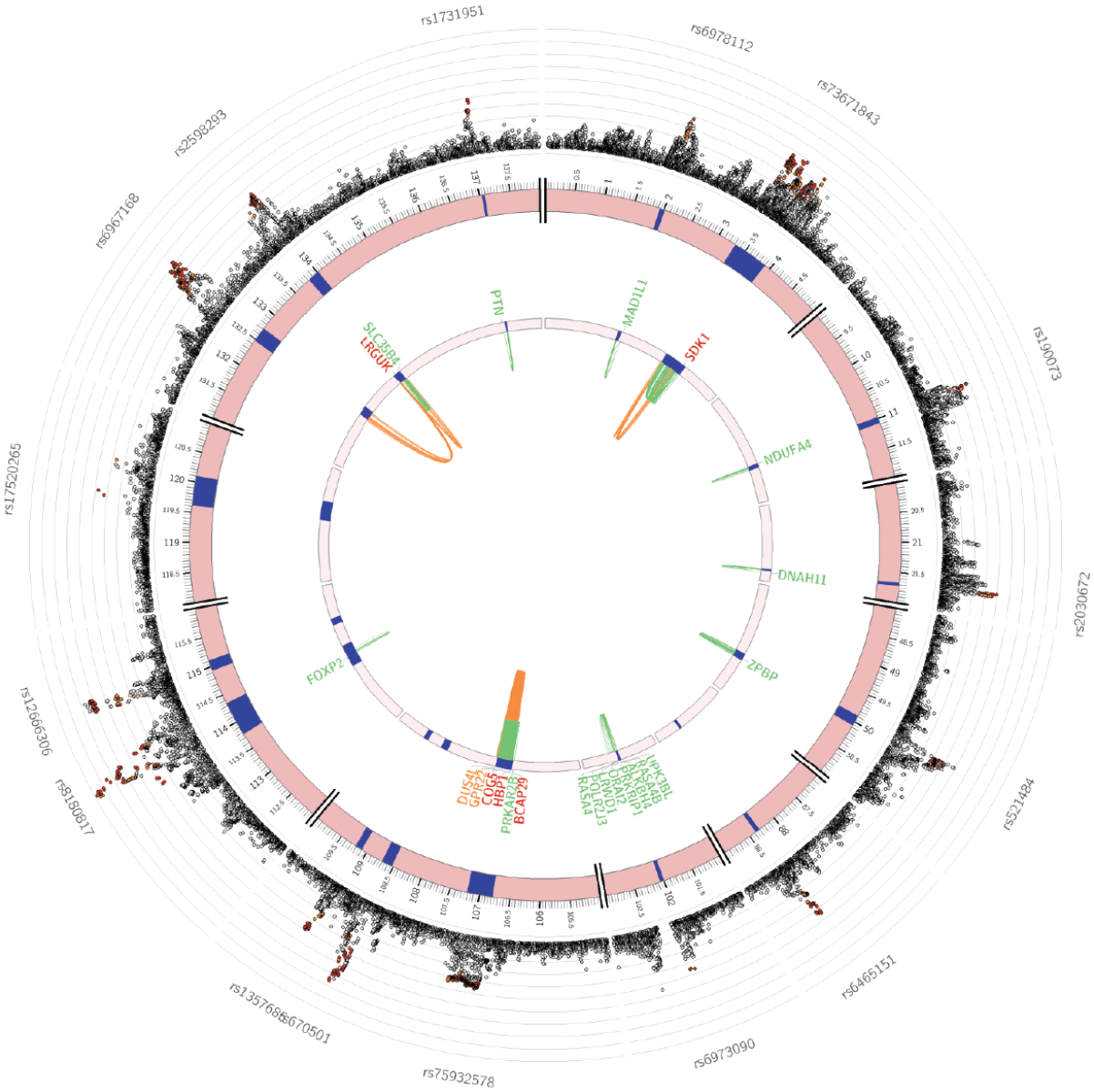
bioRxiv preprint doi: https://doi.org/10.1101/214973; this version posted February 1, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Chromosome 6



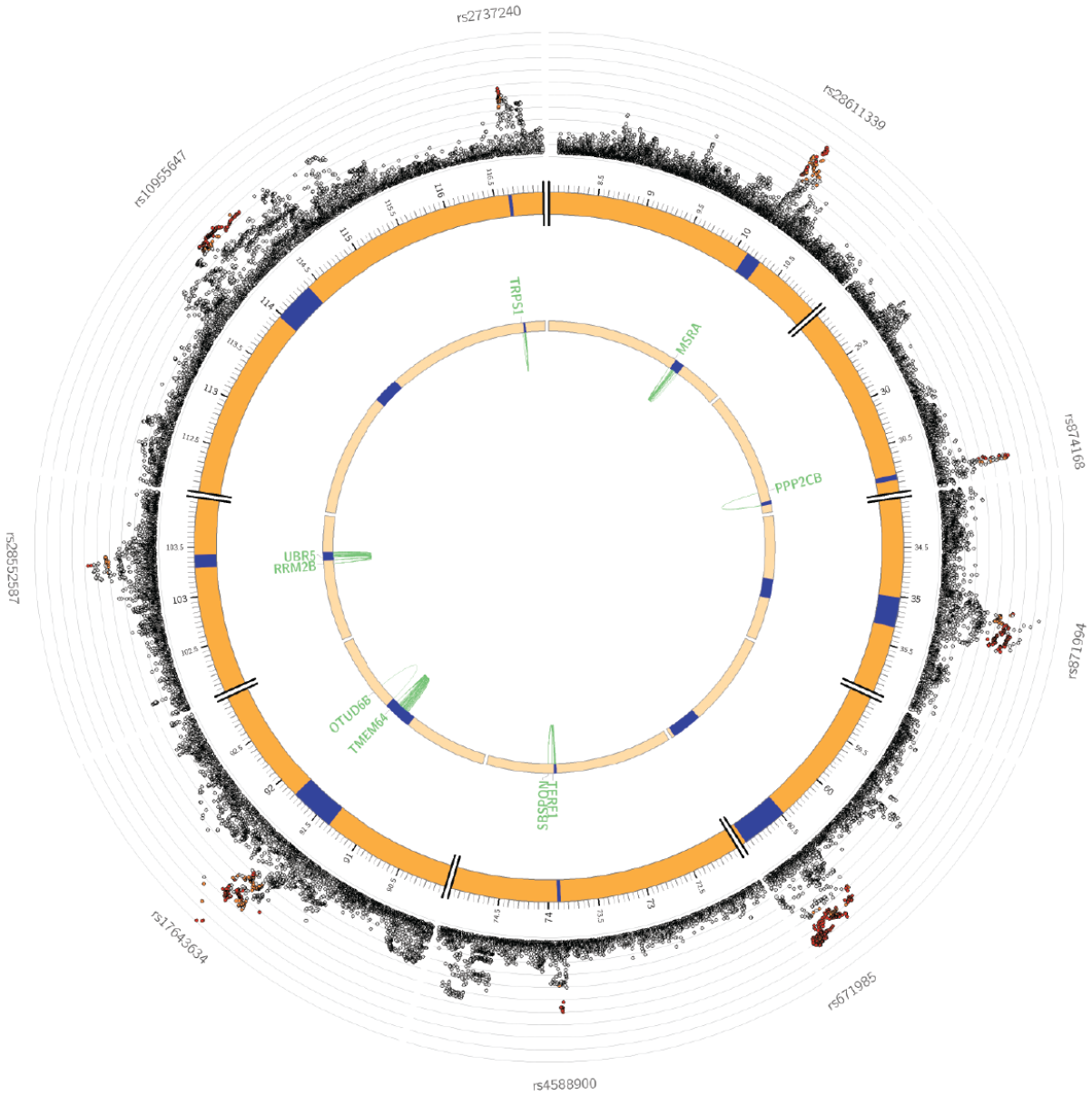
bioRxiv preprint doi: https://doi.org/10.1101/214973; this version posted February 1, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Chromosome 7



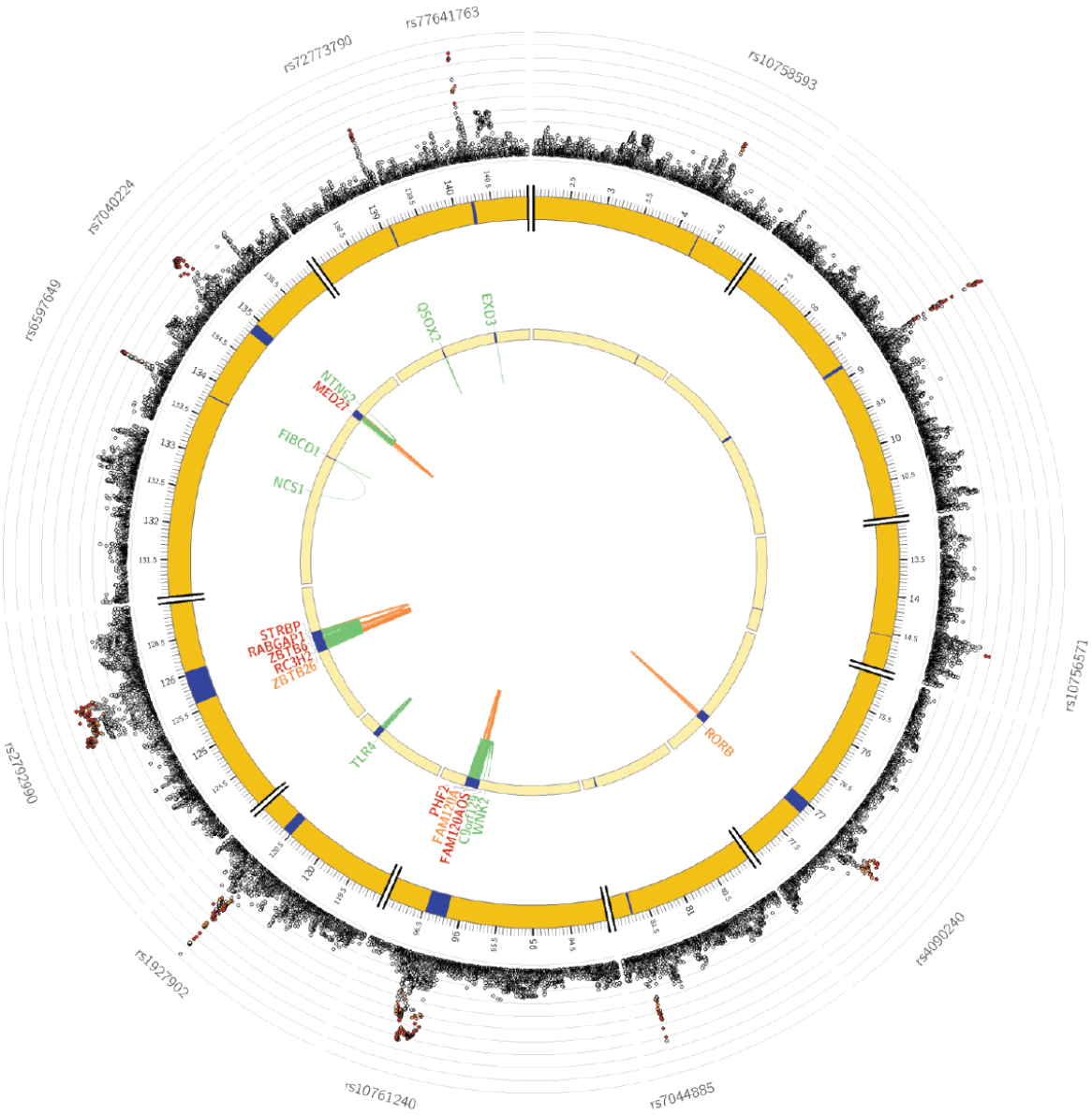
bioRxiv preprint doi: https://doi.org/10.1101/214973; this version posted February 1, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Chromosome 8



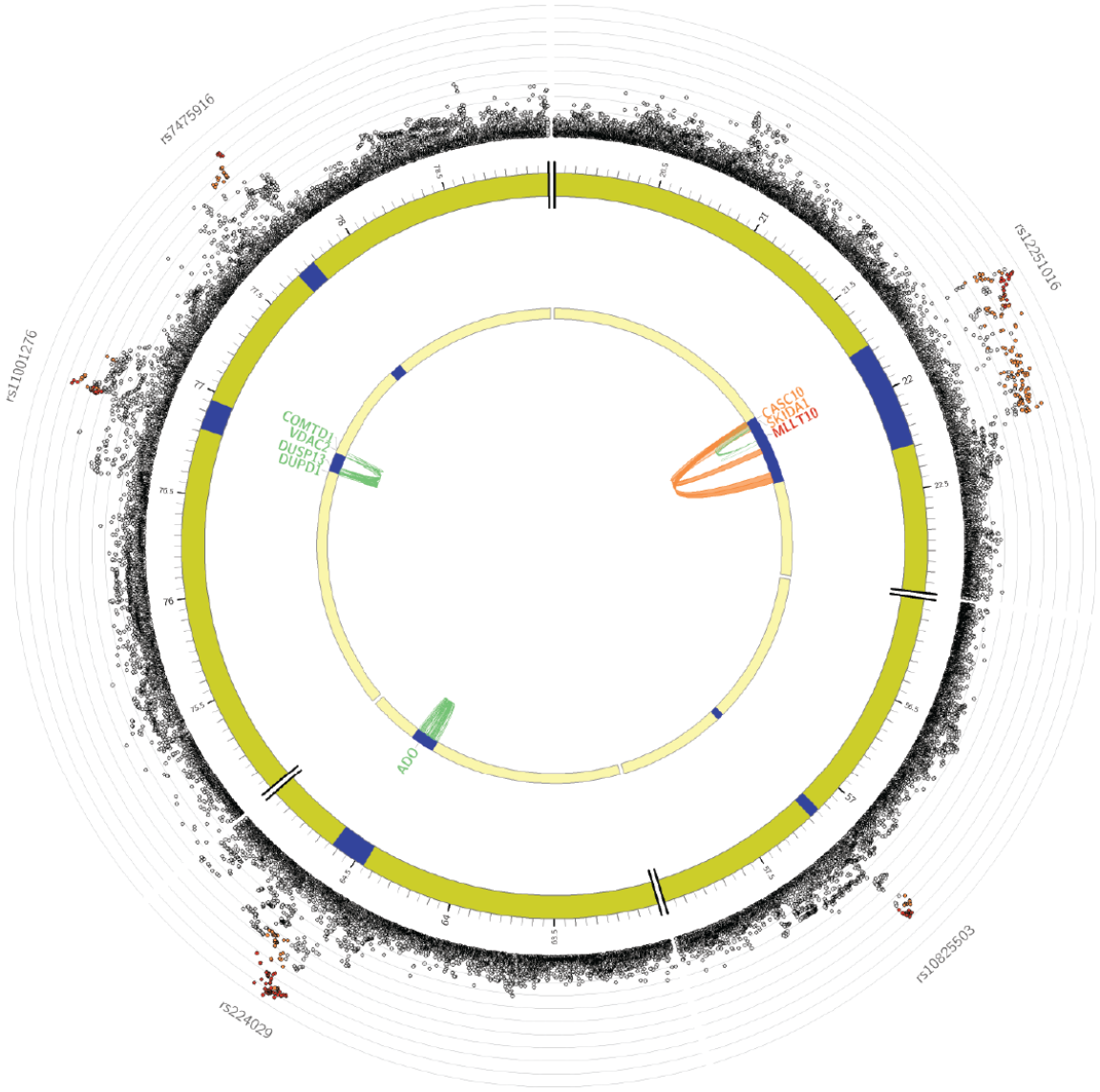
bioRxiv preprint doi: https://doi.org/10.1101/214973; this version posted February 1, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Chromosome 9

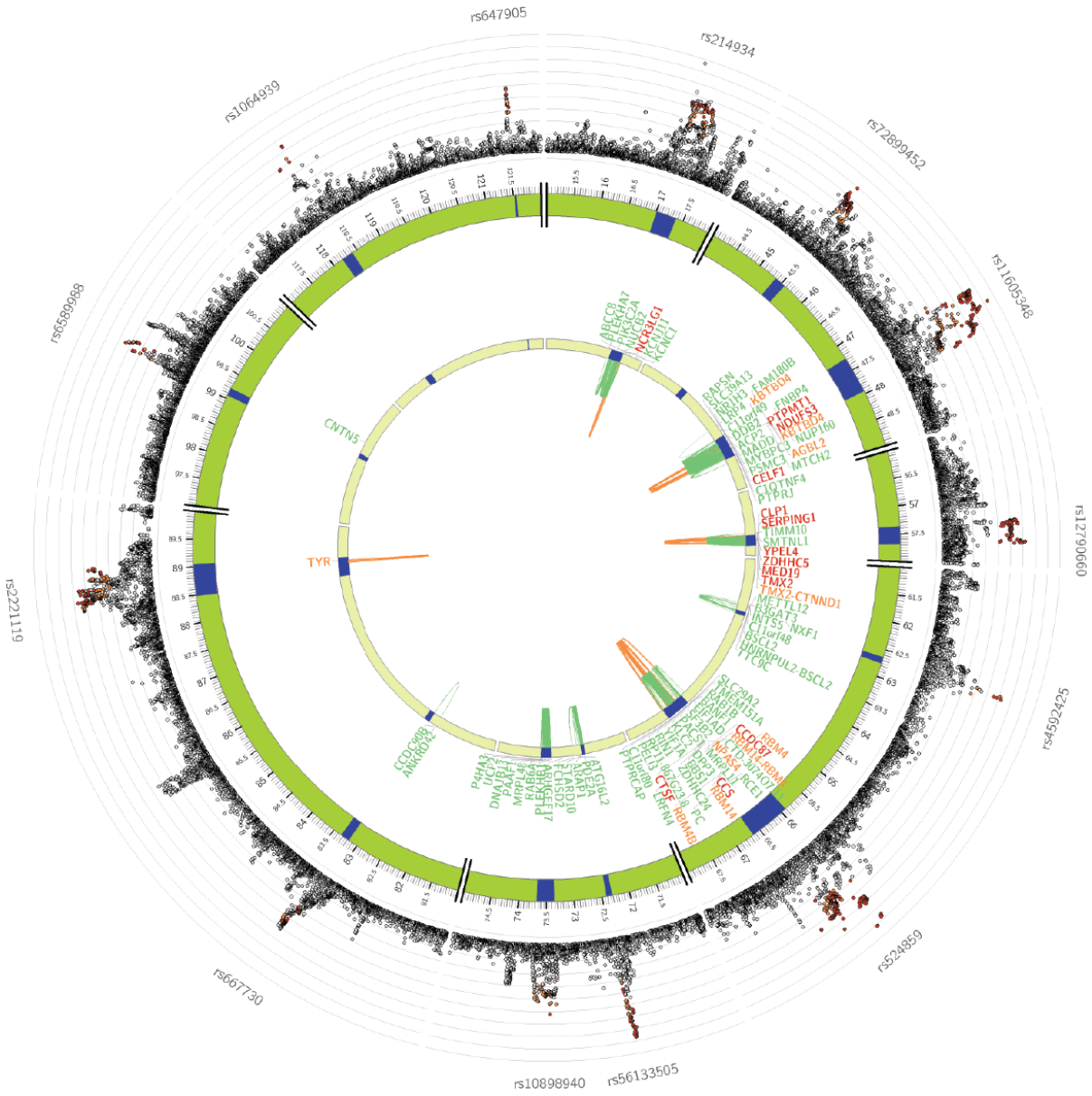


bioRxiv preprint doi: https://doi.org/10.1101/214973; this version posted February 1, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Chromosome 10

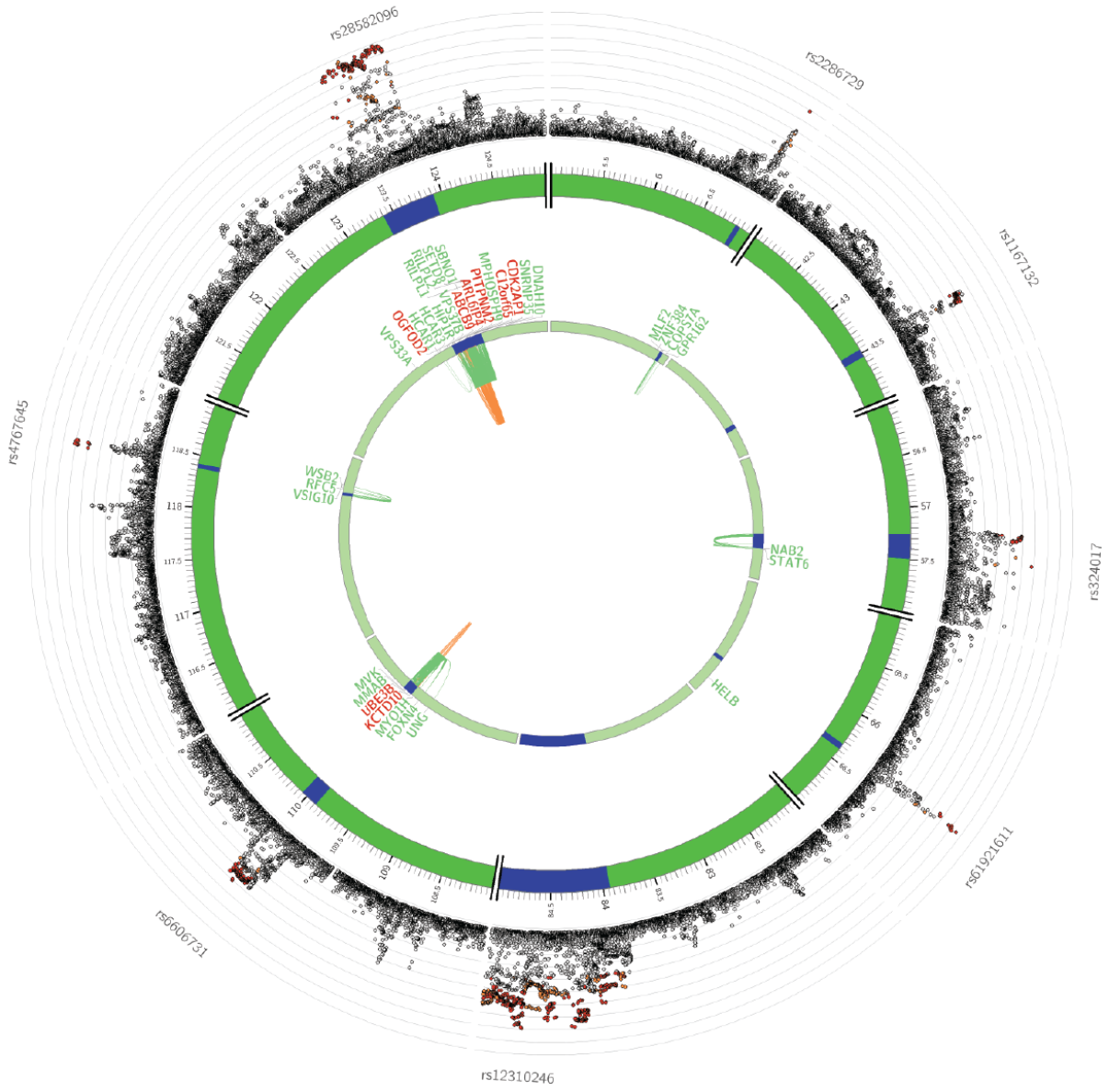


bioRxiv preprint doi: https://doi.org/10.1101/214973; this version posted February 1, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.



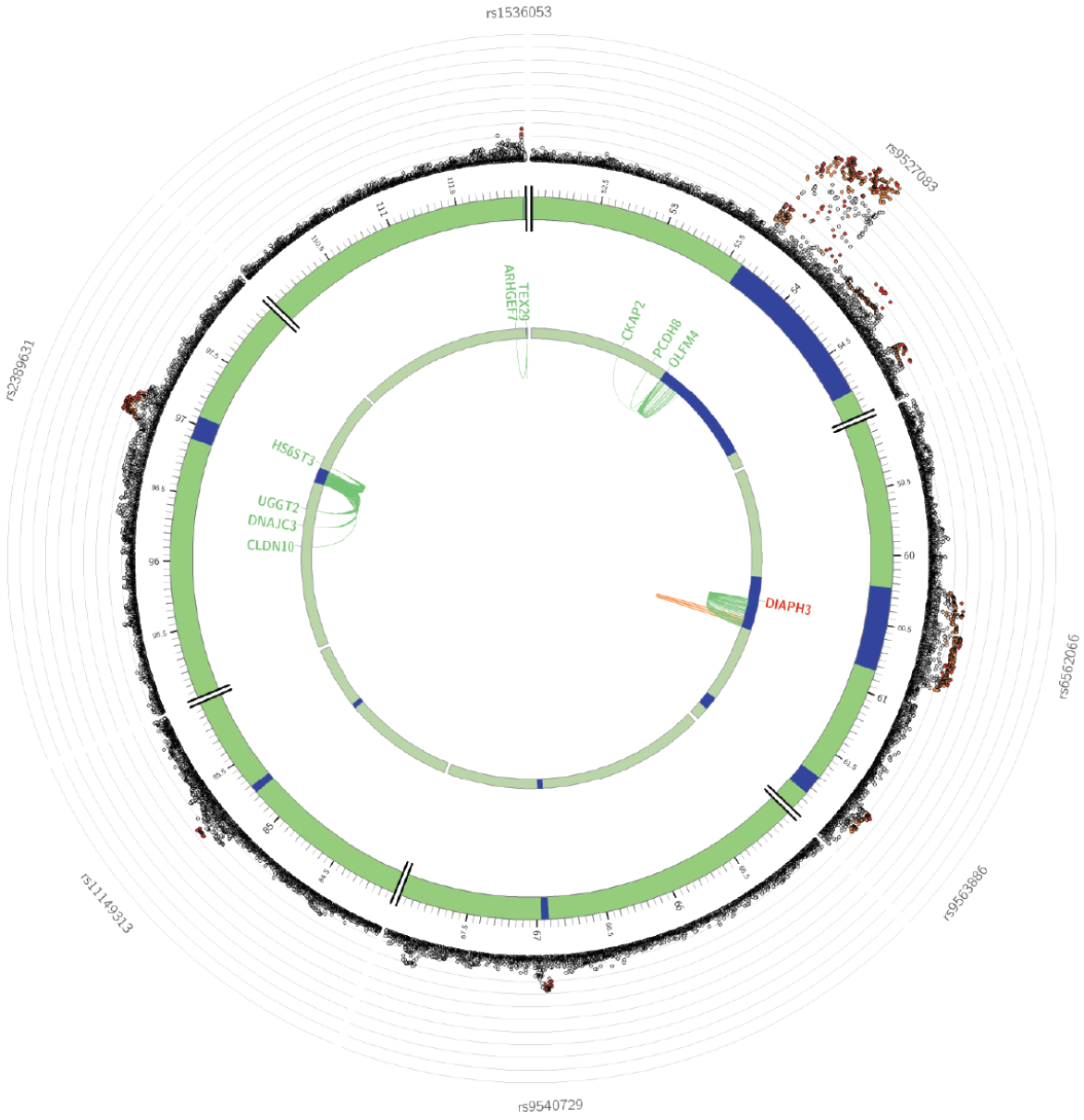
bioRxiv preprint doi: https://doi.org/10.1101/214973; this version posted February 1, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Chromosome 12



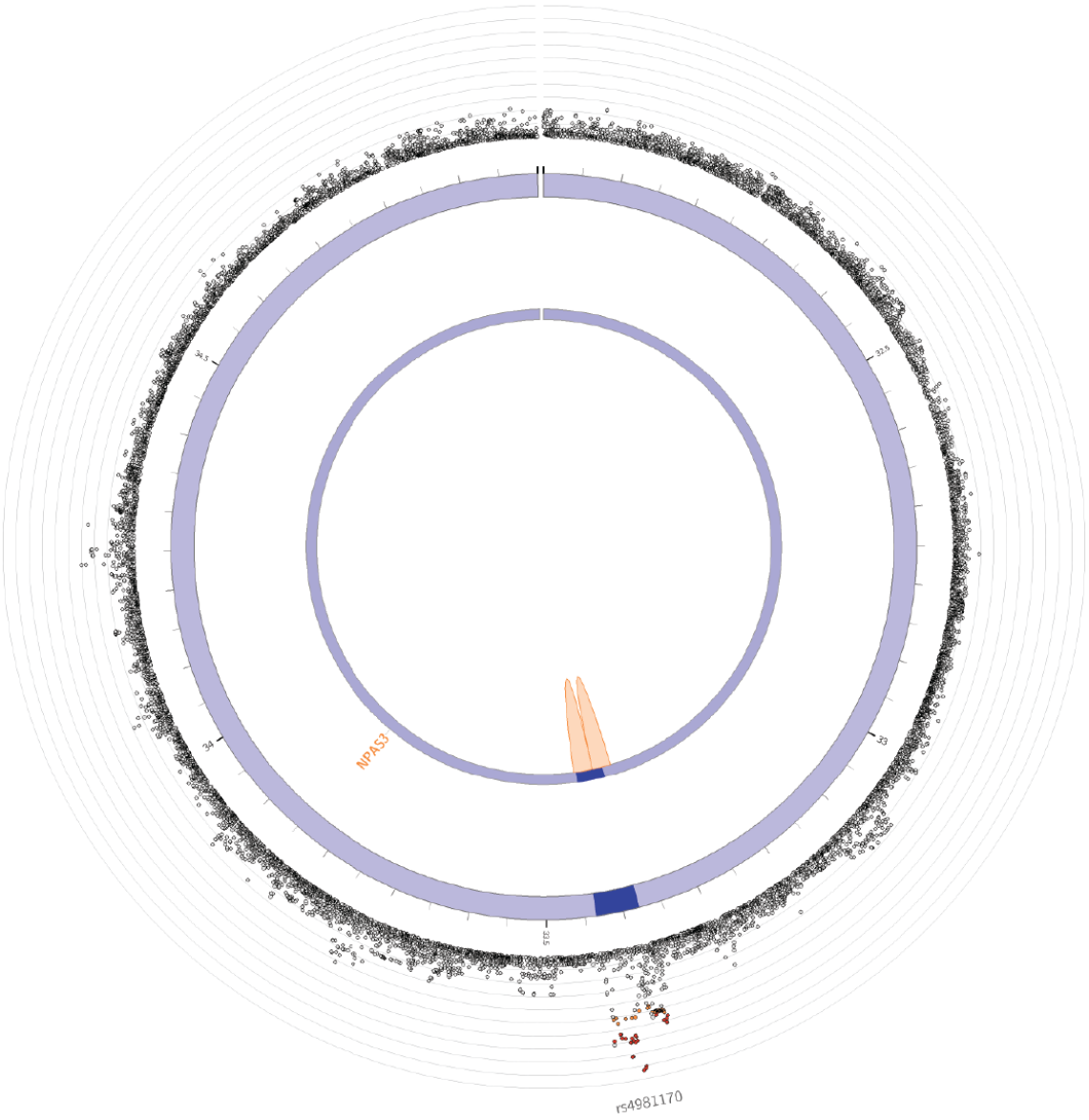
bioRxiv preprint doi: https://doi.org/10.1101/214973; this version posted February 1, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Chromosome 13



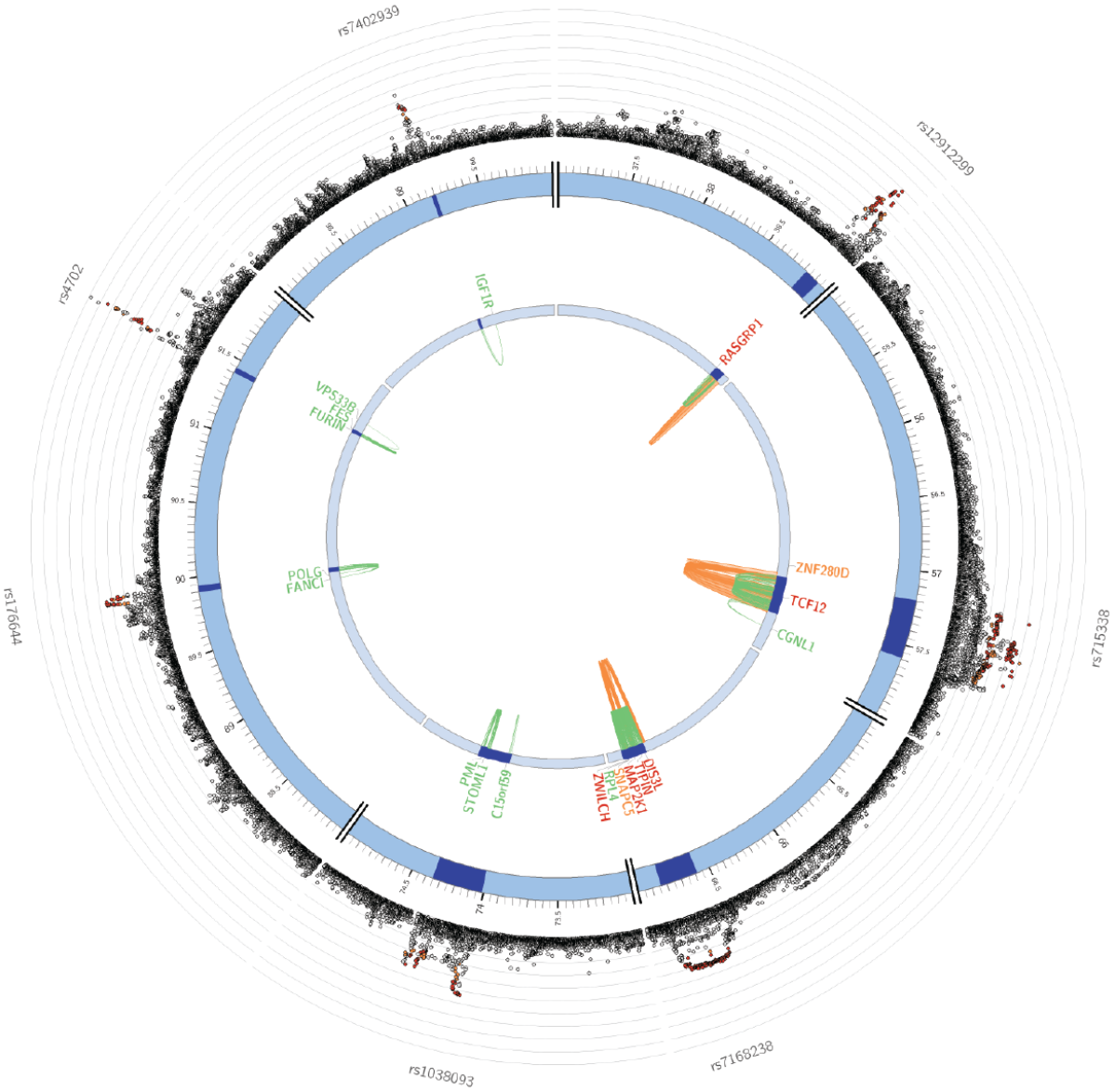
bioRxiv preprint doi: https://doi.org/10.1101/214973; this version posted February 1, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Chromosome 14



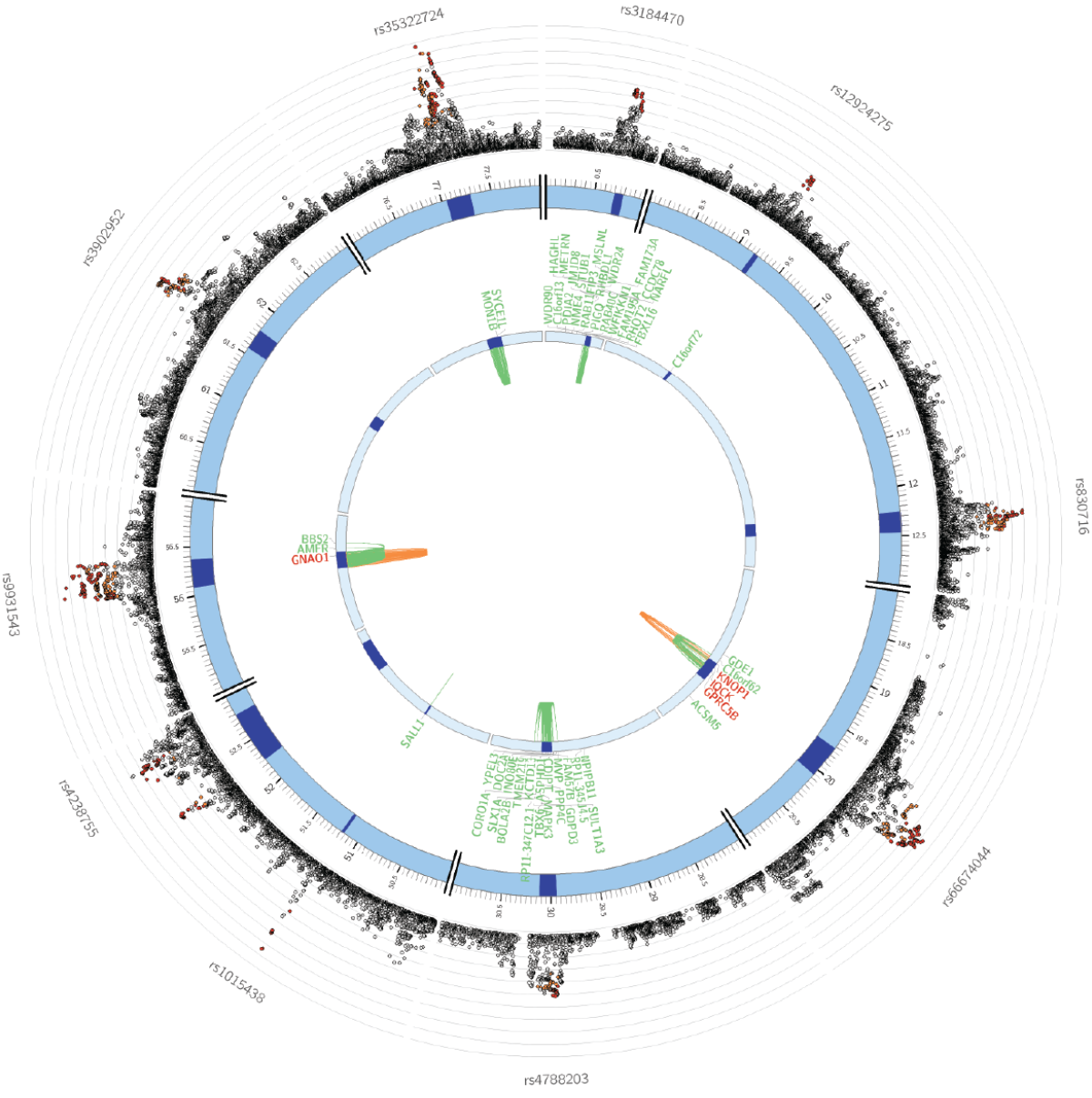
bioRxiv preprint doi: https://doi.org/10.1101/214973; this version posted February 1, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Chromosome 15



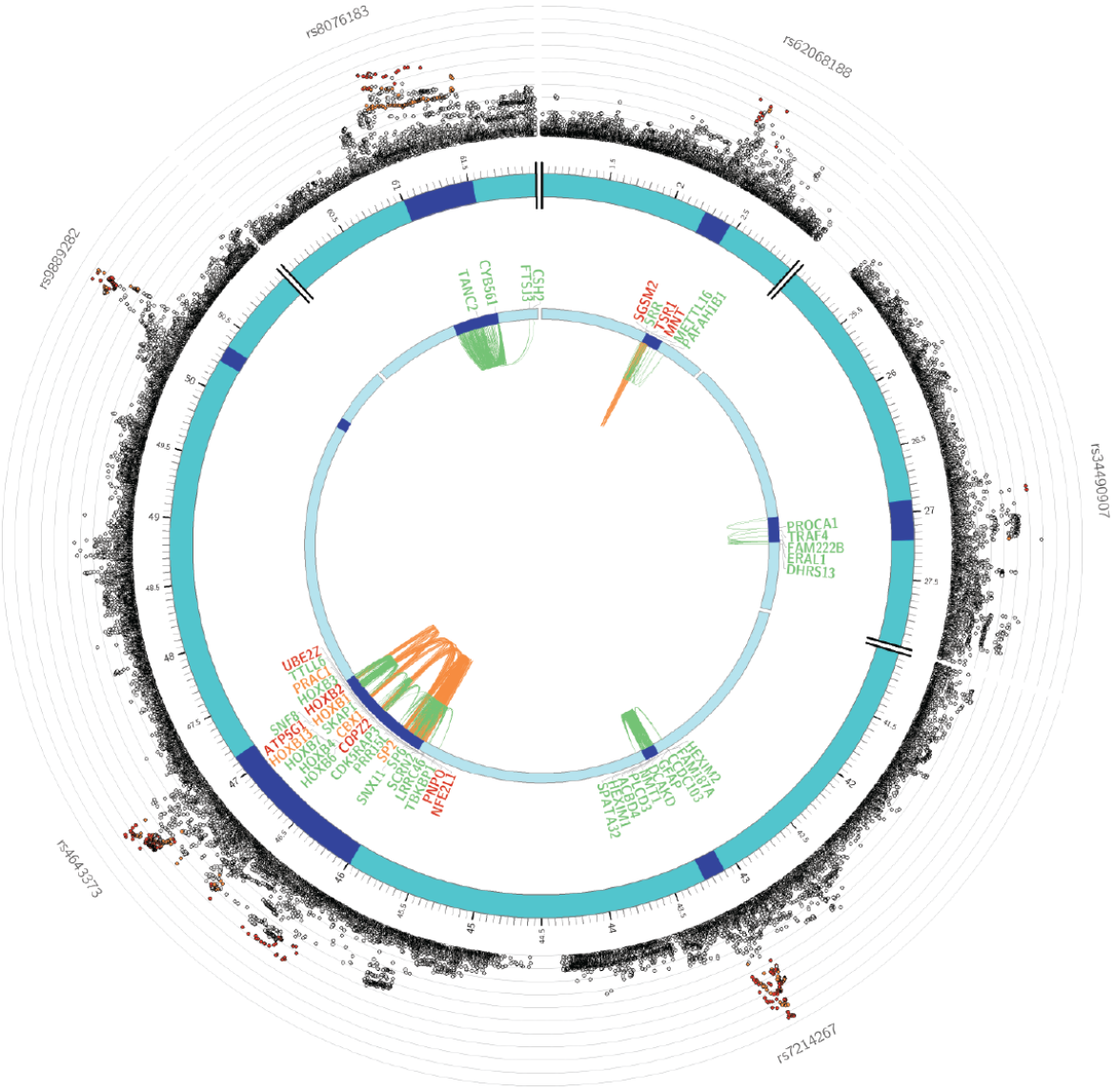
bioRxiv preprint doi: https://doi.org/10.1101/214973; this version posted February 1, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Chromosome 16



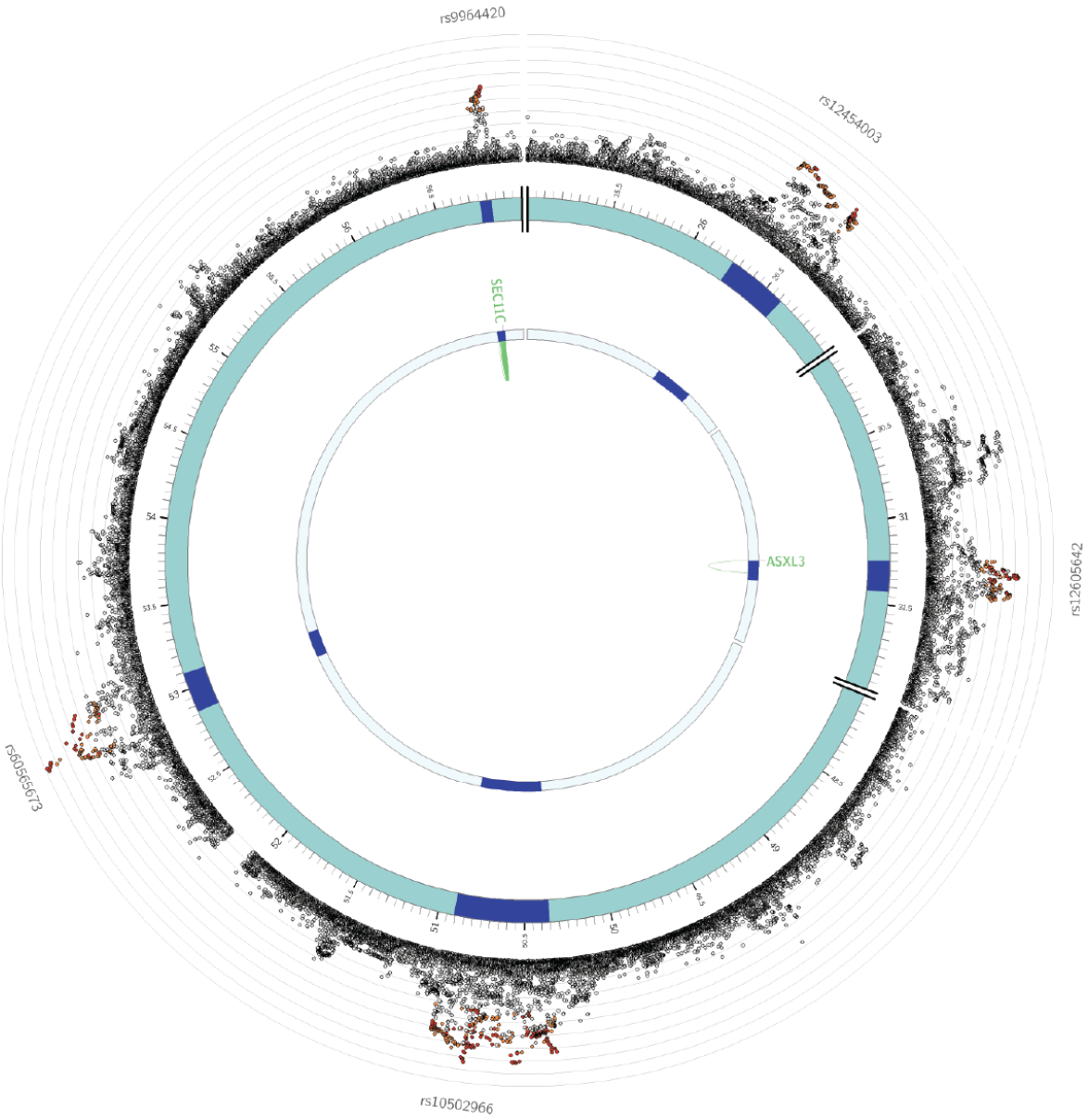
bioRxiv preprint doi: https://doi.org/10.1101/214973; this version posted February 1, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Chromosome 17



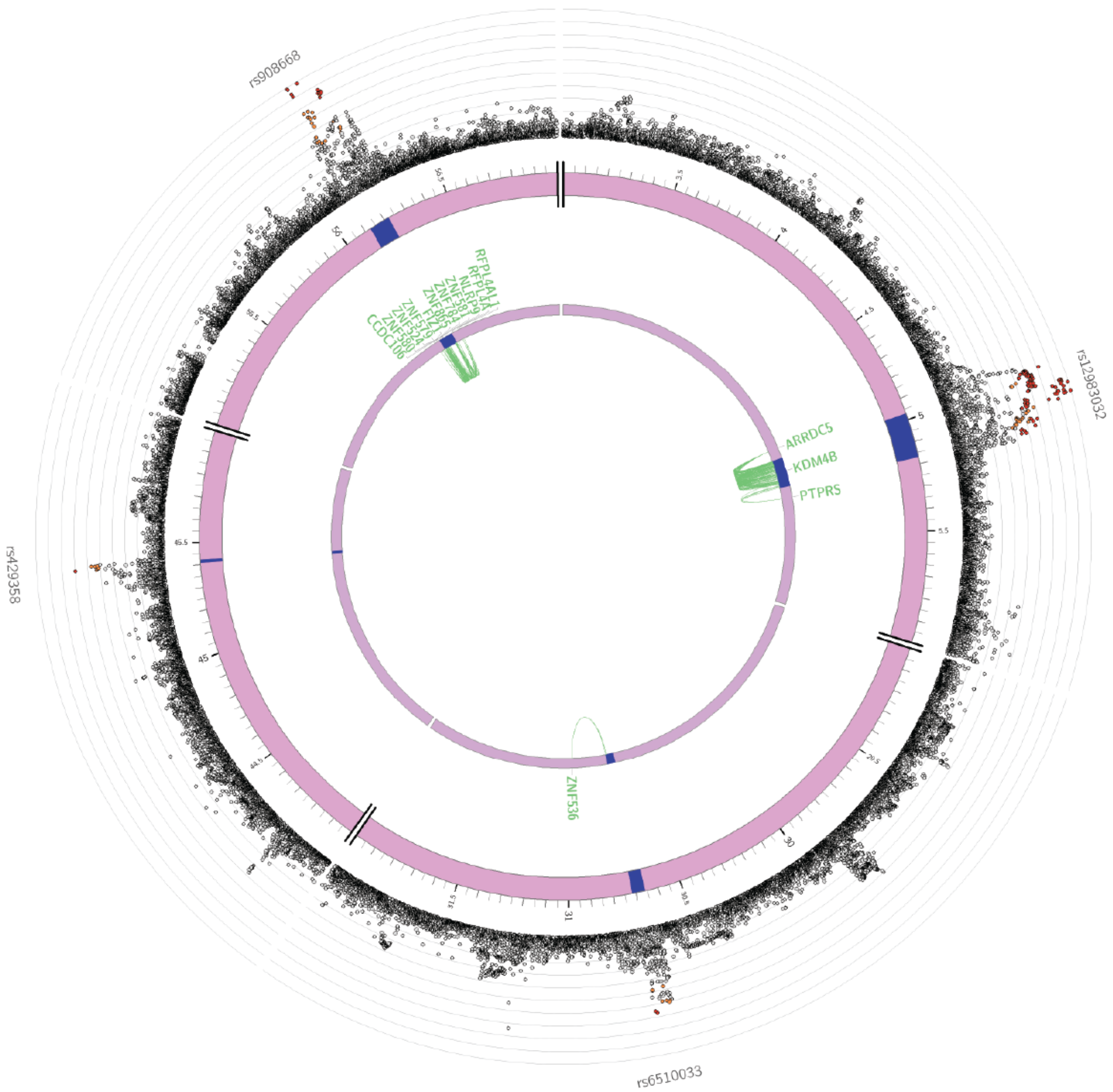
bioRxiv preprint doi: https://doi.org/10.1101/214973; this version posted February 1, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Chromosome 18



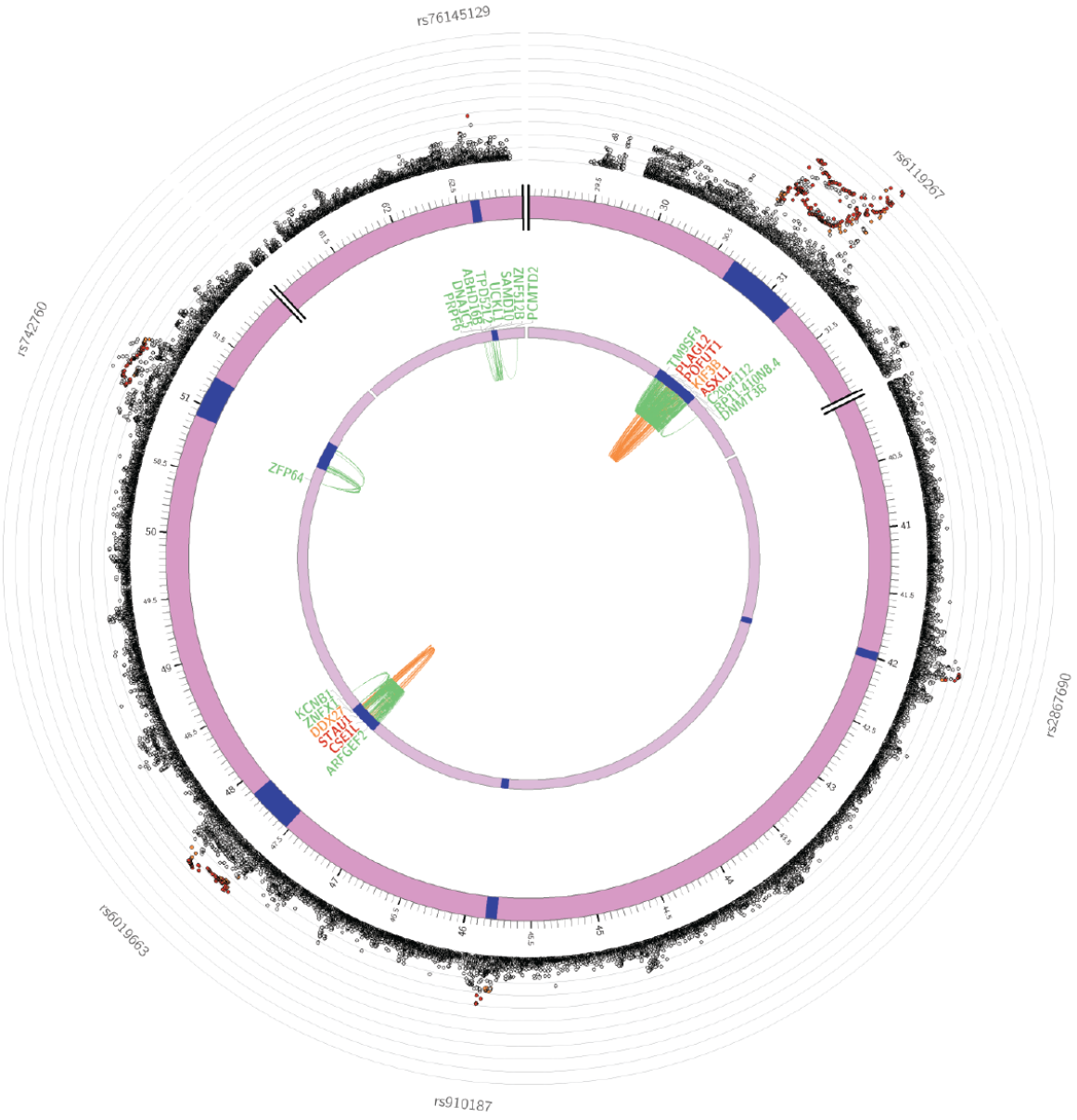
bioRxiv preprint doi: https://doi.org/10.1101/214973; this version posted February 1, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Chromosome 19



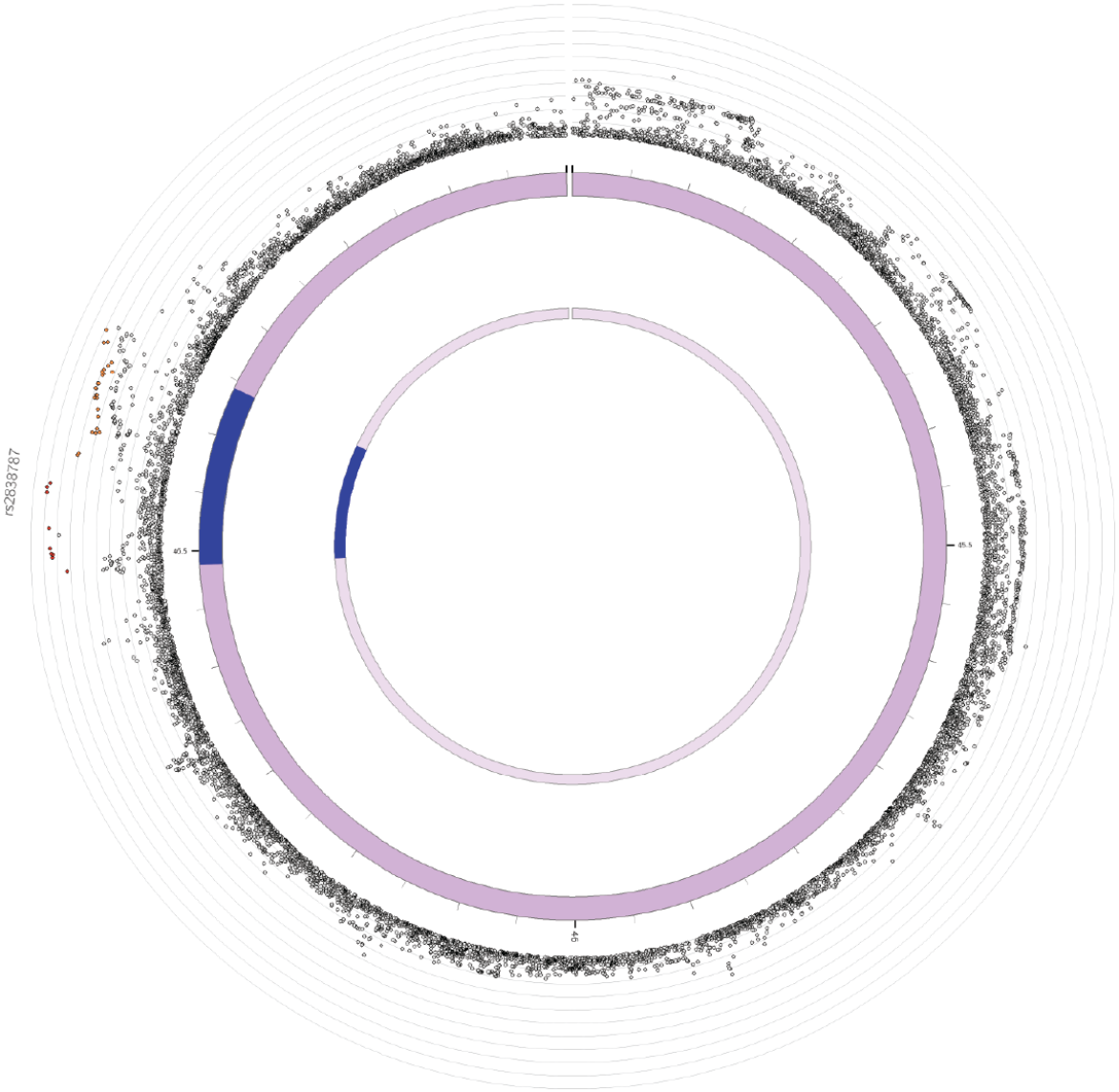
bioRxiv preprint doi: https://doi.org/10.1101/214973; this version posted February 1, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Chromosome 20



bioRxiv preprint doi: https://doi.org/10.1101/214973; this version posted February 1, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Chromosome 21



bioRxiv preprint doi: https://doi.org/10.1101/214973; this version posted February 1, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

Chromosome 22