Trans-ancestry genome-wide analysis of atrial fibrillation provides new insights into disease biology and enables polygenic prediction of cardioembolic risk

Kazuo Miyazawa[1], Kaoru Ito[1*], Zhaonan Zou[2], Hiroshi Matsunaga[1,3], Satoshi Koyama[1,4,5], Hirotaka Ieki[1,3], Seitaro Nomura[3], Masato Akiyama[6,7], Ryo Kurosawa[1], Hiroki Yoshida[1,3], Kouichi Ozaki[1,8], Yoshihiro Onouchi[1,9], BioBank Japan Project[10], Atsushi Takahashi[6,11], Koichi Matsuda[12], Yoshinori Murakami[13], Hiroyuki Aburatani[14], Michiaki Kubo[15], Yukihide Momozawa[16], Chikashi Terao[6], Shinya Oki[2], Hiroshi Akazawa[3],Yoichiro Kamatani[6,12], and Issei Komuro[3*]

[1]Laboratory for Cardiovascular Genomics and Informatics, RIKEN Center for Integrative Medical Sciences, Yokohama, Japan [2]Department of Drug Discovery Medicine, Kyoto University Graduate School of Medicine, Kyoto, Japan [3]Department of Cardiovascular Medicine, Graduate School of Medicine, The University of Tokyo, Tokyo, Japan [4]Program in Medical and Population Genetics, Broad Institute of Harvard and MIT, Cambridge, MA, USA [5]Cardiovascular Research Center, Massachusetts General Hospital, Boston, MA, USA [6]Laboratory for Statistical and Translational Genetics, RIKEN Center for Integrative Medical Sciences, Kanagawa, Japan [7]Department of Ocular Pathology and Imaging Science, Kyushu University Graduate School of Medical Sciences, Fukuoka, Japan [8]Division for Genomic Medicine, Medical Genome Center, National Center for Geriatrics and Gerontology, Obu, Japan [9]Department of Public Health, Chiba University Graduate School of Medicine, Chiba, Japan [10]Institute of Medical Science, The University of Tokyo, Tokyo, Japan [11]Department of Genomic Medicine, Research Institute, National Cerebral and Cardiovascular Center, Suita, Japan [12]Department of Computational Biology and Medical Science, Graduate School of Frontier Sciences, The University of Tokyo, Tokyo, Japan [13]Division of Molecular Pathology, Institute of Medical Science, The University of Tokyo, Tokyo, Japan [14]Genome Science Division, Research Center for Advanced Science and Technology, The University of Tokyo, Tokyo, Japan

[15]RIKEN Center for Integrative Medical Sciences, Kanagawa, Japan [16]Laboratory for Genotyping Development, RIKEN Center for Integrative Medical Sciences, Kanagawa, Japan

Correspondence should be addressed to K.I. (kaoru.ito@riken.jp), Y.K.(kamatani.yoichiro@edu.k.u-tokyo.ac.jp) and I.K. (komuro-tky@umin.ac.jp)

Abstract

To understand the genetic underpinnings of atrial fibrillation (AF) in the Japanese population, we performed a large-scale genome-wide association study comprising 9,826 cases of AF among 150,272 individuals and identified five new susceptibility loci, including East Asian-specific rare variants. A trans-ancestry meta-analysis of >1 million individuals, including 77,690 cases, identified 35 novel loci. Leveraging gene expression and epigenomic datasets to prioritize putative causal genes and their transcription factors revealed the involvement of *IL6R* gene and transcription factor ERG besides the known ones. Further, we constructed a polygenic risk score (PRS) for AF, using the trans-ancestry meta-analysis. PRS was associated with an increased risk of long-term cardiovascular and stroke mortality, and segregated individuals with cardioembolic stroke in undiagnosed AF patients. Our results provide novel biological and clinical insights into AF genetics and suggest their potential for clinical applications.

Atrial fibrillation (AF) is the most common cardiac arrhythmia, affecting approximately 46.3 million individuals worldwide[1]. The global prevalence of AF is increasing due to the rapid aging of the general population and intensified search for subclinical AF[2]. Despite progress in diagnostic and therapeutic technologies, a substantial number of patients with AF are admitted with life-threatening complications such as stroke and heart failure[3], causing a considerable burden on patients and public healthcare systems[4]. Besides conventional clinical risk factors such as ageing, obesity, hypertension, and heart failure, the genetic contribution to the development of AF is also widely recognized. Recent genome-wide association studies (GWASs) have identified more than 100 AF-associated loci, some of which are involved in cardiac developmental, electrophysiological, contractile, and structural pathways[5-8]. However, since the vast majority of AF-GWASs have been predominantly performed in European populations, the genetic pathophysiology of AF in non-European populations has not been comprehensively understood and it is difficult to apply polygenic risk scores (PRSs) derived from such GWASs to non-European populations.

Therefore, in the present study, we sought to explore the genetic architecture of AF in a non-European population and improve the statistical power of AF-GWASs by performing a large-scale Japanese GWAS, followed by conducting a trans-ancestry meta-analysis. Further, we investigated the biological role of the identified AF-associated loci by leveraging gene expression and epigenomic datasets. Additionally, we developed a PRS derived from the trans-ancestry meta-analysis and assessed the impact of the PRS on relevant phenotypes and long-term mortality, which may provide evidence for the clinical utility of AF-PRS and lay the foundation for the realization of precision medicine in AF.

Results

Five novel loci for AF identified in the Japanese GWAS

An overview of the study design is shown in Supplementary Fig. 1. We performed a GWAS on the case-control dataset from the BioBank Japan (BBJ) that comprised 9,826 AF cases and 140,446 controls, using 16,394,105 variants in the autosomes and 423,039 variants in the X chromosome with a minor allele frequency (MAF) > 0.1%. The GWAS identified 31 AF-associated loci with genome-wide significance, of which five were previously unreported (Table 1, Supplementary Table 1, Supplementary Fig. 2, and Supplementary Datasets 1 and 2). The proportion of the variation in AF (the single nucleotide polymorphism (SNP) heritability; h^2) explained by the total genome-wide genetic variation detected in the current Japanese GWAS was estimated to be 6.1% (standard error of mean [s.e.m.] 1.4%), and the liability-scale h^2 was estimated at 11.7% (s.e.m. 2.6%) using linkage disequilibrium (LD)-score regression.

We found two novel loci whose lead variants were observed only in East Asian according to the Genome Aggregation Database v2.1.1 (gnomAD)[9]: one low-frequency and one rare lead variant, rs202030113 (MAF = 1.2%) and rs778479352 (MAF = 0.25%), respectively. The lead variant on 6q25.1, rs202030113, was located in the intronic region three bases away from an exon-intron boundary of *SYNE1* and was predicted as a splice donor loss with the spliceAI delta score of 0.33[10]. *SYNE1* encodes nesprin-1 (spectrin repeat) protein, and together with Sad1p/UNC84-domain-containing proteins (SUN1/2), compose the nuclear envelope protein complex via its nucleoplasmic domains to lamin A/C. Mutations in *LMNA* and *SYNE1* genes have been identified in patients with severe muscle dystrophy and dilated cardiomyopathy[11,12]. Mutations in the *SYNE1* gene cause defects in nuclear morphology, myoblast differentiation, and heart development[13], altering the nuclear envelope protein complex that contributes to the structural substrate in atrial arrhythmogenesis. A strong signal (odds ratio [OR] for AF development = 2.00, 95% confidence interval [CI] = 1.73-2.31, $P = 1.6 \times 10^{-20}$) was displayed by rs778479352, the lead variant located in an intron of *FGF13*. *FGF13* encodes a member of the fibroblast growth factor family, which

possesses broad mitogenic and cell survival activities. *FGF13* directly binds to the C-terminus of the main cardiac sodium channel ($Na_V1.5$) in the sarcolemma; knockdown of *FGF13* in rat cardiomyocytes showed a loss of function of $Na_V1.5$-reduced Na^+ current density, decreased Na^+ channel availability, and slowed $Na_V1.5$-reduced Na^+ current recovery from inactivation[14]. This evidence of conduction disturbance in cardiomyocytes indicates that *FGF13* is an important target gene associated with AF.

To identify AF-associated variants independent of the lead variant at each locus, we performed a stepwise conditional analysis, in which 18 independent variants (locus-wide $P < 5.0\times10^{-6}$) were additionally detected, increasing the total number of AF-associated signals to 49 (Supplementary Fig. 3 and Supplementary Table 2). We identified 10 loci that had multiple independent association signals, especially in the *PITX2-C4orf32* locus with 6 association signals (N_{signal} = 2: *GORAB-PRRX1*, *CAND2*, *HAND2-AS1*, *FANCC*, *NEBL*, *AKAP6*, *ZFHX3* loci, N_{signal} = 3: *LINC02459-TBX5* locus, N_{signal} = 4: *NEURL1* locus, and N_{signal} = 6: *PITX2-C4orf32* locus). Of these additional signals, three variants were observed only in East Asian populations in gnomAD (rs577463446 at the *FANCC* locus, MAF = 0.5%, OR = 1.58; rs965277670 at *NEBL* locus, MAF = 0.6%, OR = 1.65; rs201901902 at *NEURL1* locus MAF = 0.5%, OR = 1.68).

Sex-stratified GWAS for AF

Sex-related differences in the epidemiology, clinical presentation, and prognosis of AF have been consistently reported[15,16]. For example, the age-stratified prevalence of AF is lower in female than in male, despite a higher risk for AF-related outcomes such as stroke in female[17-19]. Thus, we first perfomed sex-stratified GWAS and calculated the genetic correlation between sexes using LD-score regression to explore the genetic heterogeneity. We then found a significant correlation (r_g = 1.095, s.e.m. = 0.112, $P = 1.2 \times 10^{-22}$). Next, we examined genetic loci with sex-dependent effects on AF development in the sex-stratified GWAS. We found 21 loci with genome-wide significance in the male GWAS, but only six loci in the female GWAS (Supplementary Fig. 4a and Supplementary Table 3). Furthermore, by comparing the effect sizes of the lead variants

identified in the sex-stratified GWAS, we observed significant positive correlation and concordant allelic effects between sexes (Spearman's $\rho = 0.92$, $P = 7.4 \times 10^{-7}$; Supplementary Fig. 4b). We found several variants with significant heterogeneity in the *PITX2-C4orf32* and *CUX2* loci ($P_{het} < 1.0 \times 10^{-4}$). Although the *PITX2-C4orf32* locus reached genome-wide significance in both sex-stratified GWAS (β_{male} (s.e.m.$_{male}$) = 0.569 (0.018), $P_{male} = 1.9 \times 10^{-210}$; β_{female} (s.e.m.$_{female}$) = 0.381 (0.027), $P_{female} = 3.0 \times 10^{-46}$, for rs12644625, the lead variant in the *PITX2-C4orf32* locus), the *CUX2* locus in the female GWAS did not achieve genome-wide significance level (β_{male} (s.e.m.$_{male}$) = -0.280 (0.024), $P_{male} = 9.7 \times 10^{-33}$; β_{female} (s.e.m.$_{female}$) = -0.105 (0.034), $P_{female} = 2.1 \times 10^{-3}$, for rs12644625, the lead variant in the *CUX2* locus). *CUX2* is a transcription factor regulating sex-biased gene expression in rat and mouse liver[20,21], and the variant in the *CUX2* locus is associated with a DNA methylation site near the *ALDH2* gene, encoding aldehyde dehydrogenase, which catalyzes the oxidation of aldehydes into carboxylic acids[22]. Moreover, according to Genotype-Tissue Expression (GTEx) data[23], *CUX2* gene expression in liver was significantly different between male and female ($P = 7.8 \times 10^{-4}$, Wilcoxon rank-sum test; Supplementary Fig. 5). These results suggest that sex-biased gene expression related to the metabolism of aldehydes in the liver may partially explain the sex-related differences in the genetic basis of AF.

Trans-ancestry meta-analysis identified 33 novel loci for AF

To improve the statistical power to detect further genetic associations with AF, we conducted a trans-ancestry meta-analysis by combining the current Japanese GWAS (BBJ) and two European GWASs: a large-scale meta-analysis of European populations (EUR)[7] and a biobank data of FinnGen data release 2 (FIN). All three datasets yielded 77,690 cases (BBJ: 9,826; EUR: 60,620; FIN: 7,244) and 1,167,040 controls (BBJ: 140,446; EUR: 970,216; FIN: 56,378). We tested a total of 5,158,449 variants with MAF ≥ 1% and identified 150 AF-associated loci with genome-wide significance ($\log_{10}$ Bayes factor (BF) > 6; Fig. 1, Supplementary Table 4, and Supplementary Datasets 3 and 4). Of these loci, 33 have not been reported previously, including three novel loci detected in the current Japanese GWAS. In total, we identified 35 novel loci through the current

Japanese GWAS and trans-ancestry meta-analysis (Table 2).

Of the 5,855 variants in LD ($r^2 > 0.6$) with 150 lead variants, 27 missense variants were observed (Supplementary Table 5). Among novel loci, we found a missense variant, rs848208 (p.Ala970Val), in the *SPEN* gene, encoding a hormone-inducible transcriptional co-regulator that activates and represses downstream targets. *SPEN*-deficient zebrafish embryos developed bradycardia, atrioventricular block, and heart chamber fibrillation with downregulation of connexin 43 expression[24], which is a well-known component of gap junctions and is associated with the cardiac conduction system[25]. Another missense variant in novel loci, rs3746471 (p.Arg1045Trp), was located on the *KIAA1755* gene, which is reported to be associated with heart rate[26] and heart rate variability[27]. Given the robust relationship between autonomic nervous dysfunction and AF[28], *KIAA1755* can be a potential target gene for neural modulation contributing to AF management; however, the biological association between the *KIAA1755* gene and AF has not been fully examined.

Shared allelic effects between Japanese and European populations and fine mapping/credible set analyses in the trans-ancestry meta-analysis

We compared alternate allele frequencies and allelic effects for 150 lead variants between BBJ, EUR, and FIN. Compared to substantial concordance in allele frequencies between EUR and FIN (Spearman's $\rho = 0.974$, $P < 2.2 \times 10^{-16}$), we observed a moderate correlation in allele frequencies between BBJ and EUR as well as BBJ and FIN ($\rho = 0.592$, $P = 1.5 \times 10^{-15}$ and $\rho = 0.632$, $P < 2.2 \times 10^{-16}$, respectively; Supplementary Fig. 6a-c). Additionally, we found a significant positive correlation between concordant allelic effects of these variants ($\rho = 0.769$, $P < 2.2 \times 10^{-16}$ for BBJ vs. EUR; $\rho = 0.769$, $P < 2.2 \times 10^{-16}$ for BBJ vs. FIN; Supplementary Fig. 6d-f). To further explore the relationship of allelic effects between populations, we performed a trans-ancestry genetic correlation analysis, which showed strong correlation of BBJ with EUR and FIN (BBJ and EUR: $r_g = 0.990$, s.e.m. = 0.097, BBJ and FIN $r_g = 0.955$, s.e.m. = 0.344).

To assess the contribution of trans-ancestry meta-analyses to the refinement of the

putative causal variants, we constructed 99% credible sets for the 150 AF-associated loci detected by the current trans-ancestry meta-analysis, and compared the number of variants included in the 99% credible sets derived from three combinations of the meta-analysis (EUR + BBJ, EUR + FIN, BBJ + EUR + FIN; Supplementary Fig. 6g, h). The size of the 99% credible sets derived from EUR + BBJ significantly decreased compared with those from EUR + FIN (P = 0.004, paired Wilcoxon rank-sum test). In addition, trans-ancestry meta-analysis of three datasets (BBJ + EUR + FIN) yielded the most significant decrease in the number of variants among all the combinations of meta-analyses (median number of variants = 12; interquartile range [IQR] 5 – 36; Supplementary Table 6). In particular, a single variant, rs67329386, was identified on the *ZFHX3* locus with a high posterior probability of 0.995, only in the trans-ancestry meta-analysis of three datasets. In previous AF-GWAS, several variants were found in the *ZFHX3* locus[29,30], where a large transcription factor, *ZFHX3*, together with *PITX2*, facilitated DNA binding and transcription activity[31]. To corroborate the role of rs67329386 as a causal variant, we searched for transcription factors binding to this locus using the ChIP-seq dataset in ChIP-Atlas[32]. Compared to other AF-associated variants within this locus, rs67329386 was distinctly located on the binding site of CEBPB (Supplementary Fig. 7), aiding in the recognition and binding of target gene regulatory regions, leading to cell proliferation, differentiation, immune response, and tumor formation[33].

Pleiotropic effects and functional pathways of AF-associated loci

To characterize the AF-associated loci, we examined the pleiotropic effects of the identified lead variants and proxies using the NHGRI-EBI GWAS catalog database. Of the 150 AF-associated loci, 114 (76.0%) had at least one overlapping variant, and 739 variants overlapped with 200 diseases or traits (Supplementary Table 7). Among them, blood pressure-related traits were most frequently observed (13.0%), followed by traits related to electrocardiogram (10.1%) and heart rate (4.0%).

Given the heterogeneous study design and population in the GWAS catalog, we subsequently utilized the BBJ dataset[34-36] with a consistent study design and population, to

investigate the pleotropic effects. In particular, to explore the biological pathways related to AF development, we searched for AF-associated variants overlapping with quantitative trait loci (QTL) for clinical measurements such as anthropometric, metabolic, kidney-related, and blood pressure data (Supplementary Table 8). We found 160 loci-phenotype association pairs, where distinct functional clusters were observed (Supplementary Fig. 8); positive impact on blood pressure: *CASZ1*, *PRDM8-FGF5*, *CUX2*, and *RPH3A*, and negative impact on renal function: *MIR5096*, *MBD5*, *DNAJC12-SIRT1*, *IGF1R*, and *NFATC1*. Moreover, we observed associations on chromosome 6p21.31 harboring *HMGA1*, in which AF-associated variants overlapped with QTLs for height (rs1759628) and body mass index (BMI) (rs6913361). Both rs1759628 and rs6913361 were located 7 kilobases (kb) and 25 kb upstream of *HMGA1*, respectively, and the risk alleles of these variants were significantly associated with increased height and BMI (β_{height} (s.e.m.$_{height}$) = 0.090 (0.005), $P_{height} = 8.7 \times 10^{-93}$; β_{BMI} (s.e.m.$_{BMI}$) = 0.037 (0.005), $P_{BMI} = 1.7 \times 10^{-12}$, respectively). *HMGA1* encoded a chromatin-associated protein that attach to AT-rich regions of DNA and recruit transcription factors, resulting in upregulation of the insulin receptor gene expression[37] and the variants in *HMGA1* were associated with increased risk of metabolic syndrome[38].

To enhance the biological understanding of the AF-associated loci, we performed in silico tissue and gene-set enrichment analysis using DEPICT software[39], which prioritizes candidate genes and tissues based on gene expression data from different tissues and cell types. Consistent with a previous study[7], we found significant enrichment of AF-associated loci predominantly in heart tissues ($P < 1.2 \times 10^{-5}$ for heart and $P < 2.3 \times 10^{-5}$ for heart atria; Supplementary Fig. 9 and Supplementary Table 9). Additionally, pathway analysis demonstrated that AF-associated loci were significantly enriched in 34 out of 1,157 gene sets, most of which were involved in heart development and morphogenesis (Supplementary Table 10). Among them, the regulation of cell adhesion was a previously unreported pathway ($P = 1.6 \times 10^{-10}$), in which *ZMIZ1*, the nearest gene for rs1769758, was reported to regulate the activity of various transcription factors related to vascular development, androgen receptor coregulation, SMAD3 regulation, and coactivation of p53[40-42]. Knockdown of *ZMIZ1* in mice resulted in severe defects in the reorganization of the yolk

sac vascular plexus and cell proliferation[43].

Prioritization of AF-associated genes and transcription factors using gene expression and epigenetic features

We performed a transcriptome-wide association study (TWAS) using the identified loci in the trans-ancestry meta-analysis and GTEx data[23] to identify candidate genes associated with AF. Given the enrichment of AF-associated loci in the heart tissue, we used the gene expression data of GTEx in the atrial appendage and left ventricle as a reference. TWAS prioritized 86 and 74 candidate causal genes significantly associated with AF in the atrial appendage and left ventricle, respectively (Fig. 2a and Supplementary Table 11). We found only 14 and 15 overlapping genes between nearby genes located on the AF-associated loci and the candidate genes in the atrial appendage and left ventricle, respectively (Fig. 2b). Moreover, the median physical distances from the lead variants to the candidate genes were greater than 100 kb. Notably, 17.4% and 25.6% of the candidate genes were located 500 kb away from the lead variants (Fig. 2c). Further, to test whether the candidate genes were enriched in unique pathways, we used ingenuity pathway analysis and found several canonical pathways, including fatty acid metabolism-related pathways, in which the candidate genes prioritized in both atrial appendage and left ventricle were enriched (Fig. 2d). Intriguingly, *IL6R* was identified as a candidate gene associated with AF in the atrial appendage and was also involved in the canonical associated pathway. Despite increasing evidence for the role of inflammation in the AF pathophysiology[44], including a report on the suggestive association of the *IL6R* locus with AF ($P = 5.0 \times 10^{-4}$)[45], the genetic contribution of inflammatory process to AF development has not been fully elucidated. Our transcriptome-wide analysis revealed a significant association between *IL6R* and AF development, providing supportive evidence for the contribution of inflammation through *IL6R* to the pathophysiological mechanism underlying AF.

Next, we sought to identify transcription factors that bind to AF-associated loci which orchestrate the expression of causative genes involved in AF development. We performed enrichment analysis using the ChIP-Atlas[32] dataset, which comprised several high-throughput ChIP-

seq experiments (15,109 experiments, 1,028 transcription factors). As a result, AF-associated loci were characterized by enriched binding of 33 distinct transcription factors ($P < 0.05$) in cardiovascular cells, pluripotent stem cells-derived cardiac cells, and muscle cells (Fig. 3a and Supplementary Table 12). Of these, ERG, TAL1, and TBX5 were among the top enriched transcription factors and are known to be involved in endothelial gene expression and differentiation[46,47]. In particular, ERG binding was significantly enriched with Bonferroni-corrected significance level of $P = 3.3 \times 10^{-6}$ (0.05/15,109): ERG ChIP-seq peaks overlapped with AF-associated loci around genes encoding cardiac ion channels (*SCN10*, *KCNH2*, *HCN4*, and *CAMK2D*), where active histone markers such as H3K27ac and H3K4me3 in induced pluripotent stem cells (iPSC)-derived cardiac cells were also observed (Fig. 3b and Supplementary Fig. 10). These results suggest that ERG is a transcription factor that regulates the expression of AF-associated genes.

Predictive ability of PRS derived from single-population GWAS and trans-ancestry meta-GWAS

PRS has the potential for risk stratification of complex traits and diseases based on genetic data. However, the transferability of PRS from diverse populations to a population of another ancestry remains challenging. Therefore, we examined the performance of a PRS derived from various combinations of summary statistics in the Japanese population. We split our case-control samples into derivation, validation, and test datasets, and constructed 255 combinations of the summary statistics of three GWAS studies (BBJ, EUR, and FIN) and parameters for PRS derivation. Based on the PRS performance in the validation cohort, we determined the parameters that showed the best performance for each combination of summary statistics (BBJ, FIN, EUR, BBJ + FIN, BBJ + EUR, EUR + FIN, and BBJ + EUR + FIN) (Supplementary Table 13), and assessed the performance of the best model in the test cohort (Fig. 4, Supplementary Fig. 11, and Supplementary Table 14). Concordant with the population specificity, the PRS derived from BBJ showed a significantly better performance than those from European studies ($P < 2.2 \times 10^{-16}$ for both BBJ vs.

EUR and BBJ vs. FIN). Furthermore, the performance of the PRS derived from meta-analysis that included different ancestry groups (EUR + BBJ and FIN + BBJ) significantly exceeded that from a single study ($P < 2.2 \times 10^{-16}$ for all cases) as well as that from a meta-analysis of two European studies ($P < 2.2 \times 10^{-16}$ for both EUR + BBJ vs. EUR + FIN and BBJ vs. FIN + EUR). Among all models, the PRS derived from three studies with multi-ancestry and the largest sample size (BBJ + EUR + FIN) showed the highest performance (pseudo $R^2 = 0.144$, 95% CI = 0.130–0.154, area under the curve of receiver operating characteristic = 0.737, 95% CI = 0.726–0.748).

Impact of AF-PRS on AF-relevant phenotypes and cardiovascular outcomes

To assess the potential for clinical applications of the PRS, we investigated the association between PRS and the onset age of AF in individuals from our BBJ case samples (n = 7,459). We observed that the onset age decreased as the PRS increased, and individuals with the top 1% PRS were estimated to be approximately 4 years younger in AF onset compared to the remaining individuals (Fig. 5a, 5b, and Supplementary Fig. 12a). Next, we examined whether AF-PRS could explain the phenotypic variability of stroke in individuals without a diagnosis of AF. We performed logistic regression analysis in 140,446 control samples in our dataset, and found significant associations of the PRS with increased risks of cerebral infarction (OR [95% CI] = 1.055 [1.032 – 1.078], $P = 1.2 \times 10^{-6}$), cardioembolic stroke (OR [95% CI] = 1.391 [1.155 – 1.676], $P = 5.2 \times 10^{-4}$), atherothrombotic stroke (OR [95% CI] = 1.054 [1.001 – 1.111], $P = 0.045$), and cerebral hemorrhage (OR = 1.068, 95% CI = 1.008 – 1.133, $P = 0.027$; Fig. 5c). Importantly, we observed the largest impact of the PRS on cardioembolic stroke among those with other stroke phenotypes, indicating that AF-PRS may reveal clinically undetectable AF (i.e., subclinical AF) or AF-related conditions, such as prothrombotic or hypercoagulable state, in individuals without AF.

To further explore the clinical utility of AF-PRS, we assessed the impact of PRS on mortality using long-term follow-up data in the BBJ. The Kaplan-Meier estimates of cumulative mortality rate were increased in individuals with a high PRS, especially in cardiovascular- and stroke-related mortality (Fig. 5d and Supplementary Fig. 12b). Cox regression analysis demonstrated

that PRS was significantly associated with an increased risk of cardiovascular death (hazard ratio [HR] per 1 standard deviation [s.d.] of PRS = 1.06, 95% CI = 1.02–1.11, $P = 5.0 \times 10^{-3}$) and stroke death (HR [95% CI] = 1.14 [1.04–1.24], $P = 4.0 \times 10^{-3}$; Fig. 5e). In contrast to evidence from clinical studies, the association between PRS and heart failure death did not reach statistical significance in the present study (HR [95% CI] = 1.05 [0.95–1.16], $P = 0.39$), because the statistical power might be hampered by a small fraction of heart failure deaths with relatively wide dispersion in our cohort.

Genetic liability of AF to various cardiovascular diseases

AF is frequently associated with various cardiovascular diseases, such as valvular heart disease, heart failure, and stroke. The causal relationship between AF and cardiovascular diseases is clinically evident, but genetic causality is not comprehensively understood. Therefore, we estimated the directional effect of AF on a wide range of cardiovascular diseases using two-sample Mendelian randomization (MR), where the exposure was AF and all the distinct AF-associated variants from the trans-ancestry meta-analysis were used as instrumental variables. Consistent with clinical evidence, we observed significant causal effects on heart failure, cardiomyopathy, ischemic heart disease, valvular heart disease, and stroke (Fig. 6 and Supplementary Table 15). Notably, we found that AF was a significant predictor for sinus node dysfunction and atrioventricular block (OR [95% CI] = 1.480 [1.255–1.745], $P = 3.0 \times 10^{-6}$; OR [95% CI] = 1.253 [1.167–1.346], $P = 6.1 \times 10^{-10}$, respectively), indicating that genetic risk for AF contributes to the potential disturbance in the cardiac conduction system.

In contrast, AF is also known as the phenotype accumulated by multiple atherosclerotic- and metabolic-related traits. Therefore, we investigated whether genetic risks for such traits could affect the development of AF using two-sample MR, where the exposures represented quantitative traits, and selected the distinct variants associated with each trait as instrumental variables. As expected[48], height and BMI were significant predictors for AF (OR [95% CI] = 1.381 [1.166–1.635], $P = 1.8 \times 10^{-4}$; OR [95% CI] = 1.160 [1.099–1.225], $P = 7.6 \times 10^{-8}$, respectively). Furthermore,

blood pressure showed a significant causal effect on AF (OR [95% CI] = 1.383 [1.292–1.480], P = 9.4×10^{-21} for systolic blood pressure; OR (95% CI) = 1.393 (1.298–1.494), $P = 2.9 \times 10^{-20}$ for diastolic blood pressure; OR (95% CI) = 1.220 (1.131–1.317), $P = 3.0 \times 10^{-7}$ for pulse pressure; Supplementary Fig. 13 and Supplementary Table 15).

Discussion

We conducted a large-scale AF-GWAS with approximately 10,000 AF cases in the Japanese population and identified 31 genome-wide significant loci associated with AF. This includes five novel loci, where disease-relevant rare and highly East Asian-specific variants were found in the *SYNE1* and *FGF13* loci, suggesting the involvement of the functional alteration in the nuclear envelope and the ion channel as a mechanism underlying AF.

Furthermore, we performed the largest trans-ancestry meta-analysis for AF to date, where 150 genome-wide significant loci were identified, resulting in the discovery of 35 novel loci. We further analyzed AF-associated genes and transcription factors using gene expression and epigenetic datasets. Transcriptome-wide analysis linked AF-associated loci to target genes, including *IL6R*, which indicated that inflammation plays a key role in the pathogenesis of AF. Additionally, our approach based on the ChIP-seq dataset clearly implicated ERG as a candidate transcription factor involved in the regulation of AF-associated gene expression. ERG is highly expressed in endothelial cells and regulates a wide range of targets and pathways involved in vascular homeostasis and angiogenesis[49]. Endothelial dysfunction is observed in patients with AF[50,51], but the association between AF and ERG had not been previously examined. This may be attributable to mutational constraints on the ERG gene according to the relatively low LOEUF score of 0.33 in gnomAD and embryonic lethality in ERG knockout mice with severe defects in vascular development[52,53]. Although the contribution of ERG signaling to AF development needs further confirmation by functional experiments, the results from this study suggest that ERG may play a critical role in the pathophysiology of AF through vascular development at early developmental stages.

We found shared allelic effects of lead variants in AF-associated loci and genetic correlations between Japanese and European populations. We then derived AF-PRS from the trans-ancestry meta-analysis, which outperformed that derived from a population-specific GWAS. In addition to the predictive ability for AF itself, AF-PRS segregated individuals with AF-related phenotypes, such as early onset of AF and cardioembolic stroke, as well as those with increased risks

of long-term cardiovascular and stroke mortalities. These results indicated that the cumulative genetic risk for AF may be an indicator for early therapeutic intervention, including anticoagulation in at-risk individuals as a primary prevention of stroke. Taken together, our findings suggest the clinical utility of AF-PRS for individual risk stratification, leading to the realization of future precision medicine.

In conclusion, our large-scale Japanese and trans-ancestry genetic analyses identified 35 novel loci and provided insights into the distinct and shared genetic architecture of AF between Japanese and Europeans. Integrative analysis of transcriptome and epigenome data revealed the candidate genes and transcription factors involved in the mechanism of disease development. Furthermore, analyses of disease prediction and long-term survival demonstrated the clinical utility of the AF-PRS. These data highlight the importance of AF genetics in clinical settings and provide useful evidence for the implementation of genomic medicine.

Methods

Study populations

All subjects were Japanese who were registered in the BioBank Japan (BBJ) project (https://biobankjp.org/). The BBJ is a hospital-based national biobank project that collects DNA and serum samples and clinical information from 12 cooperative medical institutes throughout Japan (Osaka Medical Center for Cancer and Cardiovascular Diseases, Cancer Institute Hospital of Japanese Foundation for Cancer Research, Juntendo University, Tokyo Metropolitan Geriatric Hospital, Nippon Medical School, Nihon University School of Medicine, Iwate Medical University, Tokushukai Hospitals, Shiga University of Medical Science, Fukujuji Hospital, National Hospital Organization Osaka National Hospital, and Iizuka Hospital). Approximately 200,000 patients with any of the 47 target diseases were enrolled between 2003 and 2007. All samples were at least 18 years old. Informed consent was obtained from all participants, and our study was approved by the relevant ethical committees at each facility.

For the quality control of GWAS, we excluded samples with a call rate < 0.98 and related individuals with PI_HAT > 0.2, which is an index of relatedness based on pairwise identity-by-descent implemented in PLINK 2.0 (20 Aug 2018 ver.). We then excluded samples with a heterozygosity rate > +4 s.d. To identify population stratification, we performed principal component analysis using PLINK 2.0. and excluded outliers from the Japanese cluster. For the case samples in GWAS, we selected individuals with atrial fibrillation or atrial flutter diagnosed by a physician based on general medical practices or documented on a 12-lead electrocardiogram. The demographic features of the case–control cohort are shown in Supplementary Table 16.

Genotyping, imputation, and quality control

GWAS subjects were genotyped using the Illumina Human OmniExpress Genotyping BeadChip or a combination of Illumina HumanOmniExpress and HumanExome BeadChips (Illumina, San Diego, CA). For genotype quality control, we excluded variants with (i) SNP call rate

< 0.99, (ii) MAF < 0.01, and (iii) Hardy–Weinberg equilibrium P value ≤ 1.0×10^{-6}. We pre-phased the genotypes using EAGLE and imputed dosages with the 1,000 Genome Project Phase 3 (1KG Phase 3; May 2012)[54] reference panel with 1,037 Japanese in-house reference panel from BBJ using minimac3. For the X chromosome, pre-phasing was performed in both males and females, and imputation was performed separately for males and females. Dosages of variants in X chromosomes for males were assigned between 0 and 2.

GWAS

In the Japanese GWAS, association was performed by logistic regression analysis assuming an additive model with adjustment for age, age^2, sex, and top 20 principal components using PLINK 2.0. We selected variants with minimac3 imputation quality score of > 0.3 and MAF ≥ 0.001. For the X chromosome, we conducted association analyses in males and females separately, and integrated the results using an inverse-variance weighted fixed-effects model implemented in METASOFT (v2.0.1). Heterogeneity between studies was calculated using the Cochran's Q test. We filtered variants with strong heterogeneity ($P_{het} < 1.0 \times 10^{-4}$). The genome-wide significance threshold was defined at $P < 5.0 \times 10^{-8}$ for variants with MAF ≥ 1% and $P < 5.71 \times 10^{-9}$ for those with MAF < 1% (0.05/8,753,038 variants). Although the genomic inflation factor (λ_{GC}) was 1.12, LD-score regression indicated that the inflation was primarily due to polygenic effects (LD-score regression intercept = 1.02; Supplementary Fig. 14a). Adjacent genome-wide significant SNPs were grouped into one locus if they were within 1 Mb from each other. We defined a locus as the candidate region that was located within 500 Mb upstream and downstream of the genome-wide significant variants. When a locus was not in LD ($r^2 < 0.10$) with reported loci, we defined it as a novel locus. In sex-stratified association analysis of 81,050 males (6,825 cases and 74,225 controls) and 69,222 females (3,001 cases and 66,221 controls), we also performed logistic regression analysis in males and females separately using the same pipeline as described above. To investigate the difference in genetic susceptibility to AF between males and females, we compared effect sizes of the variants identified in the sex-stratified GWAS. Genotyped or imputed variants were annotated

using ANNOVAR (7 July 2017 build) (http://annovar.openbioinformatics.org).

Stepwise conditional analysis

To identify independent association signals in the loci, we conducted a stepwise conditional analysis for genome-wide significant loci defined as described in the GWAS. First, we performed logistic regression conditioning on the lead variants of each locus. We set a locus-wide significance at $P < 1.0 \times 10^{-5}$ and repeated this procedure until none of the variants reached locus-wide significance for each locus.

LD score regression and heritability

To estimate confounding biases such as cryptic relatedness and population stratification, we performed an LD score regression (Version 1.0.0). We selected SNPs with MAF ≥ 0.01, and variants within the major histocompatibility complex region were excluded. For the regression, we used the East Asian LD scores provided by the authors (https://github.com/bulik/ldsc/).

Trans-ancestry meta-analysis

Summary results from two European AF GWASs (EUR and FIN) were obtained from a previously published website (http://csg.sph.umich.edu/willer/public/afib2018)[7] and from the FinnGen research project website (https://www.finngen.fi/en), respectively. EUR is a meta-analysis of six contributing AF data sources (The Nord-Trøndelag Health Study, deCODE, the Michigan Genomics Initiative, DiscovEHR, UK Biobank, and the AFGen Consortium). FIN is composed of 102,739 Finnish participants, and the phenotypes were derived from ICD codes in Finnish national hospital registries and cause-of-death registry as a part of the FinnGen project. To inspect the biases of these GWAS, we calculated the LD-score regression intercept for each study and confirmed that these two studies were well calibrated (LD score regression intercept for EUR = 1.052 (s.e.m. = 0.012) and FIN = 1.033 (s.e.m. = 0.010); Supplementary Fig. 14b, c). We also calculated the genetic correlation and found a significant genetic correlation between EUR and FIN (r_g = 0.918, s.e.m. =

464 0.035, $P = 3.9 \times 10^{-155}$).

465 To account for the ancestral heterogeneity among the three studies, we applied the
466 MANTRA algorithm in the trans-ancestry meta-analysis[55], which allows for heterogeneity between
467 diverse ancestry groups and improves performance compared to fixed-effects meta-analysis and
468 random-effects meta-analysis. Variants with MAF ≥ 1% in both the Japanese and European
469 populations were selected for association. We considered SNPs with $\log_{10}$ BF > 6 to be genome-wide
470 significant.
471
472 **Trans-ancestry genetic correlation**
473 To estimate the trans-ancestry genetic correlation, we applied Popcorn software (v.1.0).
474 We used East Asia or European samples of 1KG for the LD reference and excluded the variants
475 within the major histocompatibility complex region according to the instructions of the software. To
476 estimate the liability-scale genetic correlation, we assumed that the AF prevalence was 0.58% in the
477 Japanese population, 1.38% in the United Kingdom population, and 0.97% in the Finnish population
478 (https://vizhub.healthdata.org/gbd-compare/).
479
480 **Credible set analysis**
481 To construct sets of variants that likely include causal variants, we constructed a 99%
482 credible set in each AF-associated locus using BF obtained from MANTRA results in the trans-
483 ancestry meta-analysis. For each locus, we calculated the posterior probability (PP) for the j^{th} SNP
484 using the following formula: $PP_j = \frac{BF_j}{\sum_k BF_k}$, where BF_j denotes the BF for the j^{th} SNP and BF_k
485 denotes all of the variants included in the locus. We then constructed the 99% credible set by adding
486 the variants in order of decreasing PP until the cumulative PP reached 0.99. To assess whether the
487 99% credible set derived from the trans-ancestry meta-analysis narrowed down the causal variants,
488 we performed two additional meta-analyses (EUR + BBJ and EUR + FIN) using the MANTRA
489 algorithm and constructed 99% credible sets for 150 AF-associated loci identified by the trans-
490 ancestry meta-analysis. We then compared the number of variants in the 99% credible set for each

491 locus. Among them, the loci where at least one variant surpassed genome-wide significance ($\log_{10}$
492 BF > 6) in all meta-analyses were included in the test. The differences in the sizes of the credible
493 sets were tested using the paired Wilcoxon rank-sum test.
494
495 **Pleiotropy**
496 We searched for AF-associated variants that overlapped with other diseases or traits using
497 the NHGRI-EBI GWAS catalog on February 8, 2019. As reported previously[34], we selected the AF-
498 associated variants in the trans-ancestry meta-analysis according to the following criteria: (i) variants
499 with $r^2 > 0.5$ in both East Asian and European samples of 1KG if the publication in GWAS catalog
500 included East Asians, (ii) variants with $r^2 > 0.5$ in European samples of 1KG if the original report
501 did not study East Asian populations; and (iii) the physical distance from AF-associated lead variants
502 in the trans-ancestry meta-analysis to variants reported in the GWAS catalog was within 25 kb.
503 For the association analysis between AF-associated loci and QTL for clinical parameters,
504 we used the summary statistics previously reported in the BBJ dataset[34-36]. We selected AF-
505 associated variants ($r^2 > 0.6$ in East Asian samples of 1KG and clinical parameters of nine distinct
506 categories: anthropometric (n=2), metabolic (n=6), serum protein (n=2), kidney-related (n=4),
507 electrolyte (n=3), liver-related (n=5), other biochemical (n=5), hematological (n=5), and blood
508 pressure (n=4). In QTL analysis, clinical measurements were corrected by taking into account the
509 effect of medications as follows: for individuals taking cholesterol-lowering drugs, total cholesterol
510 and low-density-lipoprotein cholesterol levels were corrected by dividing by 0.8 and 0.7,
511 respectively[35]; for individuals taking antihypertensive drugs, 15 mmHg was added to the systolic
512 blood pressure and 10 mmHg to the diastolic blood pressure readings[35]. Furthermore, linear
513 regression models for these quantitative traits were adjusted for sex, age, age^2, the top 10 principal
514 components, and disease status. Smoking status was also introduced into the model for white blood
515 cell count and C-reactive protein levels. We chose QTL showing a genome-wide significant
516 association ($P < 5.0 \times 10^{-8}$).
517

Tissue and gene set enrichment analysis

For tissue and gene set enrichment analysis, we used data-driven expression-prioritized integration for complex traits (DEPICT) software (https://data.broadinstitute.org/mpg/depict)[39]. To reduce the redundancy of the pathways, we applied affinity propagation clustering implemented in the apcluster package v.1.4.8 in R [56,57]. Using this algorithm, we obtained distinct clusters that contained 'exemplar' pathway and member pathways. We extracted 48 exemplar tissues and 1,157 gene sets for enrichment analysis and selected independent AF-associated variants with $\log_{10}$ BF > 5 as applied in a previous study[58]. The significance level was set at $P = 1 \times 10^{-3}$ (=0.05/48) for tissue enrichment analysis and $P = 4.3 \times 10^{-5}$ (=0.05/1,157) for gene set enrichment analysis.

Transcriptome-wide association study

We performed a transcriptome-wide association study to assess gene-AF associations using MetaXcan v0.3.512[59], which estimates the association between predicted gene expression levels and a phenotype of interest using summary statistics and gene expression prediction models. We used precomputed prediction models with LD reference data in GTEx v7 for atrial appendage and left ventricular tissues and the summary statistic of the trans-ancestry AF-GWAS as input. Bonferroni significance level was set at $P = 8.4 \times 10^{-6}$ (=0.05/5,988) for atrial appendage and $P = 9.3 \times 10^{-6}$ (=0.05/5,378) for left ventricle to account for the number of genes tested in each tissue.

Enrichment analysis of transcription factor

To assess the enrichment of transcription factor in AF-associated loci, we defined AF-associated loci as regions within 500 Mb upstream and 500 Mb downstream of the AF-associated lead variants or proxies with $r^2 > 0.8$ in European samples of 1KG. We then searched for overlaps of peak-call data archived in the ChIP-Atlas dataset with AF-associated loci and control regions selected from all the genome regions by permutation test. *P*-values were calculated with the two-tailed Fisher's exact probability test (the null hypothesis is that the two regions overlap with the ChIP-Atlas peak-call data in the same proportion). The epigenetic landscapes around cardiac ion

channel-related genes were visualized using ChIP-Atlas Peak Browser and Integrative Genomics Viewer[60].

PRS derivation and performance

We derived PRS using the pruning and thresholding methods. PRSs were created over a range of *P* value threshold (0.5, 5.0×10^{-2}, 5.0×10^{-4}, 5.0×10^{-6}, and 5.0×10^{-8}) and r^2 threshold (0.2, 0.5, and 0.8). For the survival analysis, we divided our dataset into three groups: (i) a discovery group to derive and validate PRS (6,890 cases and 49,451 controls), (ii) a test group to assess PRS performance (2,953 cases and 21,194 controls), and (iii) a group for the survival analysis (70,645 controls) (Supplementary Fig. 15). To secure independence between the PRS derivation and validation, we used a ten-fold cross-validation approach. First, we randomly split a discovery group into ten subgroups and used nine of these subgroups for PRS derivation and the remaining one for PRS validation. To determine the weight of PRS, we performed GWAS in combination with two European GWAS: (i) Japanese subgroup GWAS, (ii) two European GWAS (EUR or FIN), (iii) Japanese subgroup GWAS + EUR (fixed- or random-effect model), (iv) Japanese subgroup GWAS + FIN (fixed- or random-effect model), (v) EUR + FIN (fixed- or random-effect model, and (vi) Japanese subgroup GWAS + EUR + FIN (fixed- or random-effect model). Meta-analyses were performed using the METASOFT software. In the meta-analysis with the Japanese subgroup GWAS, LD was calculated separately based on East Asian or European samples of 1KG. We then calculated the PRSs for the withheld validation cohort using PLINK 2.0 and validated its performance. Changing the withheld validation cohort, we repeated these procedures ten times and obtained the performance measurements for each combination. Notably, the validation cohort was excluded from the derivation of GWAS. The performance of the PRS was measured as (1) Nagelkerke's pseudo R^2 obtained by modeling age, sex, and normalized PRS, and (2) the area under the curve of the receiver operator curve in the same model as Nagelkerke's pseudo R^2. The best model/parameter set for each combination model (BBJ, EUR, FIN, BBJ + EUR, BBJ + FIN, EUR + FIN, BBJ + EUR + FIN) was determined by averaging Nagelkerke's pseudo R^2 (Supplementary Table 13). Using the best models

and parameters determined in the derivation and validation cohorts, we calculated the PRSs and assessed their performance for the independent test cohort. We applied bootstrapping to assess the distribution of performance of the PRS; randomly extracted the individuals in the test cohort with replacement and calculated the performance for the samples. By repeating this procedure 5.0×10^4 times, we estimated the distribution of the performance measures.

Association of AF-PRS with age of AF onset and stroke subtypes

To assess the association between AF-PRS and age at AF onset, we extracted AF case samples with the data (n = 7,458, median age of AF onset was 63 years of age (IQR 56 – 71)), and estimated the difference in the effect of AF onset age between individuals with top PRS to those with the remaining PRS using a linear regression model adjusted for sex and top 10 principal components. For the association analysis between AF-PRS and stroke subtype, we calculated ORs and associated *P* values using a logistic regression model adjusted for age, sex, and the top 10 principal components. Among the control samples (without AF diagnosis) in our dataset (n = 140,447), we found 14,120 stroke phenotypes: 9,177 for cerebral infarction, 111 for cardioembolic stroke, 1,465 for atherothrombotic infarction, 1,266 for lacunar infarction, 1,143 for cerebral hemorrhage, and 967 for subarachnoid hemorrhage.

Survival analysis

For survival analysis, the Cox proportional hazards model was used to assess the association between AF-PRS and long-term mortality. We obtained survival follow-up data with the ICD-10 code for 132,737 individuals from the BBJ dataset[61,62]. The causes of death were classified according to ICD-10 codes: (i) cardiovascular death (I00–I99), (ii) heart failure death (I50), (iii) ischemic heart disease death (I20–I25), and (iv) stroke death (I60, I61, I63, and I64). The median follow-up period was 8.4 years (IQR 6.8 – 9.9). The Cox proportional hazards model was adjusted for sex, age, top 10 principal components, and disease status. Analyses were performed with the R package survival v.2.44, and survival curves were estimated using the R package survminer v.0.4.6,

599 with modifications.

600

601 **Mendelian randomization**

602 We performed two-sample MR to test for a causal relationship between AF and relevant
603 diseases or traits. Although the statistical power to detect the associations with AF decreased, we
604 selected AF-associated variants from the trans-ancestry meta-analysis in combination with BBJ and
605 FIN to avoid sample overlap, and extracted summary statistics from a large database of genetic
606 associations in the UK biobank (http://www.nealelab.is/uk-biobank/). To identify independent
607 variants for exposure, genome-wide significant variants ($P < 5 \times 10^{-8}$) were pruned ($r^2 < 0.05$; LD
608 window of 10,000 kb; using European samples of 1KG for LD reference) as applied previously[63].
609 First, we assessed whether AF is causally associated with cardiovascular diseases using AF-
610 associated variants as instrument variables and variants associated with cardiovascular diseases as
611 outcome variables. Next, we assessed the causal effects of quantitative traits related to
612 anthropometry, metabolites, serum protein, kidney function, liver function, hematocyte count, and
613 blood pressure on AF, where variants associated with quantitative traits were used as instrument
614 variables and AF-associated variants as outcome variables. We used the inverse-variance-weighting
615 (IVW) method. To exclude horizontal pleiotropic outliers, we performed MR-PRESSO for
616 instrument variables[64]. We also calculated Cochran's Q statistics for heterogeneity between the
617 causal effects of variants using IVW and the MR-Egger intercept for directional pleiotropy. All
618 analyses were performed using R v4.0.3, with R packages TwoSampleMR and MRPRESSO.

Data availability

Summary statistics of Japanese GWAS and polygenic risk scores derived in this study will be publicly available at the National Bioscience Database Center (https://humandbs.biosciencedbc.jp/).

Code Availability

Code used in this study is available on reasonable request to the corresponding authors.

Acknowledgements

We thank the staff of BBJ for their assistance in collecting samples and clinical information. We also thank A. P. Morris for providing us with MANTRA software and valuable advice.
This research was funded by the GRIFIN project of the Japan Agency for Medical Research and Development (AMED) (nos. JP20km0405209 and JP20ek0109487 to K. Miyazawa, K. I, S. K., S. N., H. Akazawa, and I. K.), MSD Life Science Foundation (to K. Miyazawa), the Japan Society for the Promotion of Science (to K. Miyazawa, K. I. and H. I.), and Sakakibara Memorial Research Grant from The Japan Research Promotion Society for Cardiovascular Diseases (to H. M.). BBJ was supported by the Tailor-Made Medical Treatment Program of the Ministry of Education, Culture, Sports, Science, and Technology (MEXT) and AMED under grant numbers JP17km0305002 and JP17km0305001.

Author Contributions

K. Miyazawa, K. I., C. T., Y. K., H. Akazawa, and I. K. conceived and designed the study. C. T., K. Matsuda, Y. M., M. K., and Y. K. collected and managed the BBJ sample. Y. M., A. T., M. K., and Y. K. performed the genotyping. K. Miyazawa, K. I., Z. Z., H. M., C. T., S. O., and Y. K. performed the statistical analyses. K. Miyazawa, K. I., Z. Z., H. M., H. I., M. A., K. O., Y. O., C. T., and S. O.

646 contributed to data processing, analysis, and interpretation. K. I., H. Aburatani, H. Akazawa, Y.K.
647 and I. K. supervised the study. K. Miyazawa, K. I., and M. H. wrote the manuscript, and many
648 authors have provided valuable edits.
649
650 **Competing interests**
651
652 The authors declare no competing interests associated with this manuscript.
653
654
655
656
657

References

1. Benjamin, E.J. *et al.* Heart Disease and Stroke Statistics-2019 Update: A Report From the American Heart Association. *Circulation* **139**, e56-e528 (2019).
2. Staerk, L., Sherer, J.A., Ko, D., Benjamin, E.J. & Helm, R.H. Atrial Fibrillation: Epidemiology, Pathophysiology, and Clinical Outcomes. *Circ. Res.* **120**, 1501-1517 (2017).
3. Healey, J.S. *et al.* Occurrence of death and stroke in patients in 47 countries 1 year after presenting with atrial fibrillation: a cohort study. *Lancet* **388**, 1161-1169 (2016).
4. Kim, M.H., Johnston, S.S., Chu, B.C., Dalal, M.R. & Schulman, K.L. Estimation of total incremental health care costs in patients with atrial fibrillation in the United States. *Circ. Cardiovasc. Qual. Outcomes.* **4**, 313-320 (2011).
5. Christophersen, I.E. *et al.* Large-scale analyses of common and rare variants identify 12 new loci associated with atrial fibrillation. *Nat. Genet.* **49**, 946-952 (2017).
6. Low, S.K. *et al.* Identification of six new genetic loci associated with atrial fibrillation in the Japanese population. *Nat. Genet.* **49**, 953-958 (2017).
7. Nielsen, J.B. *et al.* Biobank-driven genomic discovery yields new insight into atrial fibrillation biology. *Nat. Genet.* **50**, 1234-1239 (2018).
8. Roselli, C. *et al.* Multi-ethnic genome-wide association study for atrial fibrillation. *Nat. Genet.* **50**, 1225-1233 (2018).
9. Karczewski, K.J. *et al.* The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature* **581**, 434-443 (2020).
10. Jaganathan, K. *et al.* Predicting Splicing from Primary Sequence with Deep Learning. *Cell* **176**, 535-548.e24 (2019).
11. Bonne, G. *et al.* Mutations in the gene encoding lamin A/C cause autosomal dominant Emery-Dreifuss muscular dystrophy. *Nat. Genet.* **21**, 285-288 (1999).
12. Capell, B.C. & Collins, F.S. Human laminopathies: nuclei gone genetically awry. *Nat. Rev. Genet.* **7**, 940-952 (2006).
13. Zhou, C. *et al.* Novel nesprin-1 mutations associated with dilated cardiomyopathy cause

685 nuclear envelope disruption and defects in myogenesis. *Hum. Mol. Genet.* **26**, 2258-2276
686 (2017).
687 14. Wang, C. *et al.* Fibroblast growth factor homologous factor 13 regulates Na+ channels and
688 conduction velocity in murine hearts. *Circ. Res.* **109**, 775-782 (2011).
689 15. Lip, G.Y. *et al.* Sex-related differences in presentation, treatment, and outcome of patients
690 with atrial fibrillation in Europe: a report from the Euro Observational Research Programme
691 Pilot survey on Atrial Fibrillation. *Europace* **17**, 24-31 (2015).
692 16. Schnabel, R.B. *et al.* Gender differences in clinical presentation and 1-year outcomes in atrial
693 fibrillation. *Heart* **103**, 1024-1030 (2017).
694 17. Ball, J., Carrington, M.J., Wood, K.A., Stewart, S. & SAFETY Investigators. Women versus
695 men with chronic atrial fibrillation: insights from the Standard versus Atrial Fibrillation
696 spEcific managemenT studY (SAFETY). *PLoS One* **8**, e65795 (2013).
697 18. Benjamin, E.J. *et al.* Impact of atrial fibrillation on the risk of death: the Framingham Heart
698 Study. *Circulation* **98**, 946-952 (1998).
699 19. Scheuermeyer, F.X. *et al.* There Are Sex Differences in the Demographics and Risk Profiles
700 of Emergency Department (ED) Patients With Atrial Fibrillation and Flutter, but no Apparent
701 Differences in ED Management or Outcomes. *Acad. Emerg. Med.* **22**, 1067-1075 (2015).
702 20. Conforto, T.L., Zhang, Y., Sherman, J. & Waxman, D.J. Impact of CUX2 on the female mouse
703 liver transcriptome: activation of female-biased genes and repression of male-biased genes.
704 *Mol. Cell. Biol.* **32**, 4611-4627 (2012).
705 21. Ling, G., Sugathan, A., Mazor, T., Fraenkel, E. & Waxman, D.J. Unbiased, genome-wide in
706 vivo mapping of transcriptional regulatory elements reveals sex differences in chromatin
707 structure associated with sex-specific liver gene expression. *Mol. Cell. Biol.* **30**, 5531-5544
708 (2010).
709 22. Lin, H. *et al.* Methylome-wide Association Study of Atrial Fibrillation in Framingham Heart
710 Study. *Sci. Rep.* **7**, 40377 (2017).
711 23. Battle, A. *et al.* Genetic effects on gene expression across human tissues. *Nature* **550**, 204-

213 (2017).

24. Rattka, M., Westphal, S., Gahr, B.M., Just, S. & Rottbauer, W. Spen deficiency interferes with Connexin 43 expression and leads to heart failure in zebrafish. *J. Mol. Cell. Cardiol.* **155**, 25-35 (2021).

25. Michela, P., Velia, V., Aldo, P. & Ada, P. Role of connexin 43 in cardiovascular diseases. *Eur. J. Pharmacol.* **768**, 71-76 (2015).

26. den Hoed, M. *et al.* Identification of heart rate-associated loci and their effects on cardiac conduction and rhythm disorders. *Nat. Genet.* **45**, 621-631 (2013).

27. Nolte, I.M. *et al.* Genetic loci associated with heart rate variability and their effects on cardiac disease risk. *Nat. Commun.* **8**, 15805 (2017).

28. Chen, P.S., Chen, L.S., Fishbein, M.C., Lin, S.F. & Nattel, S. Role of the autonomic nervous system in atrial fibrillation: pathophysiology and therapy. *Circ. Res.* **114**, 1500-1515 (2014).

29. Gudbjartsson, D.F. *et al.* A sequence variant in ZFHX3 on 16q22 associates with atrial fibrillation and ischemic stroke. *Nat. Genet.* **41**, 876-878 (2009).

30. Zaw, K.T.T. *et al.* Association of ZFHX3 gene variation with atrial fibrillation, cerebral infarction, and lung thromboembolism: An autopsy study. *J. Cardiol.* **70**, 180-184 (2017).

31. Amendt, B.A., Sutherland, L.B., Semina, E.V. & Russo, A.F. The molecular basis of Rieger syndrome. Analysis of Pitx2 homeodomain protein activities. *J. Biol. Chem.* **273**, 20066-20072 (1998).

32. Oki, S. *et al.* ChIP-Atlas: a data-mining suite powered by full integration of public ChIP-seq data. *EMBO Rep.* **19**, e46255 (2018).

33. Ramji, D.P. & Foka, P. CCAAT/enhancer-binding proteins: structure, function and regulation. *Biochem. J* **365**, 561-575 (2002).

34. Akiyama, M. *et al.* Genome-wide association study identifies 112 new loci for body mass index in the Japanese population. *Nat. Genet.* **49**, 1458-1467 (2017).

35. Kanai, M. *et al.* Genetic analysis of quantitative traits in the Japanese population links cell types to complex human diseases. *Nat. Genet.* **50**, 390-400 (2018).

36. Akiyama, M. *et al.* Characterizing rare and low-frequency height-associated variants in the Japanese population. *Nat. Commun.* **10**, 4393 (2019).

37. Bustin, M. & Reeves, R. High-mobility-group chromosomal proteins: architectural components that facilitate chromatin function. *Prog. Nucleic Acid Res. Mol. Biol.* **54**, 35-100 (1996).

38. Chiefari, E. *et al.* A polymorphism of HMGA1 is associated with increased risk of metabolic syndrome and related components. *Sci. Rep.* **3**, 1491 (2013).

39. Pers, T.H. *et al.* Biological interpretation of genome-wide association studies using predicted gene functions. *Nat. Commun.* **6**, 5890 (2015).

40. Sharma, M. *et al.* hZimp10 is an androgen receptor co-activator and forms a complex with SUMO-1 at replication foci. *EMBO J.* **22**, 6101-6114 (2003).

41. Rodriguez-Magadán, H., Merino, E., Schnabel, D., Ramírez, L. & Lomelí, H. Spatial and temporal expression of Zimp7 and Zimp10 PIAS-like proteins in the developing mouse embryo. *Gene Expr. Patterns* **8**, 206-213 (2008).

42. Henderson, P., van Limbergen, J.E., Wilson, D.C., Satsangi, J. & Russell, R.K. Genetics of childhood-onset inflammatory bowel disease. *Inflamm. Bowel Dis.* **17**, 346-361 (2011).

43. Beliakoff, J. *et al.* The PIAS-like protein Zimp10 is essential for embryonic viability and proper vascular development. *Mol. Cell. Biol.* **28**, 282-292 (2008).

44. Hu, Y.F., Chen, Y.J., Lin, Y.J. & Chen, S.A. Inflammation and the pathogenesis of atrial fibrillation. *Nat. Rev. Cardiol.* **12**, 230-243 (2015).

45. Schnabel, R.B. *et al.* Large-scale candidate gene analysis in whites and African Americans identifies IL6R polymorphism in relation to atrial fibrillation: the National Heart, Lung, and Blood Institute's Candidate Gene Association Resource (CARe) project. *Circ. Cardiovasc. Genet.* **4**, 557-564 (2011).

46. Nagai, N. *et al.* Downregulation of ERG and FLI1 expression in endothelial cells triggers endothelial-to-mesenchymal transition. *PLoS Genet.* **14**, e1007826 (2018).

47. Kalna, V. *et al.* The Transcription Factor ERG Regulates Super-Enhancers Associated With

an Endothelial-Specific Gene Expression Program. *Circ. Res.* **124**, 1337-1349 (2019).

48. Levin, M.G. *et al.* Genetics of height and risk of atrial fibrillation: A Mendelian randomization study. *PLoS Med.* **17**, e1003288 (2020).

49. Shah, A.V., Birdsey, G.M. & Randi, A.M. Regulation of endothelial homeostasis, vascular development and angiogenesis by the transcription factor ERG. *Vascul. Pharmacol.* **86**, 3-13 (2016).

50. Guazzi, M. & Arena, R. Endothelial dysfunction and pathophysiological correlates in atrial fibrillation. *Heart* **95**, 102-106 (2009).

51. Khan, A.A., Thomas, G.N., Lip, G.Y.H. & Shantsila, A. Endothelial function in patients with atrial fibrillation. *Ann. Med.* **52**, 1-11 (2020).

52. Vijayaraj, P. *et al.* Erg is a crucial regulator of endocardial-mesenchymal transformation during cardiac valve morphogenesis. *Development* **139**, 3973-3985 (2012).

53. Birdsey, G.M. *et al.* The endothelial transcription factor ERG promotes vascular stability and growth through Wnt/β-catenin signaling. *Dev. Cell* **32**, 82-96 (2015).

54. Abecasis, G.R. *et al.* A map of human genome variation from population-scale sequencing. *Nature* **467**, 1061-1073 (2010).

55. Morris, A.P. Transethnic meta-analysis of genomewide association studies. *Genet. Epidemiol.* **35**, 809-822 (2011).

56. Frey, B.J. & Dueck, D. Clustering by passing messages between data points. *Science* **315**, 972-976 (2007).

57. Marouli, E. *et al.* Rare and low-frequency coding variants alter human adult height. *Nature* **542**, 186-190 (2017).

58. Malik, R. *et al.* Multiancestry genome-wide association study of 520,000 subjects identifies 32 loci associated with stroke and stroke subtypes. *Nat. Genet.* **50**, 524-537 (2018).

59. Barbeira, A.N. *et al.* Exploring the phenotypic consequences of tissue specific gene expression variation inferred from GWAS summary statistics. *Nat. Commun.* **9**, 1825 (2018).

60. Robinson, J.T. *et al.* Integrative genomics viewer. *Nat. Biotechnol.* **29**, 24-26 (2011).

61. Nagai, A. *et al.* Overview of the BioBank Japan Project: Study design and profile. *J. Epidemiol.* **27**, S2-S8 (2017).

62. Hirata, M. *et al.* Cross-sectional analysis of BioBank Japan clinical data: A large cohort of 200,000 patients with 47 common diseases. *J. Epidemiol.* **27**, S9-S21 (2017).

63. Shah, S. *et al.* Genome-wide association and Mendelian randomisation analysis provide insights into the pathogenesis of heart failure. *Nat. Commun.* **11**, 163 (2020).

64. Verbanck, M., Chen, C.Y., Neale, B. & Do, R. Detection of widespread horizontal pleiotropy in causal relationships inferred from Mendelian randomization between complex traits and diseases. *Nat. Genet.* **50**, 693-698 (2018).

Table 1 | Novel association loci in the Japanese genome-wide association study (GWAS). Sentinel variants in novel loci with a genome-wide significance in the Japanese GWAS (9,826 cases and 140,446 controls). CHR, chromosome; rsID, reference SNP cluster ID; POS, position (hg19); REF, reference allele; ALT, alternate allele; BBJ, BioBank Japan; AAF, alternate allele frequency; EUR, European populations; SE, standard error.

CHR	POS (hg19)	rsID	REF	ALT	AAF			Beta	SE	P	Gene	Annotation
					BBJ	gnomAD EUR (non-Finnish)	genomAD EUR (Finnish)					
6	152466619	rs202030113	T	C	0.012	0	0	0.352	0.063	1.90E-08	*SYNE1*	Intronic
12	104471663	rs2930856	C	T	0.579	0.861	0.854	0.090	0.015	5.80E-09	*HCFC2*	Intronic
16	30619745	rs1055894680	C	T	0.001	<0.001	0.001	1.215	0.158	1.58E-14	*ZNF689*	Intronic
X	23399501	rs73205368	T	C	0.284	0.045	0.015	0.089	0.012	7.50E-13	*PTCHD1*	Intronic
X	137790580	rs778479352	T	C	0.002	0	0	0.692	0.075	1.62E-20	*FGF13*	Intronic

Table 2 | Novel association loci in the trans-ancestry meta-analysis. Sentinel variants in novel loci with a genome-wide significance in the Japanese GWAS (77,690 cases and 1,167,040 controls). CHR, chromosome; POS, position (hg19); REF, reference allele; ALT, alternate allele; rsID, reference SNP cluster ID; BF, Bayes factor; AAF, alternate allele frequency; BBJ, BioBank Japan; SE, standard error; ncRNA, non-coding RNA.

CHR	POS (hg19)	REF	ALT	rsID	Gene	Annotation	log10BF	BBJ				EUR				FIN			
								AAF	Beta	SE	P	AAF	Beta	SE	P	AAF	Beta	SE	P
1	918617	G	A	rs4970418	*PERM1, HES4*	intergenic	7.647	0.076	0.062	0.028	2.9.E-02	0.167	0.044	0.010	7.5.E-06	0.175	0.102	0.025	4.1.E-05
1	16199051	T	C	rs9782984	*MIR5096*	ncRNA intronic	13.984	0.681	0.057	0.016	4.0.E-04	0.353	0.041	0.007	3.2.E-09	0.412	0.102	0.019	1.2.E-07
1	39385714	G	A	rs75414548	*RHBDL2*	intronic	6.970	0.723	-0.075	0.017	1.5.E-05	0.883	-0.055	0.014	8.4.E-05	0.835	-0.035	0.025	1.7.E-01
1	100149308	G	A	rs1933723	*PALMD*	intronic	13.485	0.676	0.057	0.016	3.9.E-04	0.847	0.062	0.010	2.0.E-10	0.788	0.078	0.023	7.2.E-04
4	71776935	A	C	rs12512502	*MOB1B*	intronic	6.439	0.060	0.068	0.032	3.5.E-02	0.077	0.068	0.015	4.4.E-06	0.059	0.106	0.040	8.2.E-03
4	83910712	T	G	rs6841049	*LIN54*	intronic	8.571	0.686	-0.005	0.016	7.5.E-01	0.648	-0.044	0.007	2.9.E-10	0.662	-0.056	0.020	5.1.E-03
5	139703286	T	C	rs17118812	*PFDN1, HBEGF*	intergenic	10.411	0.095	-0.008	0.025	7.5.E-01	0.033	0.145	0.020	1.7.E-13	0.040	0.019	0.047	6.9.E-01
6	22598259	C	T	rs7766436	*HDGFL1, LOC105374972*	intergenic	6.298	0.698	0.034	0.016	3.8.E-02	0.677	0.036	0.007	5.2.E-07	0.664	0.027	0.020	1.8.E-01
6	76164589	C	A	rs12209223	*LOC101928540*	ncRNA exonic	18.445	0.502	0.087	0.015	6.1.E-09	0.097	0.066	0.011	2.2.E-09	0.185	0.110	0.024	5.2.E-06
6	135119089	C	T	rs4896104	*LOC101928304, ALDH8A1*	intergenic	13.267	0.289	0.055	0.016	7.5.E-04	0.564	0.049	0.007	5.0.E-13	0.567	0.021	0.019	2.8.E-01
7	105612736	A	G	rs2727757	*CDHR3*	intronic	8.256	0.489	0.084	0.015	1.6.E-08	0.195	0.034	0.008	4.5.E-05	0.186	0.017	0.024	4.8.E-01
8	118863412	A	T	rs17430357	*EXT1*	intronic	87.114	0.085	0.071	0.026	6.2.E-03	0.333	0.135	0.007	3.1.E-79	0.361	0.125	0.020	1.6.E-10
9	119181794	G	A	rs17303101	*PAPPA, ASTN2*	intergenic	62.406	0.461	-0.150	0.015	2.0.E-23	0.535	-0.091	0.007	3.3.E-42	0.523	-0.072	0.019	1.2.E-04
10	32772734	C	T	rs11527634	*CCDC7*	intronic	39.187	0.305	-0.157	0.017	3.8.E-21	0.552	-0.060	0.007	1.6.E-19	0.643	-0.078	0.020	8.0.E-05
10	50485434	G	A	rs76460895	*C10orf128, C10orf71-AS1*	intergenic	7.418	0.363	0.020	0.015	2.0.E-01	0.416	0.038	0.007	2.1.E-08	0.364	0.041	0.020	3.6.E-02
10	80898969	G	T	rs1769758	*ZMIZ1*	intronic	14.173	0.124	0.102	0.022	3.7.E-06	0.800	0.060	0.009	3.2.E-12	0.713	0.030	0.021	1.5.E-01
11	3890059	C	T	rs7126870	*STIM1*	intronic	6.407	0.674	0.018	0.016	2.6.E-01	0.613	0.038	0.007	3.2.E-08	0.620	0.022	0.019	2.7.E-01
11	14036189	G	A	rs10500790	*SPON1*	intronic	31.262	0.325	-0.089	0.016	3.1.E-08	0.385	-0.066	0.007	2.9.E-22	0.340	-0.100	0.020	4.8.E-07
11	95089882	C	T	rs517938	*LOC100129203, FAM76B*	intergenic	15.745	0.682	-0.038	0.016	1.8.E-02	0.464	-0.055	0.007	1.1.E-16	0.530	-0.031	0.019	1.0.E-01
12	12886027	G	A	rs10845620	*APOLD1*	intronic	8.311	0.194	-0.003	0.019	8.6.E-01	0.179	-0.053	0.009	9.5.E-10	0.149	-0.083	0.027	1.8.E-03
12	104492003	A	G	rs2629755	*HCFC2*	intronic	6.641	0.664	0.031	0.016	5.3.E-02	0.100	0.059	0.011	4.9.E-08	0.085	0.035	0.033	2.9.E-01
12	110082115	T	C	rs1344543	*MVK, FAM222A*	intergenic	6.893	0.504	0.039	0.015	9.3.E-03	0.357	0.031	0.007	5.0.E-06	0.363	0.052	0.020	8.4.E-03
12	113196733	G	A	rs11614295	*RPH3A*	intronic	17.213	0.390	0.046	0.016	3.3.E-03	0.485	0.055	0.007	2.0.E-16	0.484	0.045	0.019	1.7.E-02
13	22111521	C	A	rs11841562	*MICU2*	intronic	27.396	0.435	0.052	0.015	5.1.E-04	0.140	0.094	0.009	8.3.E-24	0.153	0.119	0.026	4.4.E-06
13	74520186	T	A	rs1886512	*KLF12*	intronic	28.423	0.690	0.093	0.016	1.4.E-08	0.392	0.068	0.007	5.7.E-24	0.454	0.039	0.019	3.7.E-02
16	15902715	G	A	rs9284324	*MYH11*	intronic	9.014	0.132	-0.043	0.022	5.6.E-02	0.344	-0.044	0.007	5.0.E-10	0.372	-0.030	0.019	1.2.E-01
18	77156537	C	G	rs8096658	*NFATC1*	intronic	32.733	0.299	0.085	0.016	1.4.E-07	0.640	0.071	0.007	1.8.E-24	0.610	0.094	0.019	1.3.E-06
19	48142746	A	C	rs11881441	*BICRA*	intronic	13.406	0.308	-0.064	0.016	9.7.E-05	0.295	-0.044	0.007	2.2.E-09	0.246	-0.090	0.022	4.6.E-05
20	36841914	G	A	rs3746471	*KIAA1755*	exonic	22.109	0.784	0.066	0.018	3.2.E-04	0.602	0.062	0.007	6.9.E-20	0.546	0.052	0.019	6.3.E-03
22	21999229	C	G	rs5754508	*SDF2L1*	downstream	12.534	0.215	-0.062	0.019	9.1.E-04	0.379	-0.045	0.007	1.3.E-10	0.323	-0.057	0.020	4.5.E-03
22	42189407	T	G	rs139557	*MEI1*	intronic	8.711	0.718	0.025	0.017	1.4.E-01	0.364	0.041	0.007	2.7.E-09	0.336	0.044	0.020	2.6.E-02
X	23399501	T	C	rs73205368	*PTCHD1*	intronic	12.261	0.149	-0.024	0.021	2.6.E-01	0.404	-0.046	0.007	2.4.E-11	0.487	-0.075	0.019	6.6.E-05
X	137418967	C	A	rs1891095	*ZIC3, LINC00889*	intergenic	17.173	0.514	0.053	0.015	3.9.E-04	0.651	0.056	0.007	1.6.E-15	0.605	0.049	0.019	1.1.E-02

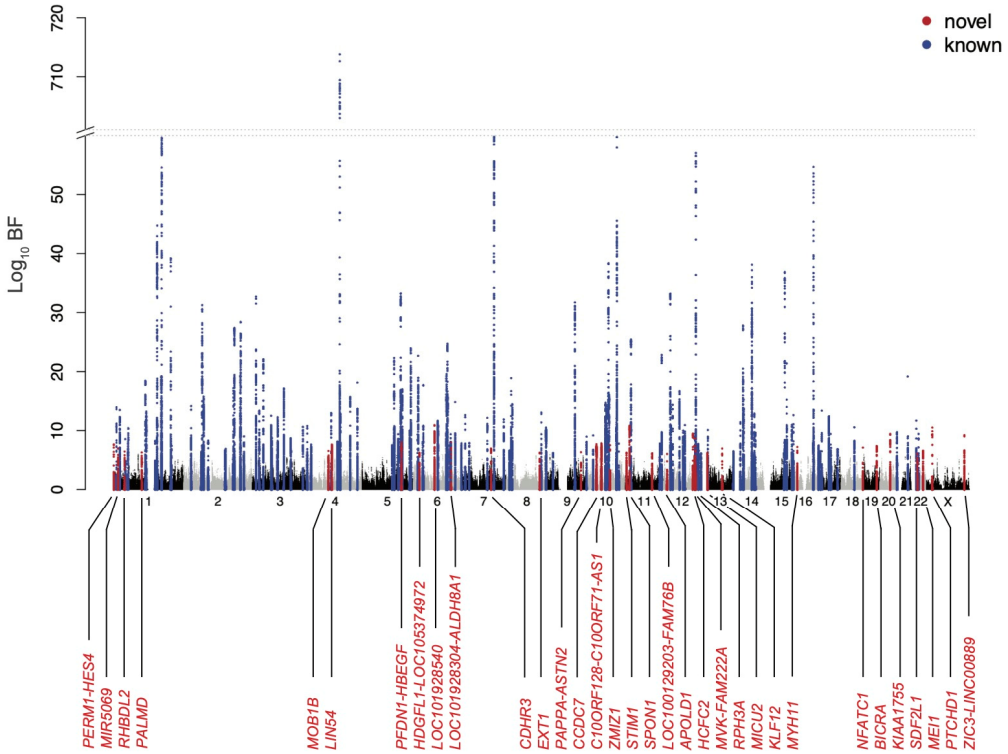


Fig. 1 | Manhattan plot for the trans-ancestry meta-analysis. The results of the trans-ancestry meta-analysis (77,690 AF cases and 1,167,040 controls) are shown. The $\log_{10}$ BFs on the *y* axis are shown against the genomic positions (hg19) on the *x* axis. Association signals that reached a genome-wide significance level ($\log_{10}$ BF > 6) are shown in blue if previously reported loci and in red if novel loci. AF, atrial fibrillation; BF, Bayes factor.

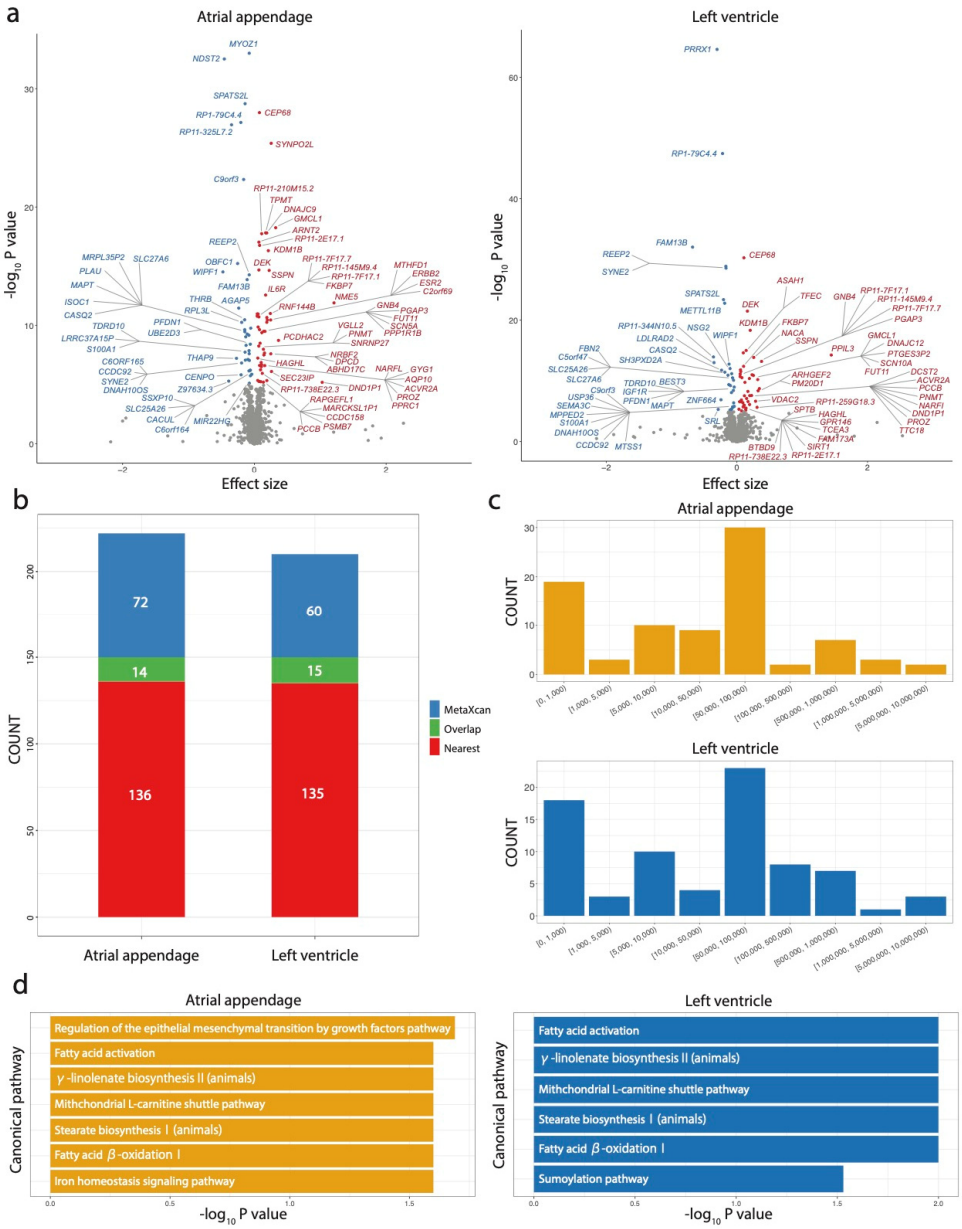


Fig. 2 | Transcriptome-wide association analysis. a, Volcano plot showing individual genes with the effect size and the -log$_{10}$ *P* value for TWAS based on atrial appendage and left ventricular tissues from GTEx project. Significant genes adjusted Bonferroni correction for all tested genes in each tissue (*P* value < 0.05/5,988 for atrial appendage and < 0.05/5,378 for left ventricle) are highlighted with red (positive effect of predicted gene expression on AF) and blue (negative effect of predicted gene expression on AF). **b**, Number of genes located close to the lead variants (red), identified by TWAS, (blue) and overlapped between them (green) in atrial appendage and left ventricular tissues. **c,**

828 Distribution of genes identified by TWAS based on physical distance to the lead variants in atrial
829 appendage and left ventricular tissues. **d,** Canonical pathways enriched in gene sets identified by
830 transcriptome-wide association study in atrial appendage (yellow) and left ventricular tissues (blue).
831 TWAS, transcriptome-wide association study; GTEx, Genotype-Tissue Expression; AF, atrial
832 fibrillation.

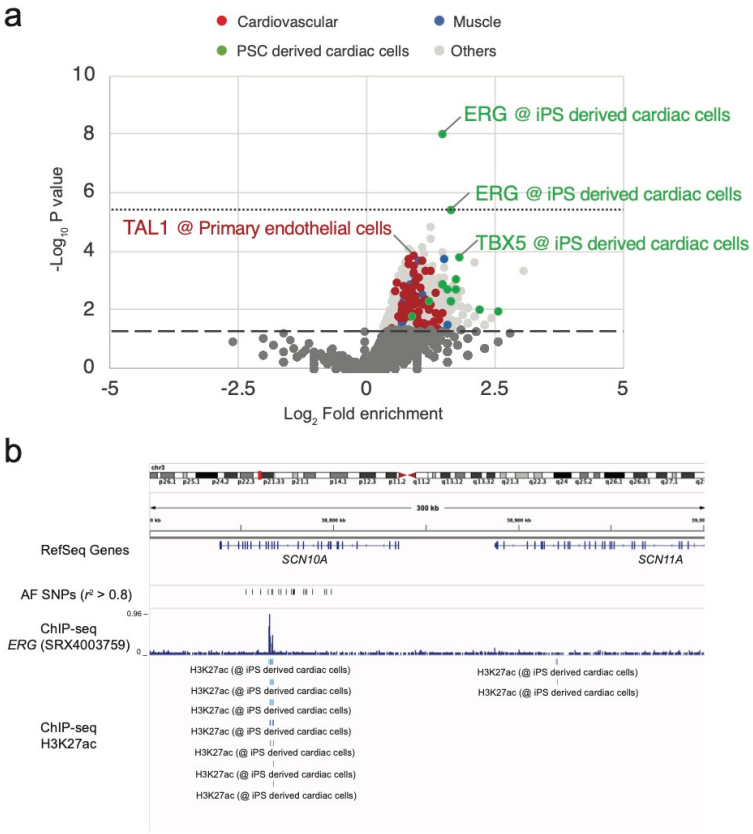


Fig. 3 | Enrichment analysis of transcription factor for AF-associated loci. a, Volcano plot analysis. Each point represents ChIP-seq experiment and is highlighted as red for cardiovascular cells, green for induced pluripotent stem cell-derived cardiac cells, blue for muscle cells, and grey for other cell types. The *x* axis shows $\log_2$-transformed hold enrichment of transcription factor in 150 AF-associated loci, compared to randomly selected 150 genome regions. The *y* axis shows $\log_{10}$-transformed *P* value for enrichment. The dashed line indicated significance threshold level of $P = 0.05$, and the dotted line indicates $P = 0.05/15{,}109$. **b**, *SCN10A* locus showing AF-associated variants with $r^2 > 0.8$ in European samples of 1KG and ChIP-seq track of ERG experiment ($P = 1.0 \times 10^{-8}$) and H3K27ac in iPSC-derived cardiac cells. AF, atrial fibrillation; iPSC, induced pluripotent stem cell.

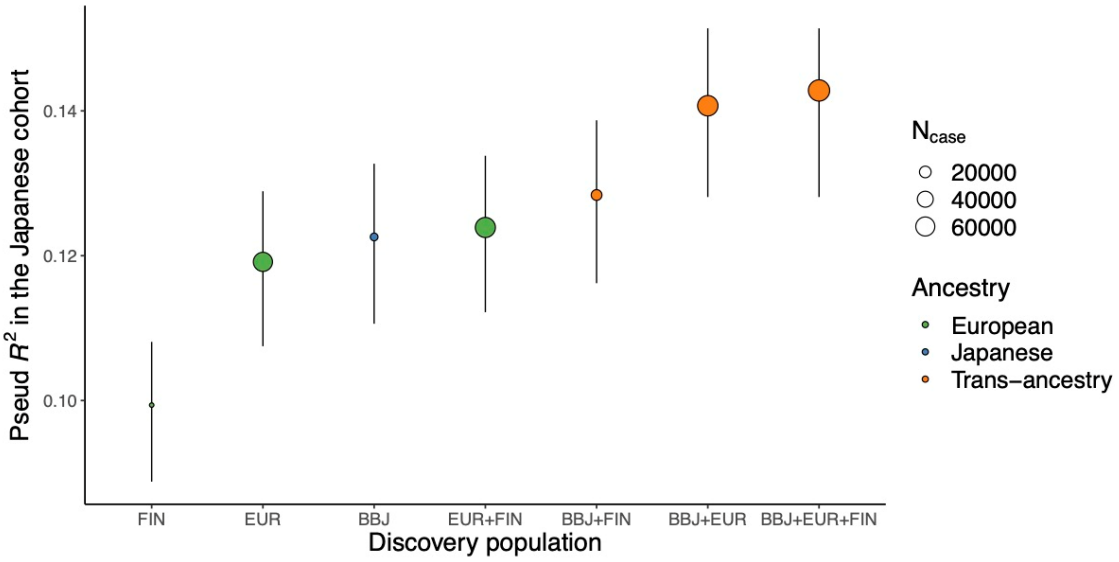


Fig. 4 | Performance of polygenic risk score (PRS) derived from the transethnic meta-analysis. Each point indicates the median Nagelkerke's pseudo R^2 for PRS in the Japanese test cohort (2,953 cases and 21,194 controls). Error bars represent the 95% CI. Median pseudo R^2 and CI were estimated from 5×10^4 times bootstrapping. The color of each point indicates the ancestry that was included in the discovery population. The size of each point indicates the number of cases in the discovery population. CI, confidence interval.

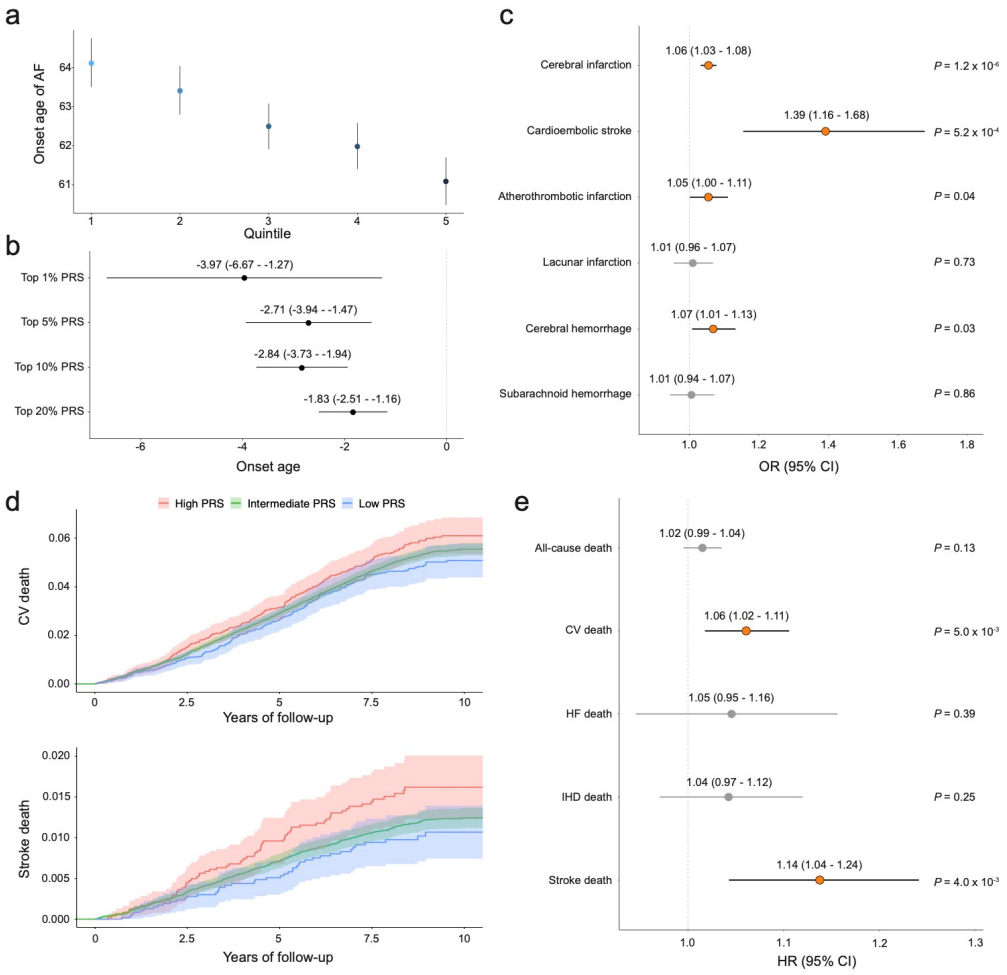


Fig. 5 | Impact of AF-PRS on long-term cardiovascular mortality. A, **b**, Association analysis between AF-PRS and age of AF onset. **A**, Each point represents median age of onset according to the PRS quintile. **B**, Each point and error bar represent estimated β and 95% CI obtained from the linear regression model. **C**, Association between AF-PRS and stroke subtypes in individuals without AF. Data are presented as estimated ORs and 95% CI for 1 s.d. increase in AF-PRS. **c**, **d**, Kaplan-Meier estimates of cumulative events from cardiovascular mortality (**a**) and stroke death (**b**) with 95% CI. Individuals are classified into high PRS (top 10 percentile, red), low PRS (bottom 10 percentile, blue), and intermediate (others, green). **e**, Effects of AF-PRS on long-term mortality. Data are presented as estimated HRs and 95% CIs for a 1 s.d. increase in AF-PRS. AF, atrial fibrillation; PRS, polygenic risk score; CI, confidence interval; OR, odds ratio; HR, hazard ratio; s.d., standard deviation; CV, cardiovascular death; HF, congestive heart failure death; IHD, ischemic heart disease death.

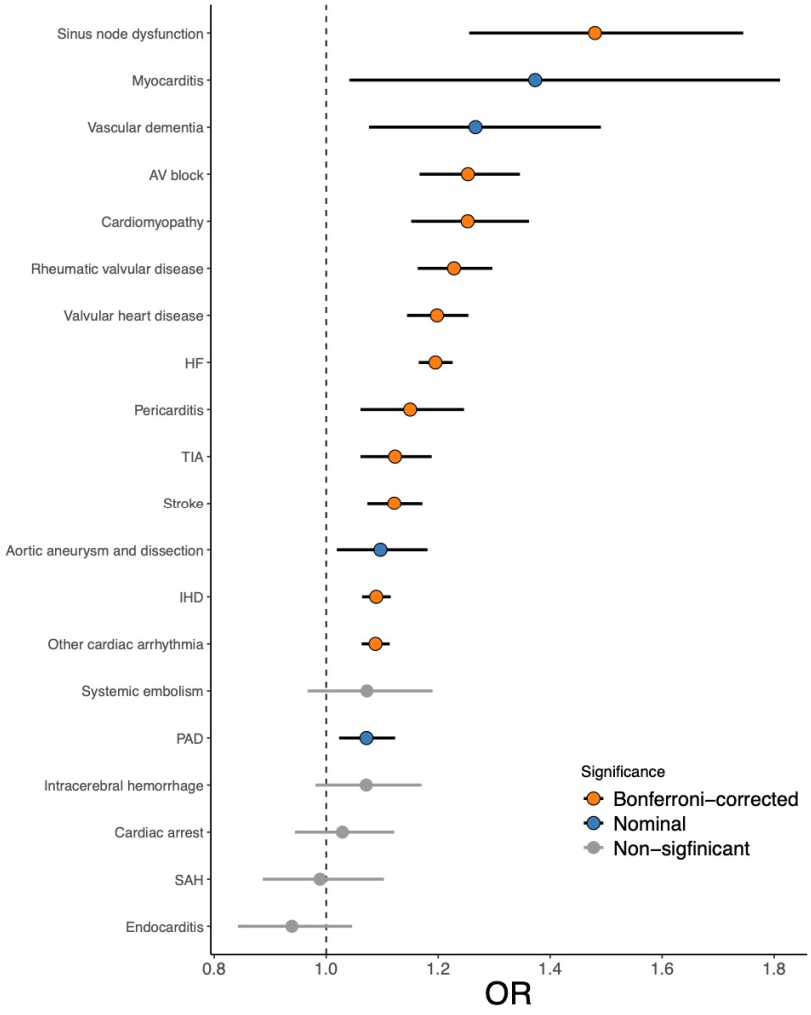
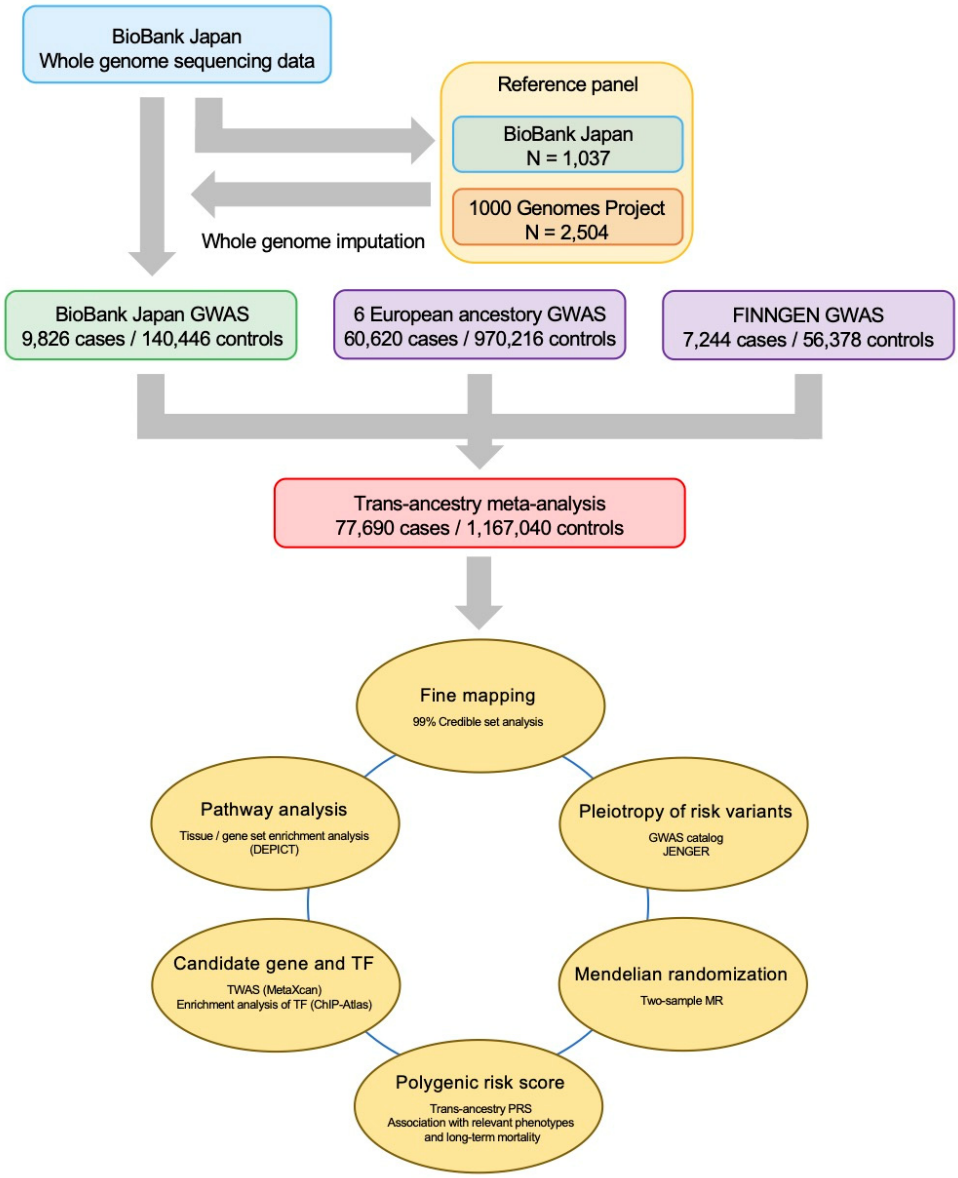
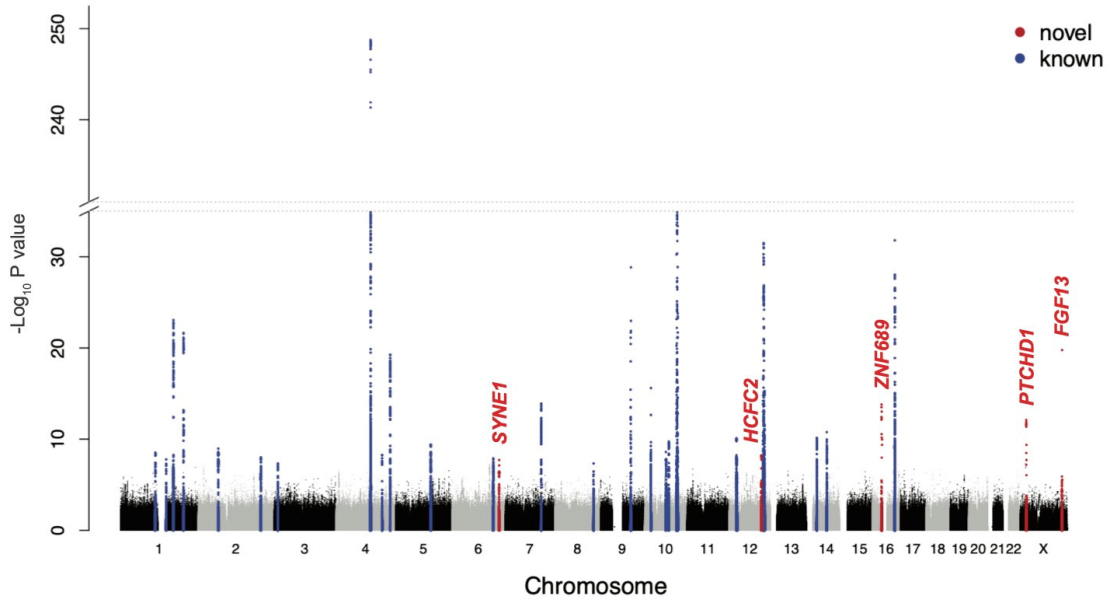


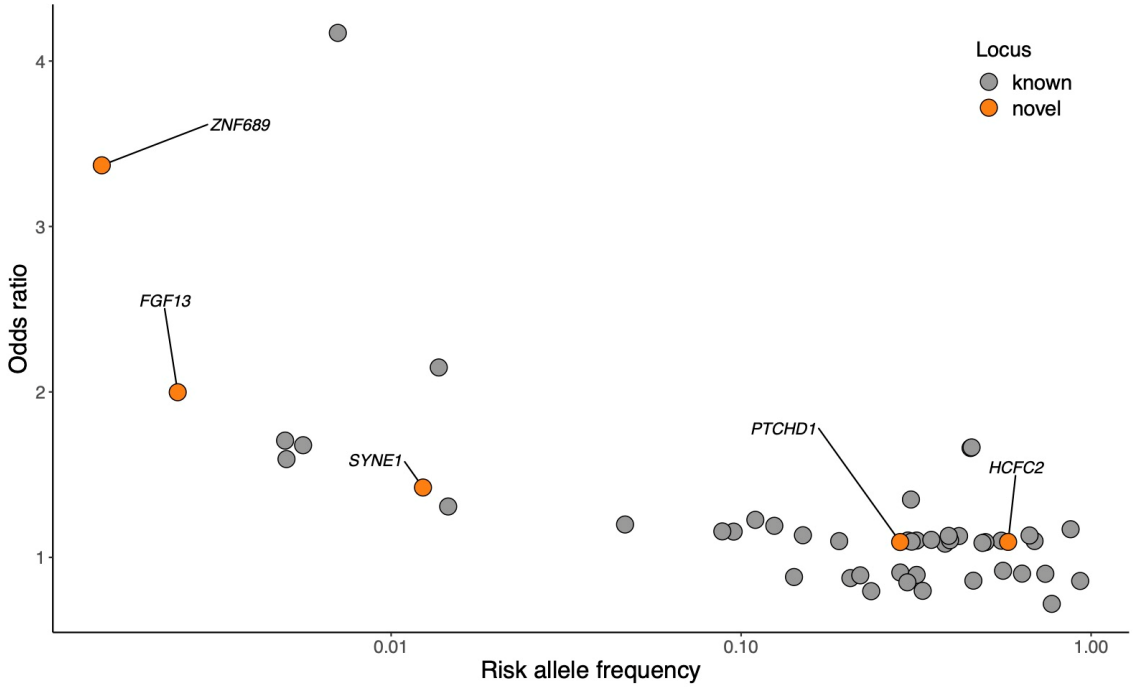
Fig. 6 | Causal effect of AF on cardiovascular and cerebrovascular diseases. Each point and error bar represent the causal estimates on the OR scale and the 95% CI for these ORs respectively. We analyzed MR analysis for the estimation of the causal effects using variants after excluding horizontal pleiotropic outliers. *P* values were determined from the IVW two-sample MR method. AF, atrial fibrillation; OR, odds ratio; CI, confidence interval; MR, sample Mendelian randomization; IVW, inverse-variance weighting; AV block, atrioventricular block; HF, heart failure; TIA, transient ischemic attack; IHD, ischemic heart disease; PAD, peripheral artery disease; SAH, subarachnoid hemorrhage.



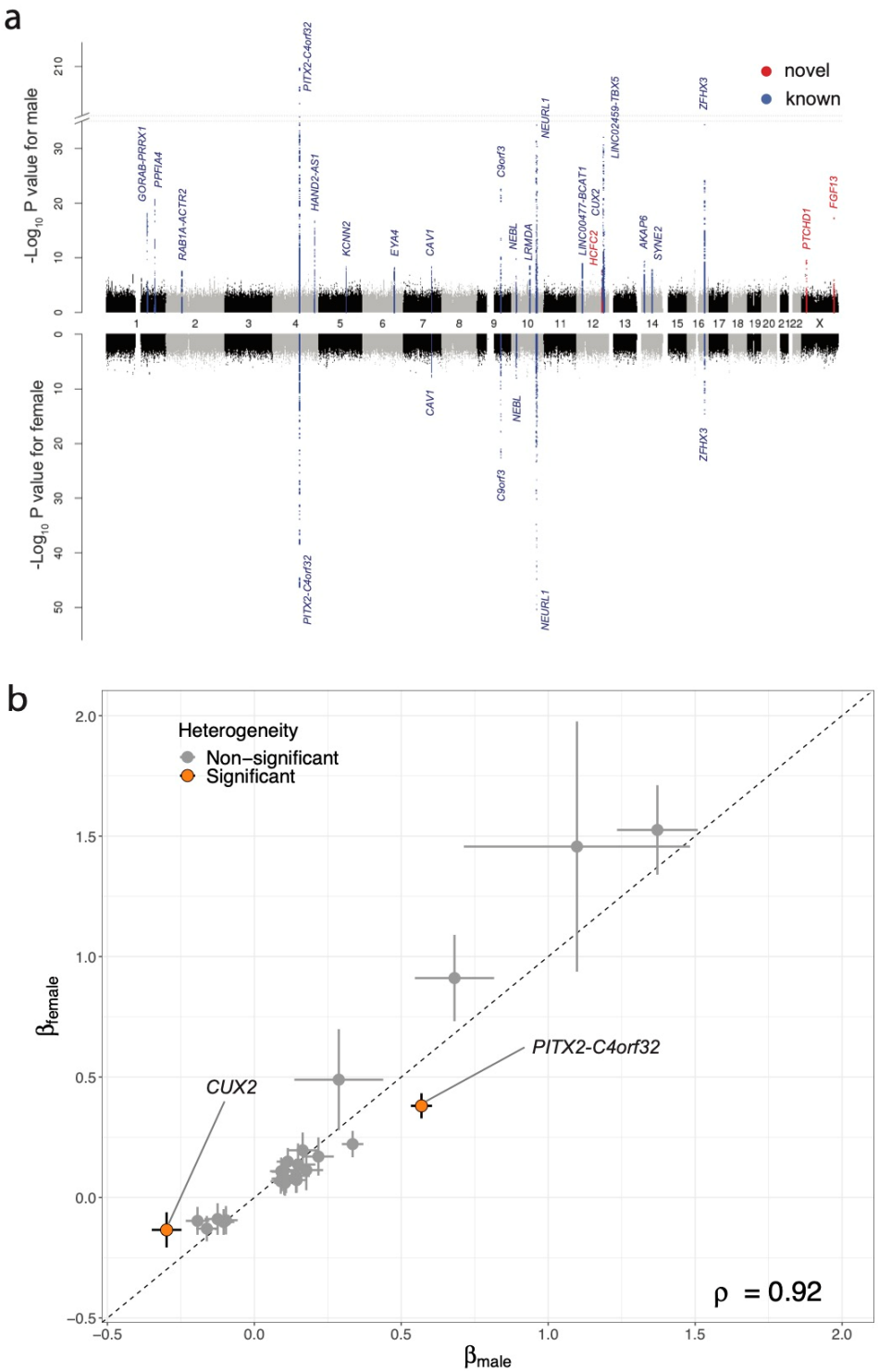
Supplementary Fig. 1 | Overview of the study design. Top, overview of the Japanese cohort of 9,826 AF cases and 140,446 controls with reference panel. Middle, GWAS summary of each study and the trans-ancestry meta-analysis with the number of cases and controls. Bottom, overview of the downstream analyses for AF-associated variants in the trans-ancestry meta-analysis. GWAS, genome-wide association study; JENGER, Japanese ENcyclopedia GEnetic associations by Riken; MR, Mendelian randomization; PRS, polygenic risk score; TWAS, transcriptome-wide association study.



Supplementary Fig. 2 | Manhattan plot for the Japanese GWAS. The results of the Japanese GWAS (9,826 AF cases and 140,446 controls) are shown. The negative $\log_{10} P$ values on the *y-axis* are shown against the genomic positions (hg19) on the *x-axis*. Association signals that reached a genome-wide significance level ($P < 5.0 \times 10^{-8}$) are shown in blue if previously reported loci and in red if novel loci. GWAS, genome-wide association study; AF, atrial fibrillation.

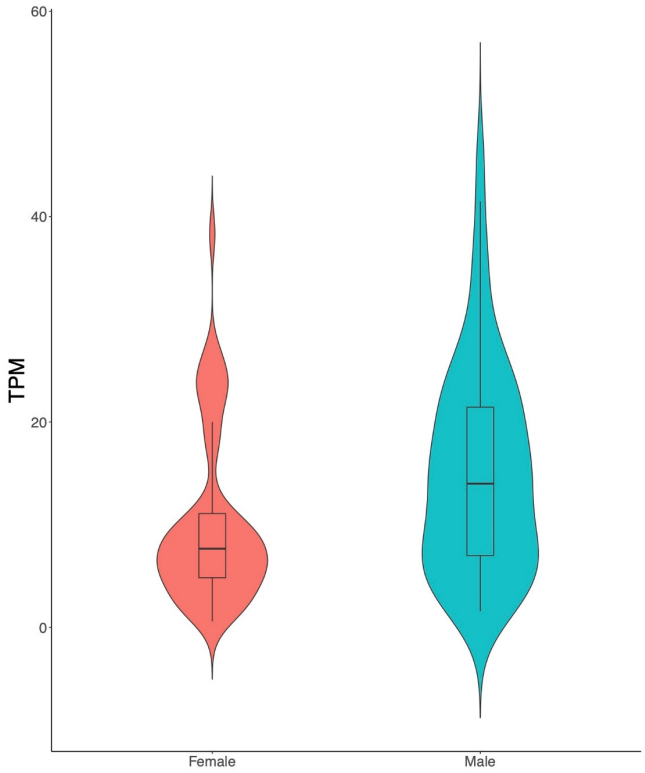


Supplementary Fig. 3 | Independent association signals in AF development. The ORs for AF development of AF 49 independent signals in the Japanese GWAS (31 lead variants and 18 independent variants) were plotted against the risk allele frequencies. Novel variants are highlighted in orange with annotated genes. AF, atrial fibrillation; OR, odds ratio; GWAS, genome-wide association study.

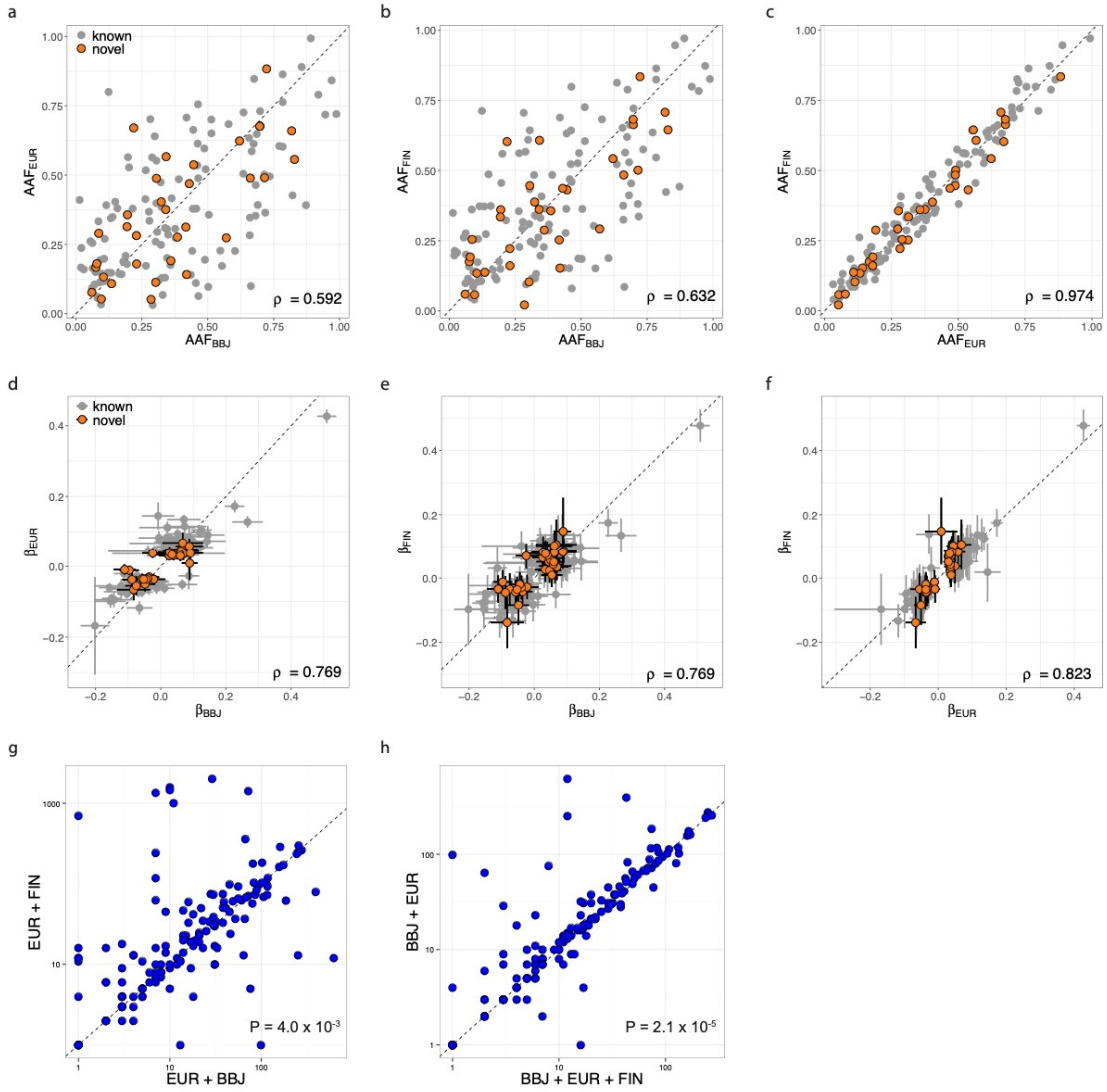


Supplementary Fig. 4 | Sex-stratified GWAS. a, Miami plot for sex-stratified GWAS. Upper panel for male GWAS (N = 81,050: 6,825 cases and 74,225 controls) and lower panel for female GWAS (N = 69,222: 3,001 cases and 66,221 controls). Association signals that reached a genome-wide

significance level ($P < 5.0 \times 10^{-8}$) are shown in blue if previously reported loci and in red if novel loci. **b**, Comparison of estimated effect sizes of the lead variants identified in both male and female GWAS. Data are presented as estimated β and 95% confidence. The lead variants with significant heterogeneity ($P_{het} < 0.0001$) are highlighted in orange, and ρ indicates Spearman's correlation coefficient. GWAS, genome-wide association study.

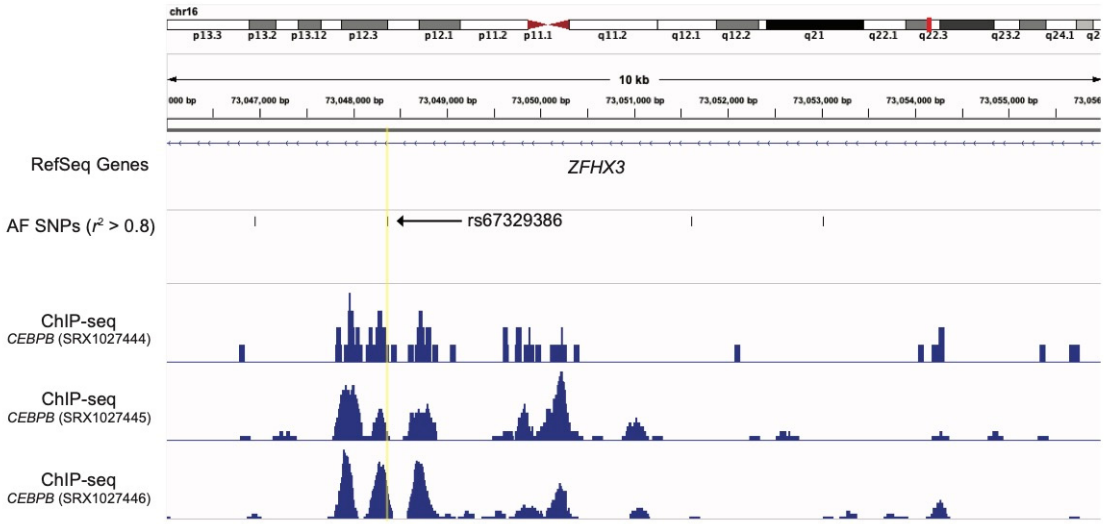


Supplementary Fig. 5 | *CUX2* gene expression in liver. Violin plots showing the distribution of *CUX2* gene expression in the liver based on GTEx data.

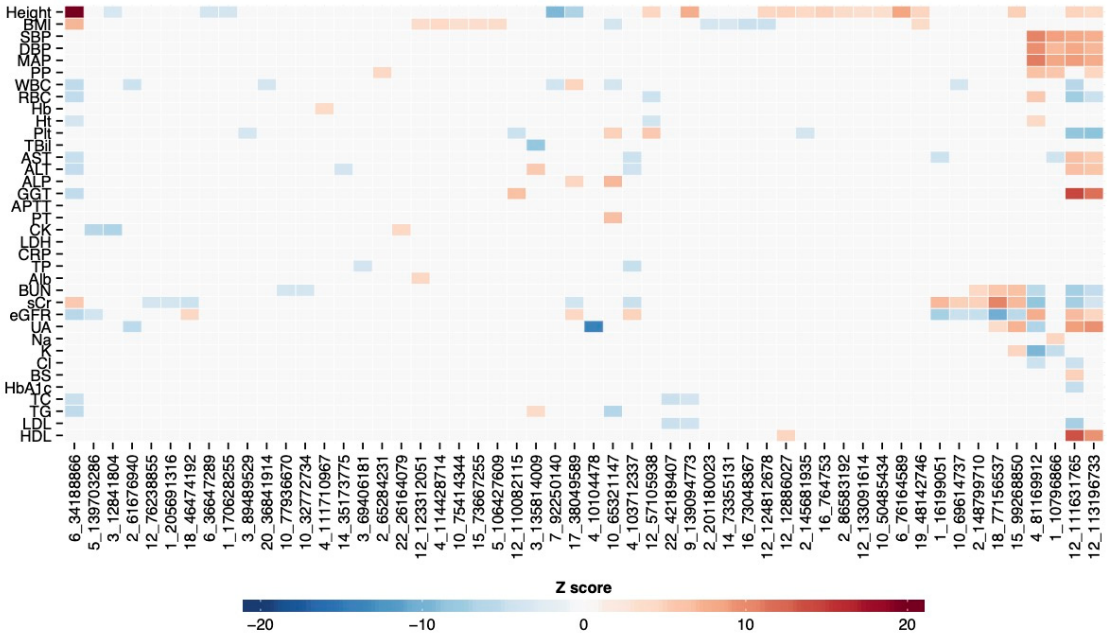


Supplementary Fig. 6 | Comparison of allele frequencies and allelic effects among ancestries and fine mapping derived from credible set analyses a-c, Comparisons of alternate allele frequencies of the 150 lead variants identified in the trans-ancestry meta-analysis (**a**, BBJ versus EUR; **b**, BBJ versus FIN; **c**, EUR versus FIN). **d-f**, Comparison of estimated effect sizes of 150 lead variants. Data are presented as estimated β and 95% CI in each study (**d**, BBJ versus EUR; **e**, BBJ versus FIN; **f**, EUR versus FIN). Grey points indicate previously reported loci, and orange points indicate newly identified loci in this study. ρ indicates Spearman's correlation coefficient. **g**, **h**, Comparisons of the number of variants included in the 99% credible sets derived from each combination of the meta-analysis (**g**, EUR + BBJ versus EUR + FIN; **h**, BBJ + EUR + FIN versus

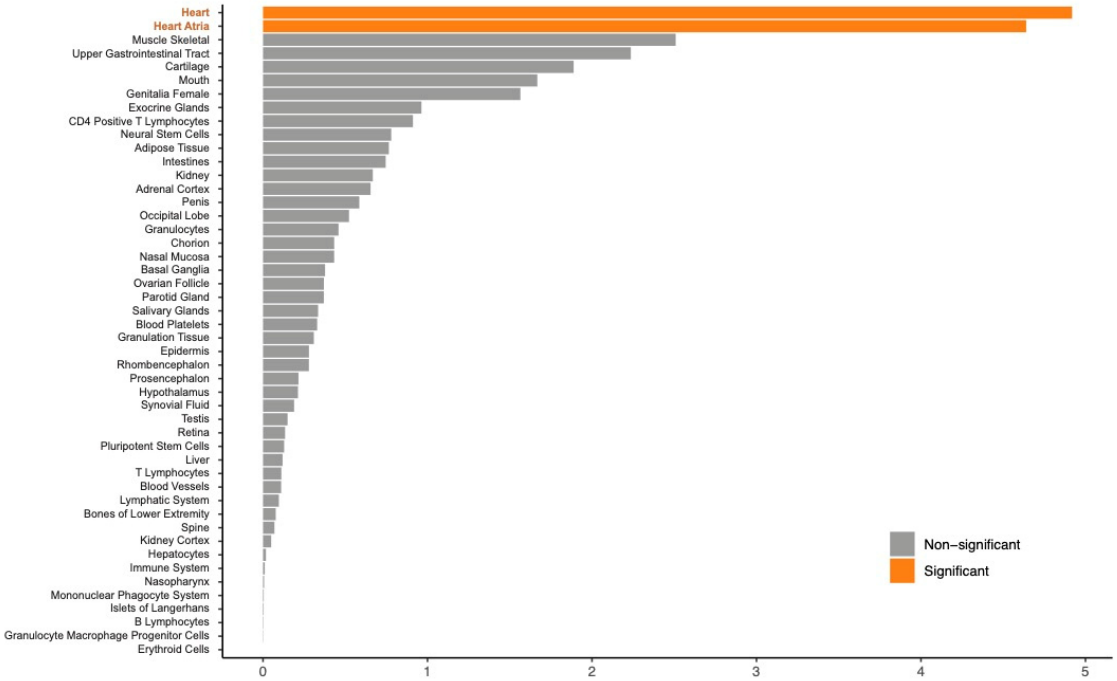
BBJ + EUR). Both *the x* and *y* axes represent the log-transformed number of variants in the 99% credible sets. BBJ, BioBank Japan; EUR, European population; FIN, FinnGen data release 2; CI, confidence interval.



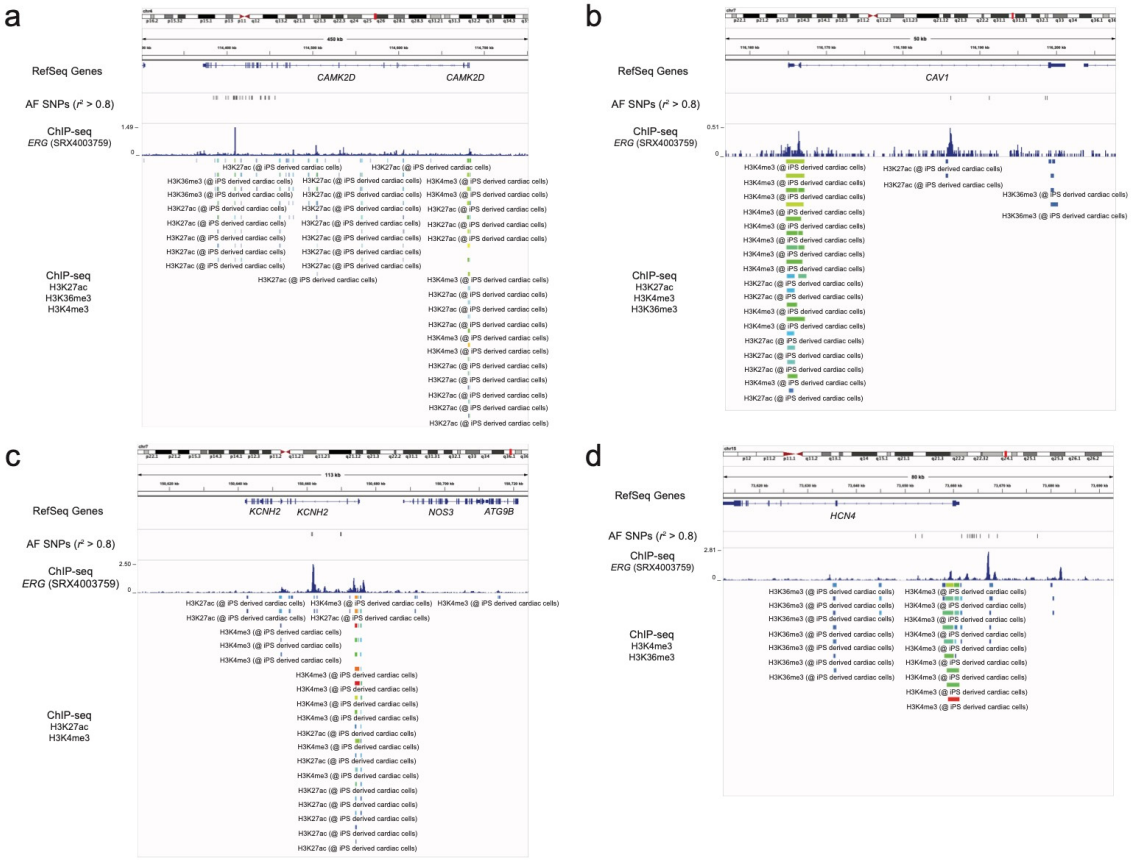
Supplementary Fig. 7 | rs67329386 and transcription factor. *ZFHX3* locus showing rs67329386 (highlighted in yellow) and proxies with $r^2 > 0.8$ in European samples in 1KG along with ChIP-seq track of CEBPB experiments in mesenchymal stem cells.



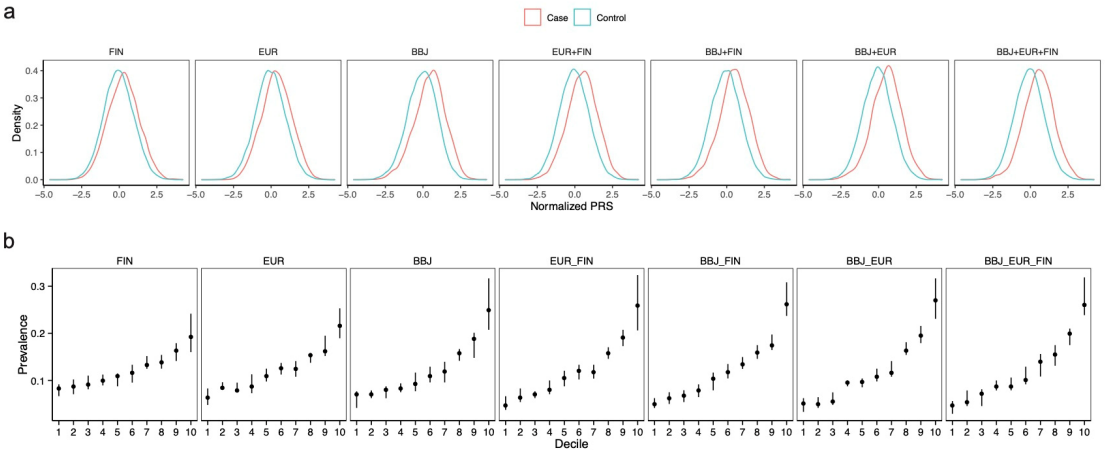
Supplementary Fig. 8 | Overlap between AF-associated signals and variants associated with quantitative traits Heat-map representation of the AF-associated loci, in which AF-associated variants in LD with $r^2 > 0.6$ in East Asian samples of 1KG overlapped with variants associated with at least one of the quantitative traits. The Z score of QTL alleles that were in accordance with the AF risk allele are shown. Positive values are shown in red, and negative values are shown in blue. The rows show quantitative traits, and the column shows the chromosome and base position of AF-associated signals. AF, atrial fibrillation; LD, linkage disequilibrium; QTL, quantitative trait loci.



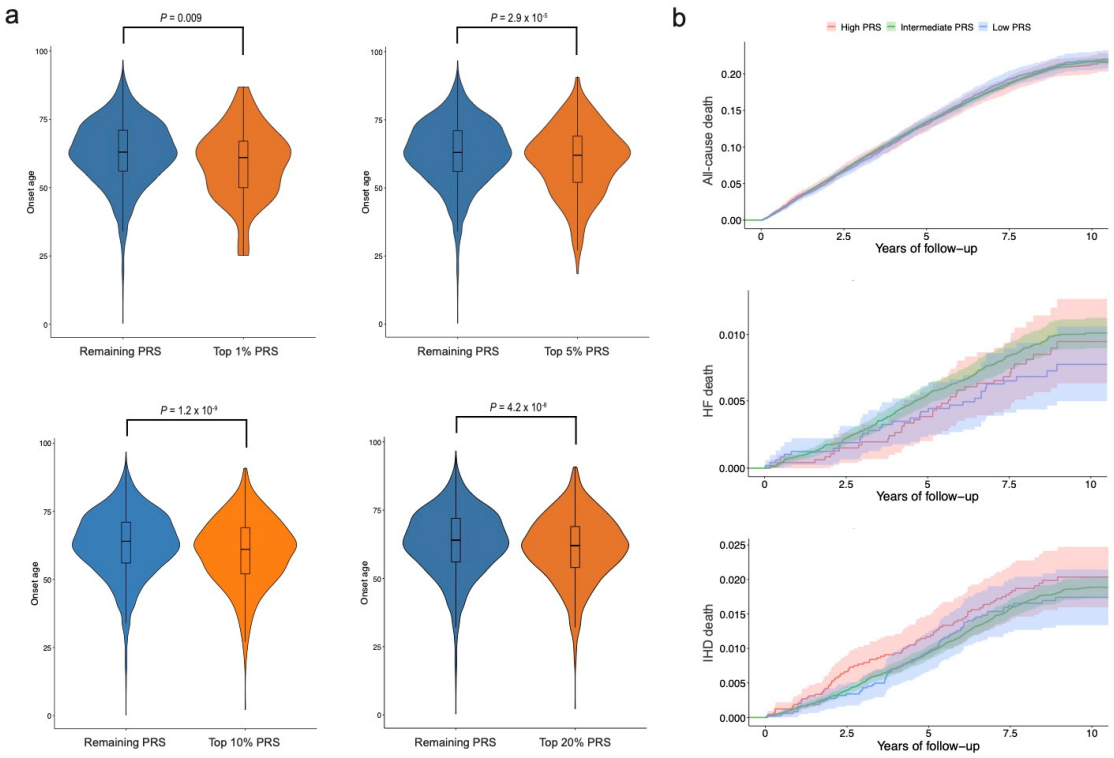
Supplementary Fig. 9 Tissue enrichment and eQTL analysis The results of tissue enrichment analysis ($\log_{10}$ BF > 5 and 16,181 variants). The *x-axis* indicates the -$\log_{10}$ *P* value. Forty-eight exemplar tissues (Methods) were tested, and the significance level was set at $P = 1 \times 10^{-3}$ (0.05/48). eQTL, expression quantitative trait locus; BF, Bayes factor.



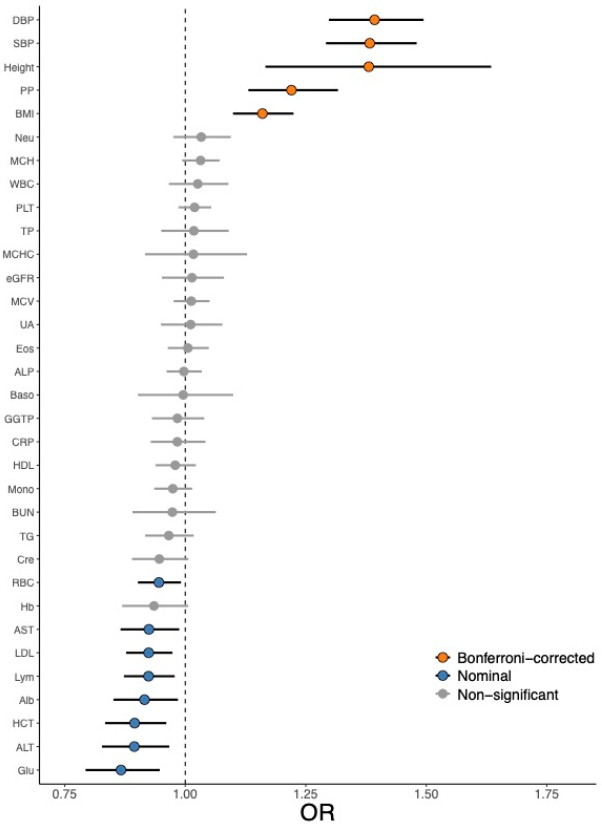
Supplementary Fig. 10 | Ion channel-related gene and ERG. a, *CAMK2D* locus **b**, *CAV1* locus, **c**, *KCNH2* locus **d**, *HCN4* locus showing AF-associated variants ($r^2 > 0.8$ in European samples in 1KG) and ChIP-seq track of ERG experiment ($P = 1.0 \times 10^{-8}$) and histone modification markers in iPSC-derived cardiac cells. iPSC, induced pluripotent stem cells.



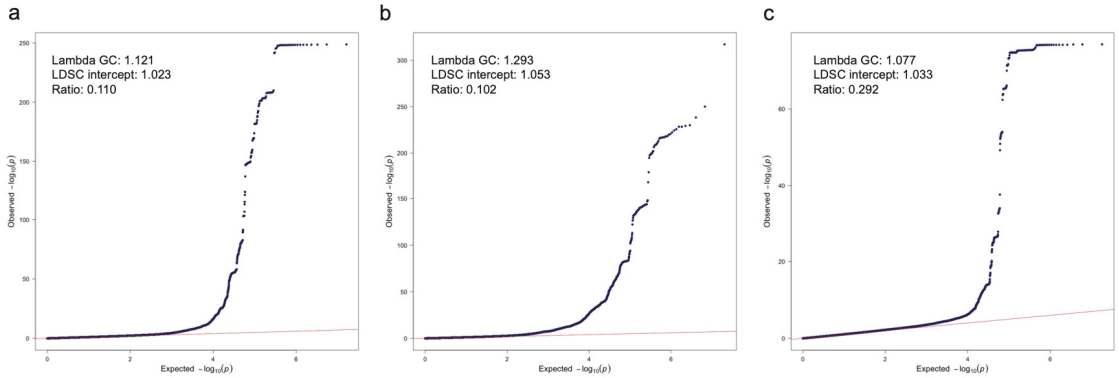
Supplementary Fig. 11 | PRS distribution and AF prevalence. a, Distribution of PRS in the case and control samples for each combination of GWAS. **b** Prevalence of AF based on the AF-PRS deciles in each combination of GWAS. Data are presented as medians and 95% CI. PRS, polygenic risk score; AF, atrial fibrillation; GWAS, genome-wide association study; CI, confidence interval.



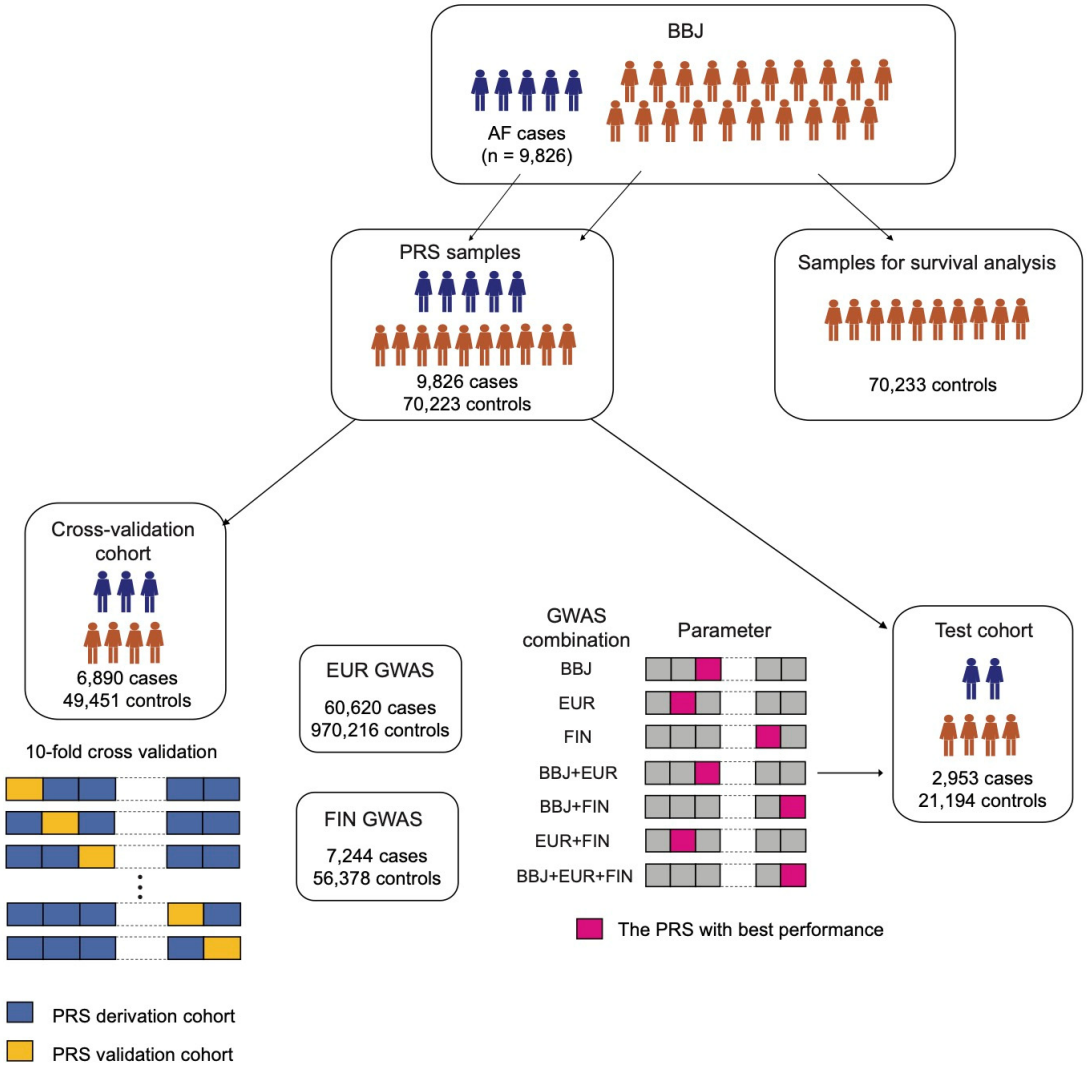
Supplementary Fig. 12 | Impact of AF-PRS on the age of AF onset and long-term mortality a, Comparison of the age of AF onset between samples with high PRS and those with remaining PRS. **b**, Kaplan-Meier estimates of cumulative events from all-cause death, heart failure death, and ischemic heart disease death with 95% CI. Individuals are classified into high PRS (top 10 percentile, red), low PRS (bottom 10 percentile, blue), and intermediate (others, green). AF, atrial fibrillation; PRS, polygenic risk score; CI, confidence interval.



Supplementary Fig. 13 | Causal effect of quantitative traits on AF development Each point and error bar represents the causal estimates on the OR scale and the 95% CI for these ORs, respectively. We analyzed the MR analysis to estimate the causal effects using variants after excluding horizontal pleiotropic outliers. The *P* values were determined using the IVW two-sample MR method. OR, odds ratio; MR, Mendelian randomization; IVW, inverse-variance-weighting; DBP, diastolic blood pressure; SBP, systolic blood pressure; PP, pulse pressure; BMI, body mass index; Neu, neutrophil count; MCH, mean corpuscular hemoglobin; WBC, white blood cell count; PLT, platelet count; TP, total protein; MCHC, mean corpuscular hemoglobin concentration; eGFR, estimated glomerular filtration rate; MCV, mean corpuscular volume; UA uric acid; Eos, eosinophil count; ALP, alkaline phosphatase; Baso, basophil count; GGT, gamma-glutamyl transferase; CRP, C-reactive protein; HDL, high-density lipoprotein cholesterol; Mono, monocyte count; BUN, blood urea nitrogen; TG, triglyceride; Cre, serum creatinine; RBC, red blood cell; Hb, hemoglobin; AST, aspartate aminotransferase; LDL, low-density lipoprotein cholesterol; Lym, lymphocyte count; Alb, albumin; HCT, hematocrit; ALT, alanine aminotransferase; Glu, blood sugar



Supplementary Fig. 14 | Quantile-quantile plot for GWAS of three studies. a, the Japanese GWAS. **b**, GWAS of a meta-analysis in European population (EUR). **c**, GWAS in data from FinnGen project (FIN). GWAS, genome-wide association study.

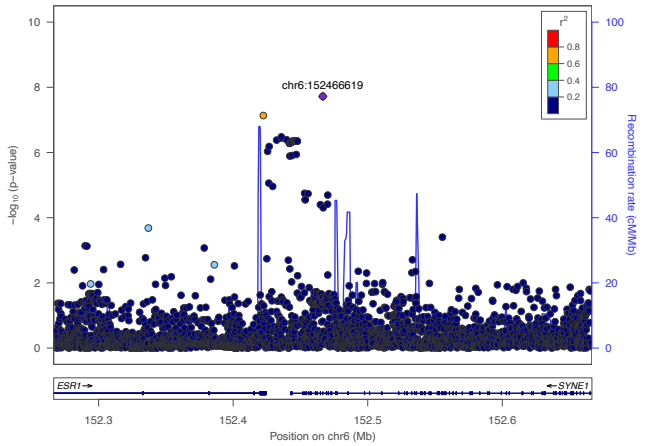


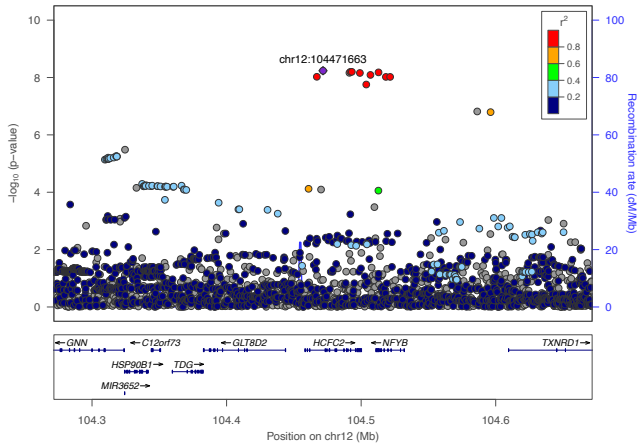
Supplementary Fig. 15 | Analytical scheme for PRS derivation, validation, and test cohorts and survival analysis. Schematic representation for cross-validation, performance testing of the AF-PRS in the independent test cohort, and survival analysis for AF-PRS. PRS, polygenic risk score; AF, atrial fibrillation.

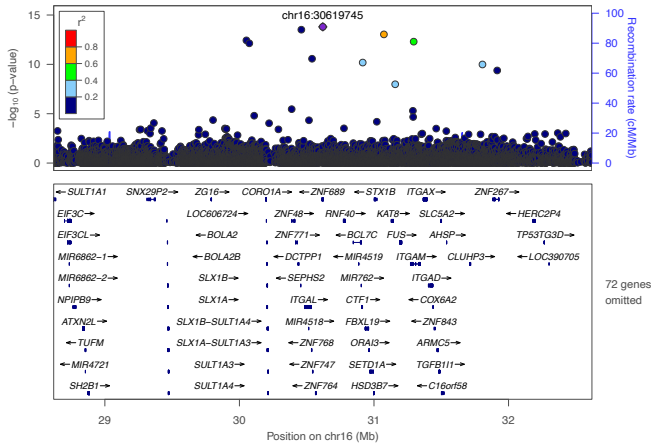
Supplementary Dataset

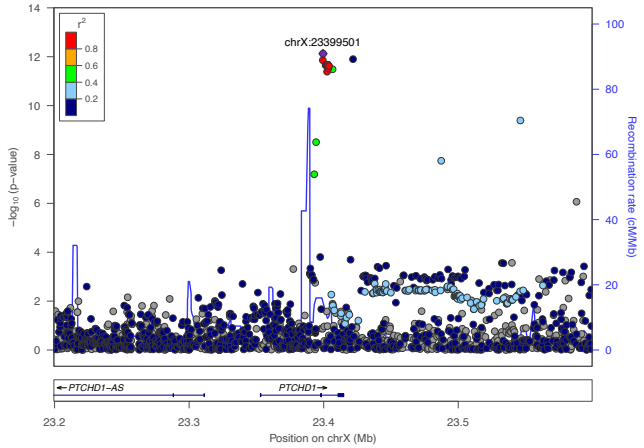
1. Novel 5 loci in the Japanese GWAS
2. Previously reported 26 loci in the Japanese GWAS
3. Novel 33 loci in the trans-ancestry meta-GWAS
4. Previously reported 127 loci in the trans-ancestry meta-GWAS

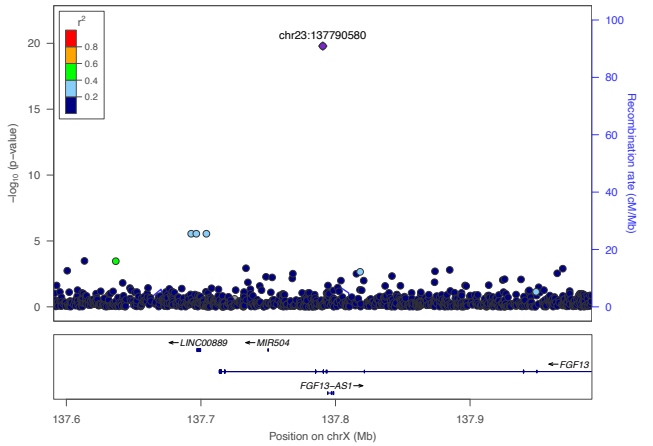
1. Novel 5 loci in the Japanese GWAS



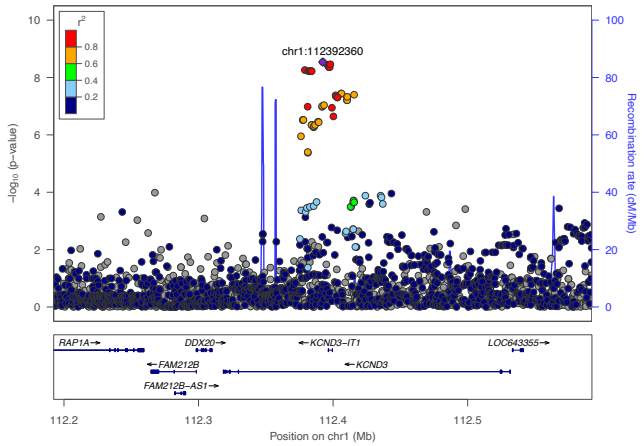


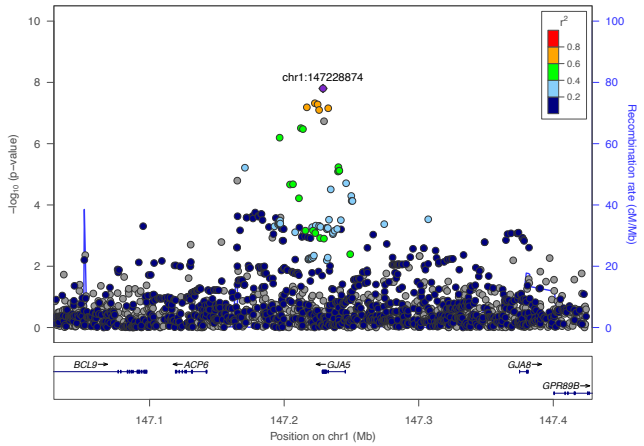


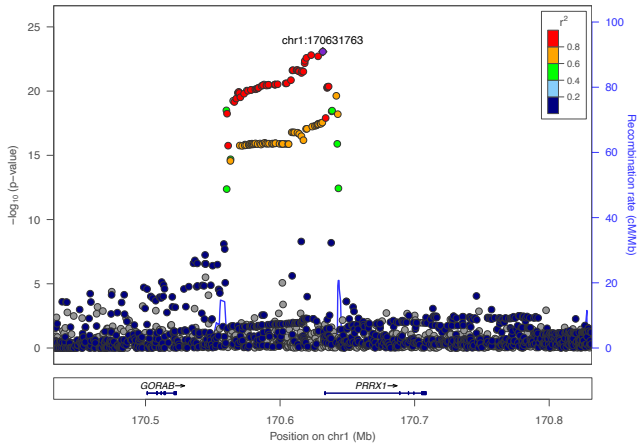


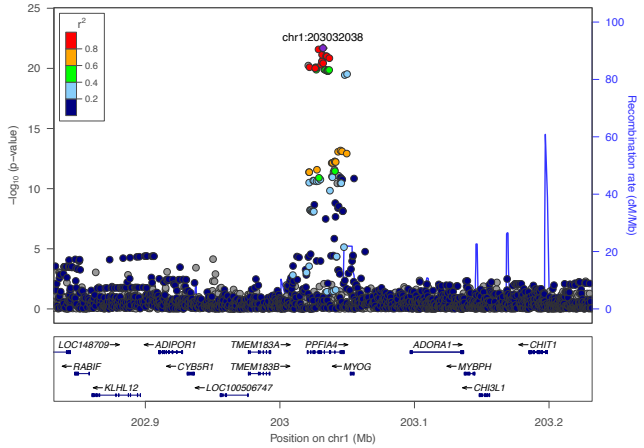


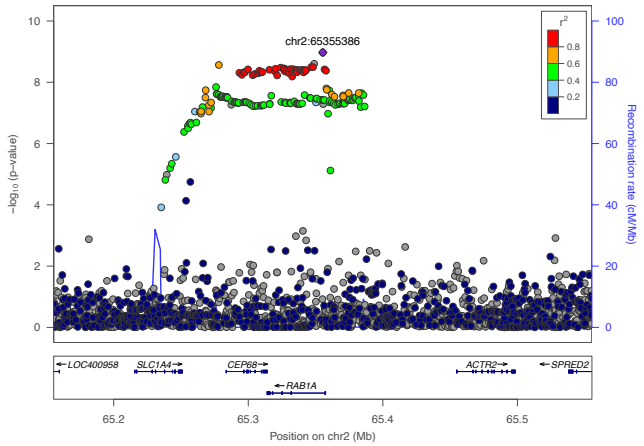
2. Previously reported 26 loci in the Japanese GWAS

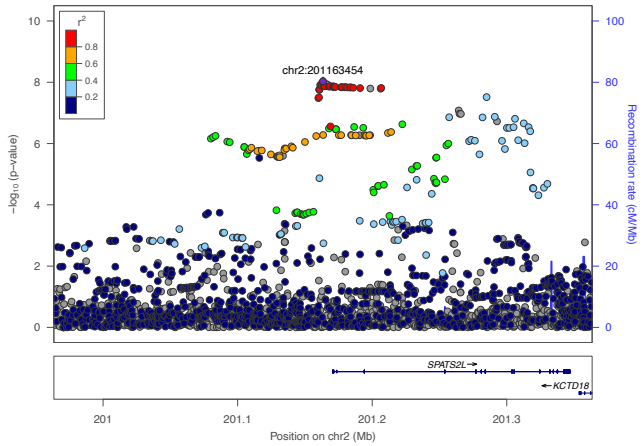


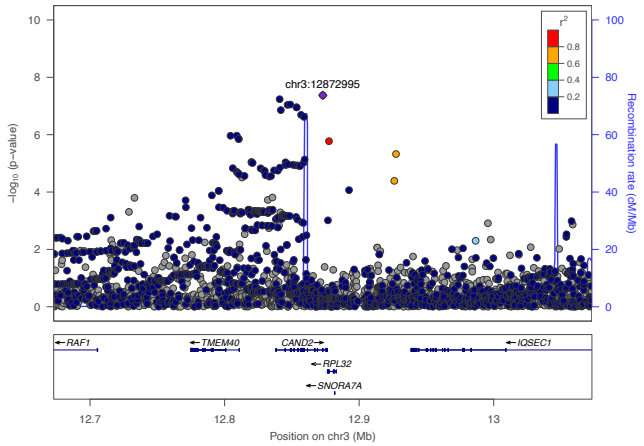


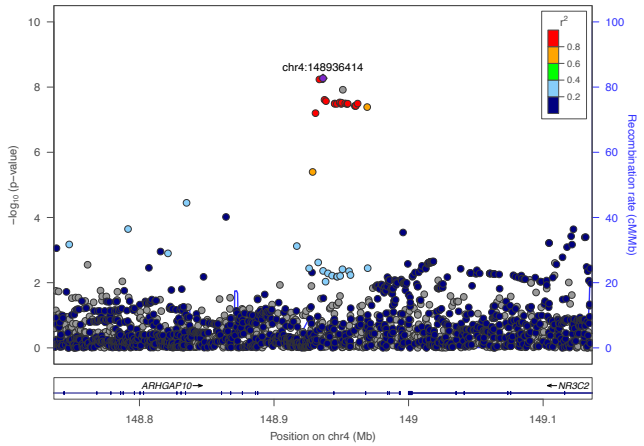


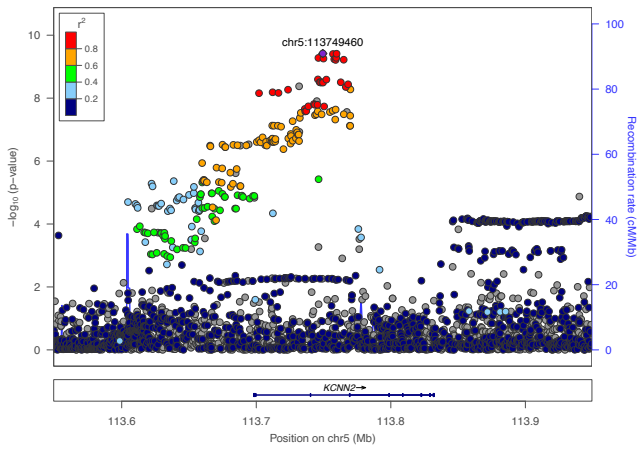


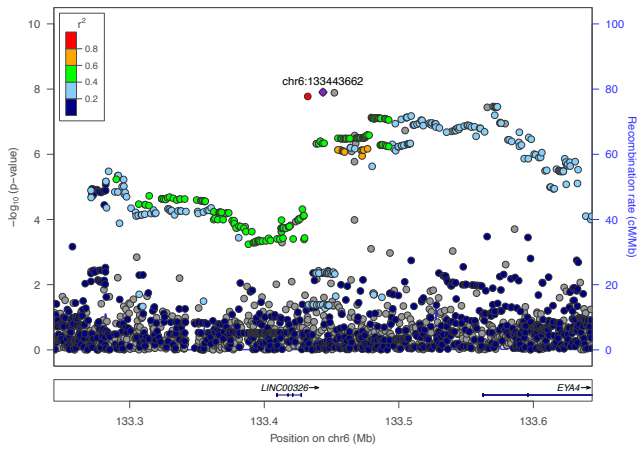


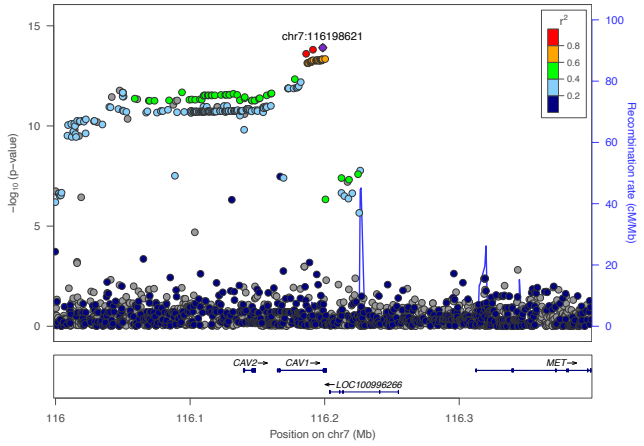


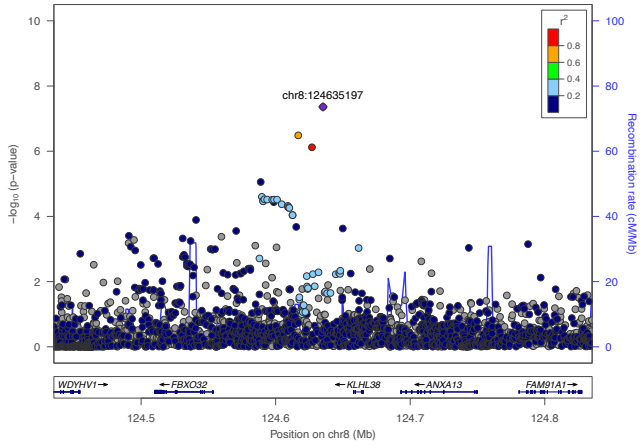


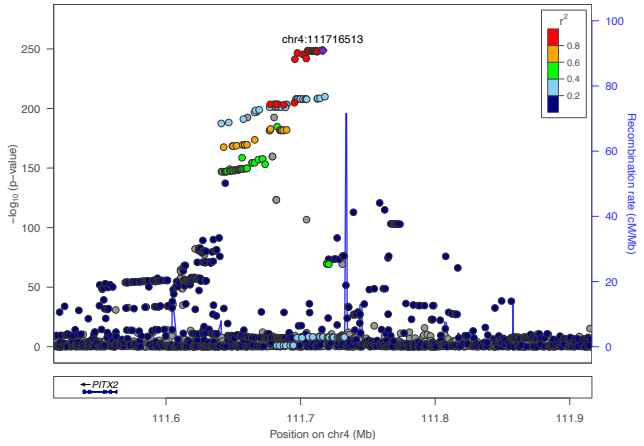


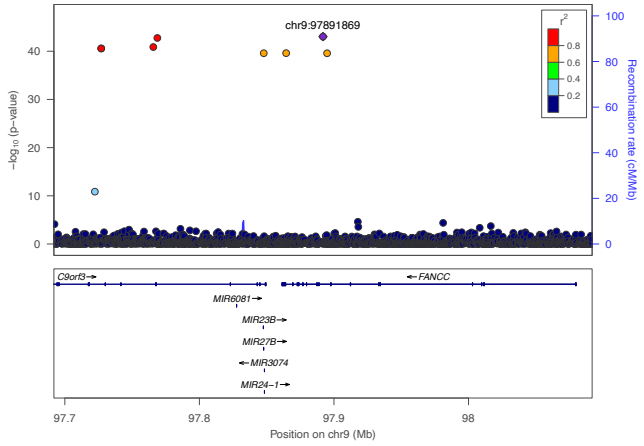

chr5:113749460
KCNN2
Position on chr5 (Mb)
Recombination rate (cM/Mb)

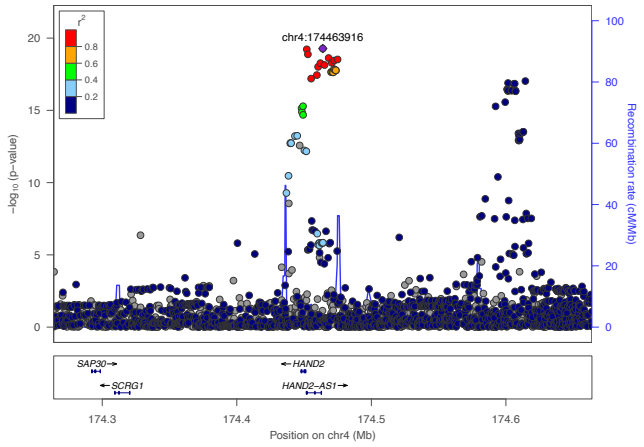

chr6:133443662
LINC00326
EYA4
Position on chr6 (Mb)
Recombination rate (cM/Mb)

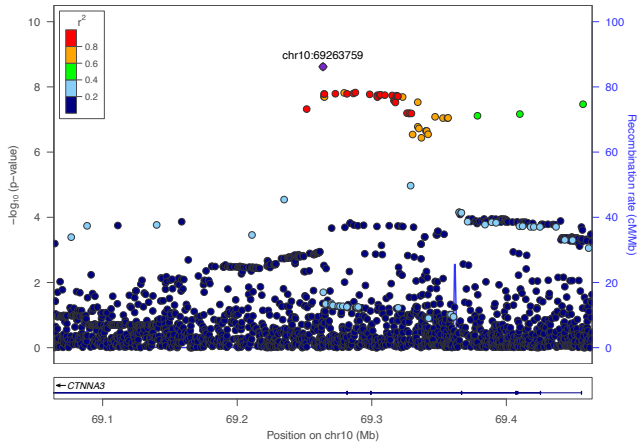

chr7:116198621
CAV2
CAV1
MET
LOC100996266
Position on chr7 (Mb)
Recombination rate (cM/Mb)

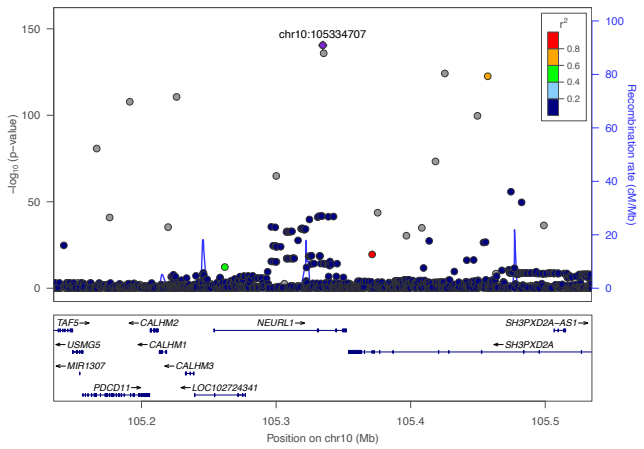

chr8:124635197
WDYHV1
FBXO32
KLHL38
ANXA13
FAM91A1
Position on chr8 (Mb)
Recombination rate (cM/Mb)

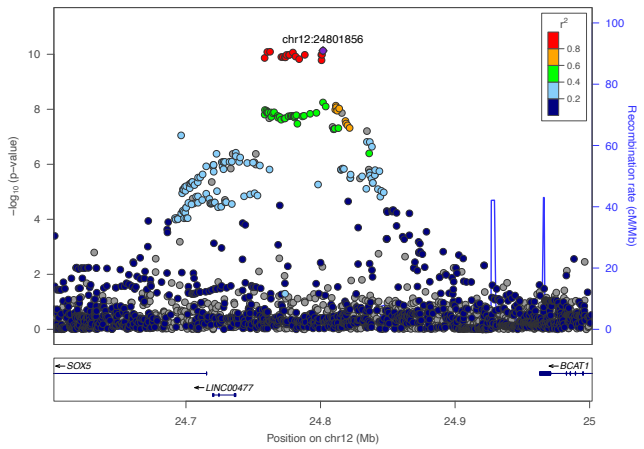

chr4:111716513
PITX2
Position on chr4 (Mb)
Recombination rate (cM/Mb)

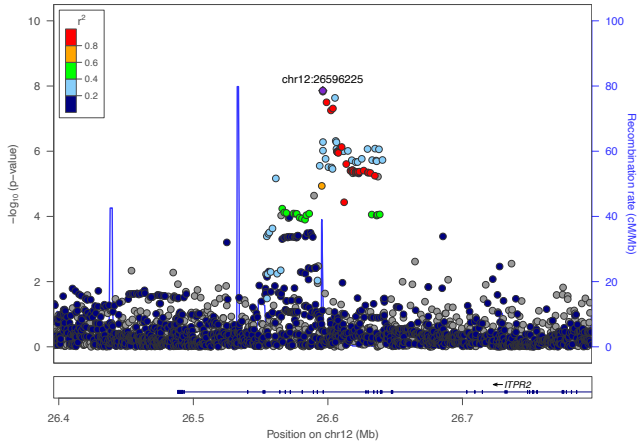

chr9:97891869
C9orf3
FANCC
MIR6081
MIR23B
MIR27B
MIR3074
MIR24-1
Position on chr9 (Mb)
Recombination rate (cM/Mb)

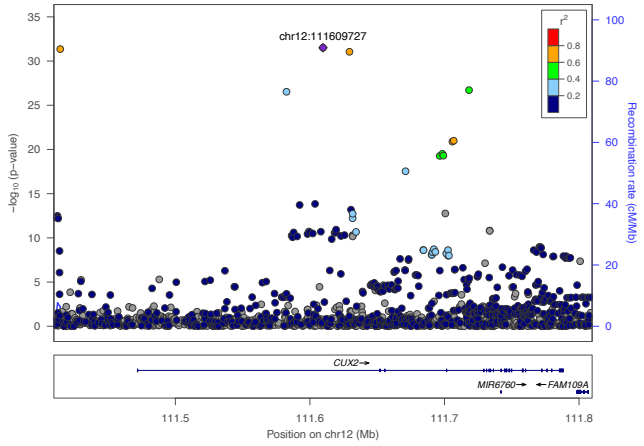

chr4:174463916
SAP30
SCRG1
HAND2
HAND2-AS1
Position on chr4 (Mb)
Recombination rate (cM/Mb)

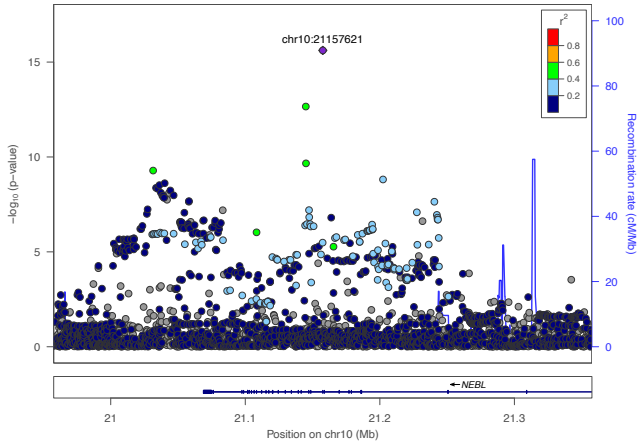

chr10:69263759
CTNNA3
Position on chr10 (Mb)
Recombination rate (cM/Mb)

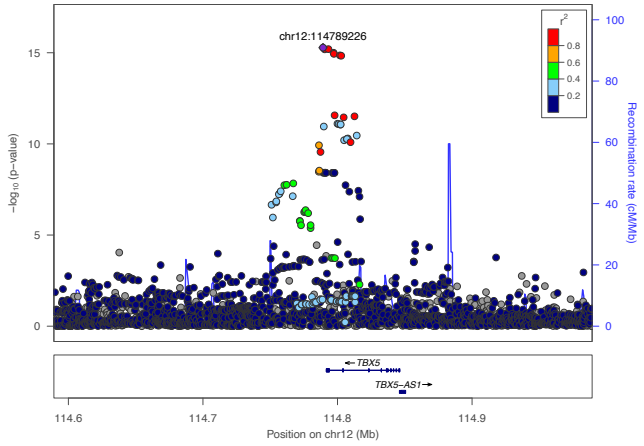

chr10:105334707
TAF5
CALHM2
NEURL1
SH3PXD2A-AS1
USMG5
CALHM1
SH3PXD2A
MIR1307
CALHM3
PDCD11
LOC102724341
Position on chr10 (Mb)
Recombination rate (cM/Mb)

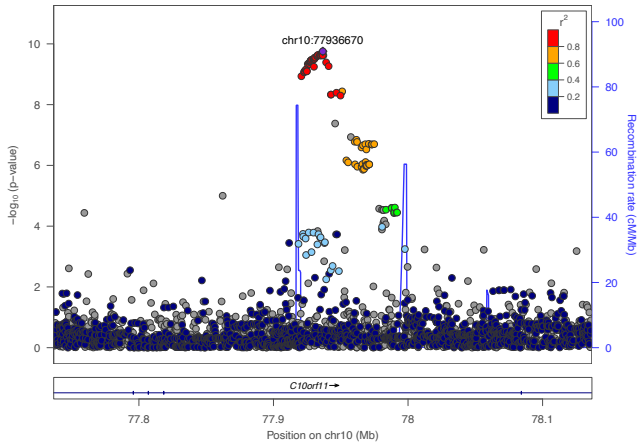

chr12:24801856
SOX5
LINC00477
BCAT1
Position on chr12 (Mb)
Recombination rate (cM/Mb)

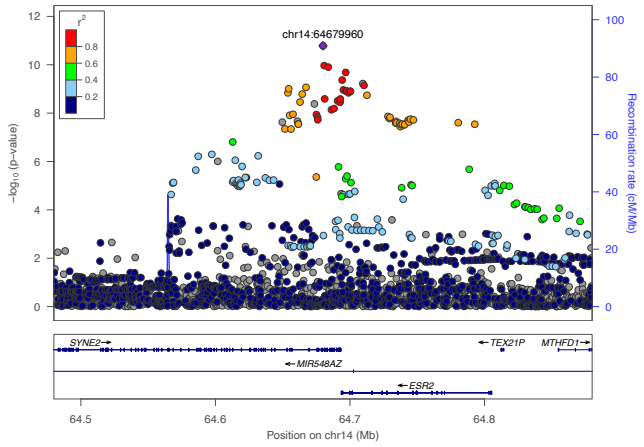

chr12:26596225
ITPR2
Position on chr12 (Mb)
Recombination rate (cM/Mb)

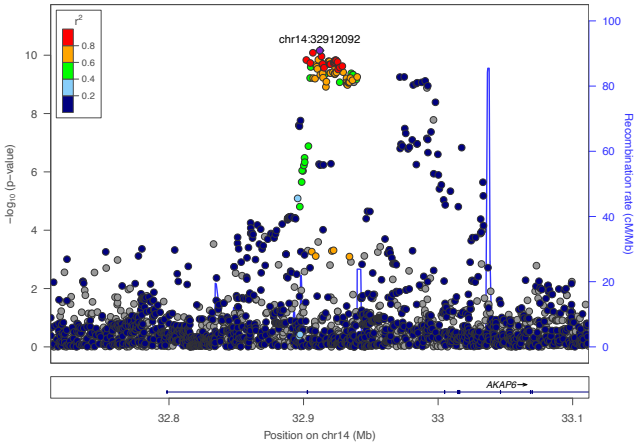

chr12:111609727
CUX2
MIR6760
FAM109A
Position on chr12 (Mb)
Recombination rate (cM/Mb)

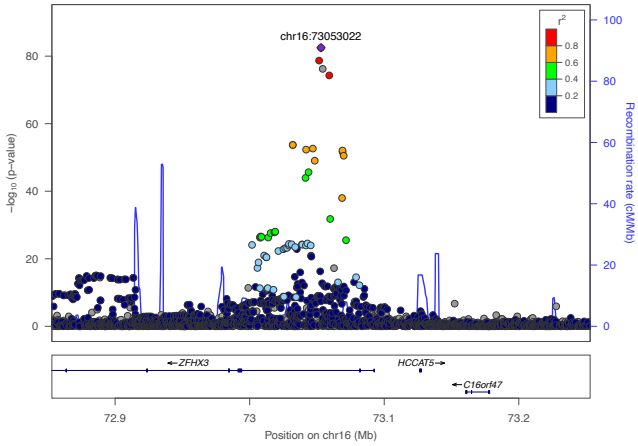

chr10:21157621
NEBL
Position on chr10 (Mb)
Recombination rate (cM/Mb)


chr12:114789226
TBX5
TBX5-AS1
Position on chr12 (Mb)
Recombination rate (cM/Mb)

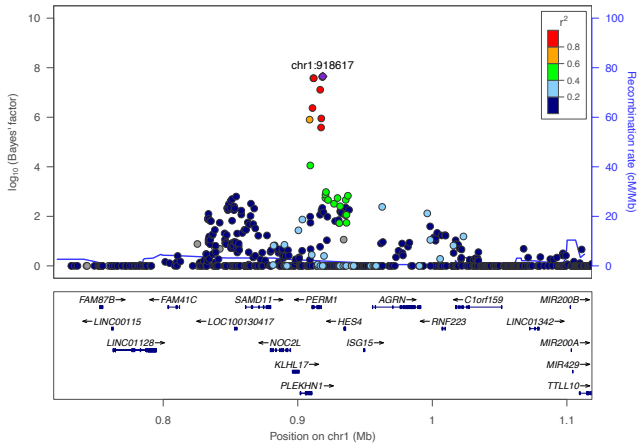

chr10:77936670
C10orf11
Position on chr10 (Mb)
Recombination rate (cM/Mb)

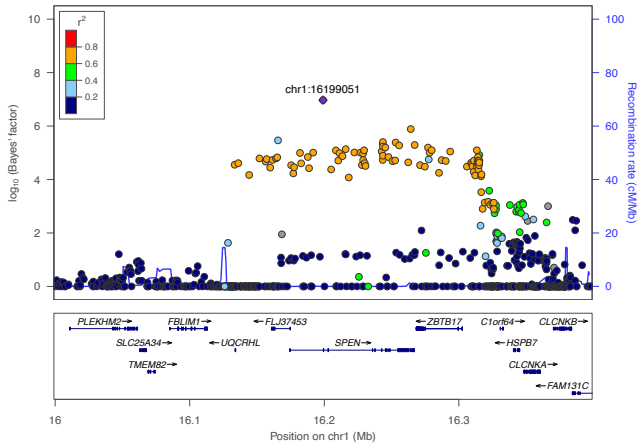

chr14:64679960
SYNE2
TEX21P
MTHFD1
MIR548AZ
ESR2
Position on chr14 (Mb)
Recombination rate (cM/Mb)

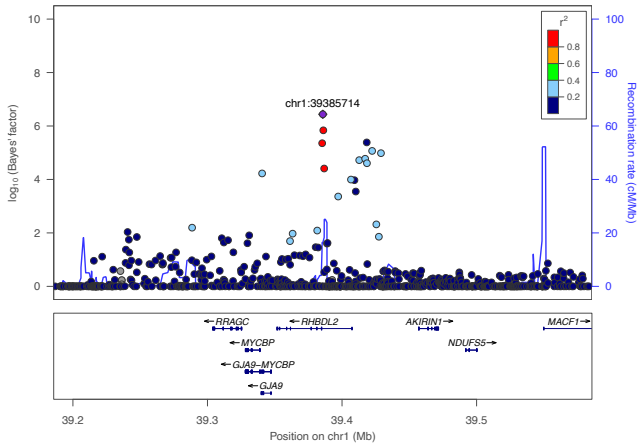

chr14:32912092
r²
0.8
0.6
0.4
0.2
−log10 (p−value)
Recombination rate (cM/Mb)
AKAP6
32.8
32.9
33
33.1
Position on chr14 (Mb)

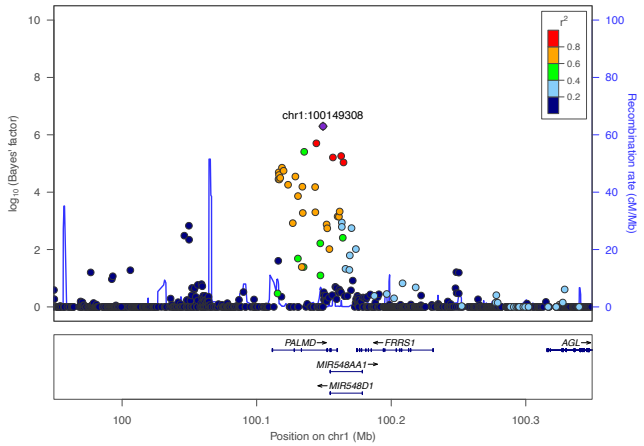

chr16:73053022
r²
0.8
0.6
0.4
0.2
−log10 (p−value)
Recombination rate (cM/Mb)
ZFHX3
HCCAT5
C16orf47
72.9
73
73.1
73.2
Position on chr16 (Mb)

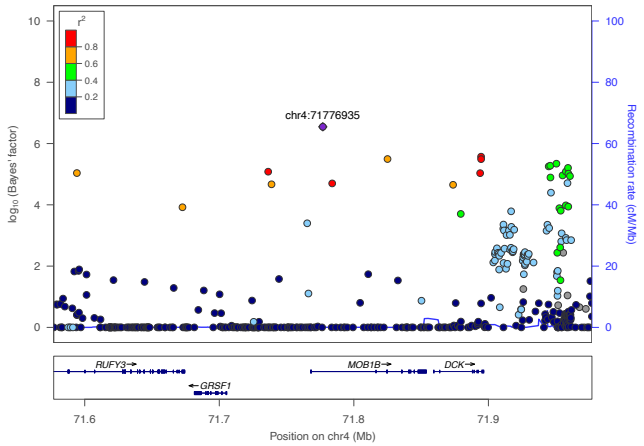
3. Novel 33 loci in the trans-ancestry meta-GWAS

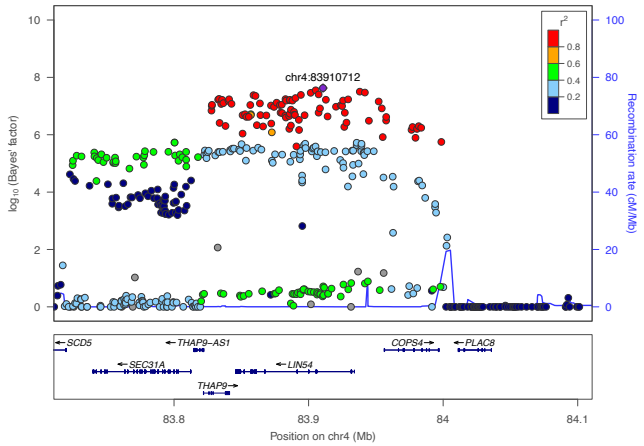


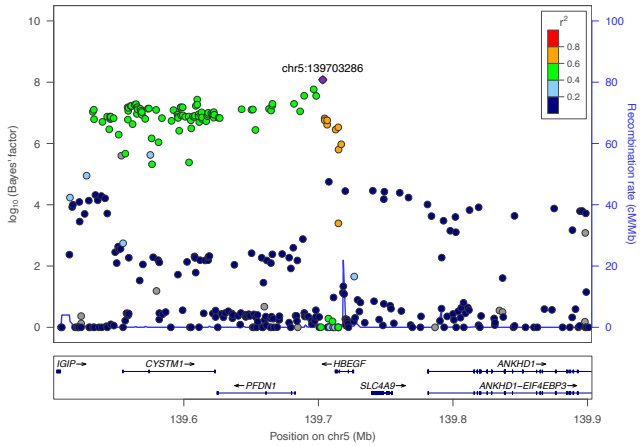


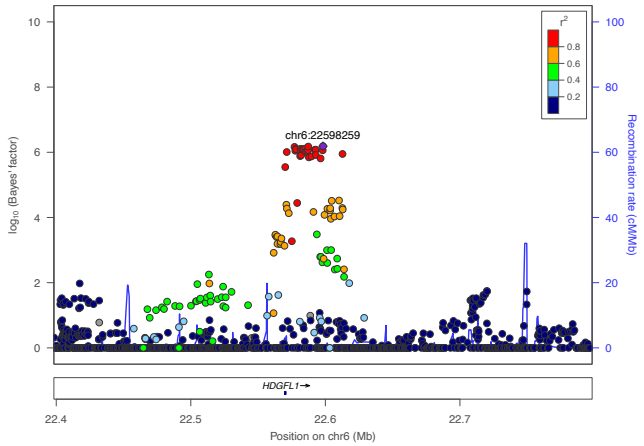


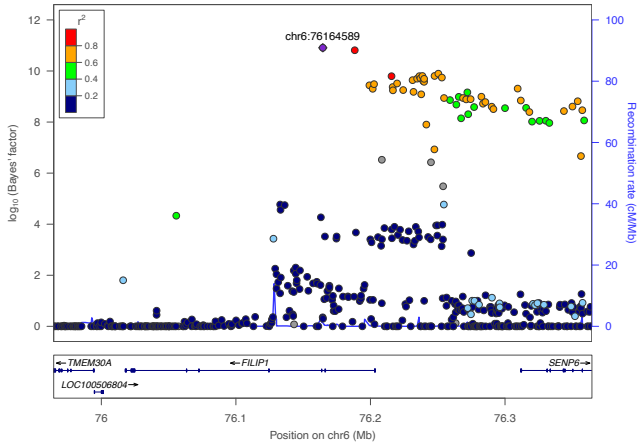


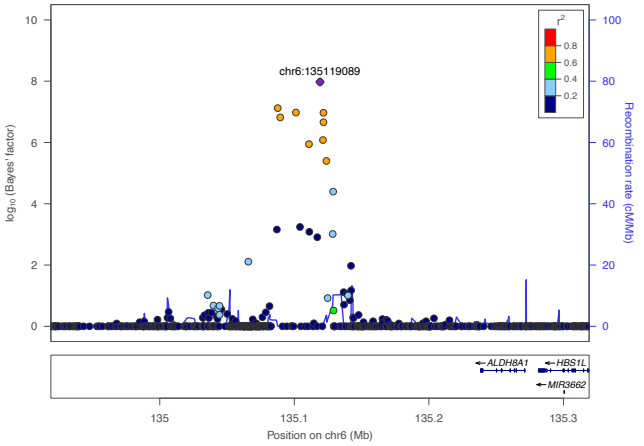


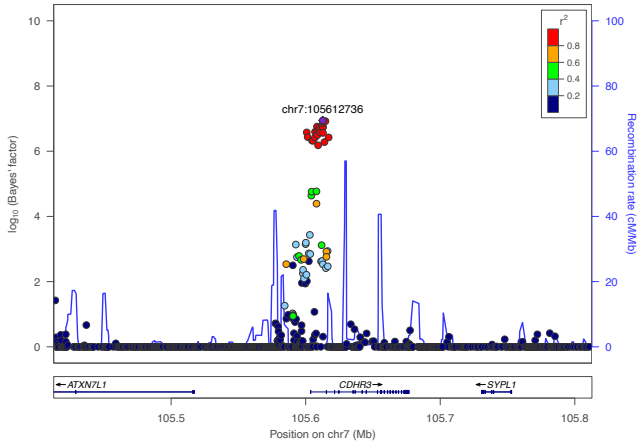


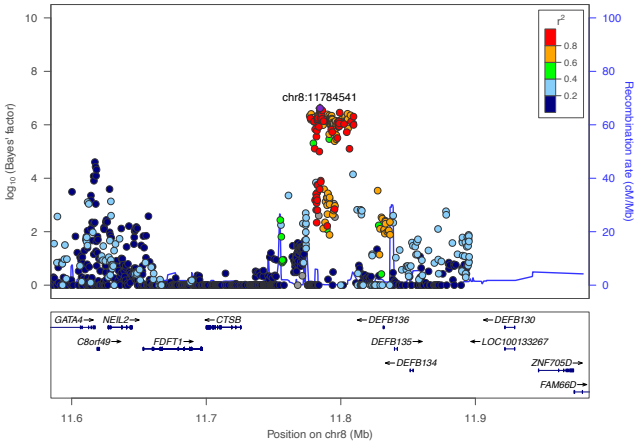


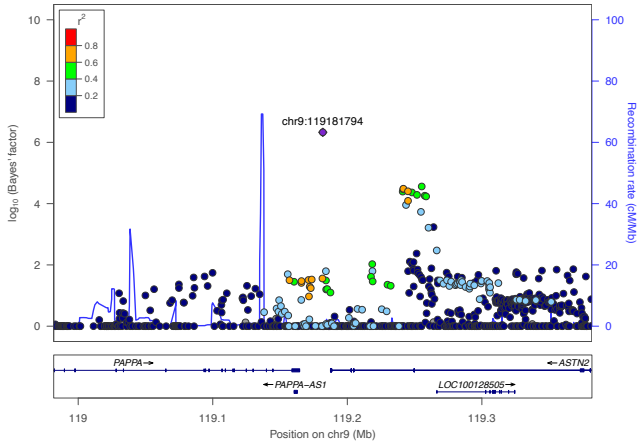


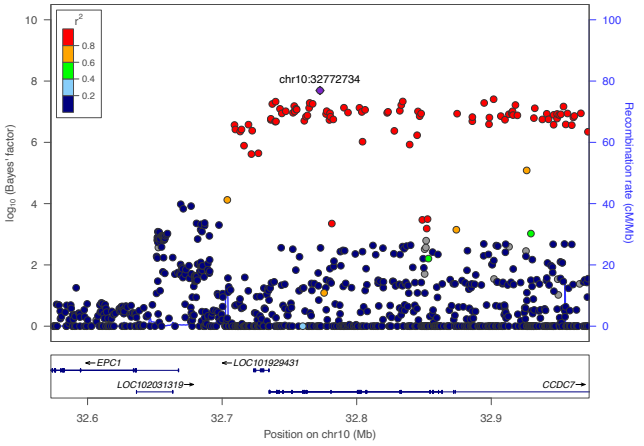

chr6:76164589
log10 (Bayes' factor)
Recombination rate (cM/Mb)
r2
0.8
0.6
0.4
0.2
TMEM30A
FILIP1
SENP6
LOC100506804
76
76.1
76.2
76.3
Position on chr6 (Mb)

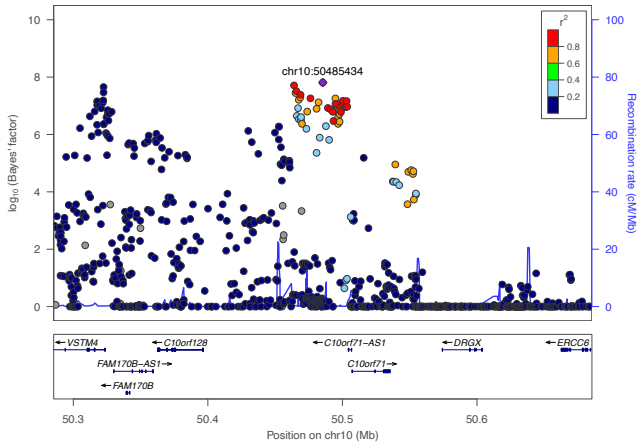

chr6:135119089
log10 (Bayes' factor)
Recombination rate (cM/Mb)
r2
0.8
0.6
0.4
0.2
ALDH8A1
HBS1L
MIR3662
135
135.1
135.2
135.3
Position on chr6 (Mb)

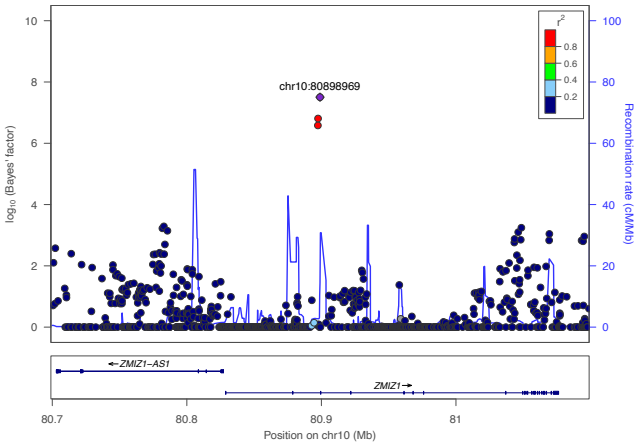

chr7:105612736
log10 (Bayes' factor)
Recombination rate (cM/Mb)
r2
0.8
0.6
0.4
0.2
ATXN7L1
CDHR3
SYPL1
105.5
105.6
105.7
105.8
Position on chr7 (Mb)

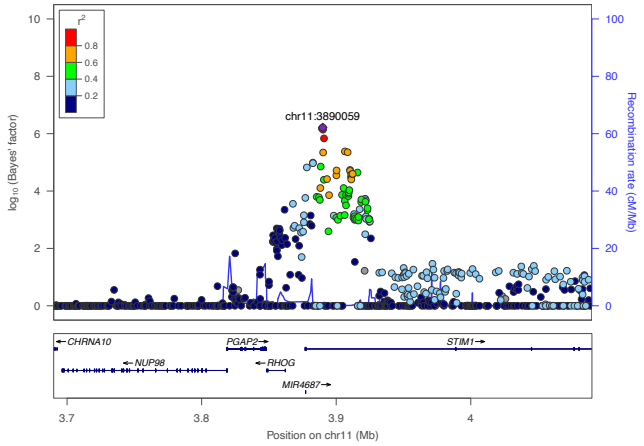

chr8:11784541
log10 (Bayes' factor)
Recombination rate (cM/Mb)
r2
0.8
0.6
0.4
0.2
GATA4
NEIL2
CTSB
DEFB136
DEFB130
C8orf49
FDFT1
DEFB135
LOC100133267
DEFB134
ZNF705D
FAM66D
11.6
11.7
11.8
11.9
Position on chr8 (Mb)

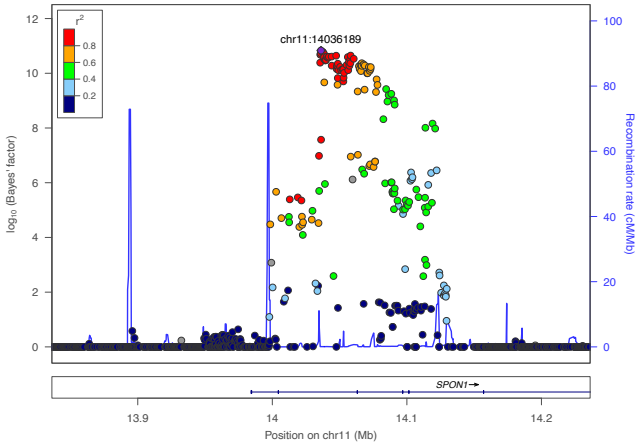

chr9:119181794
log10 (Bayes' factor)
Recombination rate (cM/Mb)
r2
0.8
0.6
0.4
0.2
PAPPA
ASTN2
PAPPA-AS1
LOC100128505
119
119.1
119.2
119.3
Position on chr9 (Mb)

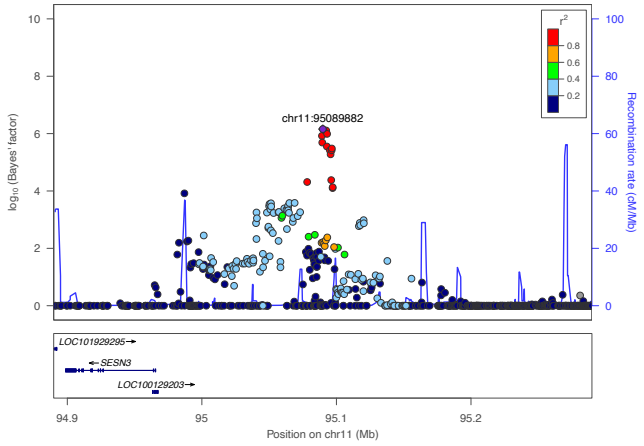

chr10:32772734
log10 (Bayes' factor)
Recombination rate (cM/Mb)
r2
0.8
0.6
0.4
0.2
EPC1
LOC101929431
LOC102031319
CCDC7
32.6
32.7
32.8
32.9
Position on chr10 (Mb)

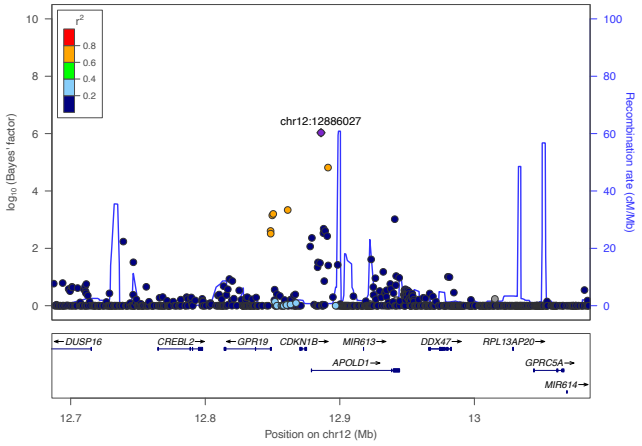

chr10:50485434
log10 (Bayes' factor)
Recombination rate (cM/Mb)
r2
0.8
0.6
0.4
0.2
VSTM4
C10orf128
C10orf71-AS1
DRGX
ERCC6
FAM170B-AS1
C10orf71
FAM170B
50.3
50.4
50.5
50.6
Position on chr10 (Mb)

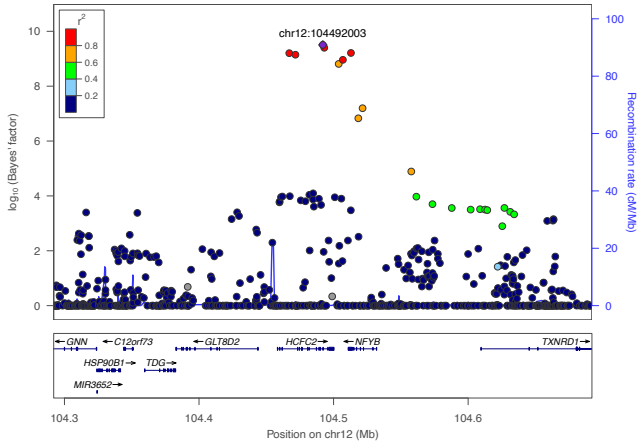

chr10:80898969
log10 (Bayes' factor)
Recombination rate (cM/Mb)
r2
0.8
0.6
0.4
0.2
ZMIZ1-AS1
ZMIZ1
80.7
80.8
80.9
81
Position on chr10 (Mb)

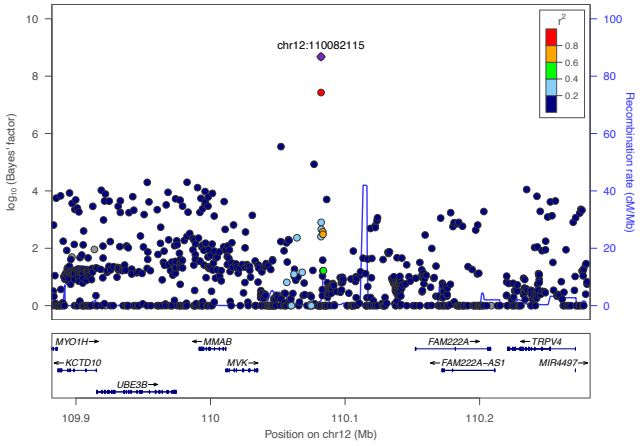

chr11:3890059
log10 (Bayes' factor)
Recombination rate (cM/Mb)
r2
0.8
0.6
0.4
0.2
CHRNA10
PGAP2
STIM1
NUP98
RHOG
MIR4687
Position on chr11 (Mb)

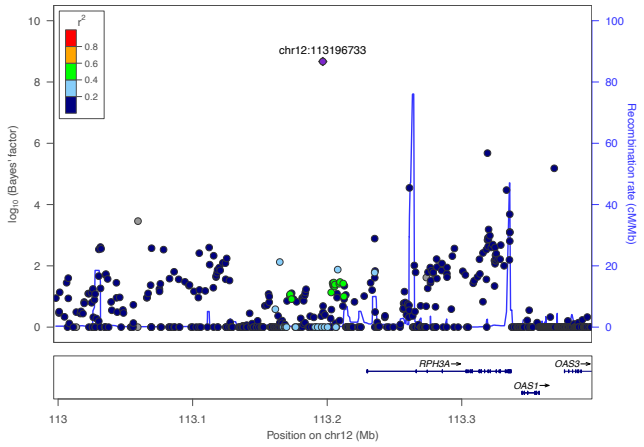

chr11:14036189
log10 (Bayes' factor)
Recombination rate (cM/Mb)
r2
0.8
0.6
0.4
0.2
SPON1
Position on chr11 (Mb)

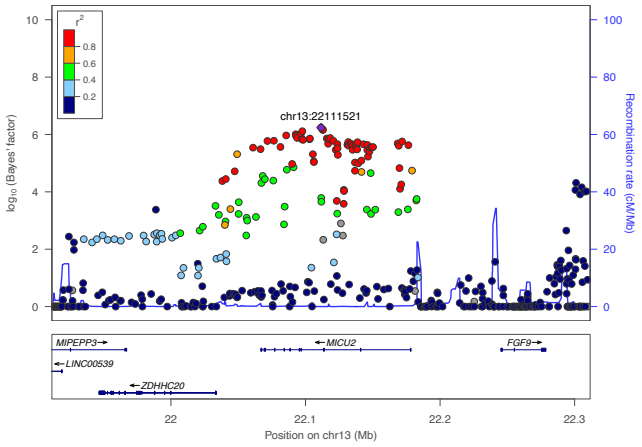

chr11:95089882
log10 (Bayes' factor)
Recombination rate (cM/Mb)
r2
0.8
0.6
0.4
0.2
LOC101929295
SESN3
LOC100129203
Position on chr11 (Mb)

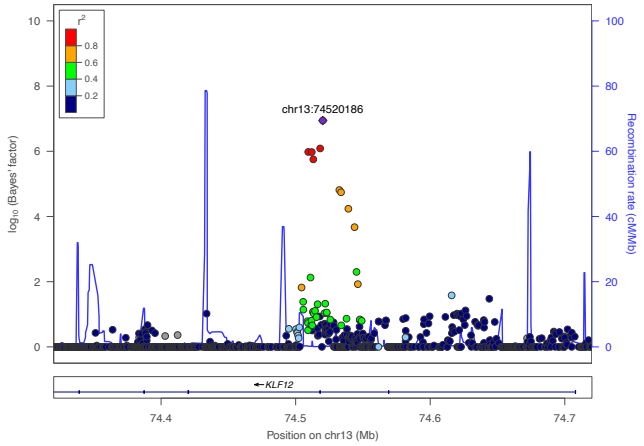

chr12:12886027
log10 (Bayes' factor)
Recombination rate (cM/Mb)
r2
0.8
0.6
0.4
0.2
DUSP16
CREBL2
GPR19
CDKN1B
MIR613
DDX47
RPL13AP20
APOLD1
GPRC5A
MIR614
Position on chr12 (Mb)

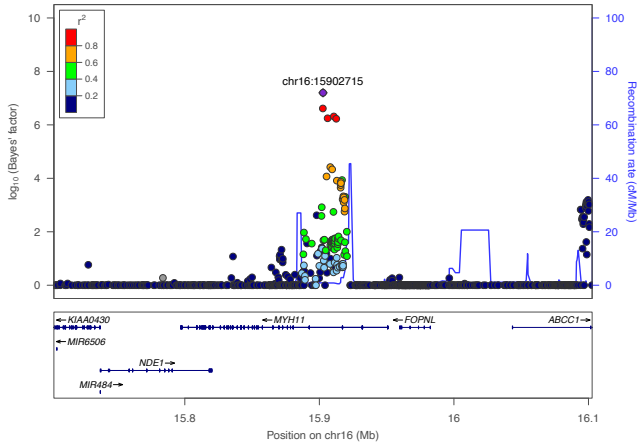

chr12:104492003
log10 (Bayes' factor)
Recombination rate (cM/Mb)
r2
0.8
0.6
0.4
0.2
GNN
C12orf73
GLT8D2
HCFC2
NFYB
TXNRD1
HSP90B1
TDG
MIR3652
Position on chr12 (Mb)

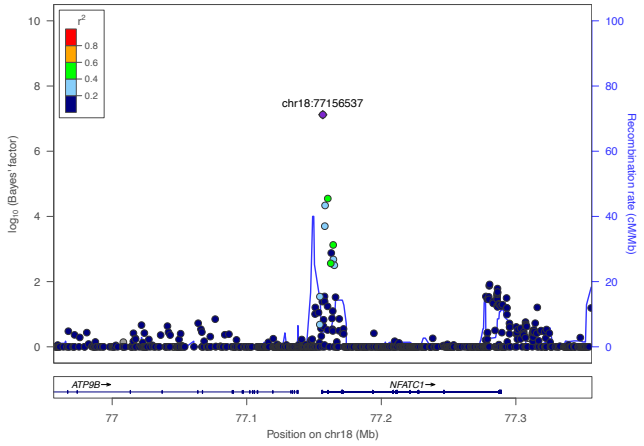

chr12:110082115
log10 (Bayes' factor)
Recombination rate (cM/Mb)
r2
0.8
0.6
0.4
0.2
MYO1H
MMAB
FAM222A
TRPV4
KCTD10
MVK
FAM222A-AS1
MIR4497
UBE3B
Position on chr12 (Mb)

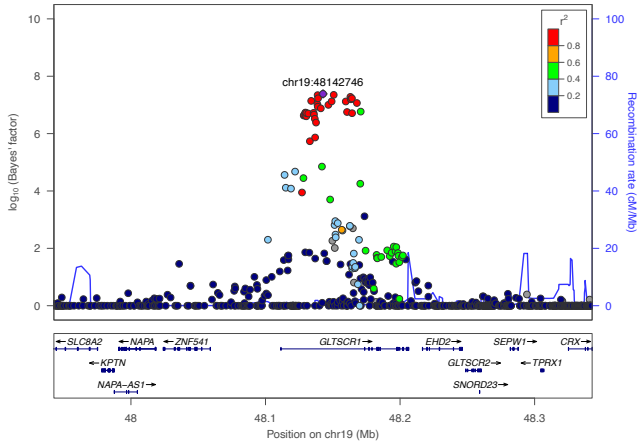

chr12:113196733
log10 (Bayes' factor)
Recombination rate (cM/Mb)
r2
0.8
0.6
0.4
0.2
RPH3A
OAS3
OAS1
Position on chr12 (Mb)

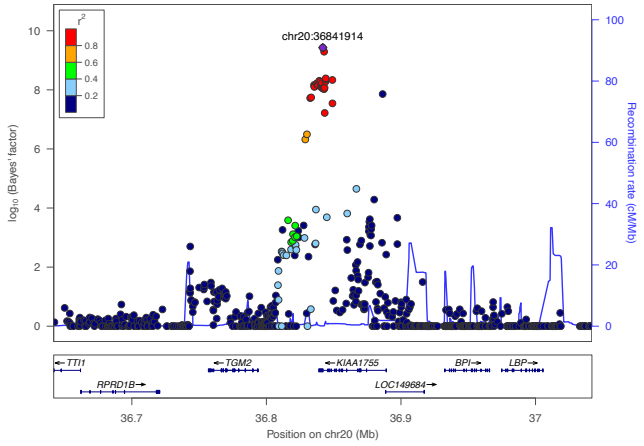

chr13:22111521
log10 (Bayes' factor)
Recombination rate (cM/Mb)
r2
0.8
0.6
0.4
0.2
MIPEPP3
MICU2
FGF9
LINC00539
ZDHHC20
Position on chr13 (Mb)

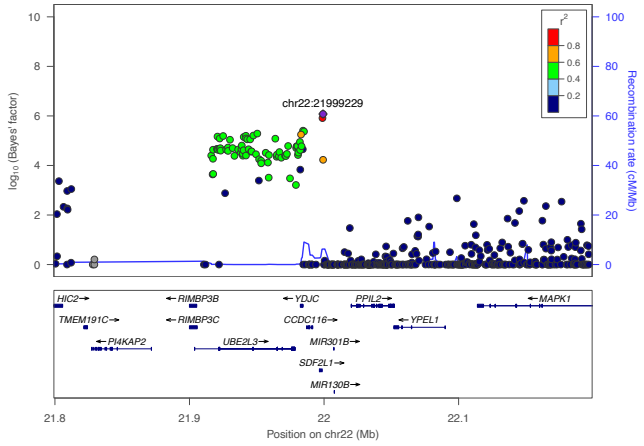

chr13:74520186
KLF12
Position on chr13 (Mb)
log10 (Bayes' factor)
Recombination rate (cM/Mb)

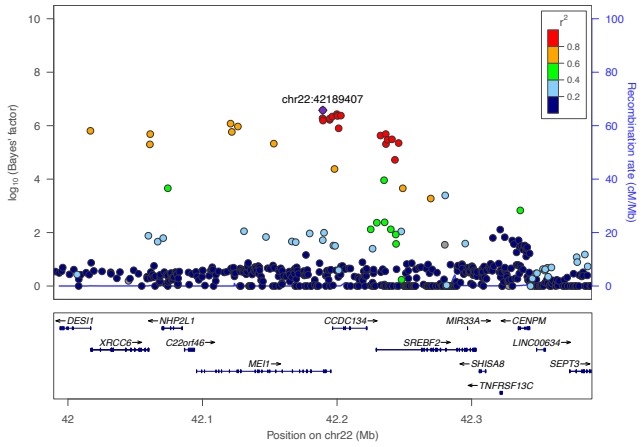

chr16:15902715
KIAA0430
MYH11
FOPNL
ABCC1
MIR6506
NDE1
MIR484
Position on chr16 (Mb)
log10 (Bayes' factor)
Recombination rate (cM/Mb)

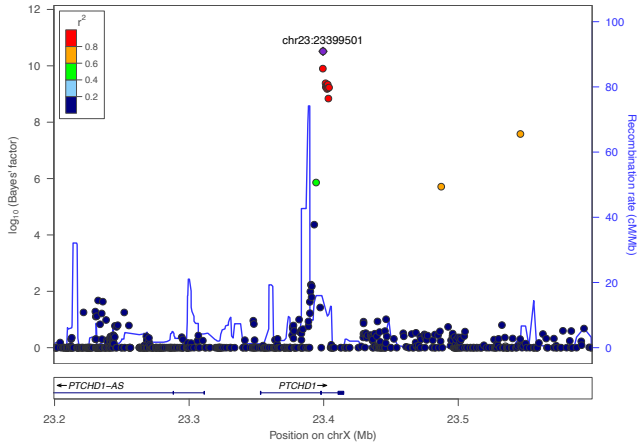

chr18:77156537
ATP9B
NFATC1
Position on chr18 (Mb)
log10 (Bayes' factor)
Recombination rate (cM/Mb)

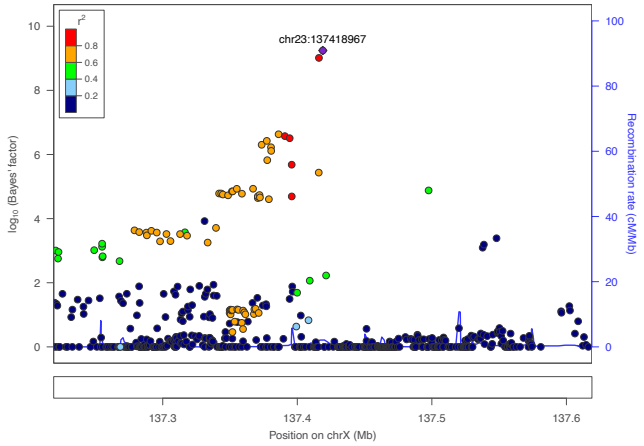

chr19:48142746
SLC8A2
NAPA
ZNF541
GLTSCR1
EHD2
SEPW1
CRX
KPTN
GLTSCR2
TPRX1
NAPA-AS1
SNORD23
Position on chr19 (Mb)
log10 (Bayes' factor)
Recombination rate (cM/Mb)


chr20:36841914
TTI1
TGM2
KIAA1755
BPI
LBP
RPRD1B
LOC149684
Position on chr20 (Mb)
log10 (Bayes' factor)
Recombination rate (cM/Mb)

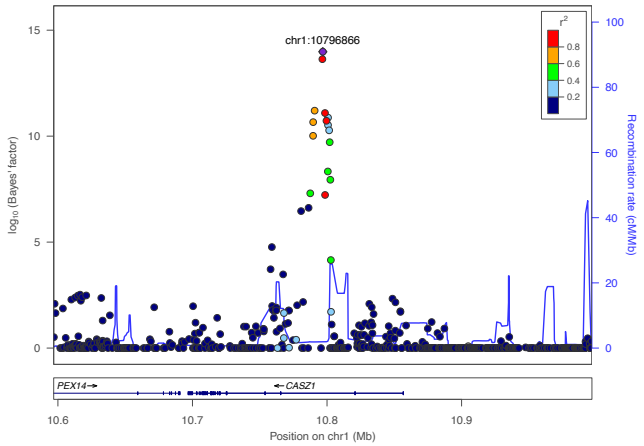

chr22:21999229
HIC2
RIMBP3B
YDJC
PPIL2
MAPK1
TMEM191C
RIMBP3C
CCDC116
YPEL1
PI4KAP2
UBE2L3
MIR301B
SDF2L1
MIR130B
Position on chr22 (Mb)
log10 (Bayes' factor)
Recombination rate (cM/Mb)

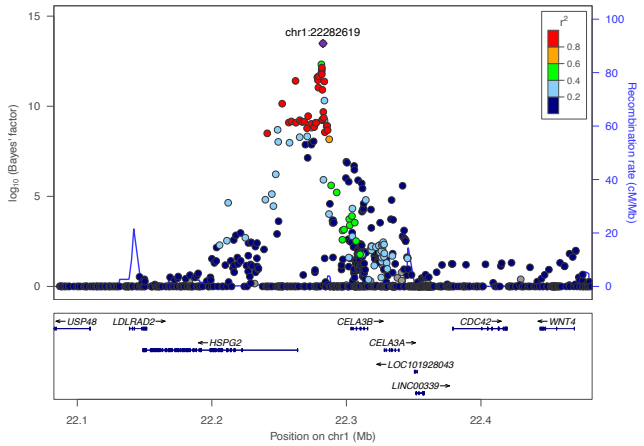

chr22:42189407
DESI1
NHP2L1
CCDC134
MIR33A
CENPM
XRCC6
C22orf46
SREBF2
LINC00634
MEI1
SHISA8
SEPT3
TNFRSF13C
Position on chr22 (Mb)
log10 (Bayes' factor)
Recombination rate (cM/Mb)

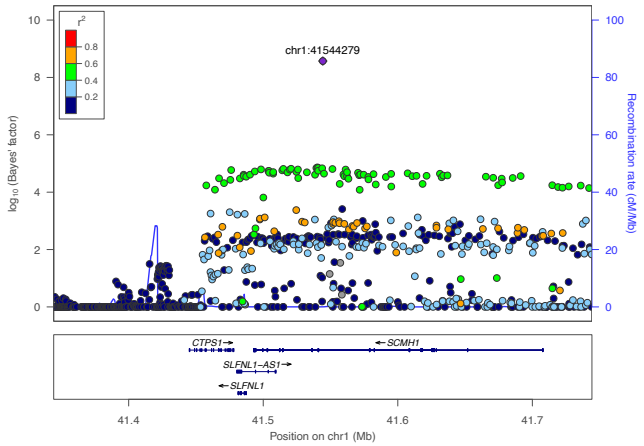

chr23:23399501
PTCHD1-AS
PTCHD1
Position on chrX (Mb)
log10 (Bayes' factor)
Recombination rate (cM/Mb)

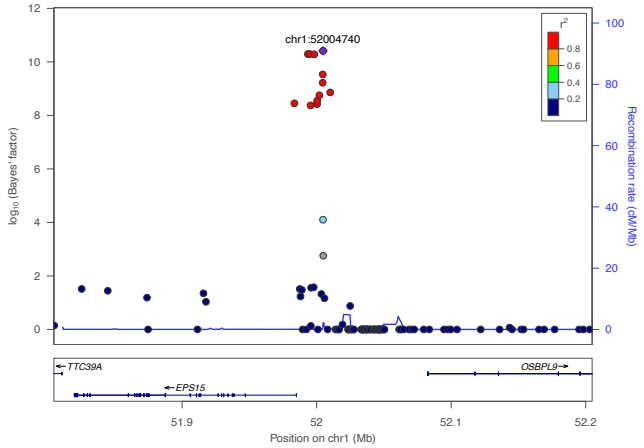

chr23:137418967
r²
0.8
0.6
0.4
0.2
log10 (Bayes' factor)
Recombination rate (cM/Mb)
Position on chrX (Mb)
137.3
137.4
137.5
137.6

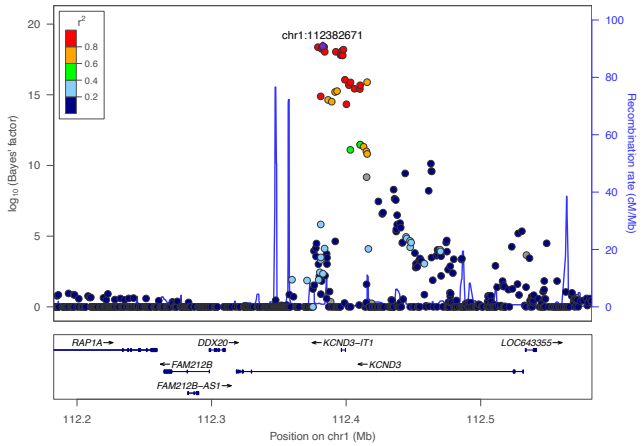
4. Previously reported 127 loci in the trans-ancestry meta-GWAS

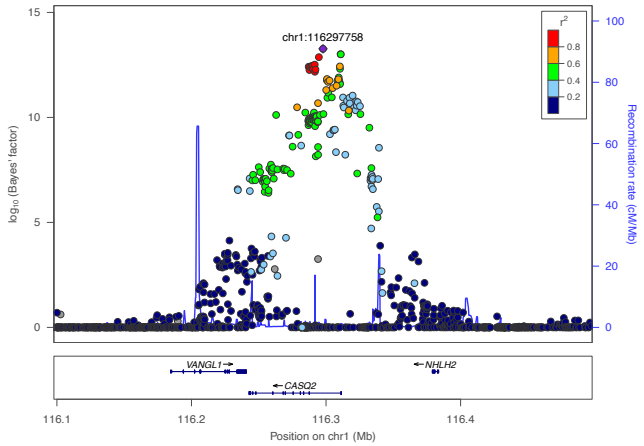


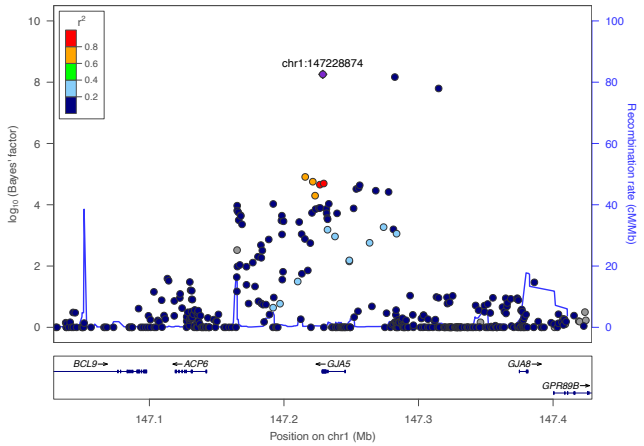


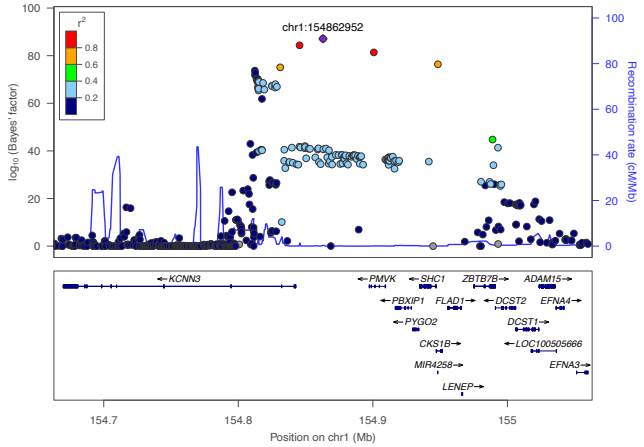


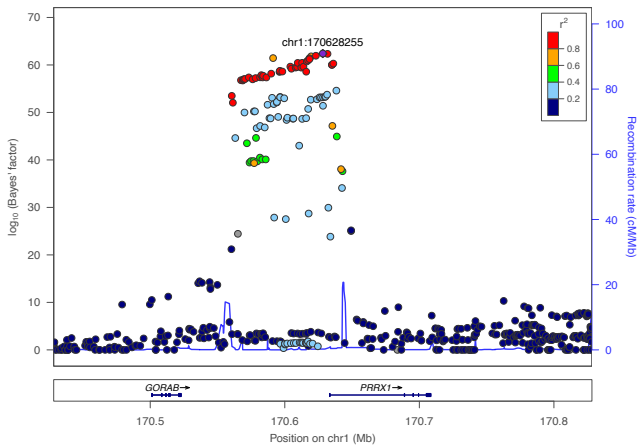


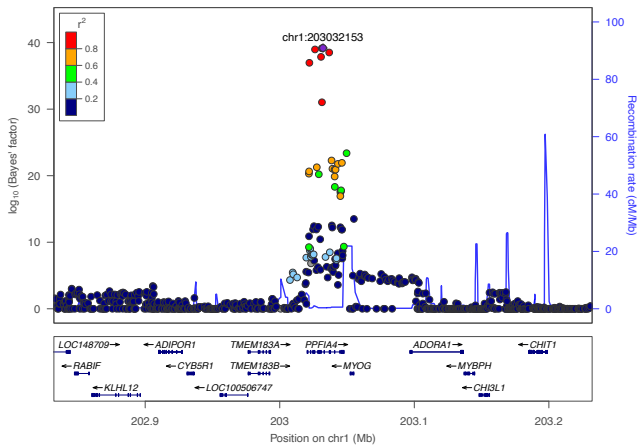


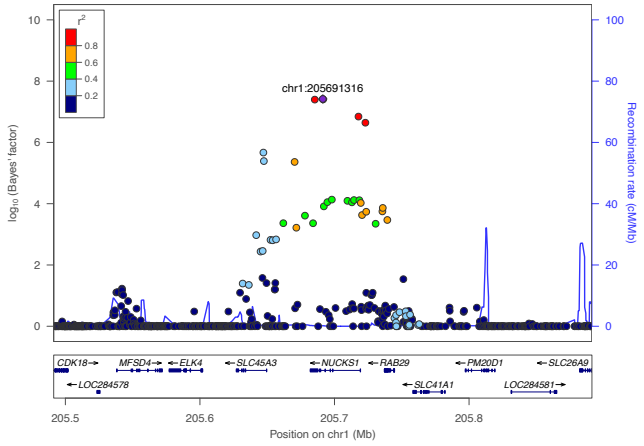


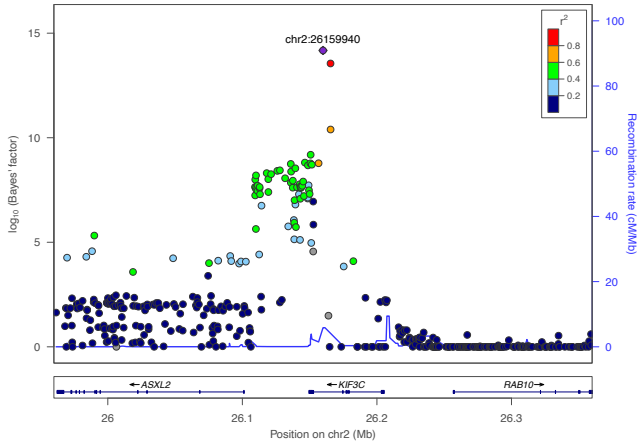


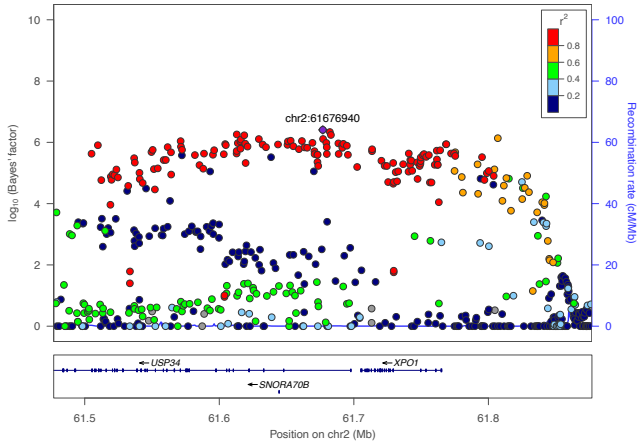


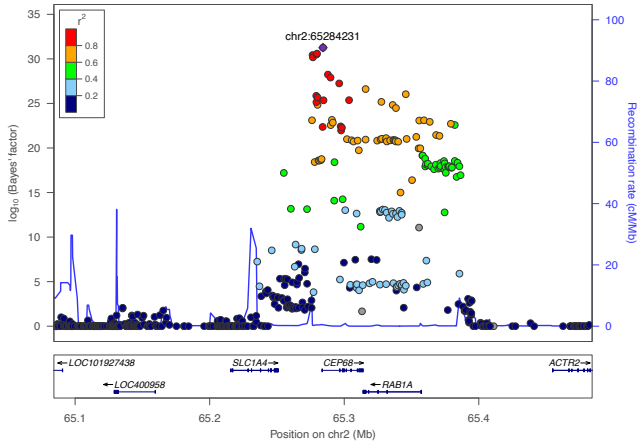

chr1:170628255
log10 (Bayes' factor)
Recombination rate (cM/Mb)
r2
0.8
0.6
0.4
0.2
GORAB
PRRX1
170.5
170.6
170.7
170.8
Position on chr1 (Mb)

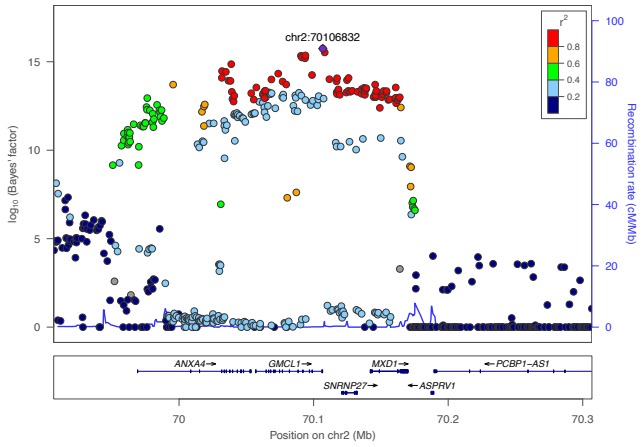

chr1:203032153
log10 (Bayes' factor)
Recombination rate (cM/Mb)
r2
0.8
0.6
0.4
0.2
LOC148709
ADIPOR1
TMEM183A
PPFIA4
ADORA1
CHIT1
RABIF
CYB5R1
TMEM183B
MYOG
MYBPH
KLHL12
LOC100506747
CHI3L1
202.9
203
203.1
203.2
Position on chr1 (Mb)

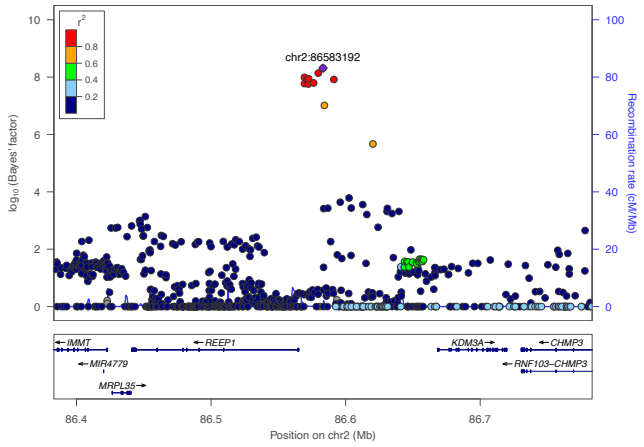

chr1:205691316
log10 (Bayes' factor)
Recombination rate (cM/Mb)
r2
0.8
0.6
0.4
0.2
CDK18
MFSD4
ELK4
SLC45A3
NUCKS1
RAB29
PM20D1
SLC26A9
LOC284578
SLC41A1
LOC284581
205.5
205.6
205.7
205.8
Position on chr1 (Mb)

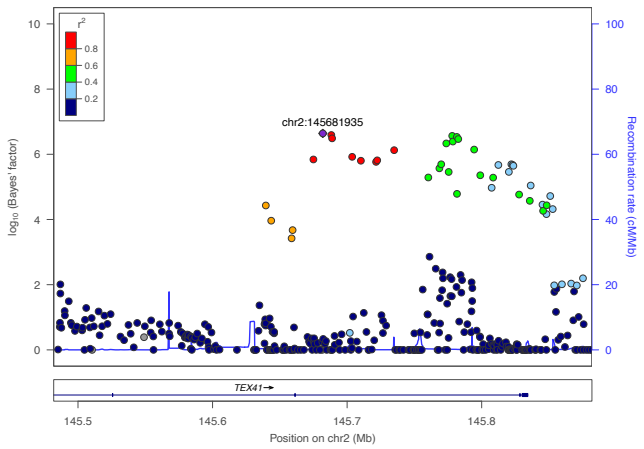

chr2:26159940
log10 (Bayes' factor)
Recombination rate (cM/Mb)
r2
0.8
0.6
0.4
0.2
ASXL2
KIF3C
RAB10
26
26.1
26.2
26.3
Position on chr2 (Mb)

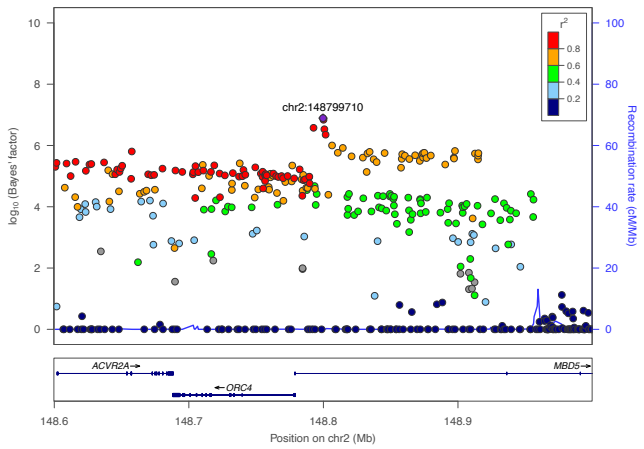

chr2:61676940
log10 (Bayes' factor)
Recombination rate (cM/Mb)
r2
0.8
0.6
0.4
0.2
USP34
XPO1
SNORA70B
61.5
61.6
61.7
61.8
Position on chr2 (Mb)

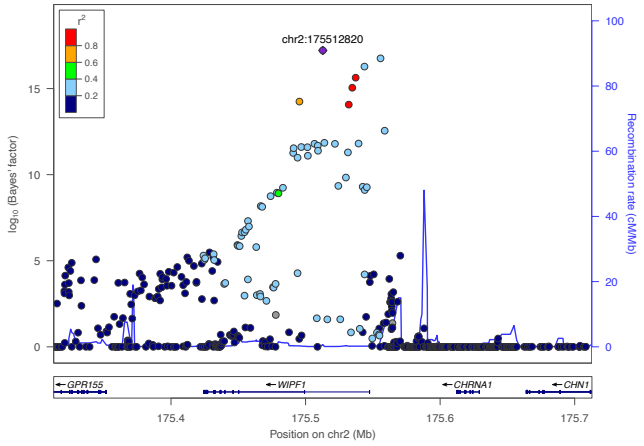

chr2:65284231
log10 (Bayes' factor)
Recombination rate (cM/Mb)
r2
0.8
0.6
0.4
0.2
LOC101927438
SLC1A4
CEP68
ACTR2
LOC400958
RAB1A
65.1
65.2
65.3
65.4
Position on chr2 (Mb)

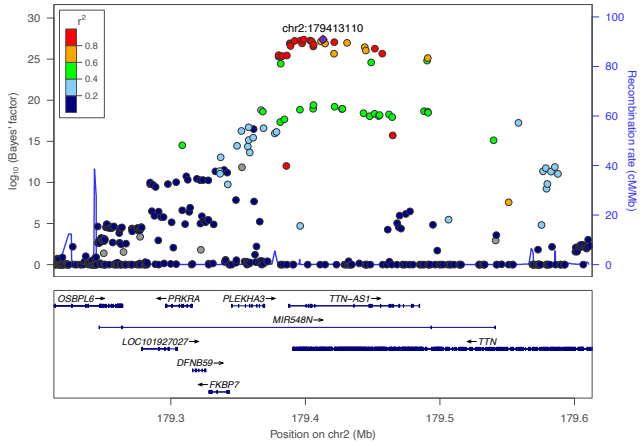

chr2:70106832
log10 (Bayes' factor)
Recombination rate (cM/Mb)
r2
0.8
0.6
0.4
0.2
ANXA4
GMCL1
MXD1
PCBP1-AS1
SNRNP27
ASPRV1
70
70.1
70.2
70.3
Position on chr2 (Mb)

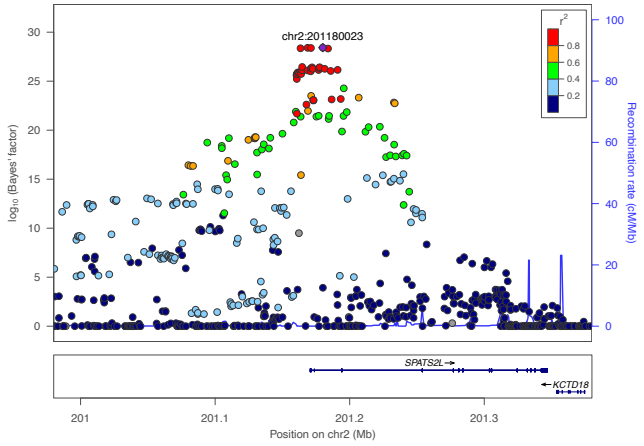

chr2:86583192
log10 (Bayes' factor)
Recombination rate (cM/Mb)
r2
0.8
0.6
0.4
0.2
IMMT
REEP1
KDM3A
CHMP3
MIR4779
RNF103-CHMP3
MRPL35
86.4
86.5
86.6
86.7
Position on chr2 (Mb)

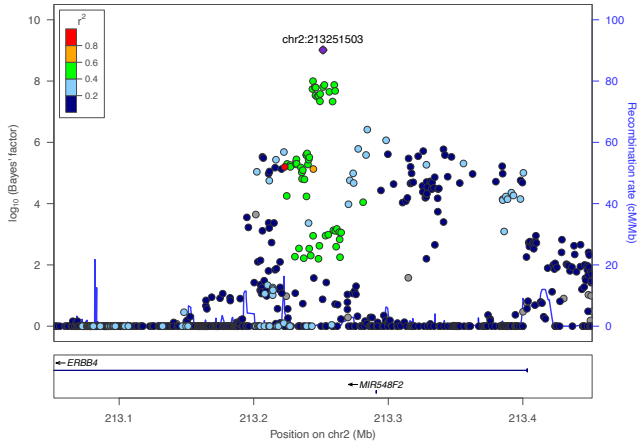

chr2:145681935
log10 (Bayes' factor)
Recombination rate (cM/Mb)
r2
0.8
0.6
0.4
0.2
TEX41
145.5
145.6
145.7
145.8
Position on chr2 (Mb)

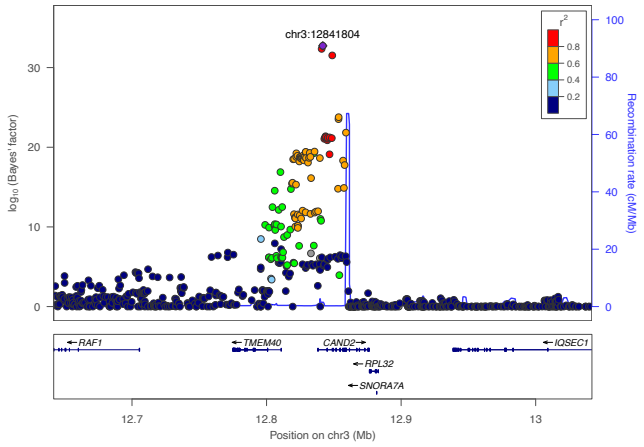

chr2:148799710
log10 (Bayes' factor)
Recombination rate (cM/Mb)
r2
0.8
0.6
0.4
0.2
ACVR2A
ORC4
MBD5
148.6
148.7
148.8
148.9
Position on chr2 (Mb)

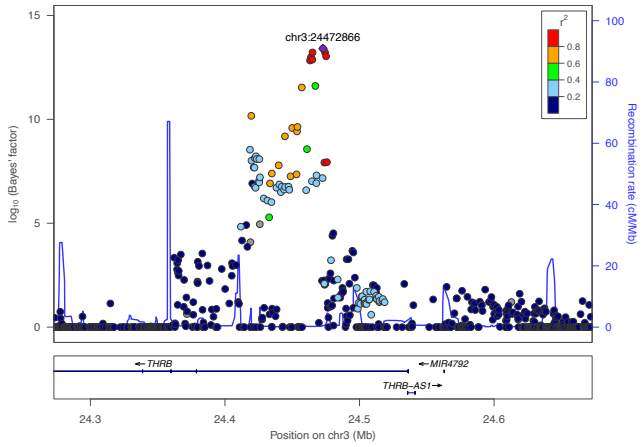

chr2:175512820
log10 (Bayes' factor)
Recombination rate (cM/Mb)
r2
0.8
0.6
0.4
0.2
GPR155
WIPF1
CHRNA1
CHN1
175.4
175.5
175.6
175.7
Position on chr2 (Mb)

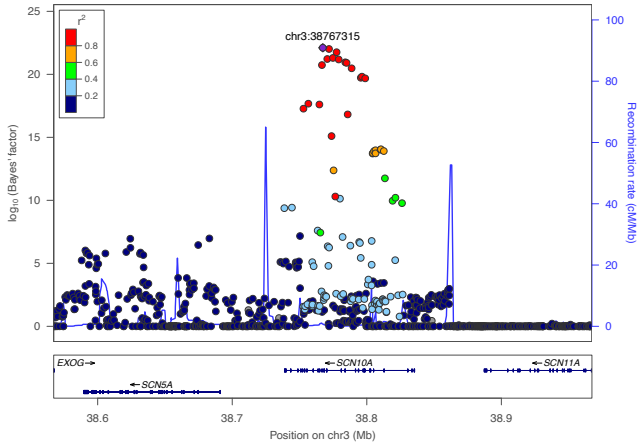

chr2:179413110
log10 (Bayes' factor)
Recombination rate (cM/Mb)
r2
0.8
0.6
0.4
0.2
OSBPL6
PRKRA
PLEKHA3
TTN-AS1
MIR548N
LOC101927027
TTN
DFNB59
FKBP7
179.3
179.4
179.5
179.6
Position on chr2 (Mb)

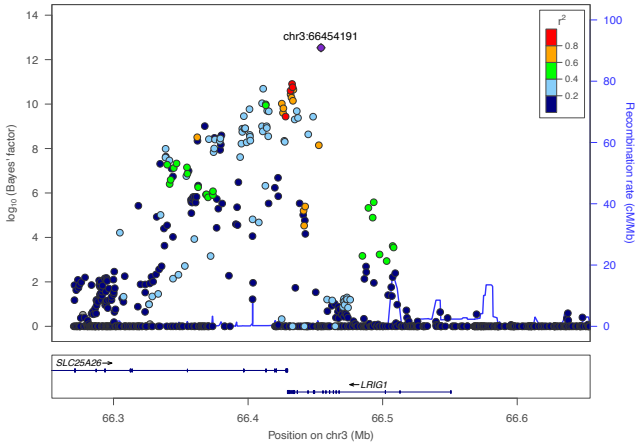

chr2:201180023
log10 (Bayes' factor)
Recombination rate (cM/Mb)
r2
0.8
0.6
0.4
0.2
SPATS2L
KCTD18
201
201.1
201.2
201.3
Position on chr2 (Mb)

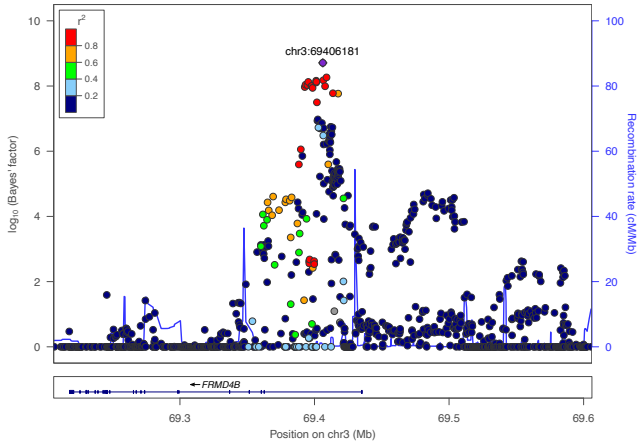

chr2:213251503
log10 (Bayes' factor)
Recombination rate (cM/Mb)
r2
0.8
0.6
0.4
0.2
ERBB4
MIR548F2
213.1
213.2
213.3
213.4
Position on chr2 (Mb)

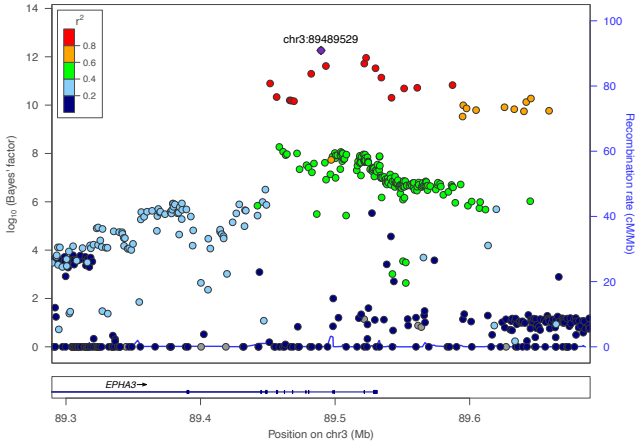

chr3:12841804
log10 (Bayes' factor)
Recombination rate (cM/Mb)
r2
0.8
0.6
0.4
0.2
RAF1
TMEM40
CAND2
IQSEC1
RPL32
SNORA7A
12.7
12.8
12.9
13
Position on chr3 (Mb)

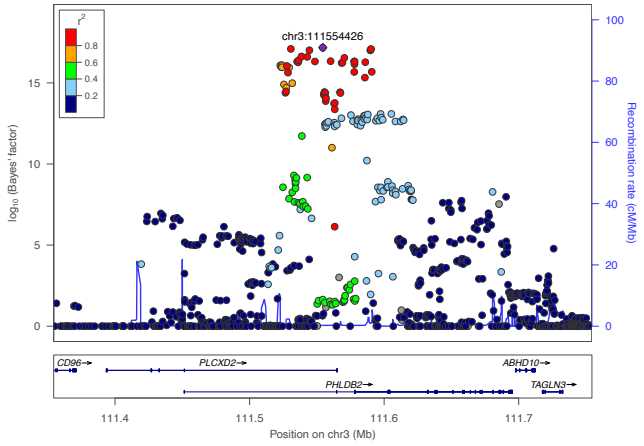

chr3:24472866
log10 (Bayes' factor)
Recombination rate (cM/Mb)
r2
0.8
0.6
0.4
0.2
THRB
MIR4792
THRB-AS1
24.3
24.4
24.5
24.6
Position on chr3 (Mb)

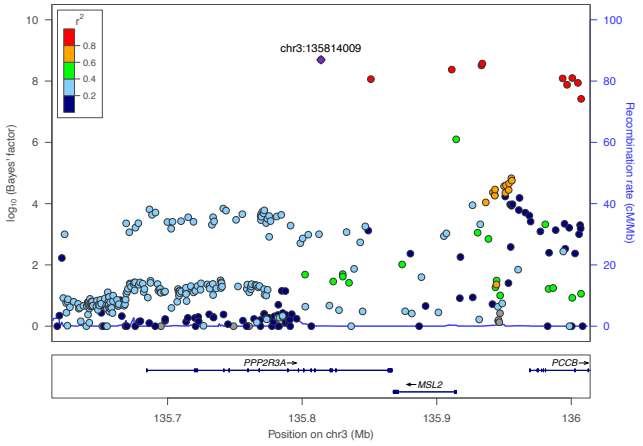

chr3:38767315
r²
0.8
0.6
0.4
0.2
log10 (Bayes' factor)
Recombination rate (cM/Mb)
EXOG
SCN10A
SCN11A
SCN5A
38.6
38.7
38.8
38.9
Position on chr3 (Mb)

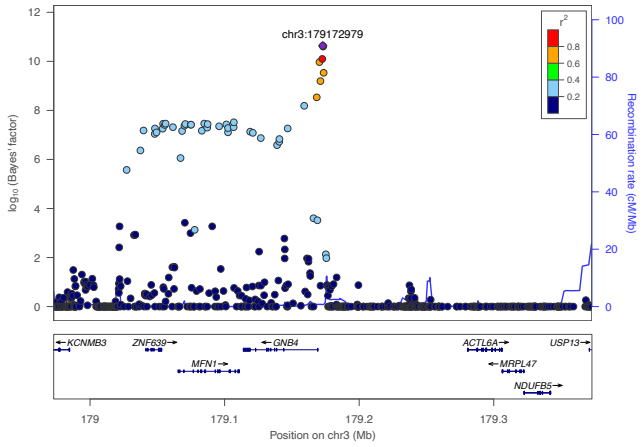

chr3:66454191
r²
0.8
0.6
0.4
0.2
log10 (Bayes' factor)
Recombination rate (cM/Mb)
SLC25A26
LRIG1
66.3
66.4
66.5
66.6
Position on chr3 (Mb)

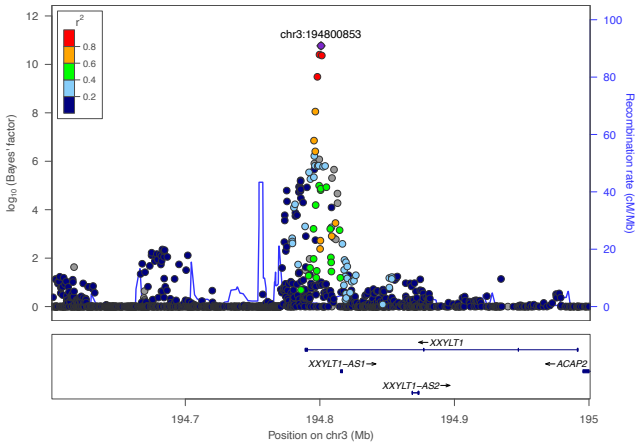

chr3:69406181
r²
0.8
0.6
0.4
0.2
log10 (Bayes' factor)
Recombination rate (cM/Mb)
FRMD4B
69.3
69.4
69.5
69.6
Position on chr3 (Mb)

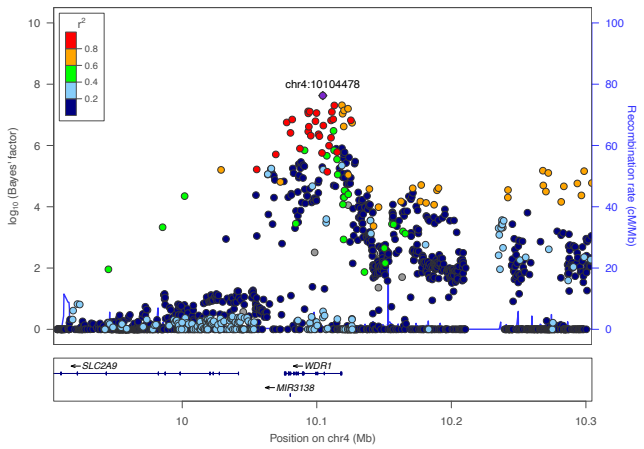

chr3:89489529
r²
0.8
0.6
0.4
0.2
log10 (Bayes' factor)
Recombination rate (cM/Mb)
EPHA3
89.3
89.4
89.5
89.6
Position on chr3 (Mb)

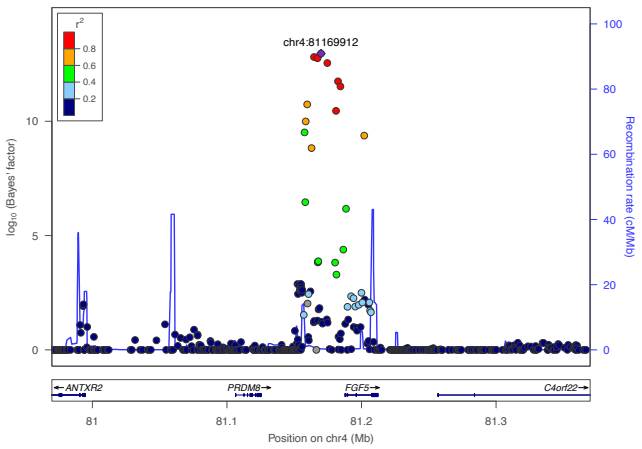

chr3:111554426
r²
0.8
0.6
0.4
0.2
log10 (Bayes' factor)
Recombination rate (cM/Mb)
CD96
PLCXD2
ABHD10
PHLDB2
TAGLN3
111.4
111.5
111.6
111.7
Position on chr3 (Mb)

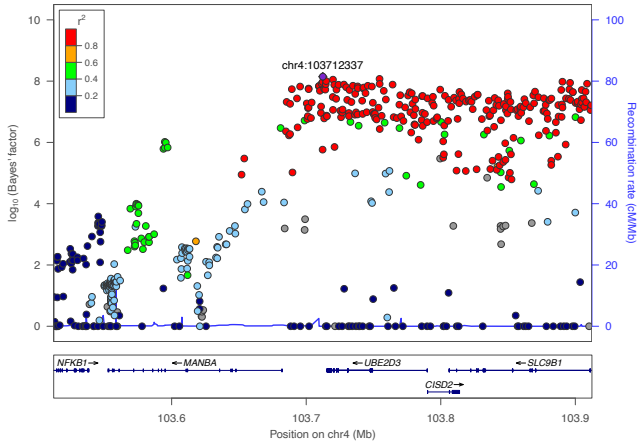

chr3:135814009
r²
0.8
0.6
0.4
0.2
log10 (Bayes' factor)
Recombination rate (cM/Mb)
PPP2R3A
PCCB
MSL2
135.7
135.8
135.9
136
Position on chr3 (Mb)

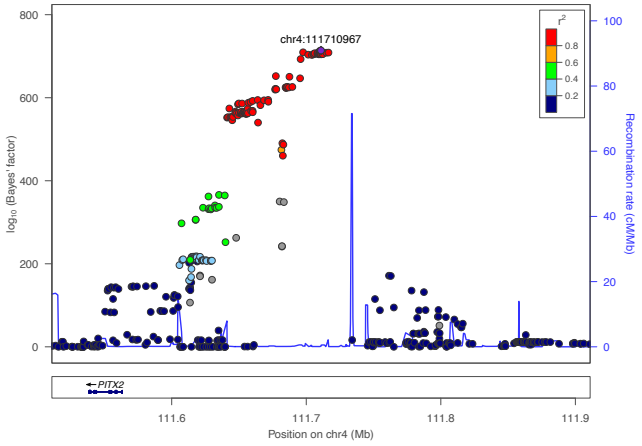

chr3:179172979
r²
0.8
0.6
0.4
0.2
log10 (Bayes' factor)
Recombination rate (cM/Mb)
KCNMB3
ZNF639
GNB4
ACTL6A
USP13
MFN1
MRPL47
NDUFB5
179
179.1
179.2
179.3
Position on chr3 (Mb)

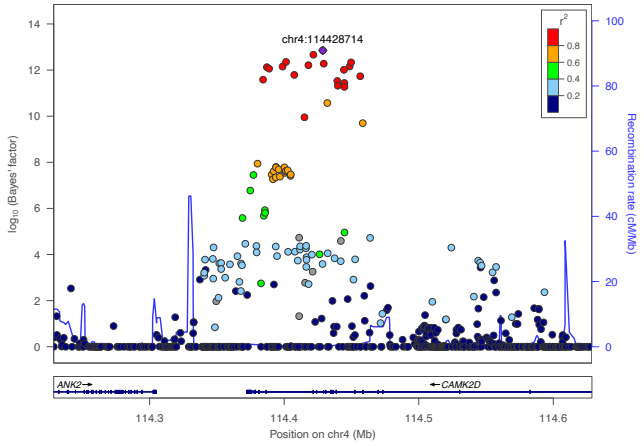

chr3:194800853
r²
0.8
0.6
0.4
0.2
log10 (Bayes' factor)
Recombination rate (cM/Mb)
XXYLT1
XXYLT1-AS1
ACAP2
XXYLT1-AS2
194.7
194.8
194.9
195
Position on chr3 (Mb)

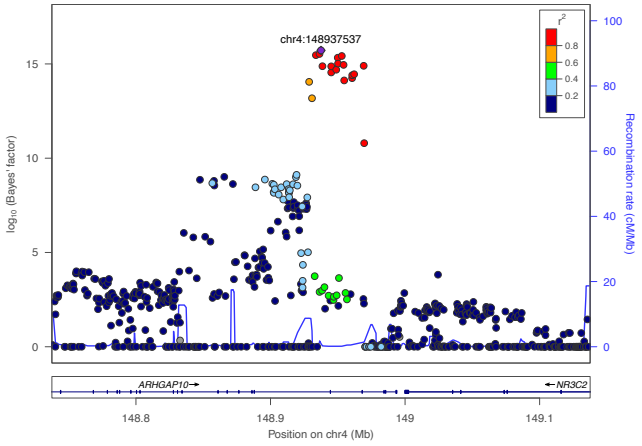

chr4:10104478
r²
0.8
0.6
0.4
0.2
log10 (Bayes' factor)
Recombination rate (cM/Mb)
SLC2A9
WDR1
MIR3138
10
10.1
10.2
10.3
Position on chr4 (Mb)

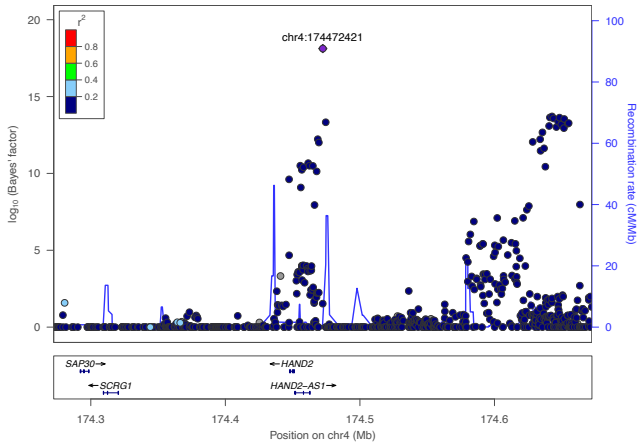

chr4:81169912
r²
0.8
0.6
0.4
0.2
log10 (Bayes' factor)
Recombination rate (cM/Mb)
ANTXR2
PRDM8
FGF5
C4orf22
81
81.1
81.2
81.3
Position on chr4 (Mb)

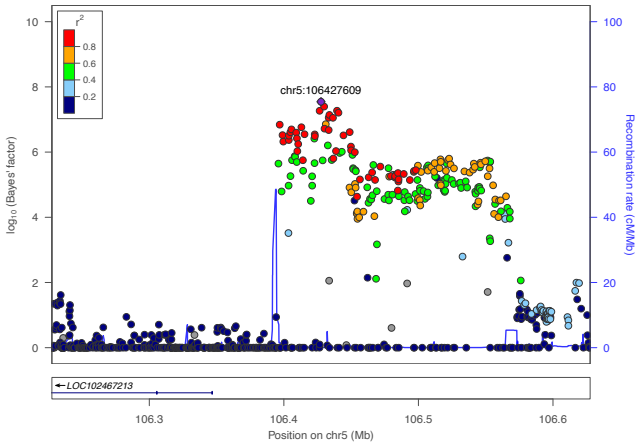

chr4:103712337
r²
0.8
0.6
0.4
0.2
log10 (Bayes' factor)
Recombination rate (cM/Mb)
NFKB1
MANBA
UBE2D3
SLC9B1
CISD2
103.6
103.7
103.8
103.9
Position on chr4 (Mb)

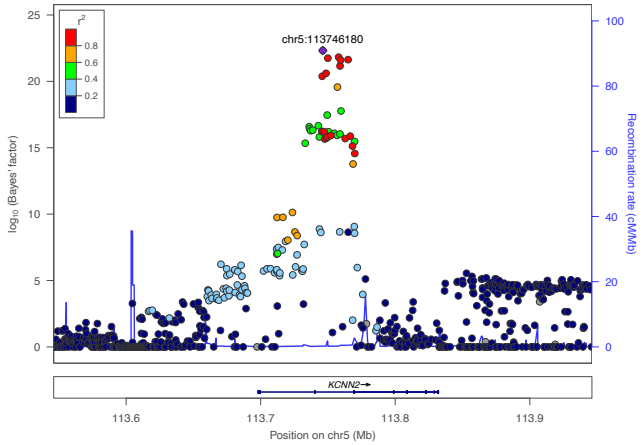

chr4:111710967
r²
0.8
0.6
0.4
0.2
log10 (Bayes' factor)
Recombination rate (cM/Mb)
PITX2
111.6
111.7
111.8
111.9
Position on chr4 (Mb)

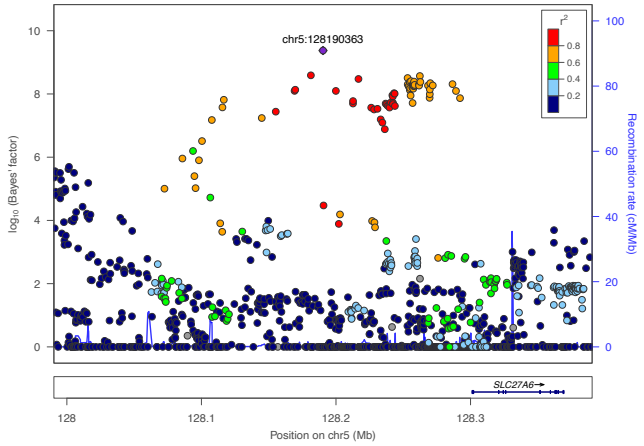

chr4:114428714
r²
0.8
0.6
0.4
0.2
log10 (Bayes' factor)
Recombination rate (cM/Mb)
ANK2
CAMK2D
114.3
114.4
114.5
114.6
Position on chr4 (Mb)

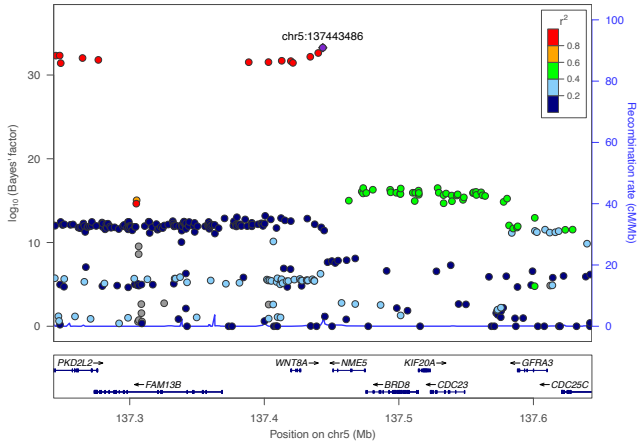

chr4:148937537
r²
0.8
0.6
0.4
0.2
log10 (Bayes' factor)
Recombination rate (cM/Mb)
ARHGAP10
NR3C2
148.8
148.9
149
149.1
Position on chr4 (Mb)

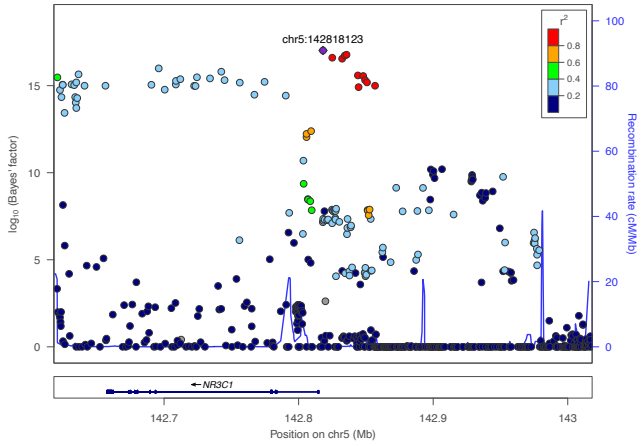

chr4:174472421
r²
0.8
0.6
0.4
0.2
log10 (Bayes' factor)
Recombination rate (cM/Mb)
SAP30
HAND2
SCRG1
HAND2-AS1
174.3
174.4
174.5
174.6
Position on chr4 (Mb)

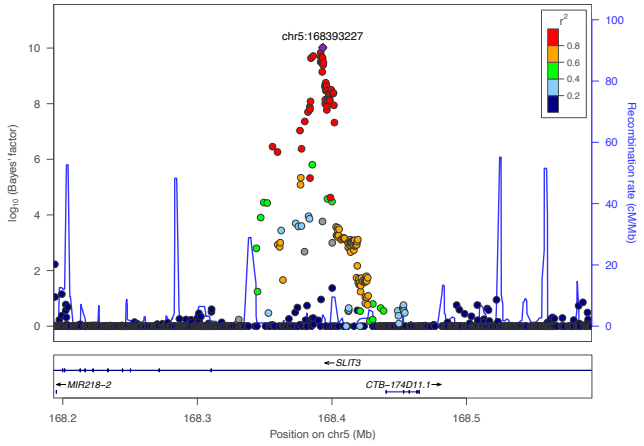

chr5:106427609
r²
0.8
0.6
0.4
0.2
log10 (Bayes' factor)
Recombination rate (cM/Mb)
LOC102467213
106.3
106.4
106.5
106.6
Position on chr5 (Mb)

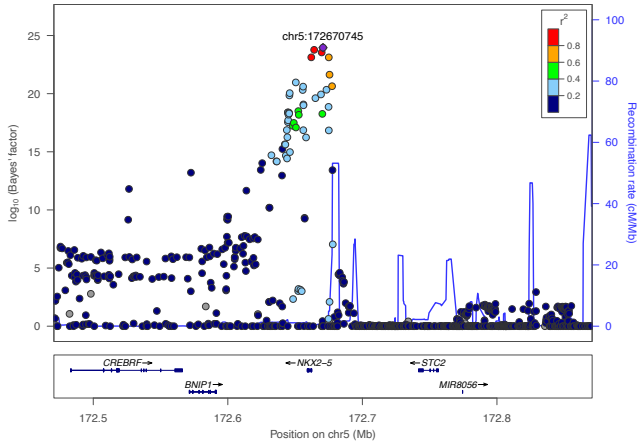

chr5:113746180
log10 (Bayes' factor)
Recombination rate (cM/Mb)
r2
0.8
0.6
0.4
0.2
KCNN2
113.6
113.7
113.8
113.9
Position on chr5 (Mb)

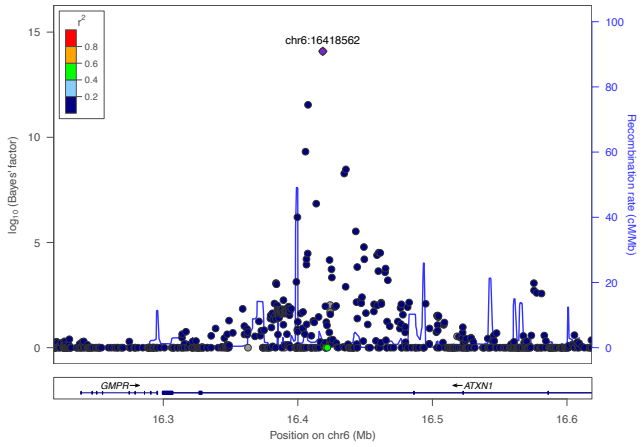

chr5:128190363
log10 (Bayes' factor)
Recombination rate (cM/Mb)
r2
0.8
0.6
0.4
0.2
SLC27A6
128
128.1
128.2
128.3
Position on chr5 (Mb)

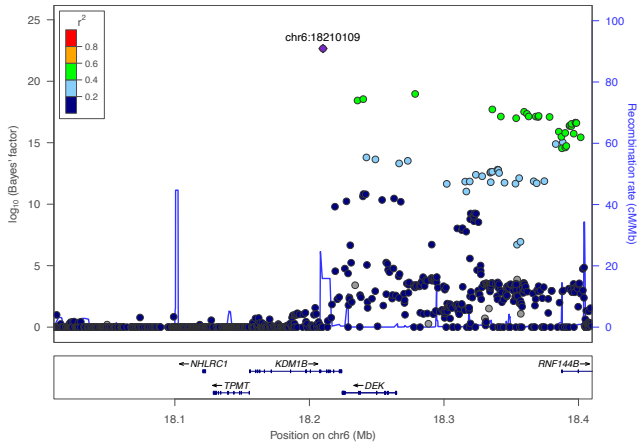

chr5:137443486
log10 (Bayes' factor)
Recombination rate (cM/Mb)
r2
0.8
0.6
0.4
0.2
PKD2L2
FAM13B
WNT8A
NME5
KIF20A
BRD8
CDC23
GFRA3
CDC25C
137.3
137.4
137.5
137.6
Position on chr5 (Mb)

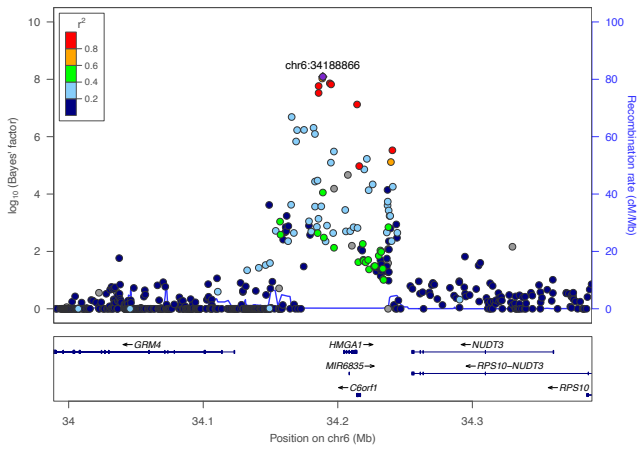

chr5:142818123
log10 (Bayes' factor)
Recombination rate (cM/Mb)
r2
0.8
0.6
0.4
0.2
NR3C1
142.7
142.8
142.9
143
Position on chr5 (Mb)

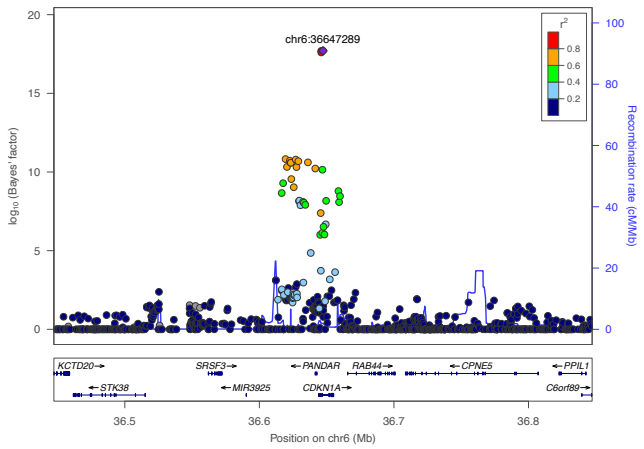

chr5:168393227
log10 (Bayes' factor)
Recombination rate (cM/Mb)
r2
0.8
0.6
0.4
0.2
SLIT3
MIR218-2
CTB-174D11.1
168.2
168.3
168.4
168.5
Position on chr5 (Mb)

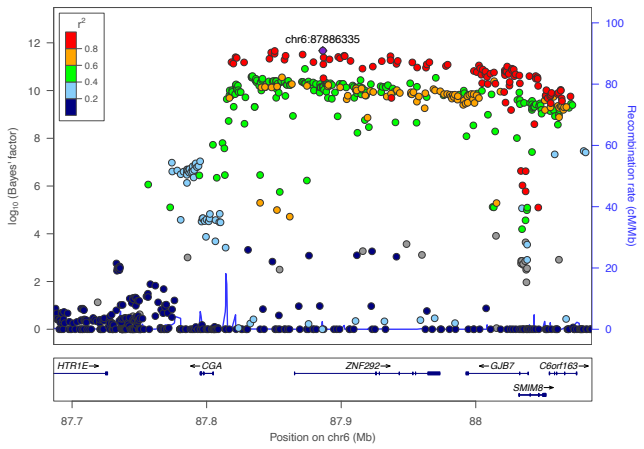

chr5:172670745
log10 (Bayes' factor)
Recombination rate (cM/Mb)
r2
0.8
0.6
0.4
0.2
CREBRF
BNIP1
NKX2-5
STC2
MIR8056
172.5
172.6
172.7
172.8
Position on chr5 (Mb)

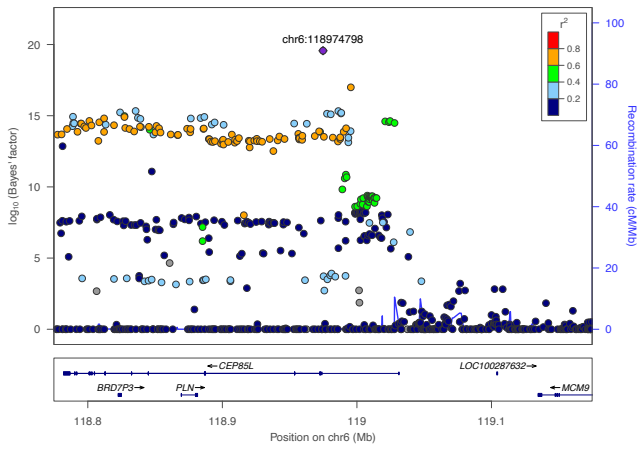

chr6:16418562
log10 (Bayes' factor)
Recombination rate (cM/Mb)
r2
0.8
0.6
0.4
0.2
GMPR
ATXN1
16.3
16.4
16.5
16.6
Position on chr6 (Mb)

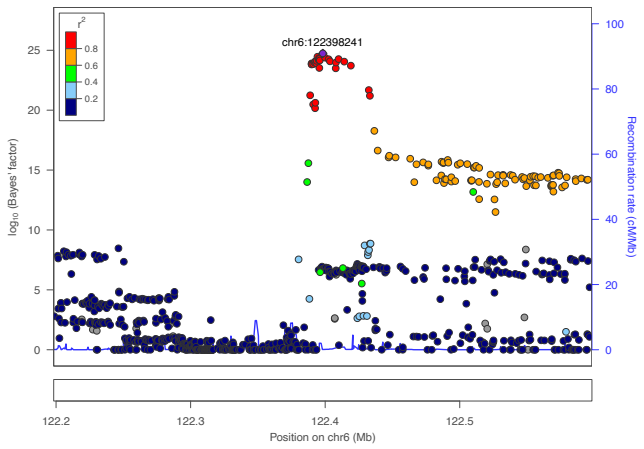

chr6:18210109
log10 (Bayes' factor)
Recombination rate (cM/Mb)
r2
0.8
0.6
0.4
0.2
NHLRC1
KDM1B
RNF144B
TPMT
DEK
18.1
18.2
18.3
18.4
Position on chr6 (Mb)

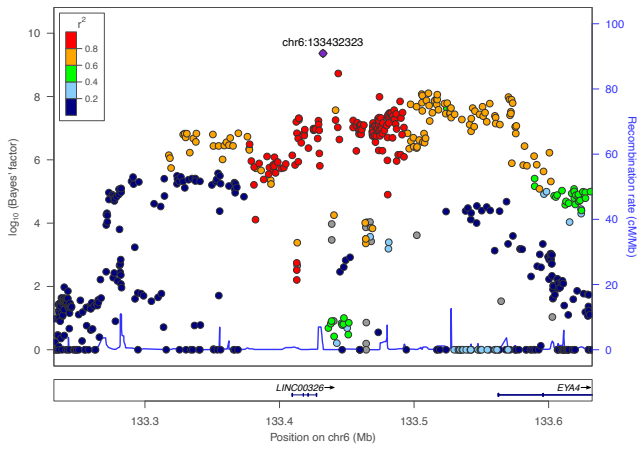

chr6:34188866
log10 (Bayes' factor)
Recombination rate (cM/Mb)
GRM4
HMGA1
NUDT3
MIR6835
RPS10–NUDT3
C6orf1
RPS10
Position on chr6 (Mb)

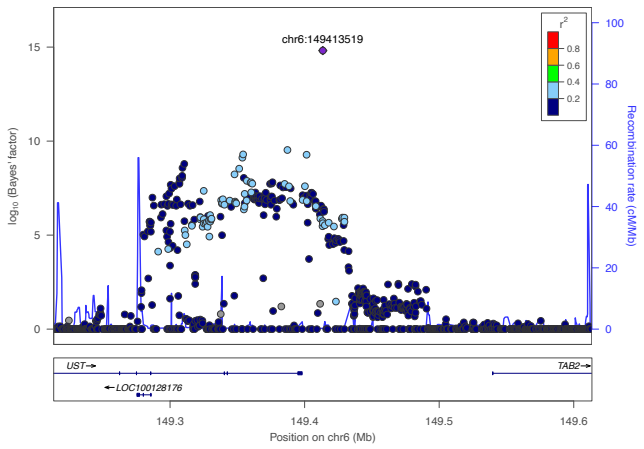

chr6:36647289
log10 (Bayes' factor)
Recombination rate (cM/Mb)
KCTD20
SRSF3
PANDAR
RAB44
CPNE5
PPIL1
STK38
MIR3925
CDKN1A
C6orf89
Position on chr6 (Mb)

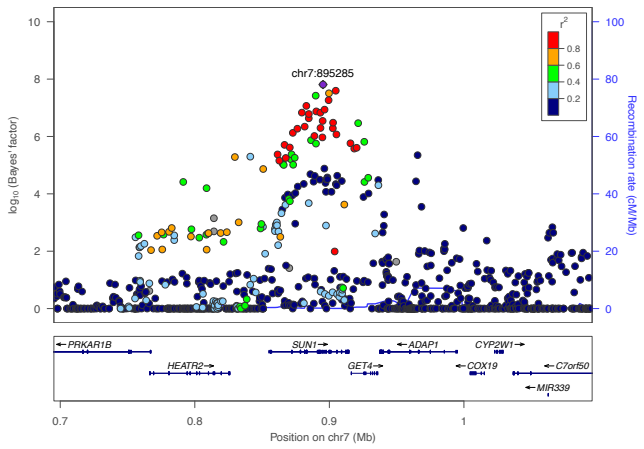

chr6:87886335
log10 (Bayes' factor)
Recombination rate (cM/Mb)
HTR1E
CGA
ZNF292
GJB7
C6orf163
SMIM8
Position on chr6 (Mb)

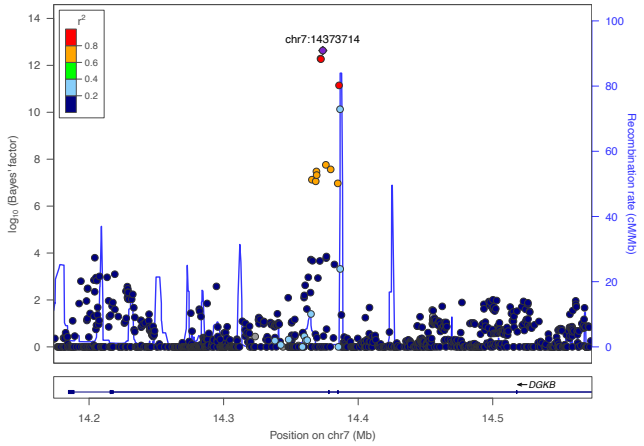

chr6:118974798
log10 (Bayes' factor)
Recombination rate (cM/Mb)
CEP85L
LOC100287632
BRD7P3
PLN
MCM9
Position on chr6 (Mb)

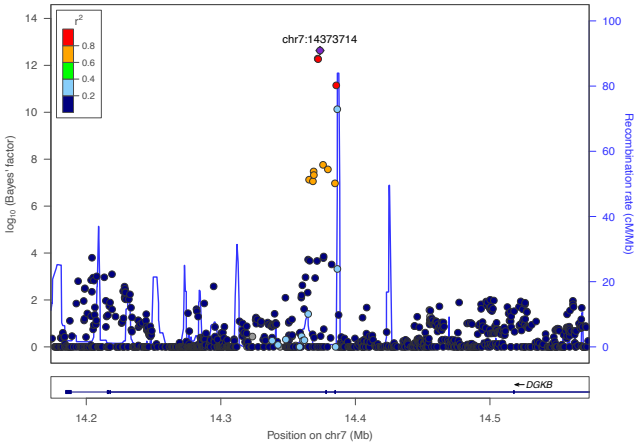

chr6:122398241
log10 (Bayes' factor)
Recombination rate (cM/Mb)
Position on chr6 (Mb)

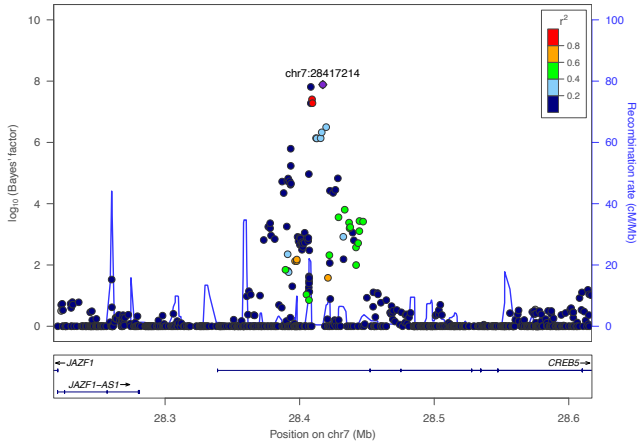

chr6:133432323
log10 (Bayes' factor)
Recombination rate (cM/Mb)
LINC00326
EYA4
Position on chr6 (Mb)

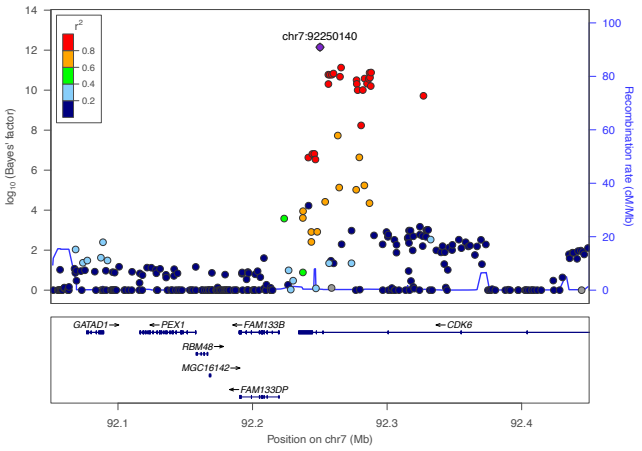

chr6:149413519
log10 (Bayes' factor)
Recombination rate (cM/Mb)
UST
TAB2
LOC100128176
Position on chr6 (Mb)

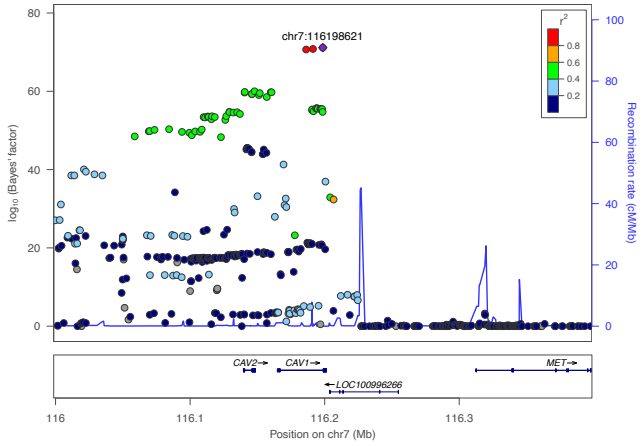

chr7:895285
log10 (Bayes' factor)
Recombination rate (cM/Mb)
PRKAR1B
SUN1
ADAP1
CYP2W1
HEATR2
GET4
COX19
C7orf50
MIR339
Position on chr7 (Mb)

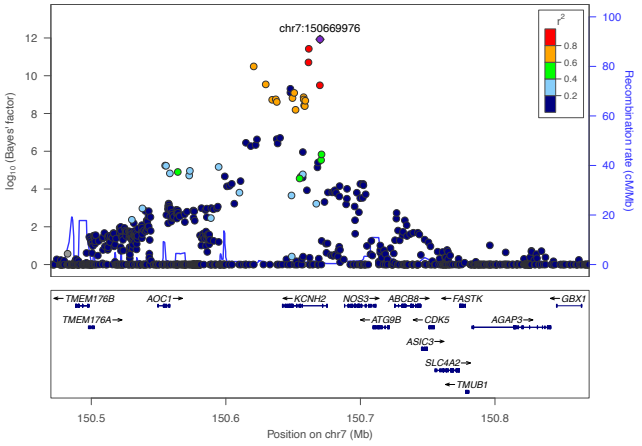

chr7:14373714
DGKB
Position on chr7 (Mb)
Recombination rate (cM/Mb)

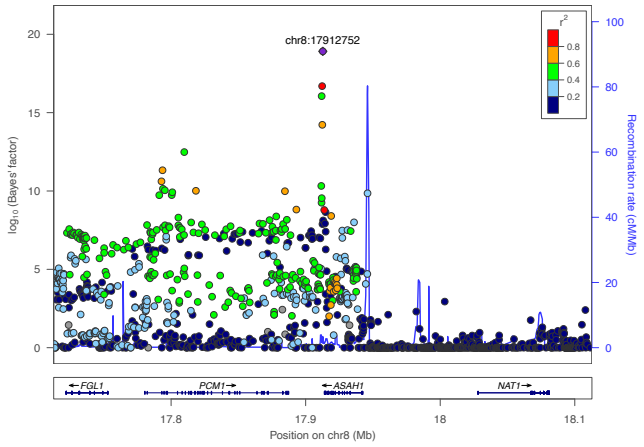

chr7:14373714
DGKB
Position on chr7 (Mb)
Recombination rate (cM/Mb)

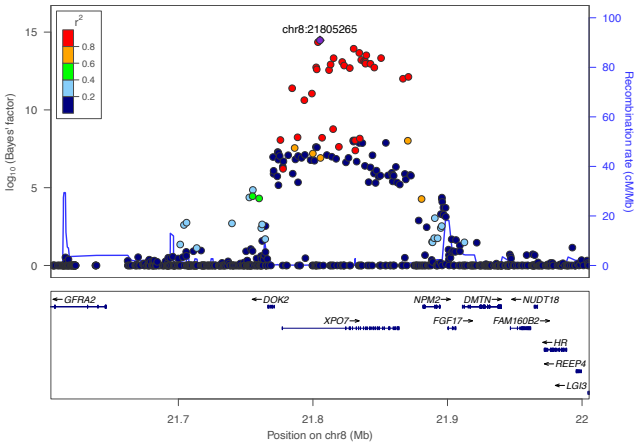

chr7:28417214
JAZF1
JAZF1-AS1
CREB5
Position on chr7 (Mb)
Recombination rate (cM/Mb)

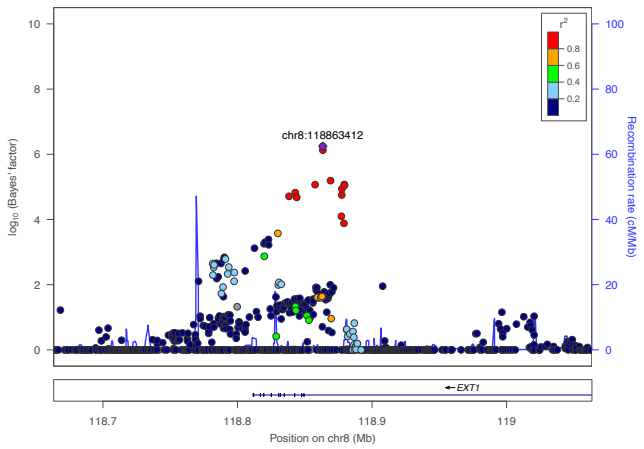

chr7:92250140
GATAD1
PEX1
FAM133B
CDK6
RBM48
MGC16142
FAM133DP
Position on chr7 (Mb)
Recombination rate (cM/Mb)

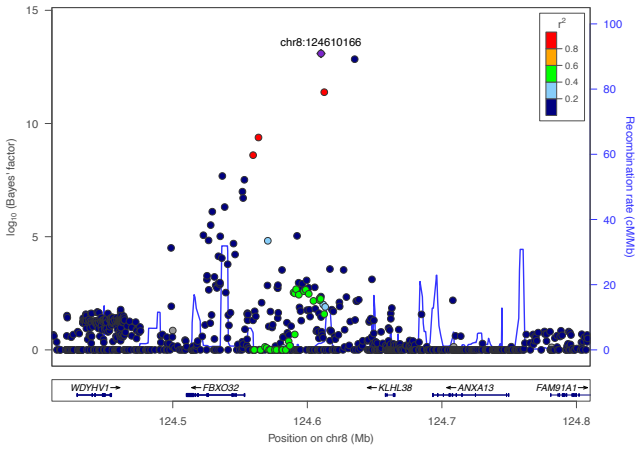

chr7:116198621
CAV2
CAV1
MET
LOC100996266
Position on chr7 (Mb)
Recombination rate (cM/Mb)

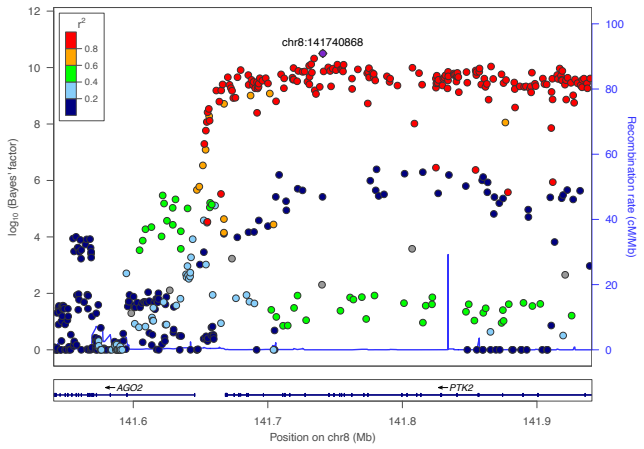

chr7:150669976
TMEM176B
TMEM176A
AOC1
KCNH2
NOS3
ABCB8
FASTK
GBX1
ATG9B
CDK5
AGAP3
ASIC3
SLC4A2
TMUB1
Position on chr7 (Mb)
Recombination rate (cM/Mb)

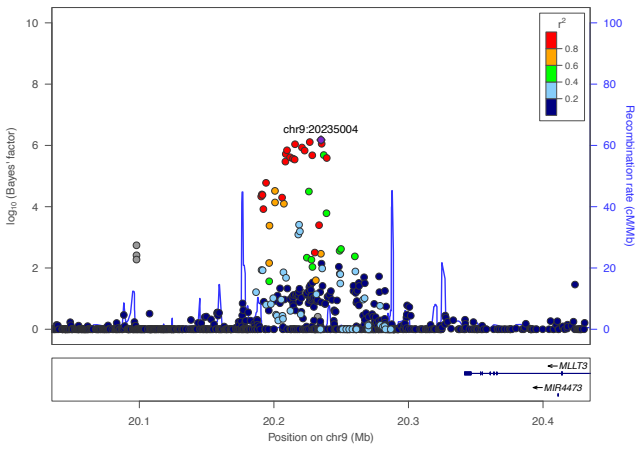

chr8:17912752
FGL1
PCM1
ASAH1
NAT1
Position on chr8 (Mb)
Recombination rate (cM/Mb)

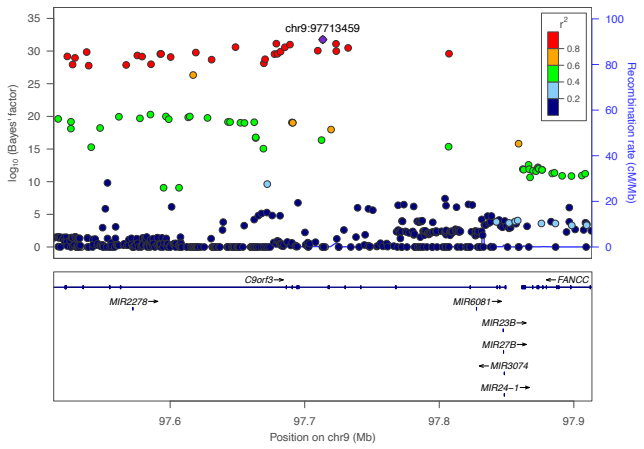

chr8:21805265
GFRA2
DOK2
XPO7
NPM2
DMTN
NUDT18
FGF17
FAM160B2
HR
REEP4
LGI3
Position on chr8 (Mb)
Recombination rate (cM/Mb)

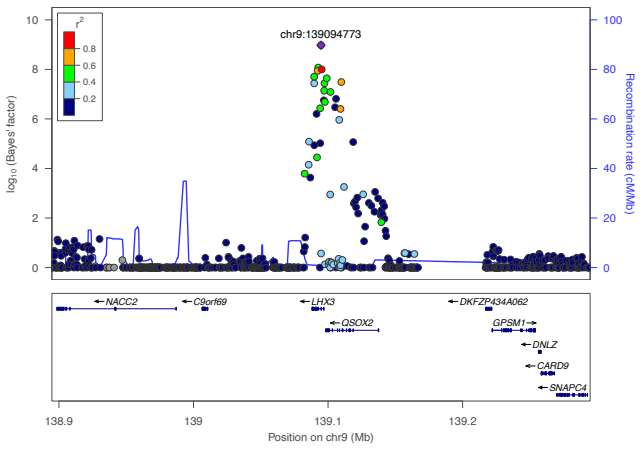

chr8:118863412
EXT1
Position on chr8 (Mb)
Recombination rate (cM/Mb)

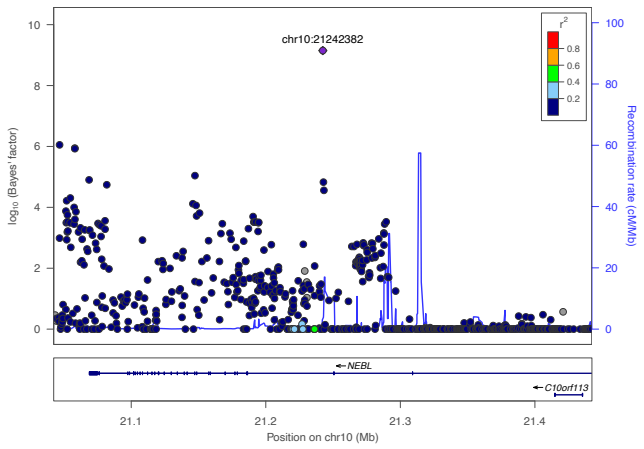

chr8:124610166
WDYHV1
FBXO32
KLHL38
ANXA13
FAM91A1
Position on chr8 (Mb)
Recombination rate (cM/Mb)

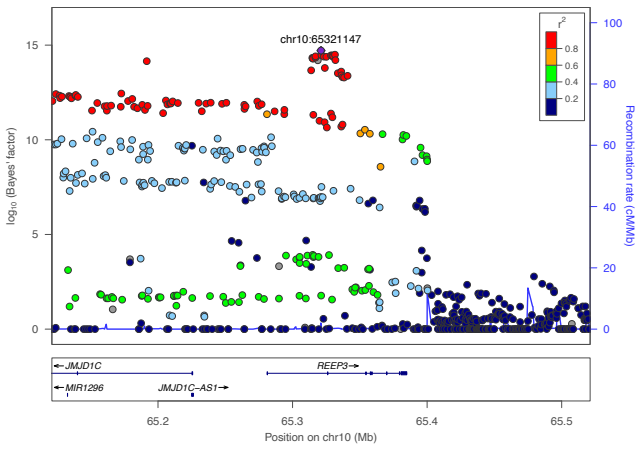

chr8:141740868
AGO2
PTK2
Position on chr8 (Mb)
Recombination rate (cM/Mb)

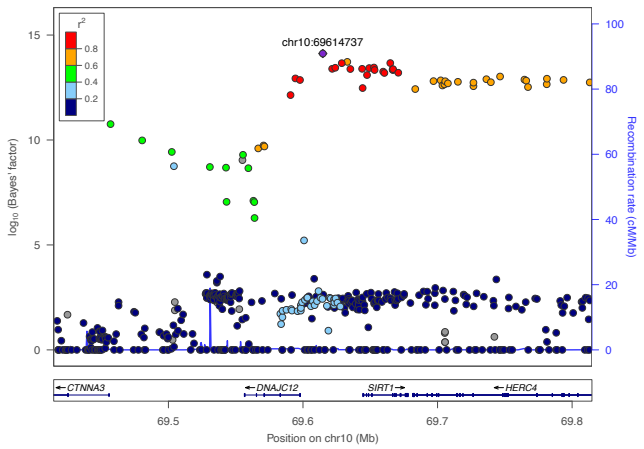

chr9:20235004
MLLT3
MIR4473
Position on chr9 (Mb)
Recombination rate (cM/Mb)

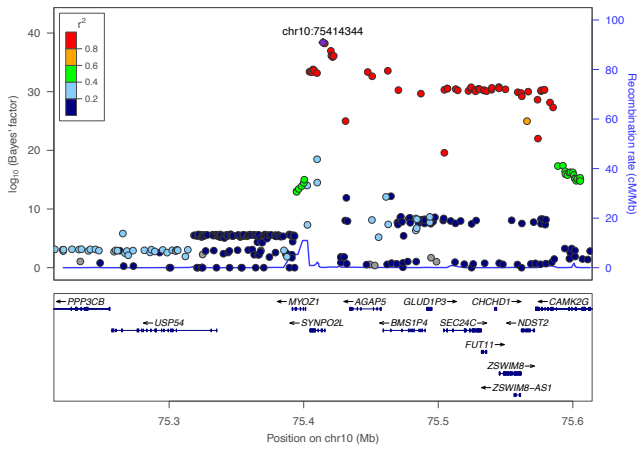

chr9:97713459
C9orf3
FANCC
MIR2278
MIR6081
MIR23B
MIR27B
MIR3074
MIR24-1
Position on chr9 (Mb)
Recombination rate (cM/Mb)

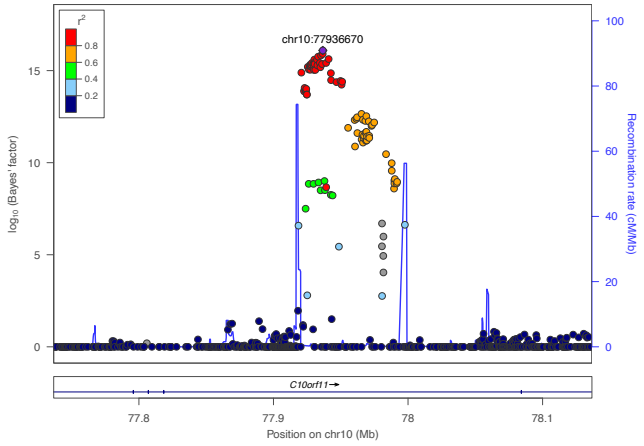

chr9:139094773
NACC2
C9orf69
LHX3
DKFZP434A062
QSOX2
GPSM1
DNLZ
CARD9
SNAPC4
Position on chr9 (Mb)
Recombination rate (cM/Mb)

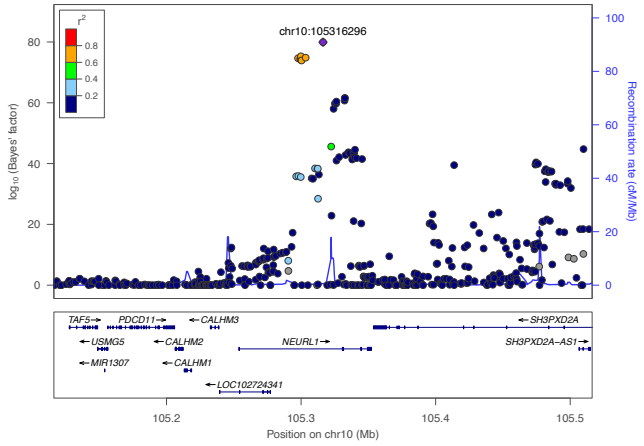

chr10:21242382
NEBL
C10orf113
Position on chr10 (Mb)
Recombination rate (cM/Mb)

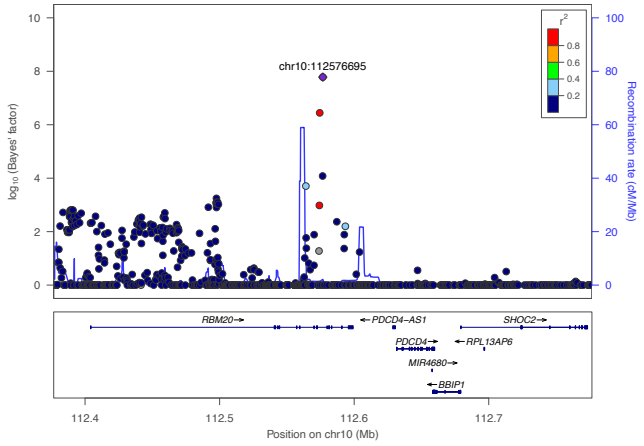

chr10:65321147
JMJD1C
REEP3
MIR1296
JMJD1C-AS1
Position on chr10 (Mb)
Recombination rate (cM/Mb)

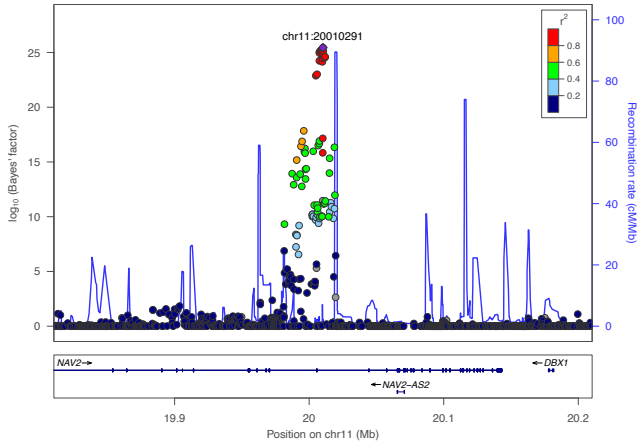

chr10:69614737
r²
0.8
0.6
0.4
0.2
log10 (Bayes' factor)
Recombination rate (cM/Mb)
CTNNA3
DNAJC12
SIRT1
HERC4
69.5
69.6
69.7
69.8
Position on chr10 (Mb)

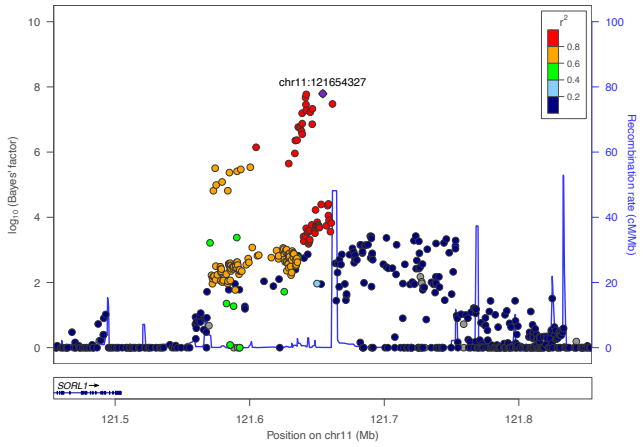

chr10:75414344
r²
0.8
0.6
0.4
0.2
log10 (Bayes' factor)
Recombination rate (cM/Mb)
PPP3CB
USP54
MYOZ1
AGAP5
GLUD1P3
CHCHD1
CAMK2G
SYNPO2L
BMS1P4
SEC24C
NDST2
FUT11
ZSWIM8
ZSWIM8-AS1
75.3
75.4
75.5
75.6
Position on chr10 (Mb)

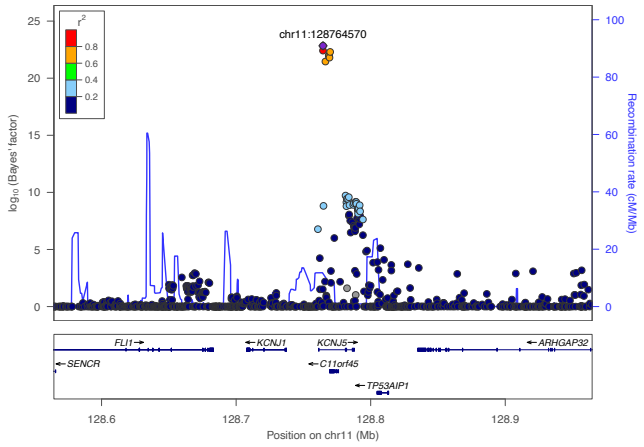

chr10:77936670
r²
0.8
0.6
0.4
0.2
log10 (Bayes' factor)
Recombination rate (cM/Mb)
C10orf11
77.8
77.9
78
78.1
Position on chr10 (Mb)

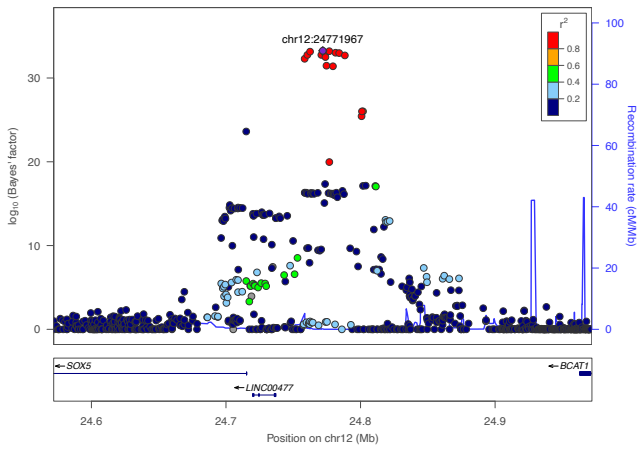

chr10:105316296
r²
0.8
0.6
0.4
0.2
log10 (Bayes' factor)
Recombination rate (cM/Mb)
TAF5
PDCD11
CALHM3
SH3PXD2A
USMG5
CALHM2
NEURL1
SH3PXD2A-AS1
MIR1307
CALHM1
LOC102724341
105.2
105.3
105.4
105.5
Position on chr10 (Mb)

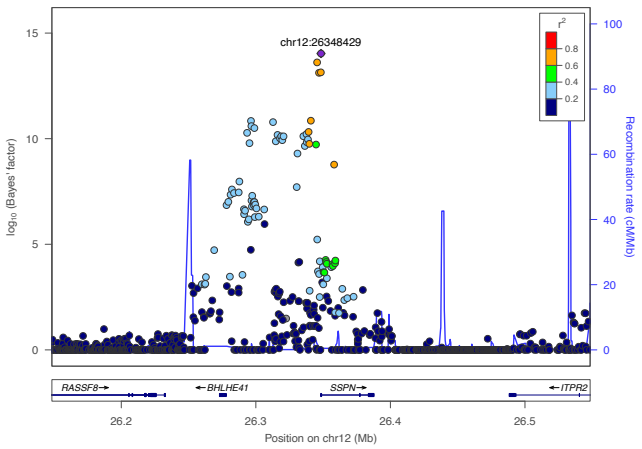

chr10:112576695
r²
0.8
0.6
0.4
0.2
log10 (Bayes' factor)
Recombination rate (cM/Mb)
RBM20
PDCD4-AS1
SHOC2
PDCD4
RPL13AP6
MIR4680
BBIP1
112.4
112.5
112.6
112.7
Position on chr10 (Mb)

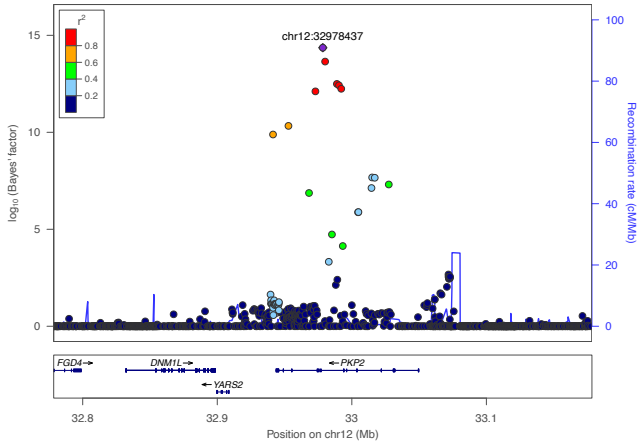

chr11:20010291
r²
0.8
0.6
0.4
0.2
log10 (Bayes' factor)
Recombination rate (cM/Mb)
NAV2
DBX1
NAV2-AS2
19.9
20
20.1
20.2
Position on chr11 (Mb)

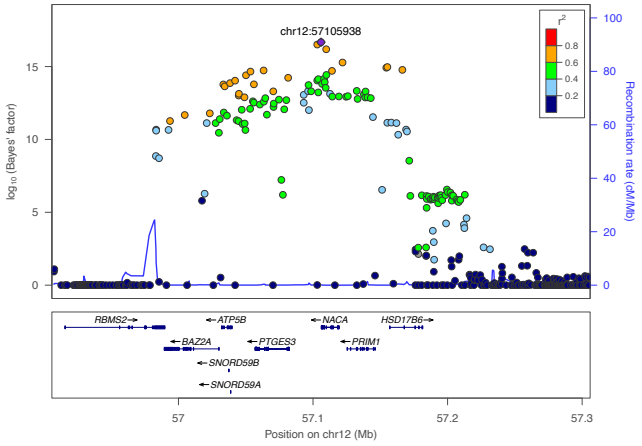

chr11:121654327
r²
0.8
0.6
0.4
0.2
log10 (Bayes' factor)
Recombination rate (cM/Mb)
SORL1
121.5
121.6
121.7
121.8
Position on chr11 (Mb)

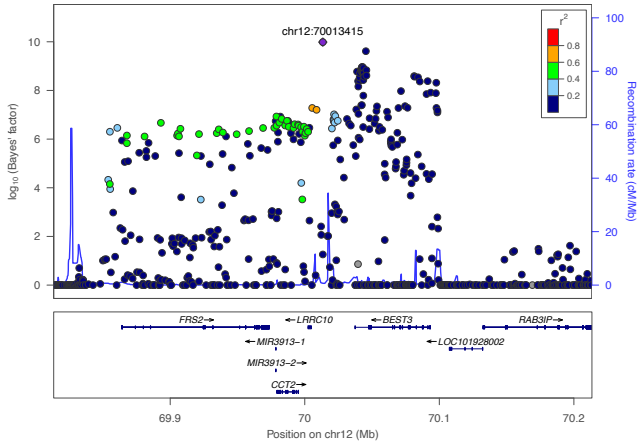

chr11:128764570
r²
0.8
0.6
0.4
0.2
log10 (Bayes' factor)
Recombination rate (cM/Mb)
FLI1
KCNJ1
KCNJ5
ARHGAP32
SENCR
C11orf45
TP53AIP1
128.6
128.7
128.8
128.9
Position on chr11 (Mb)

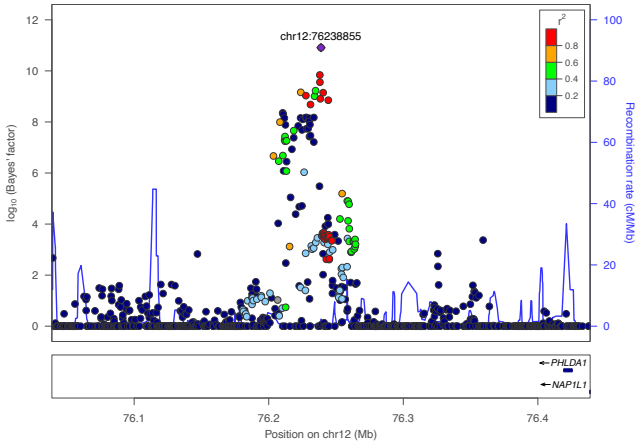

chr12:24771967
log10 (Bayes' factor)
Recombination rate (cM/Mb)
r2
0.8
0.6
0.4
0.2
SOX5
BCAT1
LINC00477
24.6
24.7
24.8
24.9
Position on chr12 (Mb)

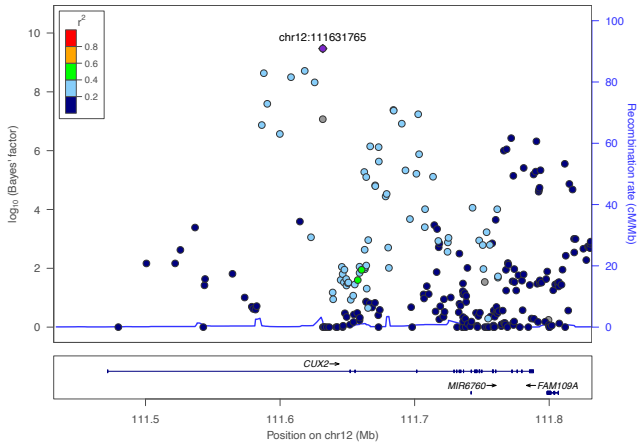

chr12:26348429
log10 (Bayes' factor)
Recombination rate (cM/Mb)
r2
0.8
0.6
0.4
0.2
RASSF8
BHLHE41
SSPN
ITPR2
26.2
26.3
26.4
26.5
Position on chr12 (Mb)

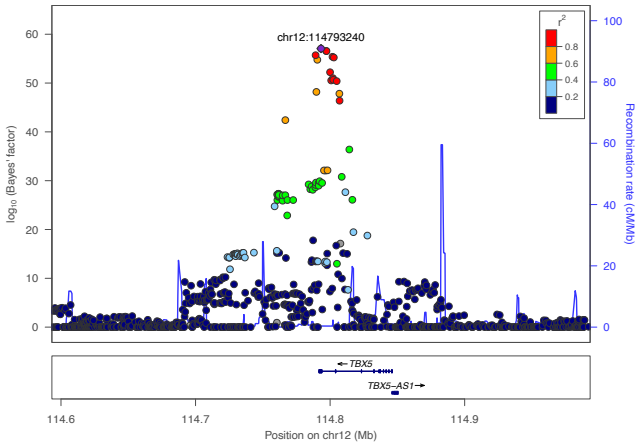

chr12:32978437
log10 (Bayes' factor)
Recombination rate (cM/Mb)
r2
0.8
0.6
0.4
0.2
FGD4
DNM1L
PKP2
YARS2
32.8
32.9
33
33.1
Position on chr12 (Mb)

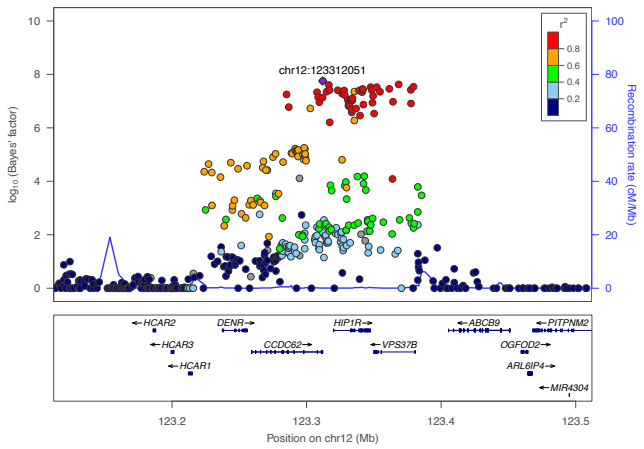

chr12:57105938
log10 (Bayes' factor)
Recombination rate (cM/Mb)
r2
0.8
0.6
0.4
0.2
RBMS2
ATP5B
NACA
HSD17B6
BAZ2A
PTGES3
PRIM1
SNORD59B
SNORD59A
57
57.1
57.2
57.3
Position on chr12 (Mb)

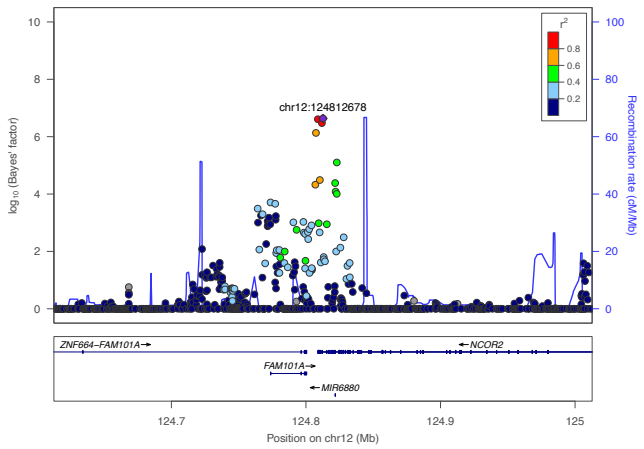

chr12:70013415
log10 (Bayes' factor)
Recombination rate (cM/Mb)
r2
0.8
0.6
0.4
0.2
FRS2
LRRC10
BEST3
RAB3IP
MIR3913-1
LOC101928002
MIR3913-2
CCT2
69.9
70
70.1
70.2
Position on chr12 (Mb)

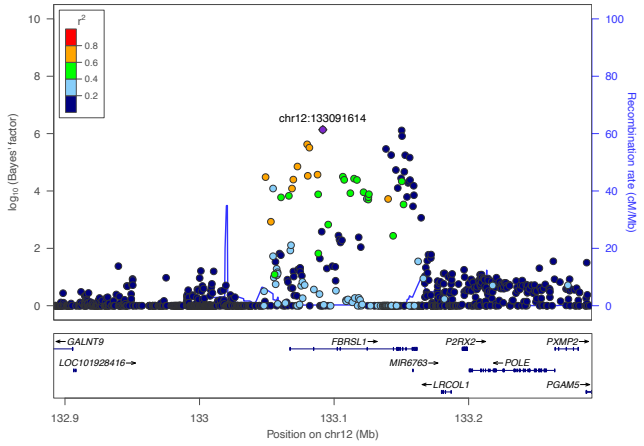

chr12:76238855
log10 (Bayes' factor)
Recombination rate (cM/Mb)
r2
0.8
0.6
0.4
0.2
PHLDA1
NAP1L1
76.1
76.2
76.3
76.4
Position on chr12 (Mb)

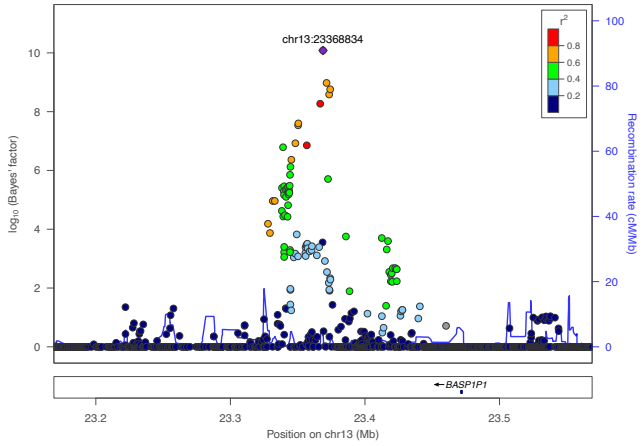

chr12:111631765
log10 (Bayes' factor)
Recombination rate (cM/Mb)
r2
0.8
0.6
0.4
0.2
CUX2
MIR6760
FAM109A
111.5
111.6
111.7
111.8
Position on chr12 (Mb)

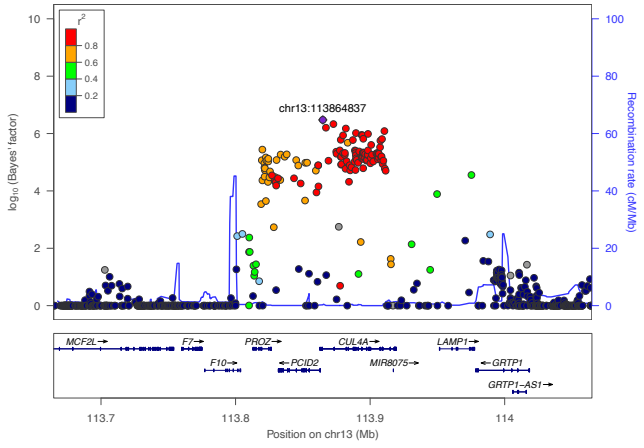

chr12:114793240
log10 (Bayes' factor)
Recombination rate (cM/Mb)
r2
0.8
0.6
0.4
0.2
TBX5
TBX5-AS1
114.6
114.7
114.8
114.9
Position on chr12 (Mb)

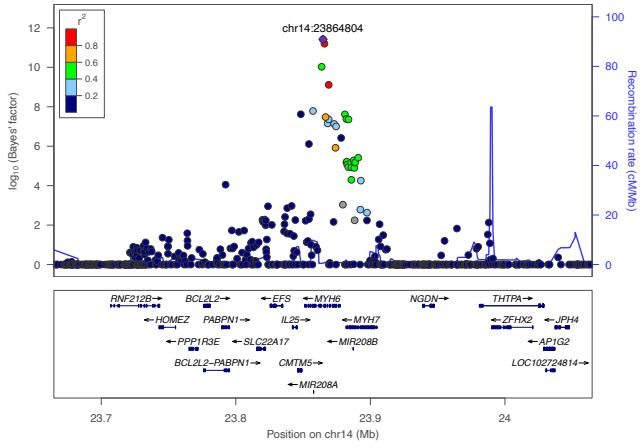

chr12:123312051
log10 (Bayes' factor)
Recombination rate (cM/Mb)
r2
HCAR2
DENR
HIP1R
ABCB9
PITPNM2
HCAR3
CCDC62
VPS37B
OGFOD2
HCAR1
ARL6IP4
MIR4304
Position on chr12 (Mb)

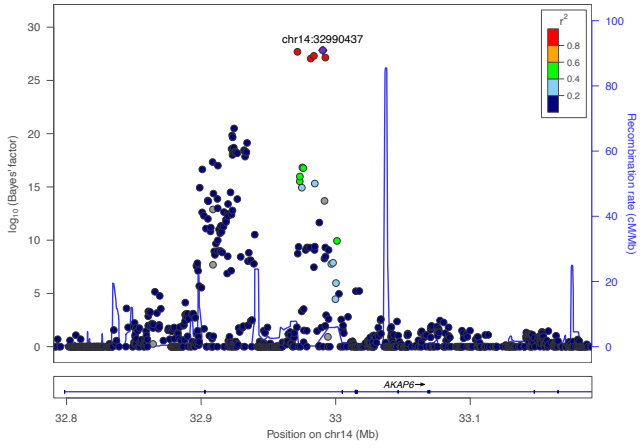

chr12:124812678
log10 (Bayes' factor)
Recombination rate (cM/Mb)
r2
ZNF664–FAM101A
NCOR2
FAM101A
MIR6880
Position on chr12 (Mb)

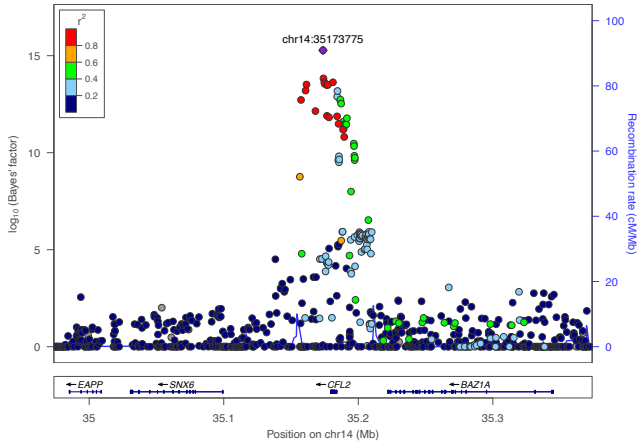

chr12:133091614
log10 (Bayes' factor)
Recombination rate (cM/Mb)
r2
GALNT9
FBRSL1
P2RX2
PXMP2
LOC101928416
MIR6763
POLE
LRCOL1
PGAM5
Position on chr12 (Mb)

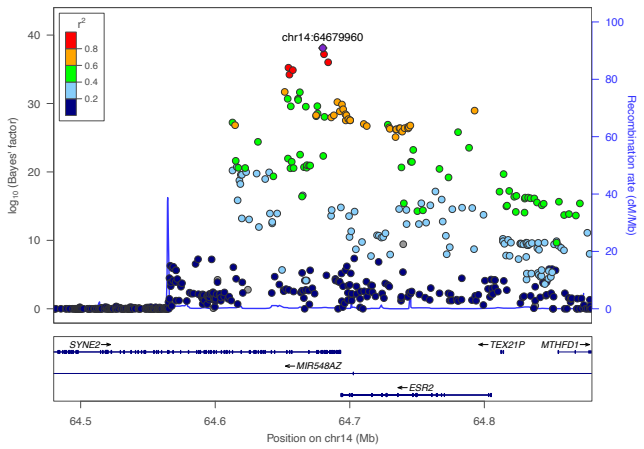

chr13:23368834
log10 (Bayes' factor)
Recombination rate (cM/Mb)
r2
BASP1P1
Position on chr13 (Mb)

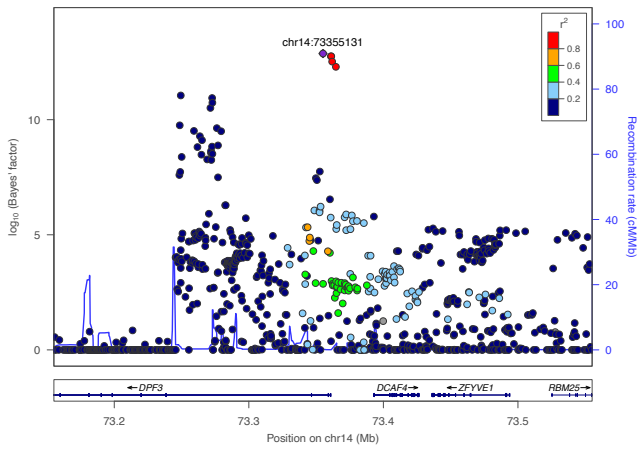

chr13:113864837
log10 (Bayes' factor)
Recombination rate (cM/Mb)
r2
MCF2L
F7
PROZ
CUL4A
LAMP1
F10
PCID2
MIR8075
GRTP1
GRTP1-AS1
Position on chr13 (Mb)

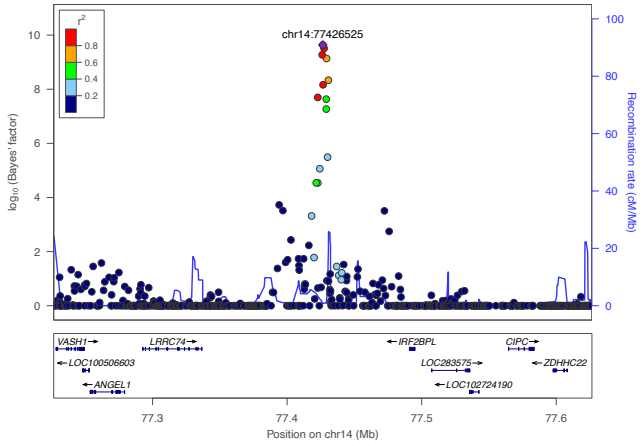

chr14:23864804
log10 (Bayes' factor)
Recombination rate (cM/Mb)
r2
RNF212B
BCL2L2
EFS
MYH6
NGDN
THTPA
HOMEZ
PABPN1
IL25
MYH7
ZFHX2
JPH4
PPP1R3E
SLC22A17
MIR208B
AP1G2
BCL2L2-PABPN1
CMTM5
LOC102724814
MIR208A
Position on chr14 (Mb)

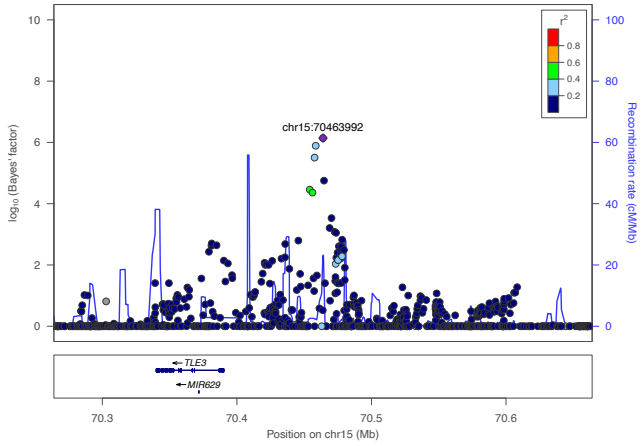

chr14:32990437
log10 (Bayes' factor)
Recombination rate (cM/Mb)
r2
AKAP6
Position on chr14 (Mb)

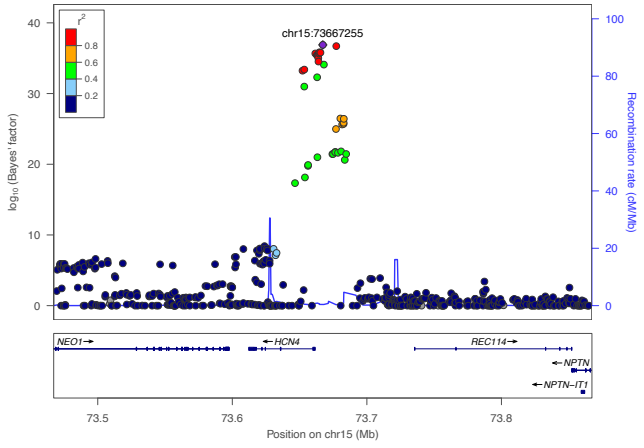

chr14:35173775
log10 (Bayes' factor)
Recombination rate (cM/Mb)
r2
EAPP
SNX6
CFL2
BAZ1A
Position on chr14 (Mb)

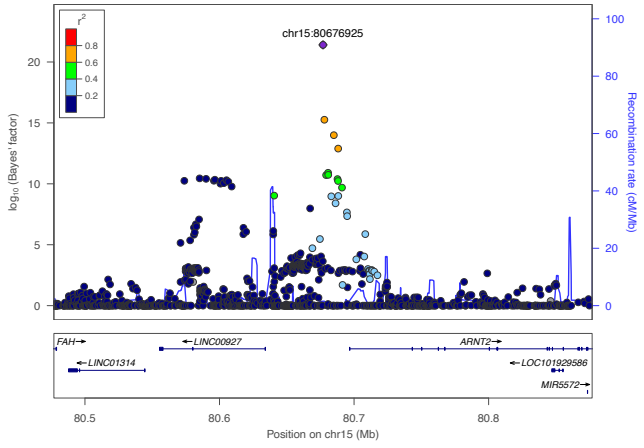

chr14:64679960
log10 (Bayes' factor)
Recombination rate (cM/Mb)
SYNE2
TEX21P
MTHFD1
MIR548AZ
ESR2
Position on chr14 (Mb)

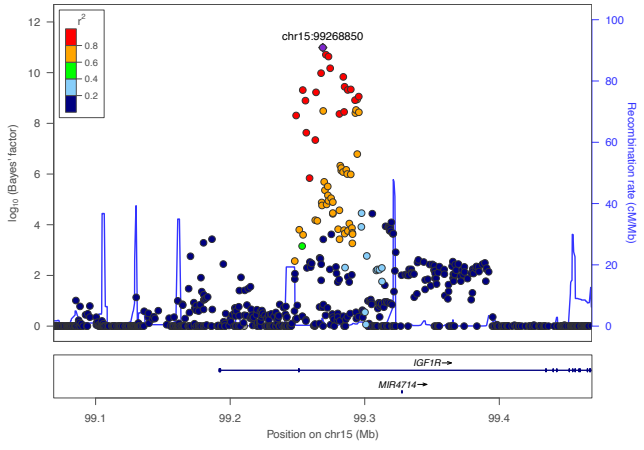

chr14:73355131
log10 (Bayes' factor)
Recombination rate (cM/Mb)
DPF3
DCAF4
ZFYVE1
RBM25
Position on chr14 (Mb)

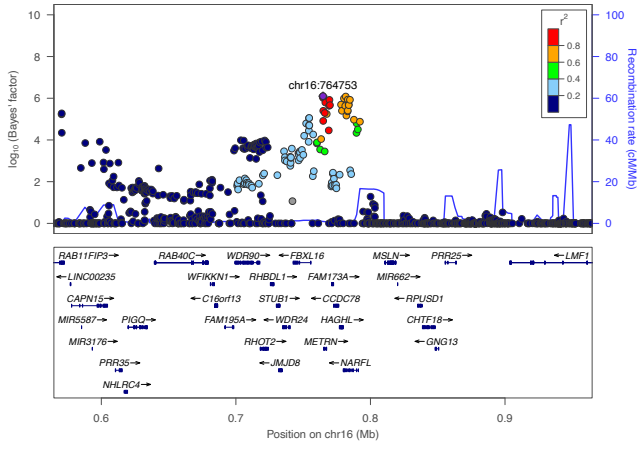

chr14:77426525
log10 (Bayes' factor)
Recombination rate (cM/Mb)
VASH1
LRRC74
IRF2BPL
CIPC
LOC100506603
LOC283575
ZDHHC22
ANGEL1
LOC102724190
Position on chr14 (Mb)

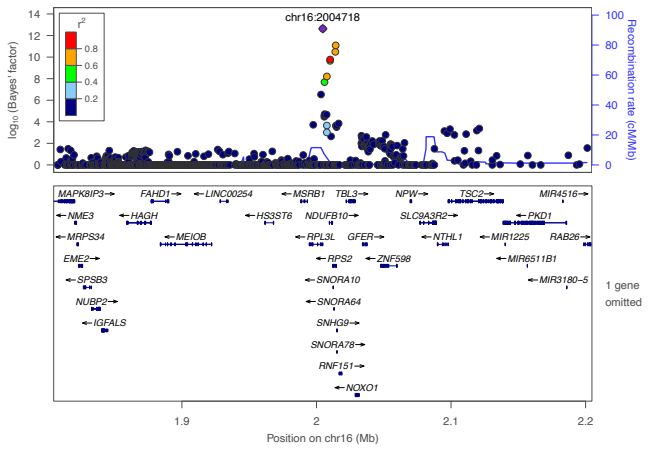

chr15:70463992
log10 (Bayes' factor)
Recombination rate (cM/Mb)
TLE3
MIR629
Position on chr15 (Mb)

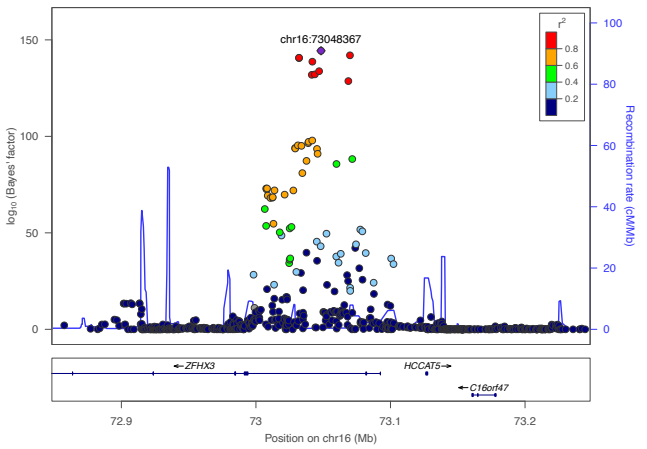

chr15:73667255
log10 (Bayes' factor)
Recombination rate (cM/Mb)
NEO1
HCN4
REC114
NPTN
NPTN-IT1
Position on chr15 (Mb)

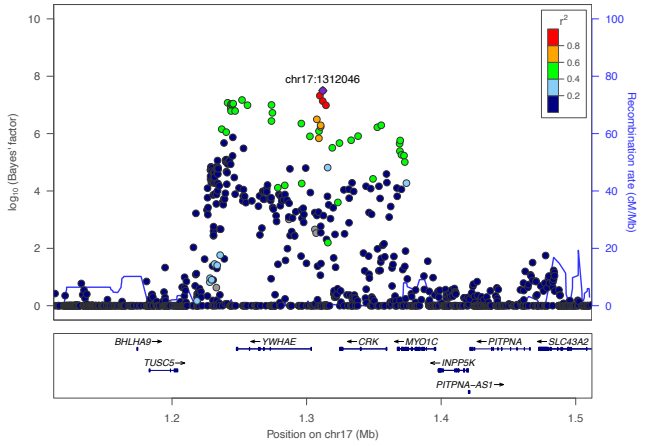

chr15:80676925
log10 (Bayes' factor)
Recombination rate (cM/Mb)
FAH
LINC00927
ARNT2
LINC01314
LOC101929586
MIR5572
Position on chr15 (Mb)

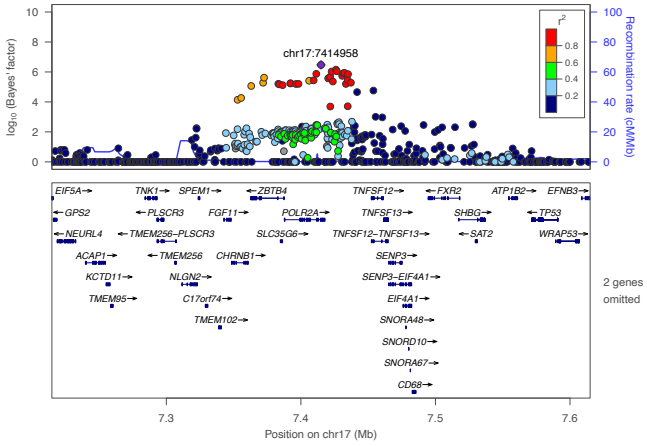

chr15:99268850
log10 (Bayes' factor)
Recombination rate (cM/Mb)
IGF1R
MIR4714
Position on chr15 (Mb)

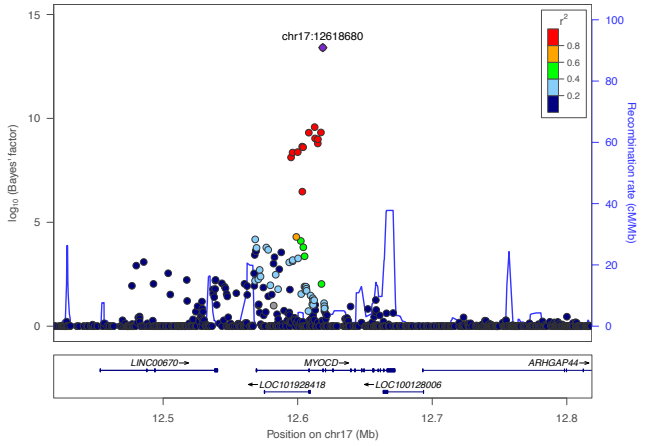

chr16:764753
log10 (Bayes' factor)
Recombination rate (cM/Mb)
RAB11FIP3
RAB40C
WDR90
FBXL16
MSLN
PRR25
LMF1
LINC00235
WFIKKN1
RHBDL1
FAM173A
MIR662
CAPN15
C16orf13
STUB1
CCDC78
RPUSD1
MIR5587
PIGQ
FAM195A
WDR24
HAGHL
CHTF18
MIR3176
RHOT2
METRN
GNG13
PRR35
JMJD8
NARFL
NHLRC4
Position on chr16 (Mb)

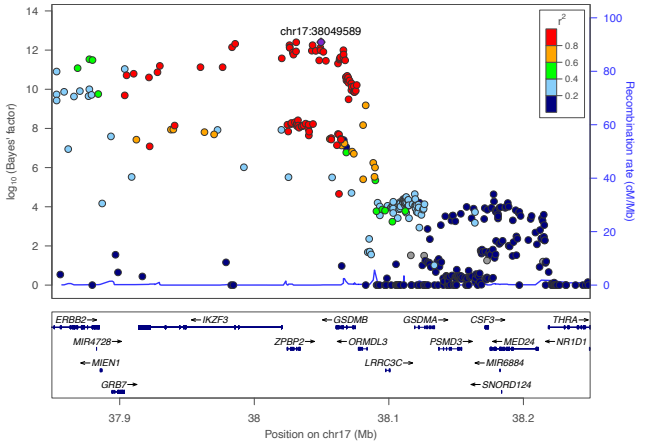

chr16:2004718
log10 (Bayes' factor)
Recombination rate (cM/Mb)
r2
0.8
0.6
0.4
0.2
MAPK8IP3
FAHD1
LINC00254
MSRB1
TBL3
NPW
TSC2
MIR4516
NME3
HAGH
HS3ST6
NDUFB10
SLC9A3R2
PKD1
MRPS34
MEIOB
RPL3L
GFER
NTHL1
MIR1225
RAB26
EME2
RPS2
ZNF598
MIR6511B1
SPSB3
SNORA10
MIR3180–5
NUBP2
SNORA64
IGFALS
SNHG9
SNORA78
RNF151
NOXO1
1 gene
omitted
1.9
2
2.1
2.2
Position on chr16 (Mb)

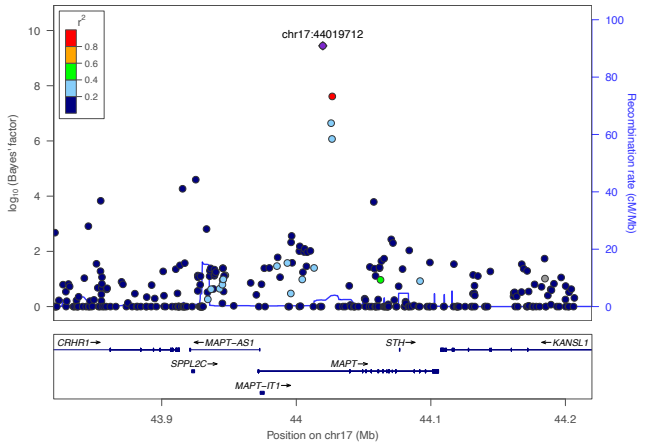

chr16:73048367
log10 (Bayes' factor)
Recombination rate (cM/Mb)
r2
0.8
0.6
0.4
0.2
ZFHX3
HCCAT5
C16orf47
72.9
73
73.1
73.2
Position on chr16 (Mb)

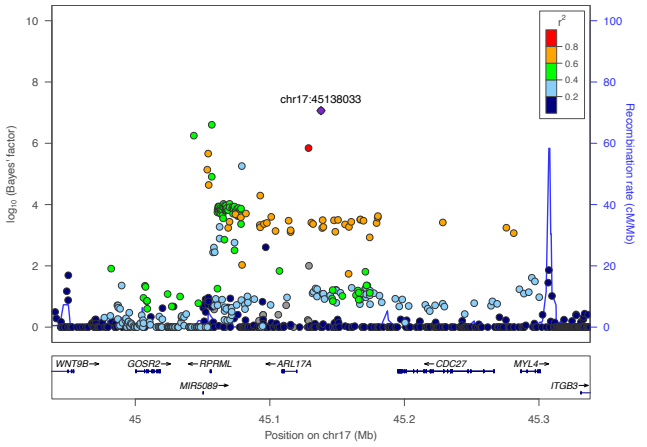

chr17:1312046
log10 (Bayes' factor)
Recombination rate (cM/Mb)
r2
0.8
0.6
0.4
0.2
BHLHA9
YWHAE
CRK
MYO1C
PITPNA
SLC43A2
TUSC5
INPP5K
PITPNA–AS1
1.2
1.3
1.4
1.5
Position on chr17 (Mb)

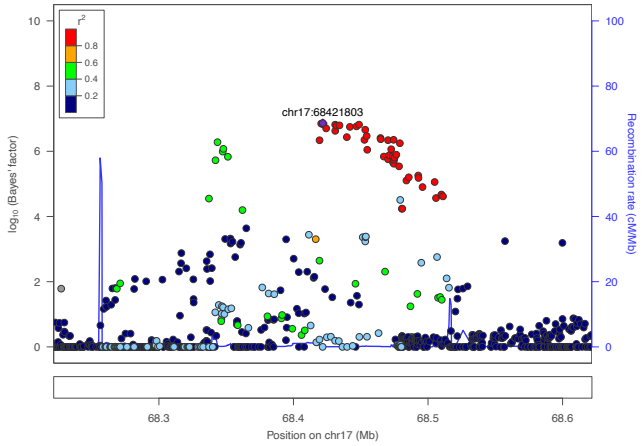

chr17:7414958
log10 (Bayes' factor)
Recombination rate (cM/Mb)
r2
0.8
0.6
0.4
0.2
EIF5A
TNK1
SPEM1
ZBTB4
TNFSF12
FXR2
ATP1B2
EFNB3
GPS2
PLSCR3
FGF11
POLR2A
TNFSF13
SHBG
TP53
NEURL4
TMEM256–PLSCR3
SLC35G6
TNFSF12–TNFSF13
SAT2
WRAP53
ACAP1
TMEM256
CHRNB1
SENP3
KCTD11
NLGN2
SENP3–EIF4A1
TMEM95
C17orf74
EIF4A1
TMEM102
SNORA48
SNORD10
SNORA67
CD68
2 genes
omitted
7.3
7.4
7.5
7.6
Position on chr17 (Mb)

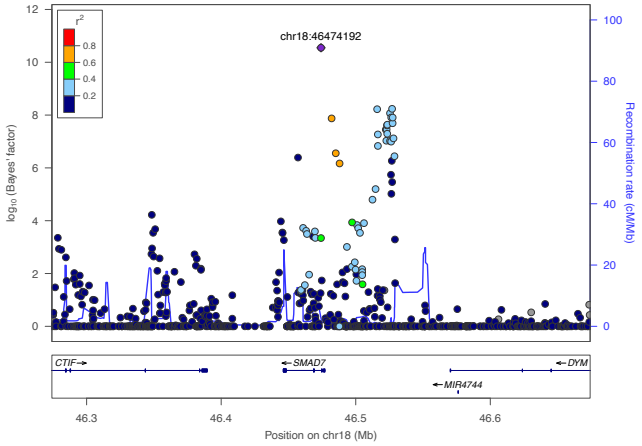

chr17:12618680
log10 (Bayes' factor)
Recombination rate (cM/Mb)
r2
0.8
0.6
0.4
0.2
LINC00670
MYOCD
ARHGAP44
LOC101928418
LOC100128006
12.5
12.6
12.7
12.8
Position on chr17 (Mb)

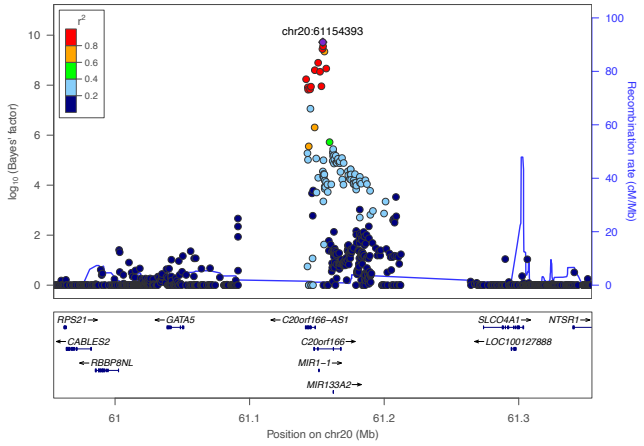

chr17:38049589
log10 (Bayes' factor)
Recombination rate (cM/Mb)
r2
0.8
0.6
0.4
0.2
ERBB2
IKZF3
GSDMB
GSDMA
CSF3
THRA
MIR4728
ZPBP2
ORMDL3
PSMD3
MED24
NR1D1
MIEN1
LRRC3C
MIR6884
GRB7
SNORD124
37.9
38
38.1
38.2
Position on chr17 (Mb)

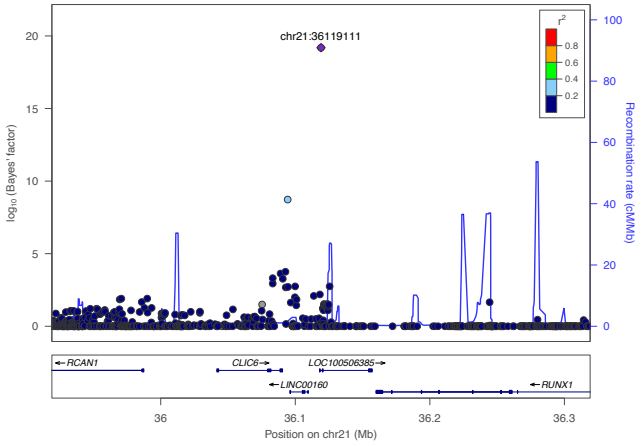

chr17:44019712
log10 (Bayes' factor)
Recombination rate (cM/Mb)
r2
0.8
0.6
0.4
0.2
CRHR1
MAPT–AS1
STH
KANSL1
SPPL2C
MAPT
MAPT–IT1
43.9
44
44.1
44.2
Position on chr17 (Mb)

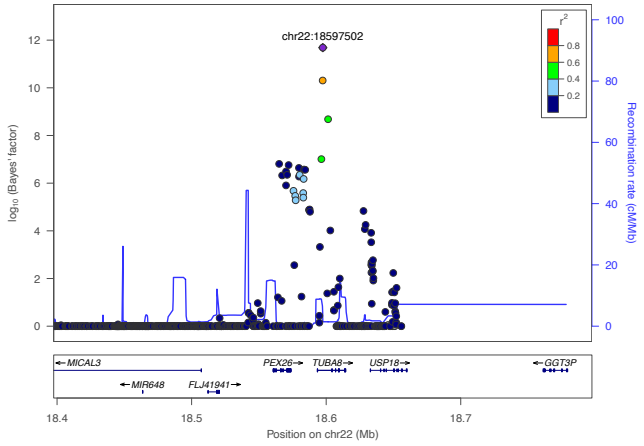

chr17:45138033
log10 (Bayes' factor)
Recombination rate (cM/Mb)
r2
0.8
0.6
0.4
0.2
WNT9B
GOSR2
RPRML
ARL17A
CDC27
MYL4
MIR5089
ITGB3
45
45.1
45.2
45.3
Position on chr17 (Mb)

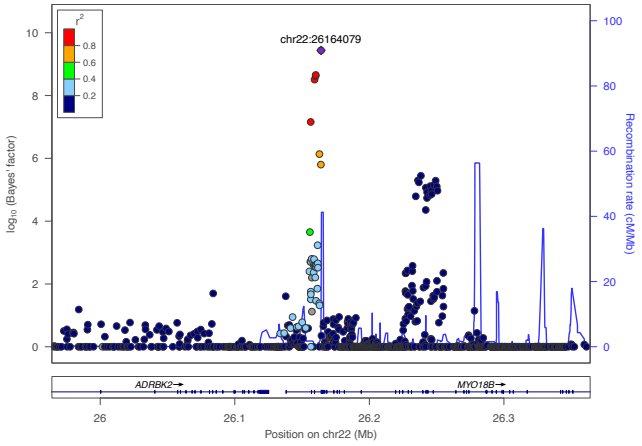

chr17:68421803
log10 (Bayes' factor)
Recombination rate (cM/Mb)
Position on chr17 (Mb)


chr18:46474192
CTIF
SMAD7
DYM
MIR4744
log10 (Bayes' factor)
Recombination rate (cM/Mb)
Position on chr18 (Mb)


chr20:61154393
RPS21
GATA5
C20orf166-AS1
SLCO4A1
NTSR1
CABLES2
C20orf166
LOC100127888
RBBP8NL
MIR1-1
MIR133A2
log10 (Bayes' factor)
Recombination rate (cM/Mb)
Position on chr20 (Mb)


chr21:36119111
RCAN1
CLIC6
LOC100506385
LINC00160
RUNX1
log10 (Bayes' factor)
Recombination rate (cM/Mb)
Position on chr21 (Mb)


chr22:18597502
MICAL3
PEX26
TUBA8
USP18
GGT3P
MIR648
FLJ41941
log10 (Bayes' factor)
Recombination rate (cM/Mb)
Position on chr22 (Mb)


chr22:26164079
ADRBK2
MYO18B
log10 (Bayes' factor)
Recombination rate (cM/Mb)
Position on chr22 (Mb)