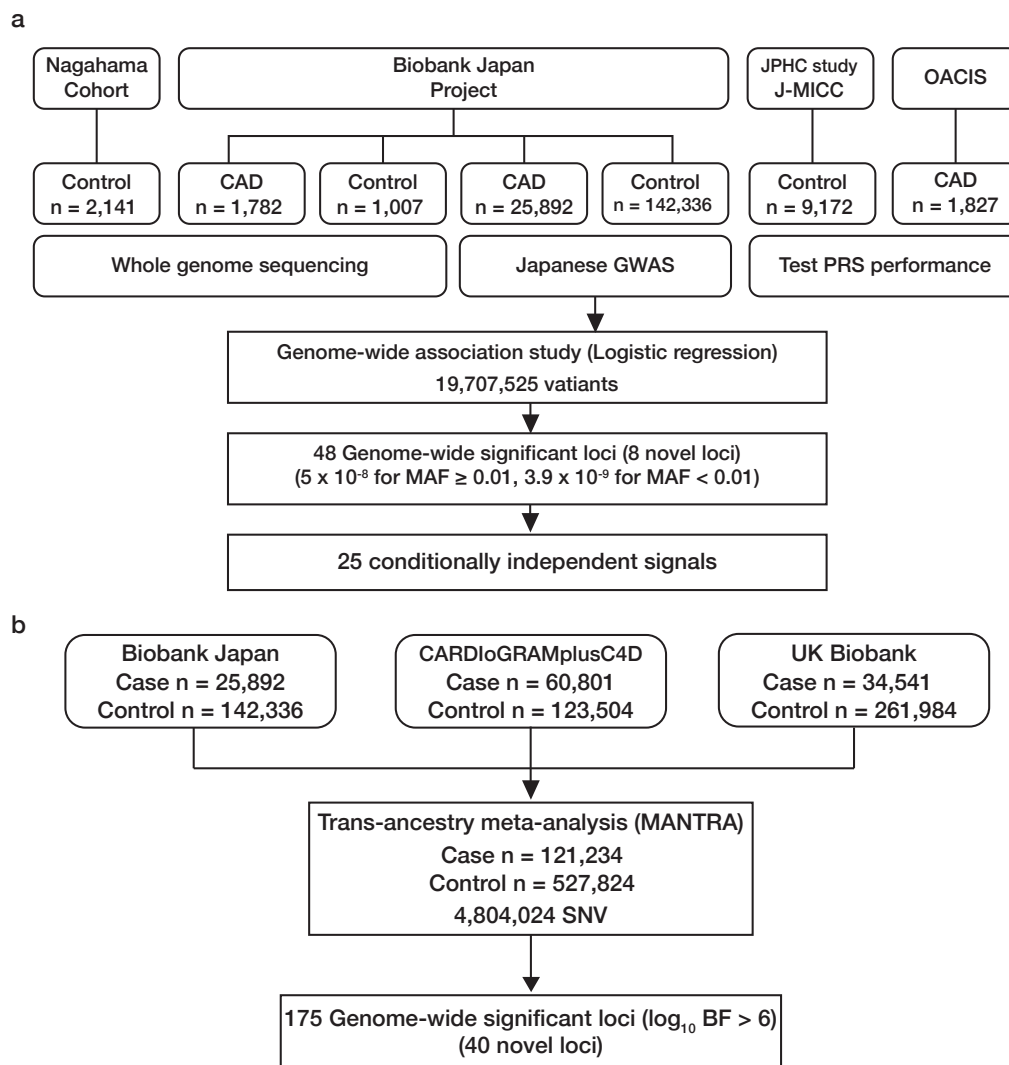

Supplementary information

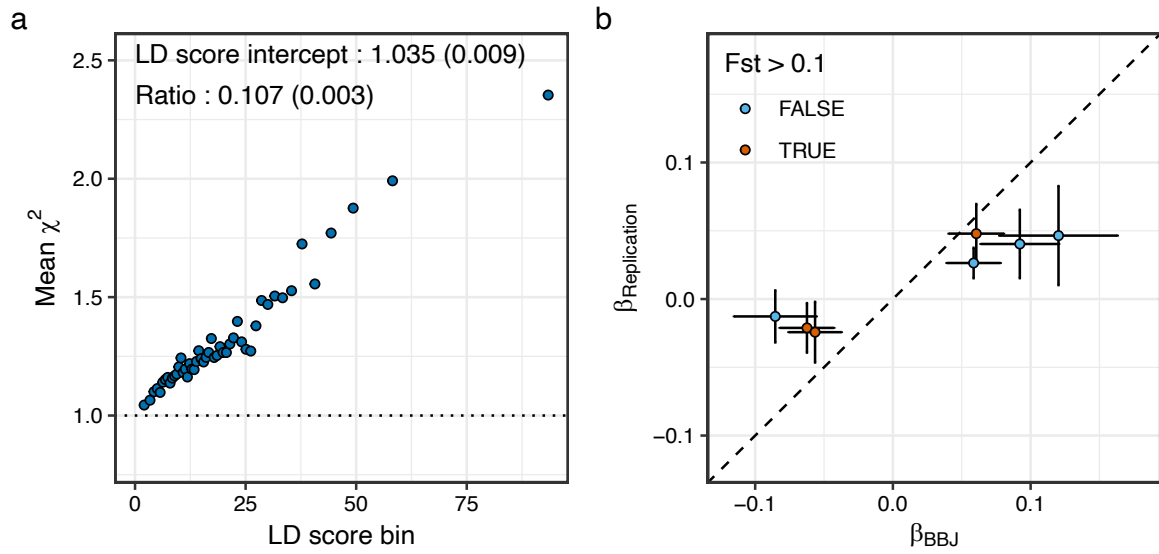
Population-specific and trans-ancestry genome-wide analyses identify distinct and shared genetic risk loci for coronary artery disease

In the format provided by the
authors and unedited

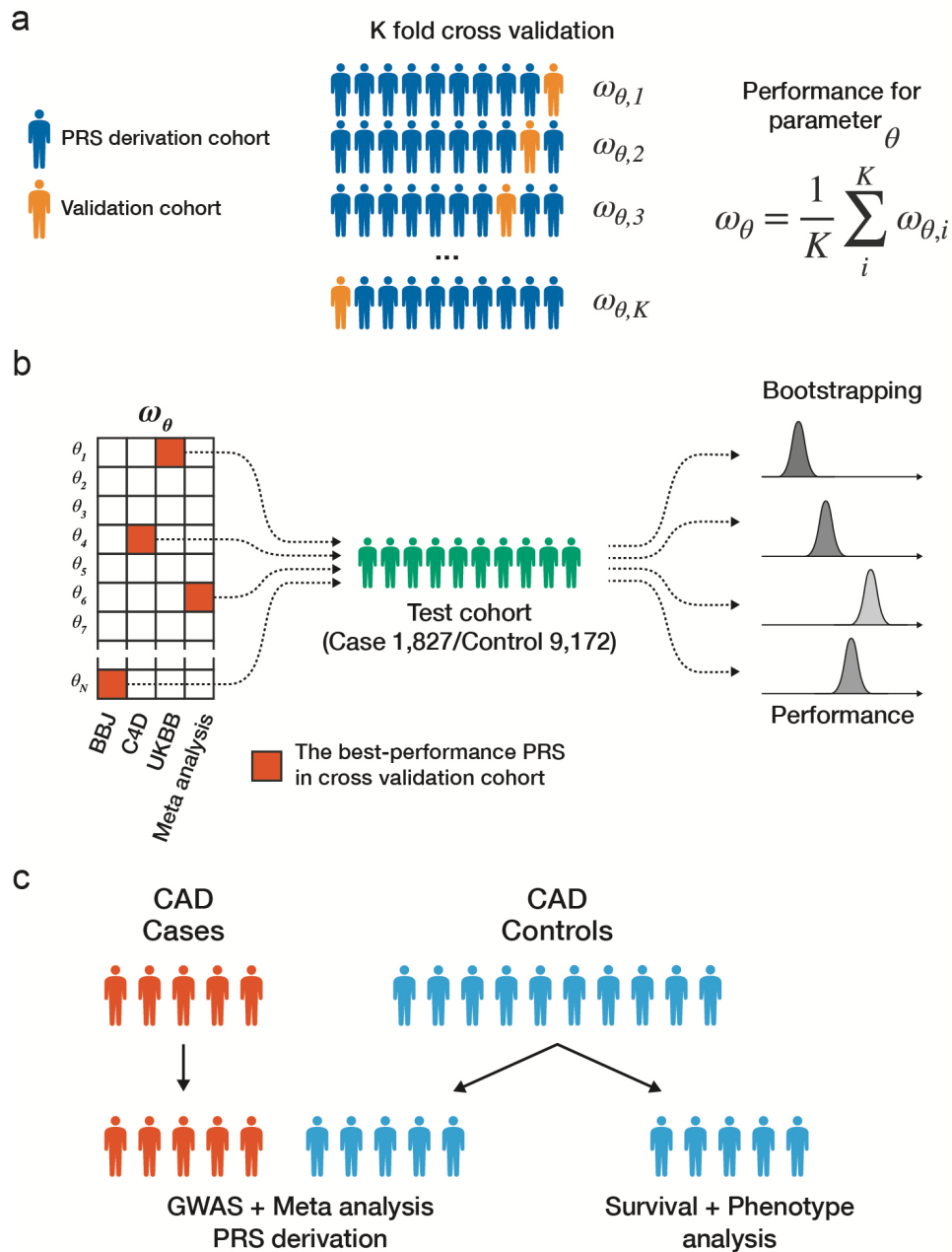
Supplementary Figures



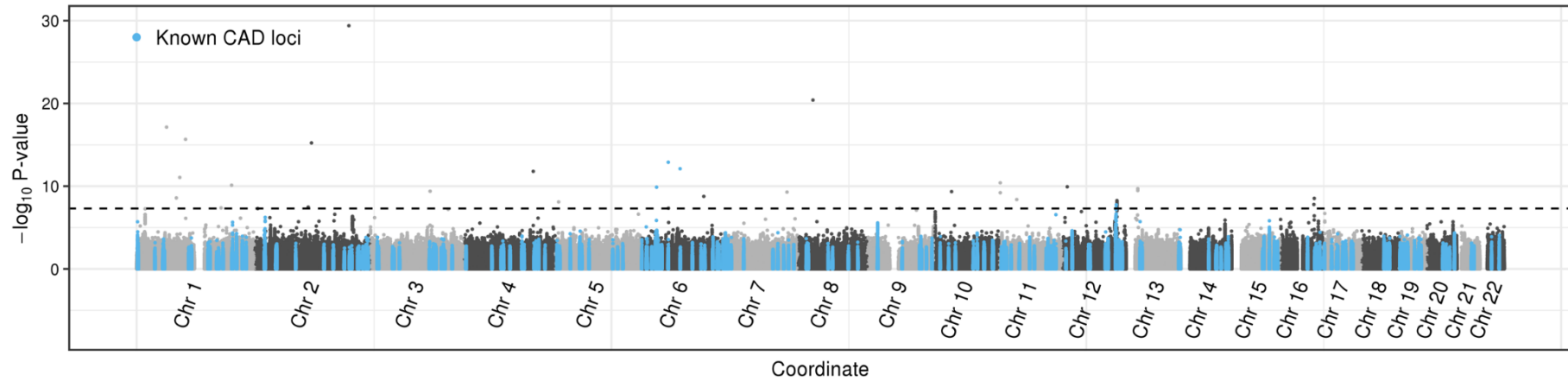
Supplementary Fig. 1 | Overview of the study. **a**, Overview of the Japanese cohorts and GWAS. **b**, Overview of trans-ancestry meta-analysis. JPHC, Japan Public Health Center-based prospective; J-MICC, Japan Multi-Institutional Collaborative Cohort; OACIS, Osaka Acute Coronary Insufficiency Study; CAD, coronary artery disease; MAF, minor allele frequency; GWAS, genome-wide association study; PRS, polygenic risk score; SNV, single nucleotide variant; BF Bayes factor.



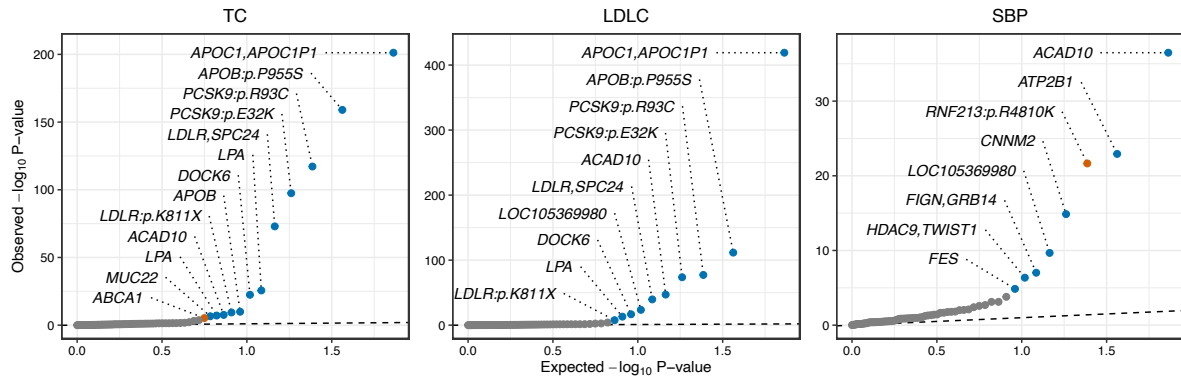
Supplementary Fig. 2 | Quality control and replication analysis for the Japanese GWAS. a. Mean χ^2 values for LD score bins are plotted. LD-score was estimated from the Japanese whole-genome sequencing data. **b.** Each point indicates a novel signal discovered in the Japanese GWAS. The X-axis shows beta coefficient in the current Japanese GWAS, and Y-axis shows beta coefficient in the replication study from the European studies. Error bars indicate the corresponding 95% confidence intervals. Variants with $F_{\text{st}} > 0.1$ were colored in orange, and others were colored in blue. LD, linkage disequilibrium; BBJ, Biobank Japan. F_{st} , fixation index.



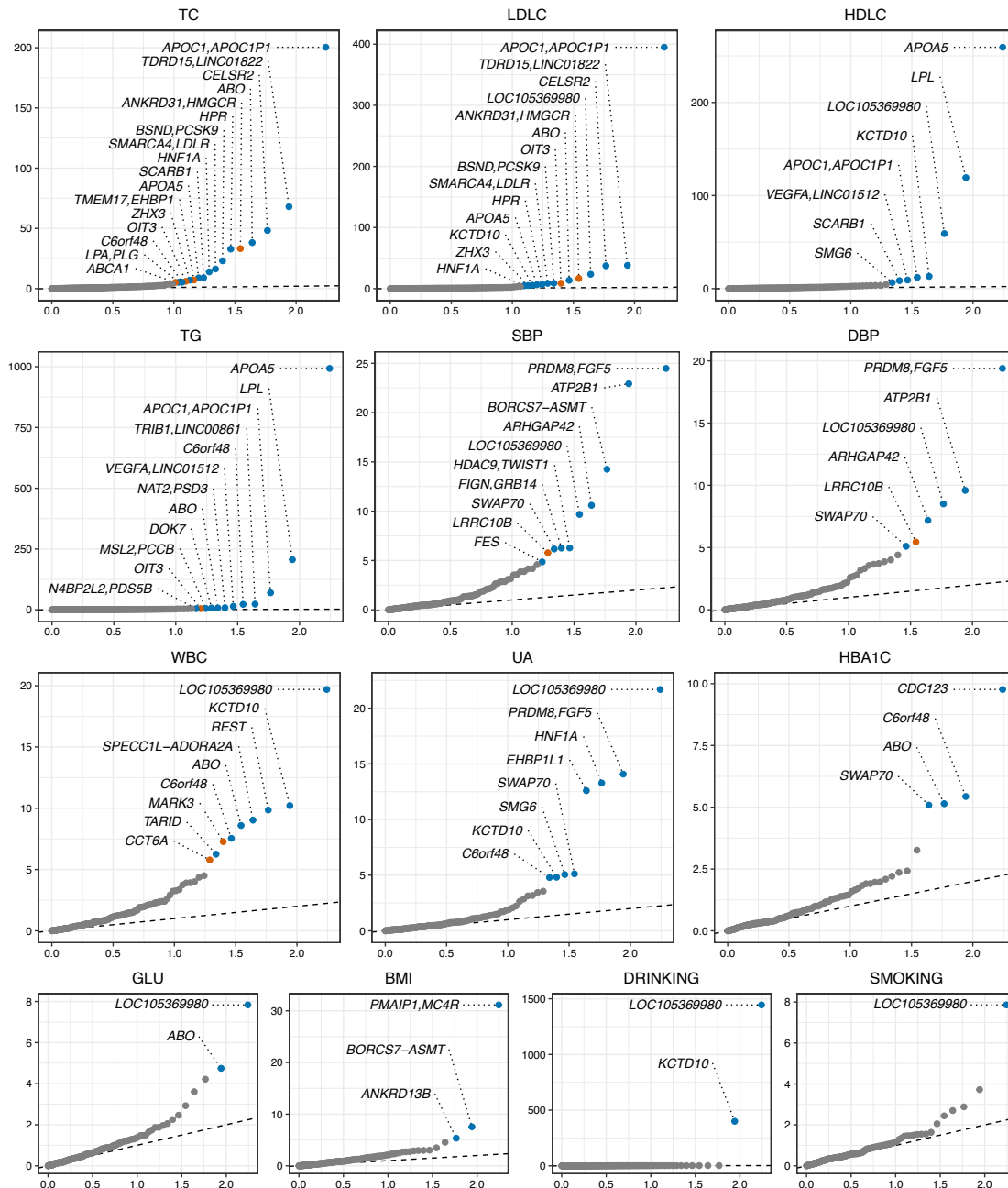
Supplementary Fig. 3 | Analytical scheme for PRS derivation, performance validation and test. Schematic representation for **a)** K-fold cross-validation, **b)** performance testing of the CAD-PRS in the independent cohort, **c)** downstream analysis of CAD-PRS. PRS, polygenic risk score; BBJ, Biobank Japan; C4D, CARDIoGRAMplusC4D; UKBB, UK Biobank. CAD, coronary artery disease; GWAS, genome-wide association study.



Supplementary Fig. 4 | Case-control association study using whole-genome sequence data. The results of the case-control association study, including 1,781 CAD cases and 2,636 controls, are shown. The negative $\log_{10} P$ -values are shown against the genomic coordinates. Variants in previously reported loci are presented in blue. Two-sided P -values were calculated using a logistic regression model. The dashed line indicates a genome-wide significant threshold ($P = 5 \times 10^{-8}$).



Supplementary Fig. 5 | The pleiotropic effects of distinct signals. Quantile-quantile plots of the association analysis of 73 independent signals determined by the Japanese GWAS for TC ($n = 134,224$), LDLC ($n = 76,477$), and SBP ($n = 139,847$) in the BBJ dataset. The X-axis indicates expected negative $\log_{10} P$ -values, and Y-axis indicates observed negative $\log_{10} P$ -values. Associations with Bonferroni adjusted significance ($P < 0.05/2,482$) in known loci are plotted in blue and in novel loci are plotted in orange. Dashed lines indicate theoretical distributions. Two-sided P -values were calculated using a linear regression model. Number of individuals included each analysis and exact P -values are found in Supplementary Table 20. TC, total cholesterol; LDLC, low-density lipoprotein cholesterol; SBP, systolic blood pressure.



Supplementary Fig. 6 | Pleiotropy of lead variants detected in trans-ancestry meta-analysis. Quantile-quantile plots of the association analysis of 175 lead variants determined by trans-ancestry meta-analysis for 13 CAD-PRS associated traits. The X-axis indicates expected negative $\log_{10}$ P -values, and Y-axis indicates observed negative $\log_{10}$ P -values. Associations with Bonferroni adjusted significance ($P < 0.05/2,275$) in known loci are plotted in blue and in novel loci are plotted in orange. Dashed lines indicate theoretical distributions. Two-sided P -values were calculated using a linear regression model. Number of individuals included each analysis and exact P -values are found in Supplementary Table 22. Abbreviations for the phenotypes are defined in Supplementary Table 19.

Supplementary Note

Supplementary Notes

GWAS quality control and replication. In the current Japanese GWAS, the lambda GC value was slightly inflated at 1.09 (all variants, 1.24 for HapMap3 SNPs). The linkage disequilibrium (LD) score regression intercept ¹ was 1.035, indicating that 89.3% of the inflation was likely due to the polygenic nature of Coronary artery disease (CAD, Supplementary Fig. 2a). Seven of the eight lead variants in the novel loci were previously reported in the European studies, and all variants demonstrated concordant direction of allelic effects with the current study (Supplementary Fig. 2b), with six of the seven variants exhibiting nominal association with CAD. Alternatively, rs112735431, the lead variant in *RNF213* locus, was not previously reported as rs112735431 is rare among European populations. Recently, the association between rs112735431 and CAD was reported by a target genotyping study on the Japanese population ².

Comparison with previous studies. Recently, two genome-wide association studies (GWAS) for CAD-related phenotypes were reported from Biobank Japan project (BBJ) ^{3,4}. Compared to Matsunaga et. al. ³ (15,302 cases and 36,140 controls in the Japanese GWAS and 91,316 cases and 300,925 controls in the trans-ancestry meta-analysis), the current study has an increased number of subjects (25,892 cases and 142,336 controls in the Japanese GWAS and 121,234 cases and 527,824 controls in the trans-ancestry meta-analysis) and discovered novel loci. On the other hand, Ishigaki et. al.'s study ⁴ included more subjects than the current study (29,139 cases and 183,134 controls for the Japanese GWAS). We identified three additional CAD-associated loci which showed borderline significance in Ishigaki et. al.'s study ⁴ (rs2380472 $P = 4.2 \times 10^{-6}$, rs35093463 $P = 8.1 \times 10^{-7}$, rs9951447 $P = 2.9 \times 10^{-7}$). In addition to the difference in the composition of the control samples (the current study contains controls from BBJ, Ishigaki et. al.'s study contains BBJ and other population-based cohorts), we consider the difference in the imputation quality attributed to these different results. One example can be found in the variant rs2380472 on the chromosome 8. We found that the imputation quality of this variant was improved by using our Japanese-specific reference panel ($r^2 = 0.978$ (the current study), and $r^2 = 0.879$ (Ishigaki et. al.'s study)). We also observed that the significance of the variant was decreased ($P = 3.5 \times 10^{-6}$) when we imputed this region in the same case-control samples as our study by 1KG reference panel used in Ishigaki et. al.'s study.

Shared loci between CAD and moyamoya disease (MMD). To elucidate the genetic similarities between CAD and MMD, we compared the results of recently reported GWAS of MMD ⁵ with those of the current study. We found that these two diseases also shared the rs2107595 in *HDAC9/TWIST1* locus at a genome-wide significance with the same direction ($\beta_{\text{MMD}} = 0.490$, $P_{\text{MMD}} = 1.5 \times 10^{-29}$, $\beta_{\text{CAD}} = 0.058$, $P_{\text{CAD}} = 1.7 \times 10^{-8}$). rs2107595 is an eQTL variant

of *TWIST1* in the Aorta ($\beta_{TWIST1} = 0.23$, $P = 5.7 \times 10^{-7}$)⁶. This association is concordant with recent research, which showed that the increased expression of *TWIST1* in the vascular wall resulted in increased atherosclerotic burden⁷. Besides, rs9916351, an intronic variant of *RNF213*, which is also significant in the MMD study, is associated with CAD with the same allelic effect direction. The risk allele for CAD/MMD (T allele) at rs9916351 cause decreased expression of *RNF213* in various tissues, including arteries (Artery – Tibial, $\beta_{RNF213} = -0.21$, $P = 9.8 \times 10^{-8}$). This association is concordant with the evidence that *rnf213* knockdown in zebrafish mimics abnormal vascular formation of MMD⁸. From this evidence, we speculate that the abnormal function or expression of *RNF213* and *TWIST1* in the vessel wall might be involved in the pathogenic pathway shared between CAD and MMD.

Pleiotropic effect of the lead variants in the Japanese GWAS. To explore the biological pathways in which conditionally-independent signals play roles in CAD development, we performed association analyses for 34 clinical indices, including clinical measurements (body mass index (BMI) and blood pressure), serum laboratory measurements (e.g., Total cholesterol (TC), low-density lipoprotein cholesterol (LDLC)) and lifestyles (cigarette smoking and alcohol drinking) in the BBJ dataset. We found 120 significant associations ($P < 0.05 / 2,482$) in 29 phenotypes among the 34 tested (Supplementary Fig. 5, Supplementary Table 20). The most frequent association was found between serum TC and these variants. The CAD risk-increasing alleles of these variants are completely matched with the TC increased allele (13/13).

High impact coding and splicing variants on the *LDLR*. In addition to rs87925521 (*LDLR* p.K811X), rs553577764, a rare intronic variant in *DOCK6* found in *LDLR* – *SPC24* locus with a high OR_{CAD} , also showed a large impact on serum TC levels (OR_{CAD} (95% CI) = 3.08 (2.22 – 4.26), $P = 1.2 \times 10^{-11}$, β_{TC-raw} (SE) = 45.9 (3.39) mg/dL per allele, $P_{TC-adjusted} = 3.7 \times 10^{-23}$). We noted that this variant was in high LD ($r^2 = 0.639$) with rs746959386 (*LDLR* p.L568V), which is registered in the ClinVar database as “Likely pathogenic” for familial hypercholesterolemia (FH) (MAF = 0.057%, OR_{CAD} (95% CI) = 3.11 (2.21 – 4.37), $P = 4.8 \times 10^{-11}$, β_{TC-raw} (SE) = 49.9 (3.54) mg/dL per allele, $P_{TC-adjusted} = 1.6 \times 10^{-22}$, Supplementary Table 21). The association between CAD and rs553577764 (*DOCK6* intron) was not significant ($P_{Conditioned} = 0.06$) when we conditioned on rs746959386 (*LDLR* p.L568V). We also noted that a splicing variant of *LDLR* showed a nominal association with CAD development (rs778408161, MAF = 0.074%, OR_{CAD} (95% CI) = 2.15 (1.44 – 3.21), $P = 1.6 \times 10^{-4}$) and a significant association with TC levels (β_{TC-raw} (SE) = 34.6 (3.83) mg/dL per allele, $P_{TC-adjusted} = 3.7 \times 10^{-11}$).

Pleiotropic effect of the lead variants in the trans-ancestry meta-analysis. To assess the related loci of 13 polygenic risk score for CAD (CAD-PRS) associated traits, we performed association analyses for the 175 lead variants detected in the trans-ancestry meta-analysis and the 13 CAD-PRS-associated traits in the BBJ dataset. Ninety-six variant-phenotype pairs showed

Bonferroni-corrected significance ($P < 0.05/2,275$, Supplementary Fig. 6, Supplementary Table 22). Serum TC level was the most frequently associated trait (17 associations), and the directions of the allelic effect (CAD-risk and TC-increasing) were completely concordant (17/17).

Gene prioritization improved the performance of causal gene inference. We assessed the performance of the gene prioritization algorithm developed in the current study by counting the number of overlapping genes with drug target ones. First, we curated genes encoding drugs target proteins ($n = 1,506$), and then detected overlaps with genes prioritized by our algorithm or genes nearest to the genome-wide signals. We found that 48 genes were overlapped with prioritized genes by our algorithm. On the other hand, only 28 drug target genes were overlapped with nearest genes ($P = 0.013$, Fisher's exact test). We also observed marginal significant increase in the number of overlapping genes with cardiovascular drug target protein-encoding ones ($n = 378$; 22 overlapped with prioritized genes, and 12 overlapped with nearest genes, $P = 0.10$). When we consider the genes encoding interacting proteins with cardiovascular drug target ones, we also observed significantly increased overlap with prioritized genes than nearest genes (88 prioritized genes and 59 nearest genes, $P = 2.3 \times 10^{-3}$). Based on these results, we consider our algorithm successfully identified biologically relevant, and potentially causal genes in given loci.

Supplementary Methods

Gene prioritization. To identify the most likely causal gene in the locus detected by the trans-ancestry meta-analysis, we conducted gene prioritization analysis. We scored each gene in the genome-wide significant loci based on the following evidence. 1) Coding evidence: A, genes which have the lead variant as a damaging variant (missense variant, in-frame/frameshift indels, stop-gain, stop-loss, and splice site). B, genes which have one or more damaging variant(s) in linkage ($r^2 > 0.9$, in the Japanese/1KGEUR/1KGEAS) with the lead variants in the Japanese/Trans-ancestry GWAS. C, genes significantly associated with CAD by the gene-based test. 2) Mouse phenotype evidence: genes with CAD-related phenotypes indicated in MGI database⁹. We extracted CAD-related phenotypes by querying the following MP category. Artery, MP:0002191; abnormal artery morphology. Lipid, MP:0010330; abnormal circulating lipoprotein level, MP:0003949; abnormal circulating lipid level, MP:0002118; abnormal lipid homeostasis. Glucose, MP:0000188; abnormal circulating glucose level, MP:0002078, abnormal glucose homeostasis. Uric acid, MP:0008820, abnormal blood uric acid level. Blood pressure, MP:0000230; abnormal systemic arterial blood pressure. BMI, MP:0001260; increased body weight. WBC, MP:0002442; abnormal leukocyte physiology. 3) Publication evidence: genes whose publications were significantly enriched in CAD. Publications related to CAD is curated by pubTator¹⁰ pipeline by queuing "MeSH:D003324". We tested the enrichment of CAD-related publication by the hypergeometric test and assessed 1,557 genes with a sufficient number of

publications (more than 100 CAD-related publications). The significance level was set at $P = 0.05/1,557$. 4) regulatory evidence: genes attained the highest regulatory score in the locus. We obtained the regulatory scores for each gene through the open target genetics, variant to gene pipeline¹¹. Briefly, in this pipeline, the score was assigned to each combination between the lead variant and gene in the locus of interest, based on the evidence of expression QTL, protein QTL, PCHI-C, DHS-promoter correlation, and enhancer-TSS correlation. 5) DEPICT gene prioritization: genes prioritized in the gene prioritization analysis implemented in DEPICT software¹² with FDR < 0.01. The gene prioritization analysis by DEPICT was performed using the summary statistics of trans-ancestry meta-analysis. Here, we also used a threshold of $\log_{10} BF > 5$, the same as that used for tissue and gene-set enrichment analyses. 6) Pathway evidence: genes belonging to pathways that demonstrated significance in gene-set analysis. We scored each gene based on the number of criteria that were met (range 0 – 6) and added a numerical value of regulatory evidence (range 0 – 1). Those with the highest score (range 0 – 7) in the locus were identified as a candidate causal gene.

Drug target genes and protein-protein interaction We extracted the drug-target genes from Therapeutic Target Database (<http://bidd.nus.edu.sg/group/cjttd/>)¹³. Cardiovascular drugs were defined as a composition of drugs targeting ICD-11 diseases of the circulatory system (BB01 – BE2Z) and disorders of lipoprotein metabolism or certain specified lipidaemias (5C80, 5C81, 5C8Y, and 5C8Z). Protein-protein interaction was defined by combined score ≥ 0.9 from STRING database (<https://string-db.org>)¹⁴.

Supplementary References

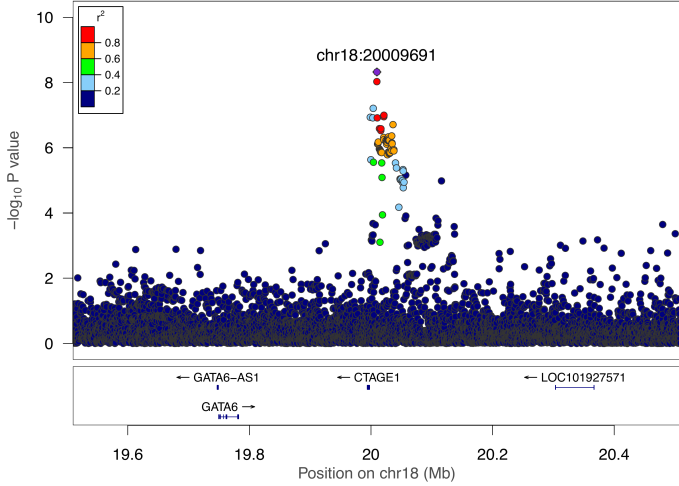
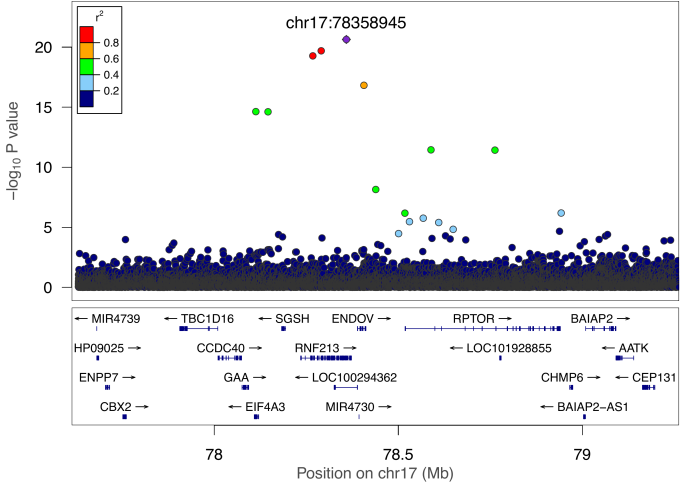
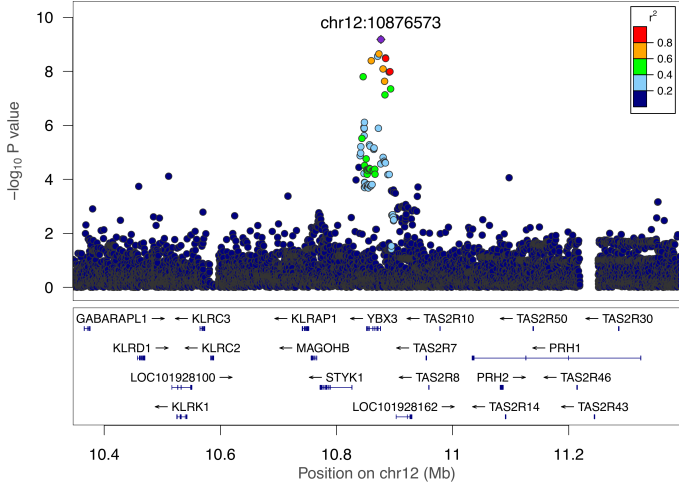
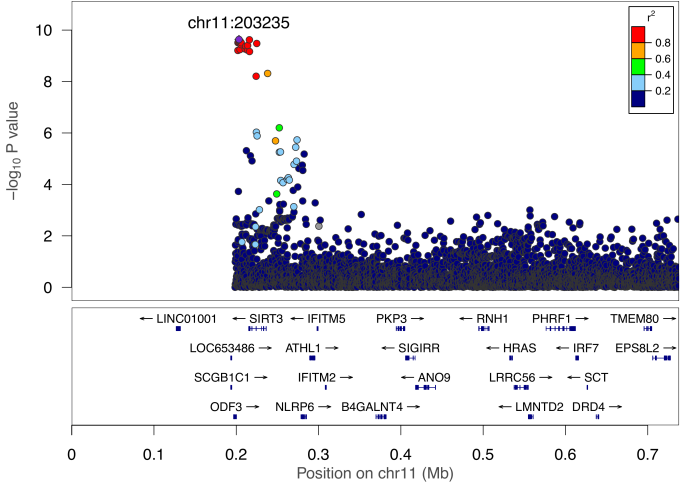
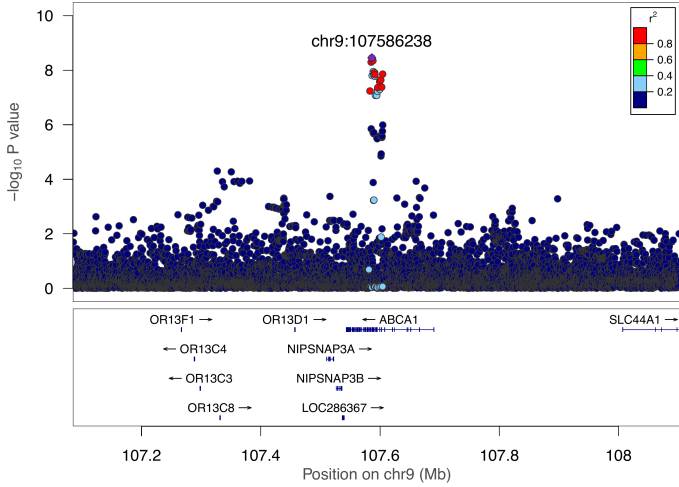
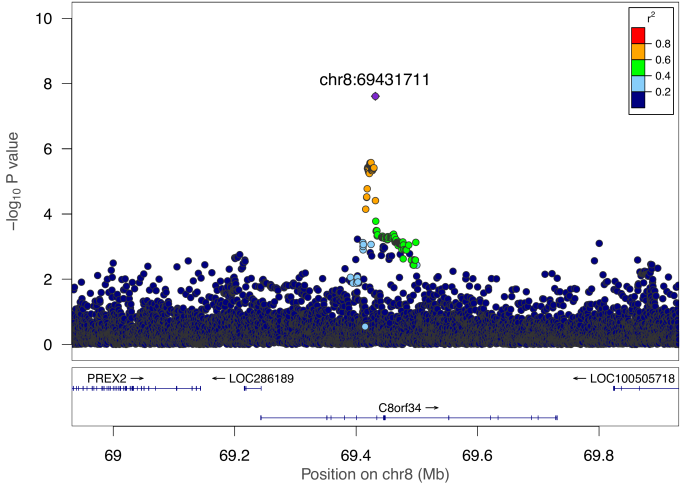
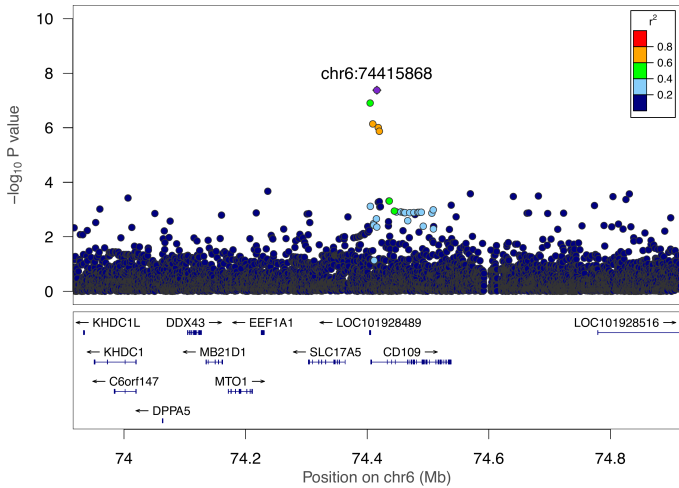
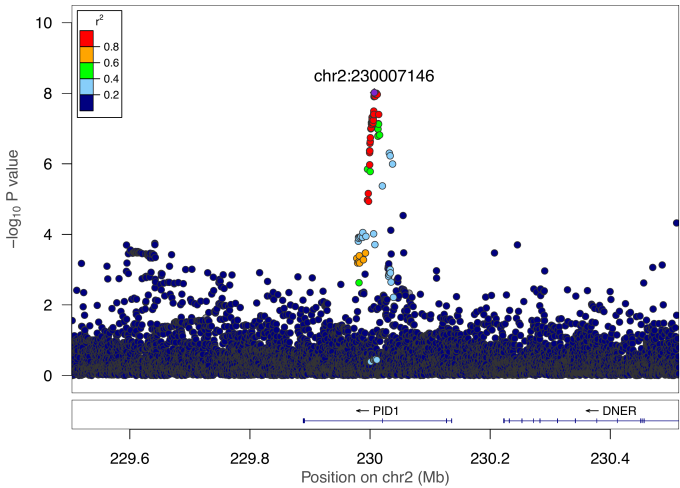
1. Bulik-Sullivan, B. *et al.* LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat. Genet.* **47**, 291–295 (2015).
2. Morimoto, T. *et al.* Significant association of RNF213 p.R4810K, a moyamoya susceptibility variant, with coronary artery disease. *PLoS One* **12**, e0175649 (2017).
3. Matsunaga, H. *et al.* Transethnic Meta-Analysis of Genome-Wide Association Studies Identifies Three New Loci and Characterizes Population-Specific Differences for Coronary Artery Disease. *Circ. Genom. Precis. Med.* **13**, e002670 (2020).
4. Ishigaki, K. *et al.* Large-scale genome-wide association study in a Japanese population identifies novel susceptibility loci across different diseases. *Nat. Genet.* **52**, 669–679 (2020).
5. Duan, L. *et al.* Novel Susceptibility Loci for Moyamoya Disease Revealed by a Genome-Wide Association Study. *Stroke* **49**, 11–18 (2018).
6. GTEx Consortium *et al.* Genetic effects on gene expression across human tissues. *Nature* **550**, 204–213 (2017).
7. Nurnberg, S. T. *et al.* Genomic profiling of human vascular cells identifies *TWIST1* as a causal gene for common vascular diseases. *PLoS Genet.* **16**, e1008538 (2020).
8. Liu, W. *et al.* Identification of *RNF213* as a susceptibility gene for moyamoya disease and its possible role in vascular development. *PLoS One* **6**, e22542 (2011).

9. Bult, C. J. *et al.* Mouse genome database (MGD) 2019. *Nucleic Acids Res.* **47**, D801–D806 (2019).
10. Wei, C.-H., Allot, A., Leaman, R. & Lu, Z. PubTator central: automated concept annotation for biomedical full text articles. *Nucleic Acids Res.* **47**, W587–W593 (2019).
11. Carvalho-Silva, D. *et al.* Open Targets Platform: new developments and updates two years on. *Nucleic Acids Res.* **47**, D1056–D1065 (2019).
12. Pers, T. H. *et al.* Biological interpretation of genome-wide association studies using predicted gene functions. *Nat. Commun.* **6**, 5890 (2015).
13. Wang, Y. *et al.* Therapeutic target database 2020: enriched resource for facilitating research and early development of targeted therapeutics. *Nucleic Acids Res.* **48**, D1031–D1041 (2020).
14. Szklarczyk, D. *et al.* STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res.* **47**, D607–D613 (2019).

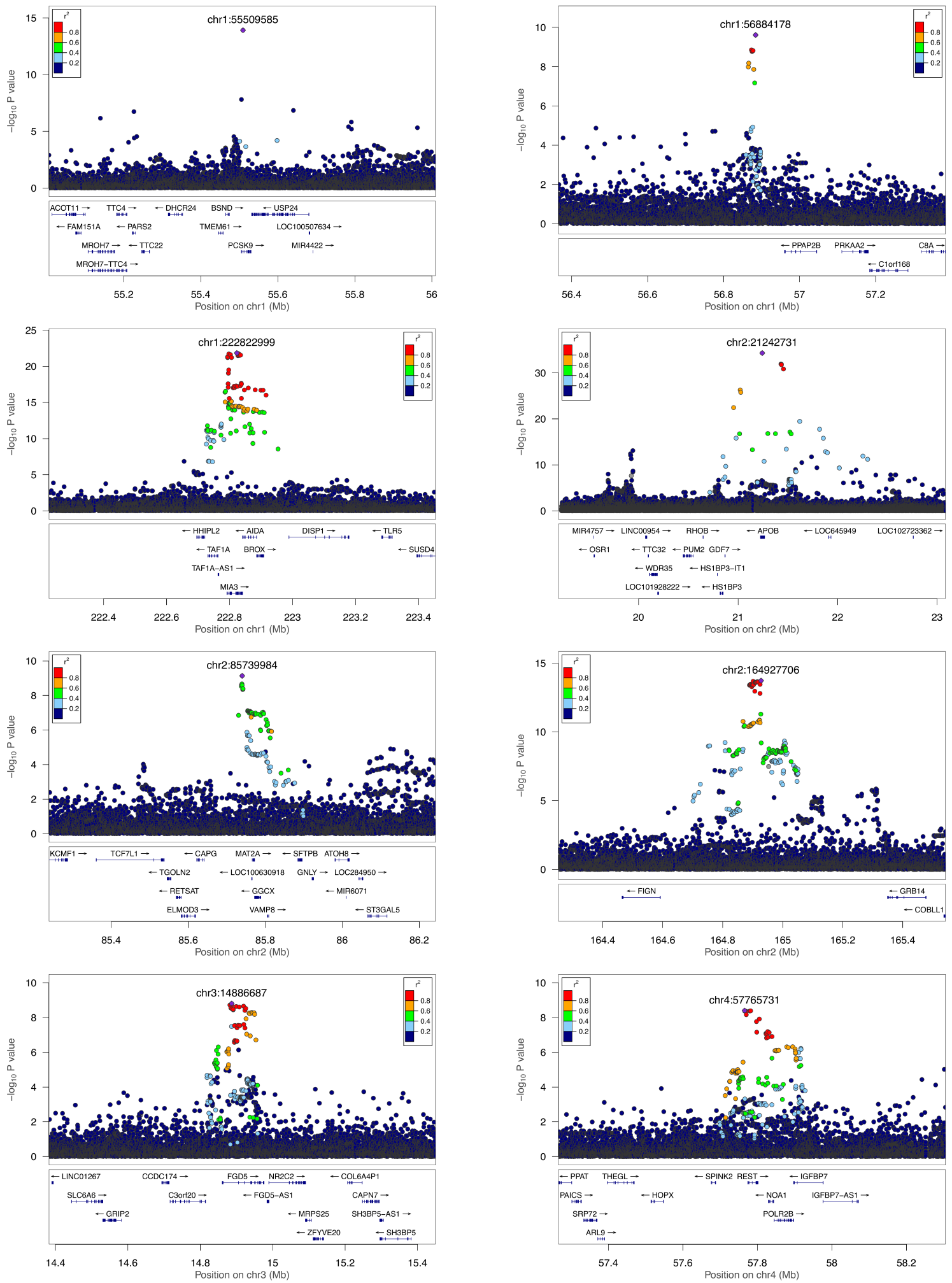
Supplementary Dataset

1. Novel loci in the Japanese GWAS
2. Previously reported loci in the Japanese GWAS
3. Novel loci in the Trans-ancestry meta-analysis
4. Previously reported loci in the Trans-ancestry meta-analysis

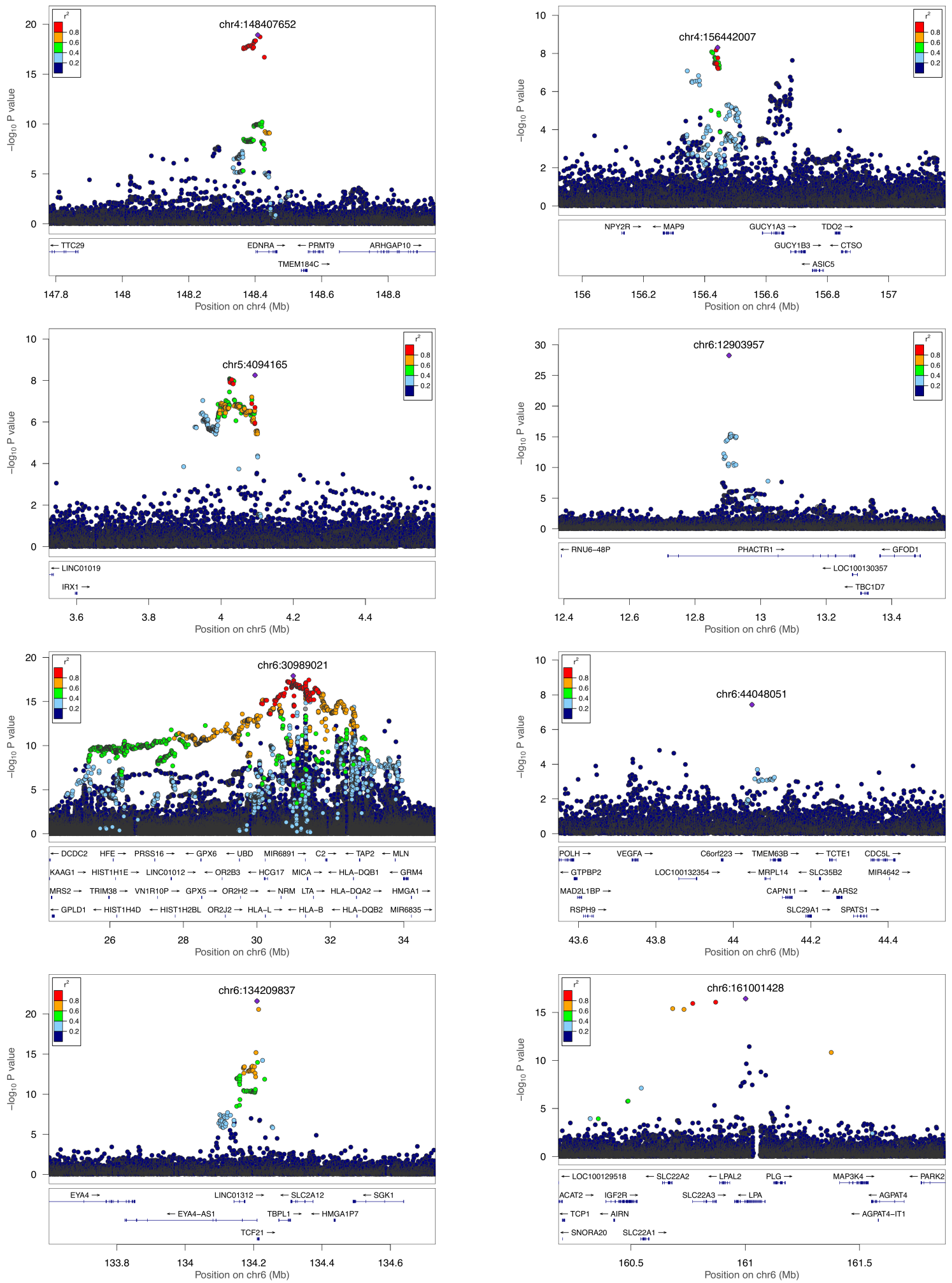
1. Novel loci in the Japanese GWAS



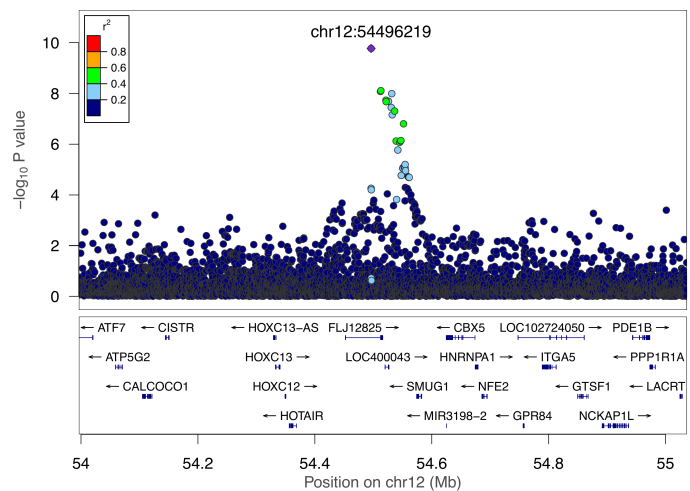
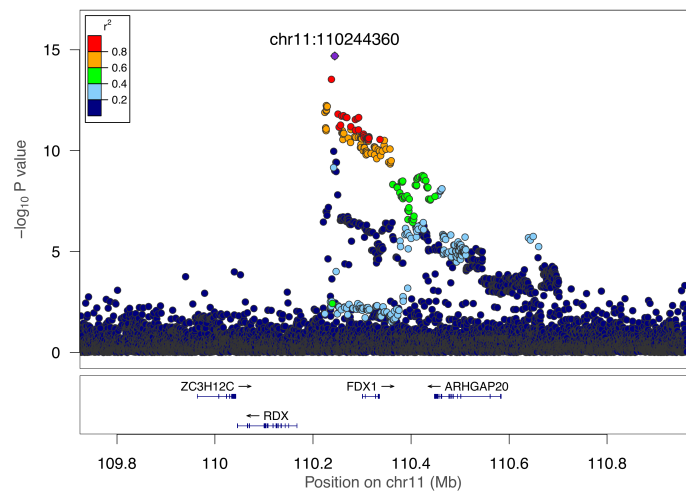
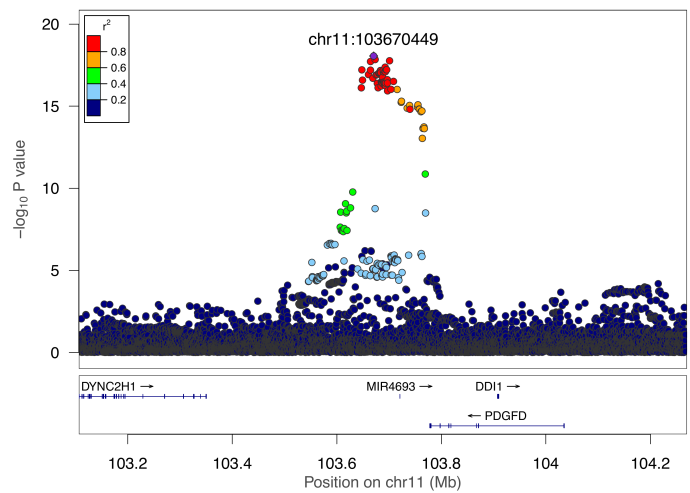
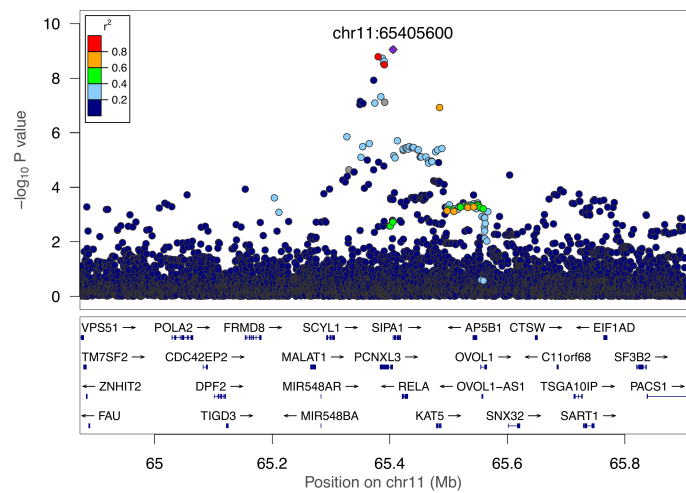
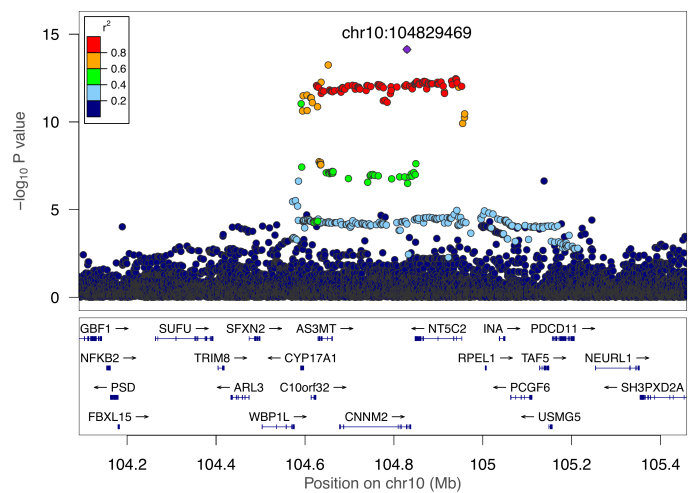
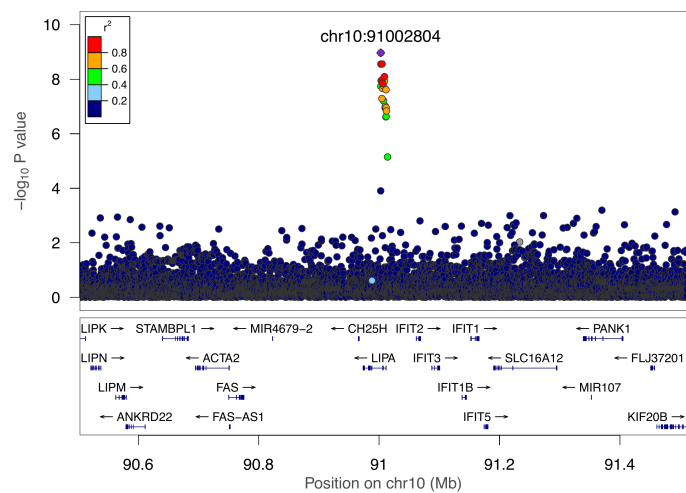
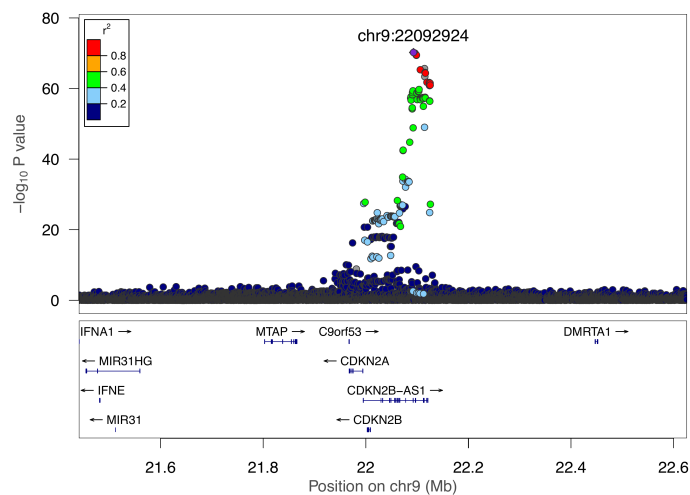
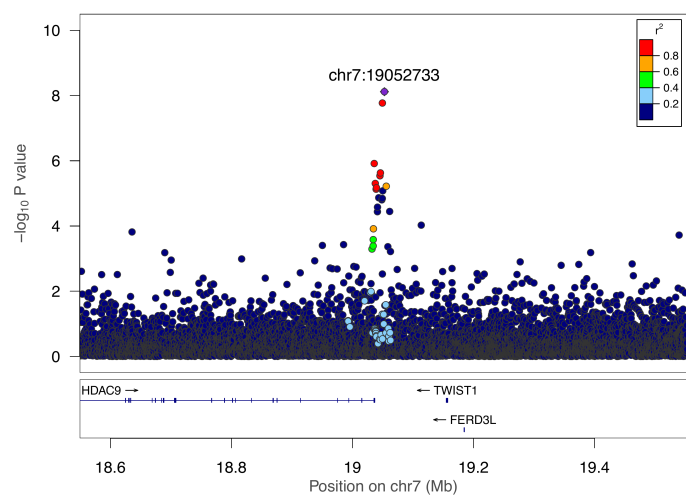
2. Previously reported loci in the Japanese GWAS



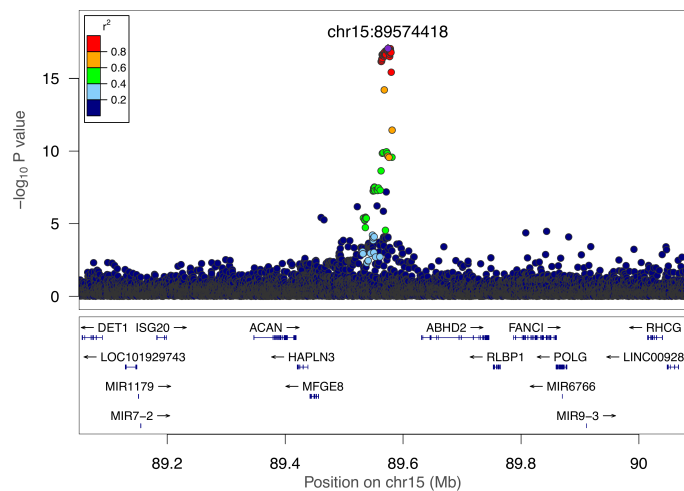
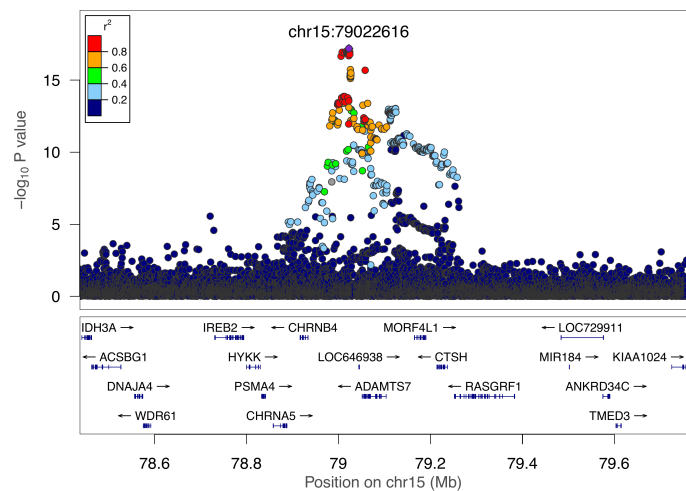
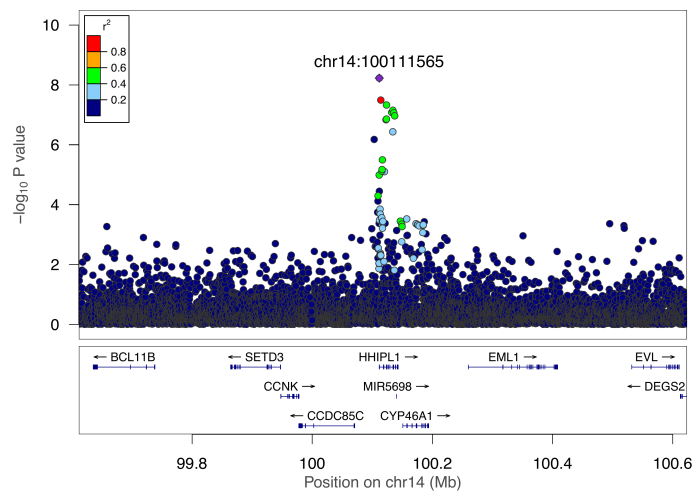
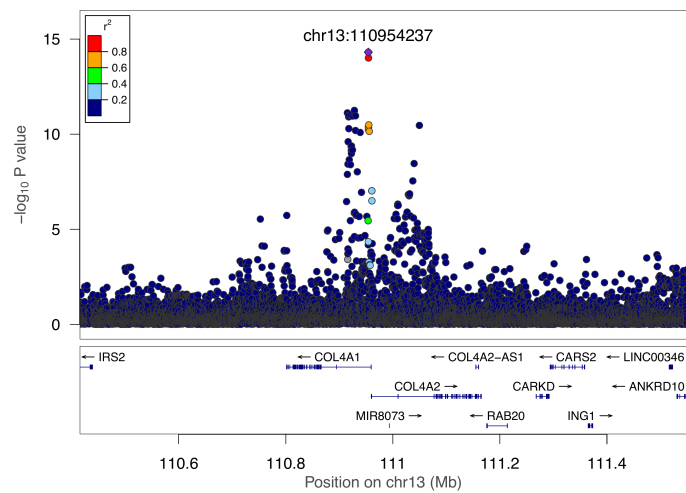
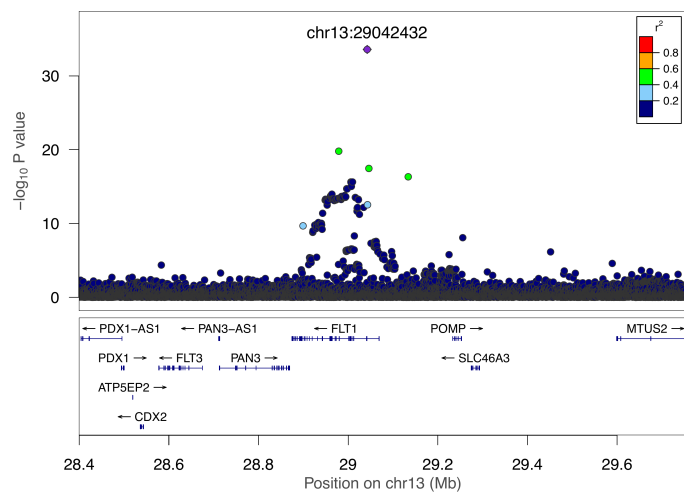
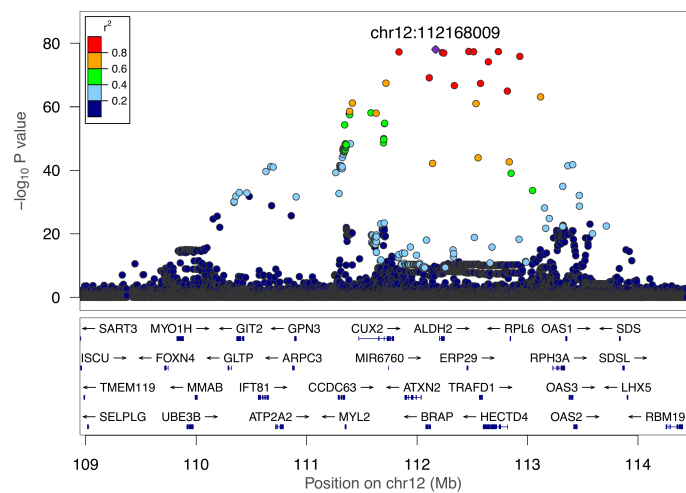
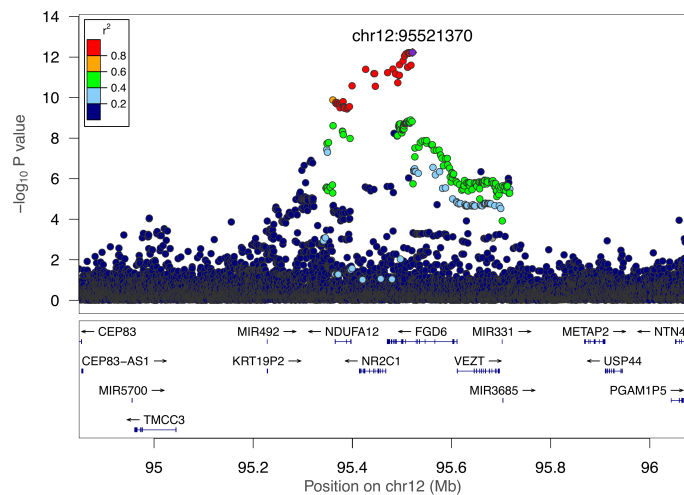
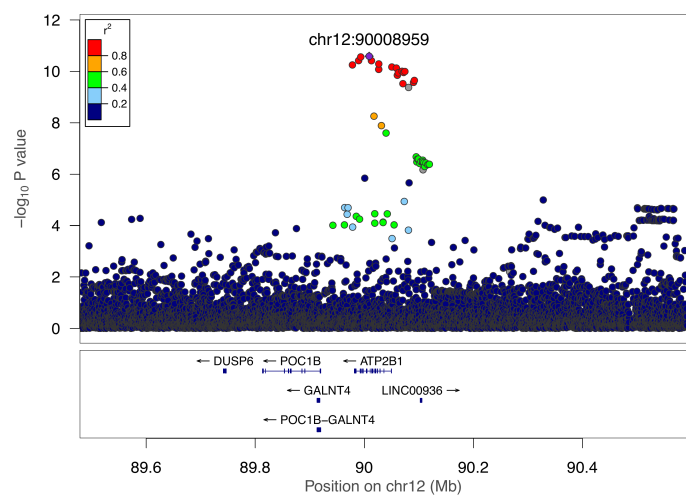
2. Previously reported loci in the Japanese GWAS



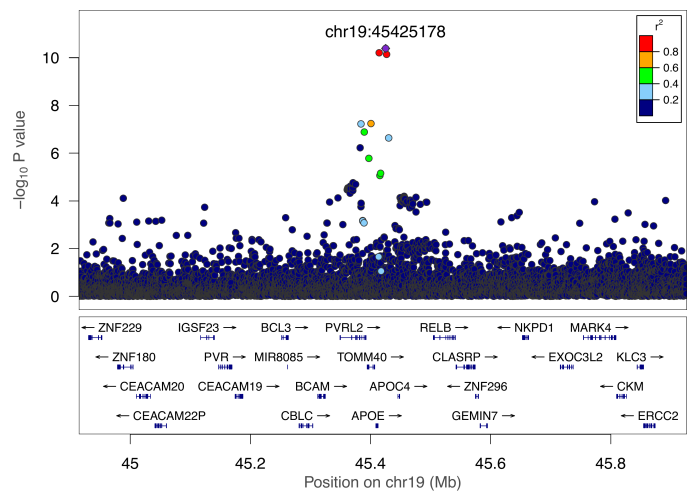
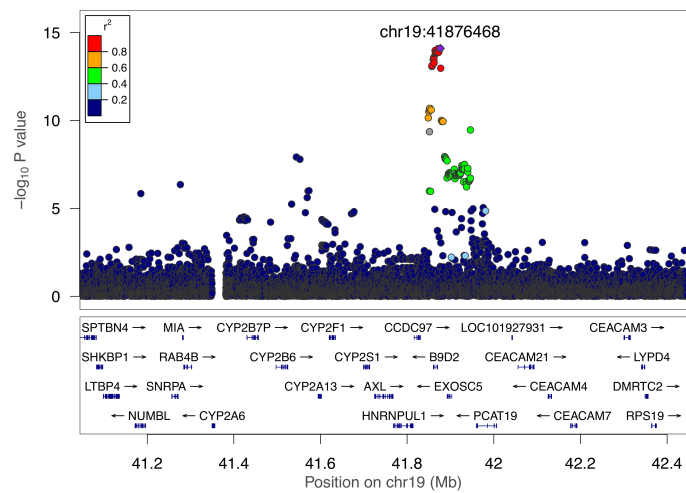
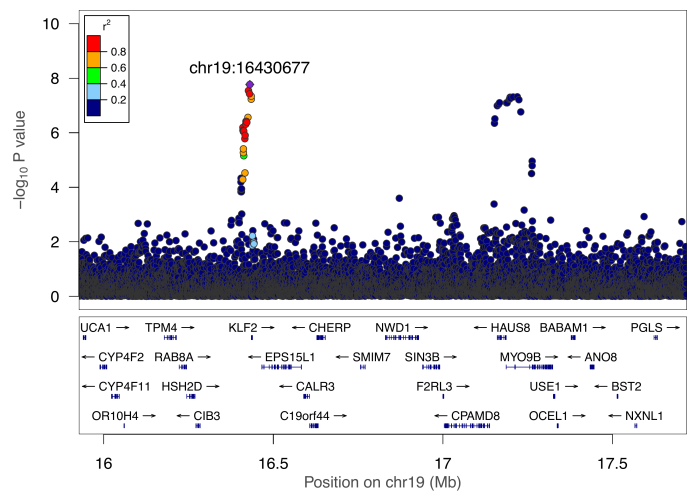
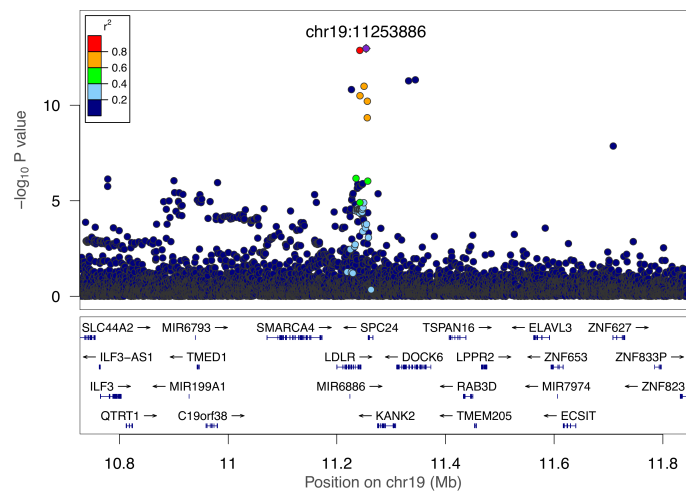
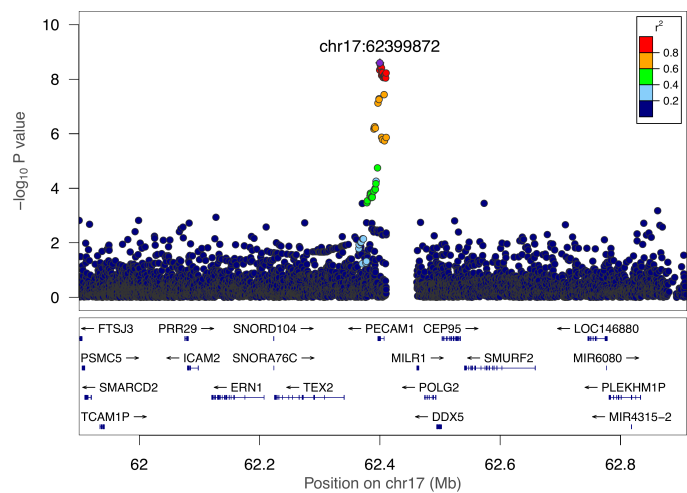
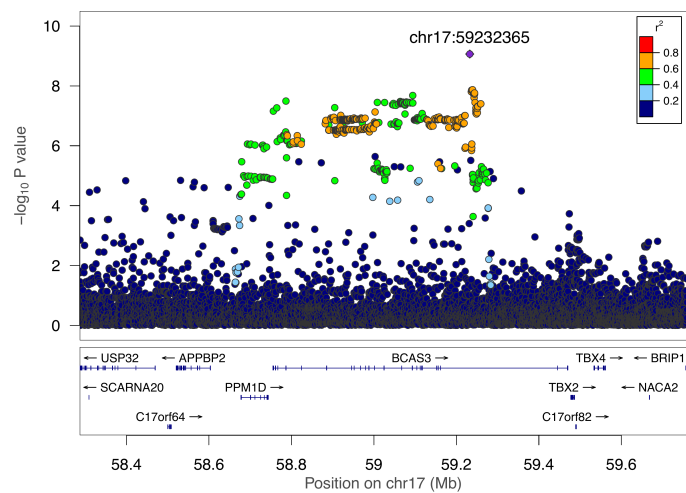
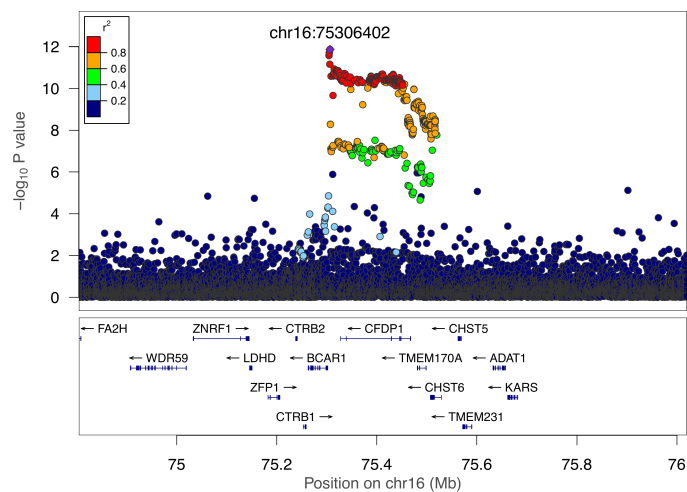
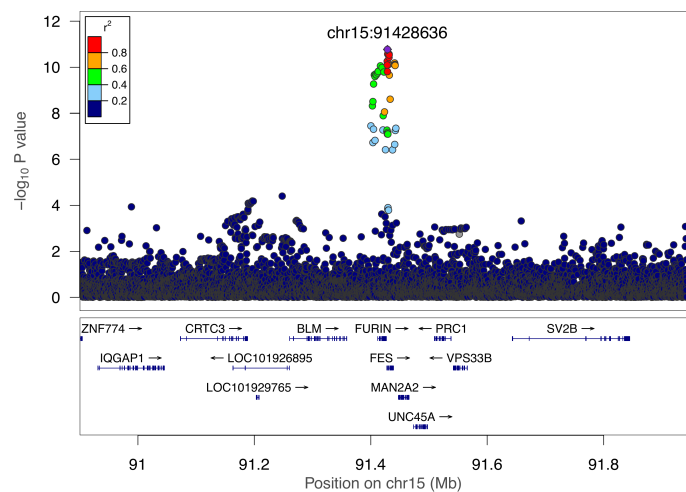
2. Previously reported loci in the Japanese GWAS



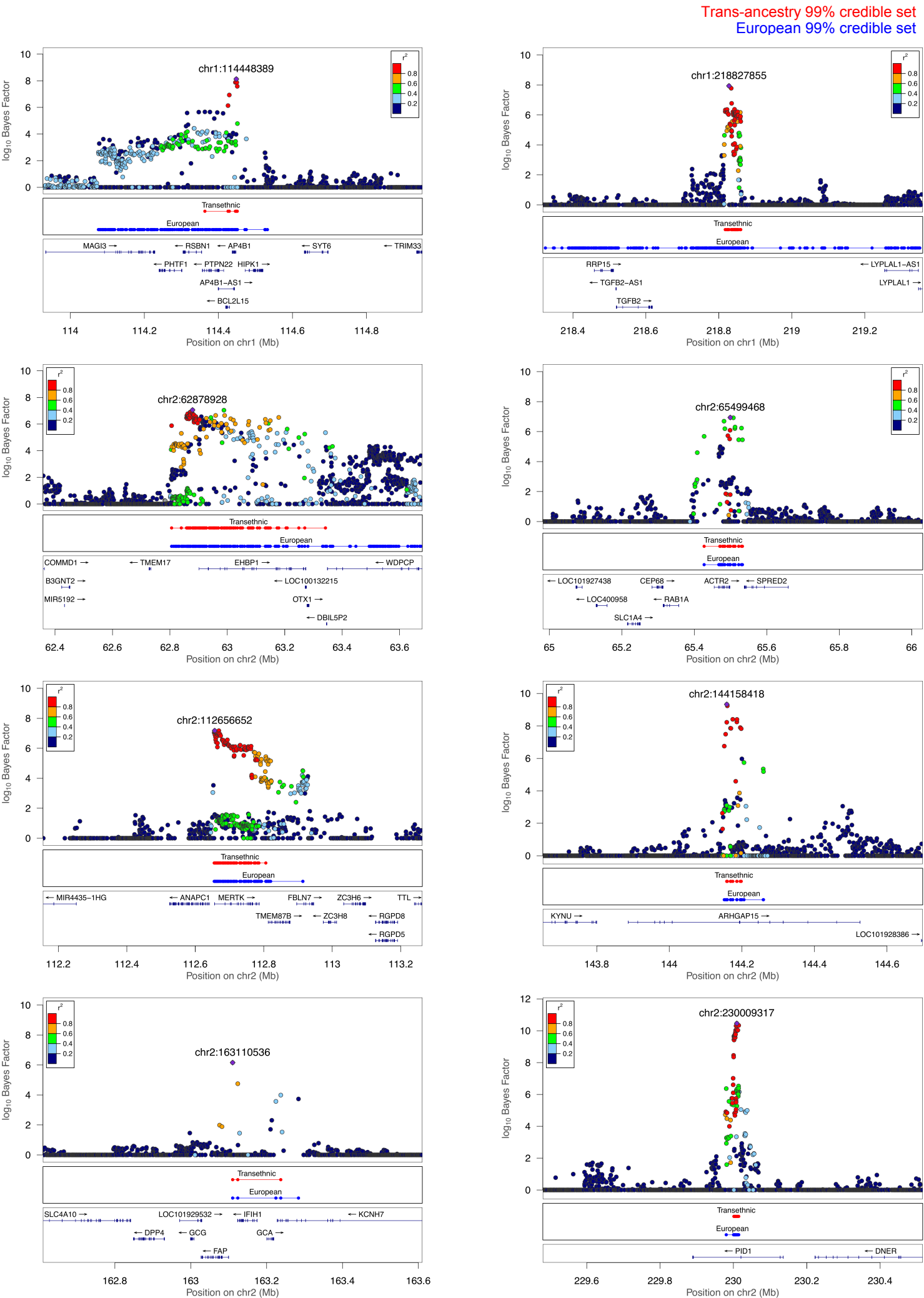
2. Previously reported loci in the Japanese GWAS



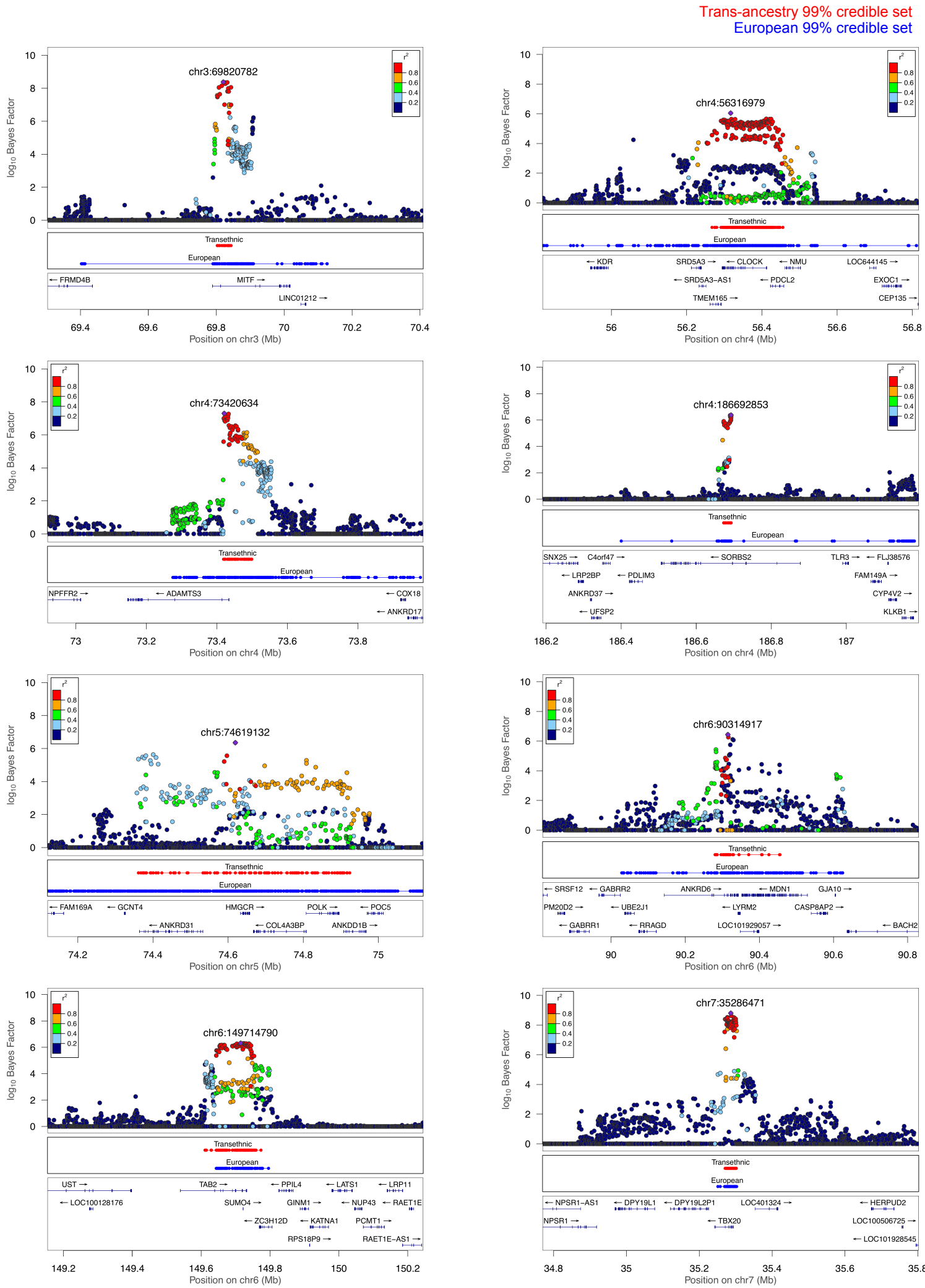
2. Previously reported loci in the Japanese GWAS



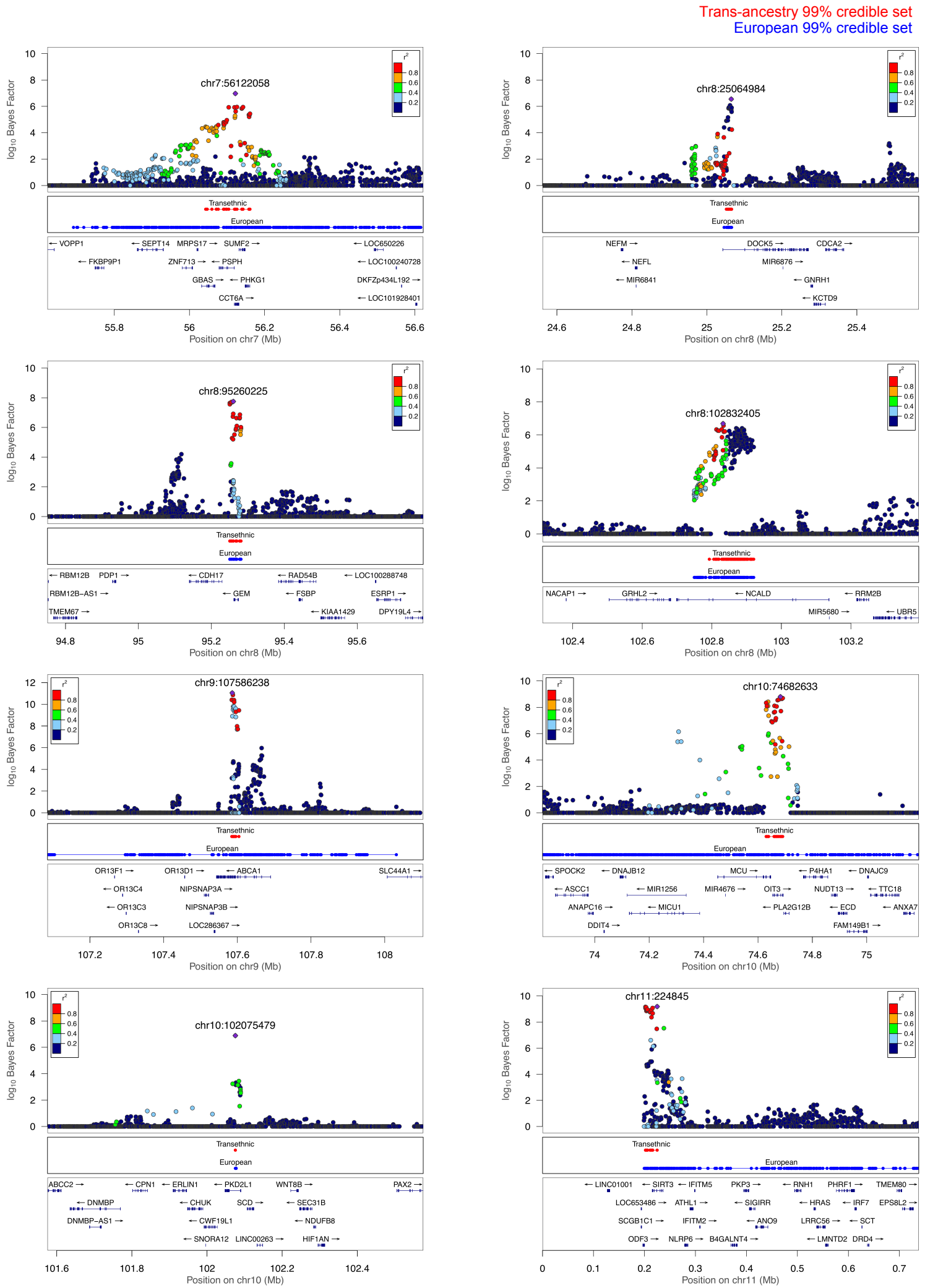
3. Novel loci in the Trans-ancestry meta-analysis



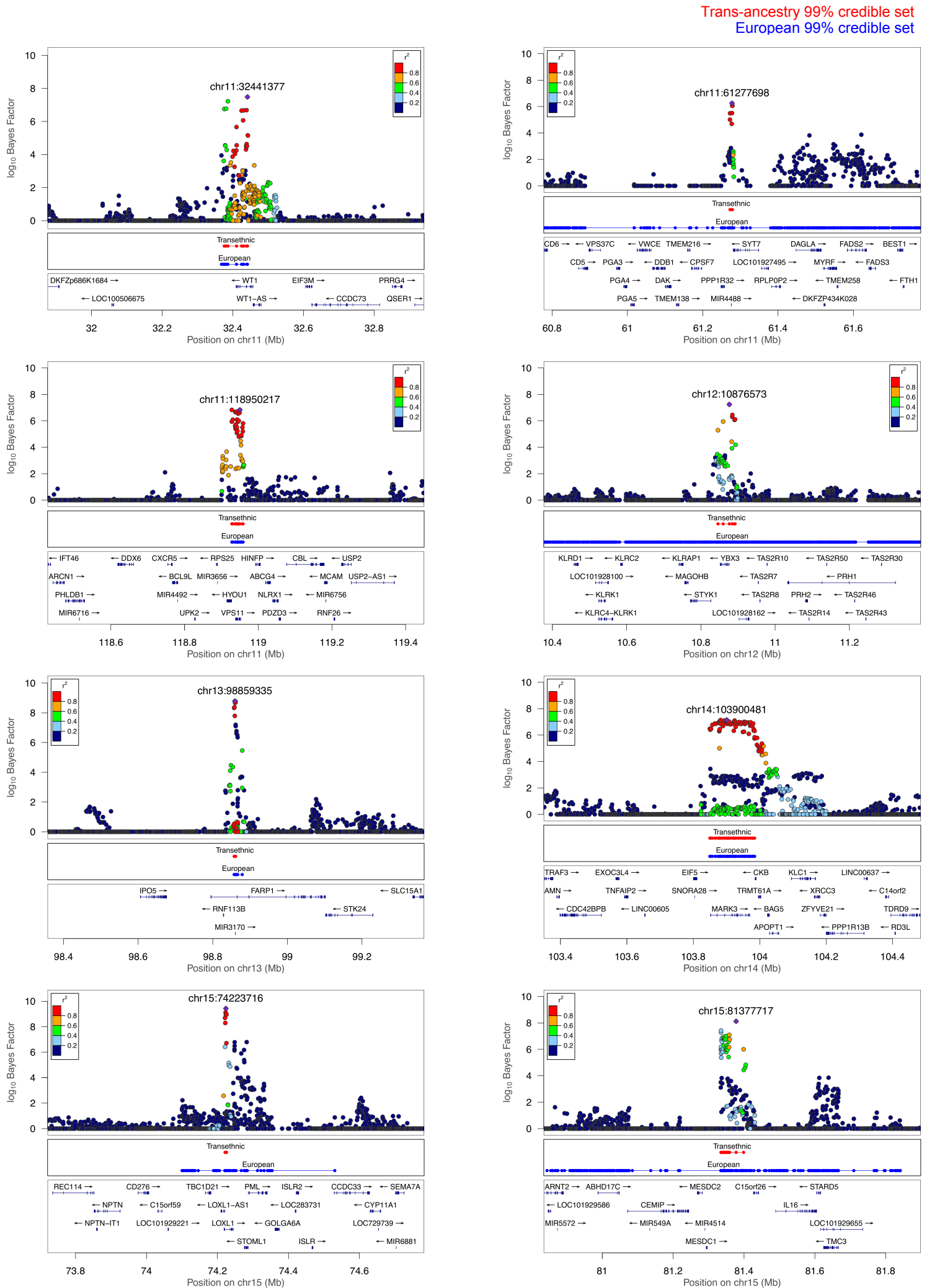
3. Novel loci in the Trans-ancestry meta-analysis



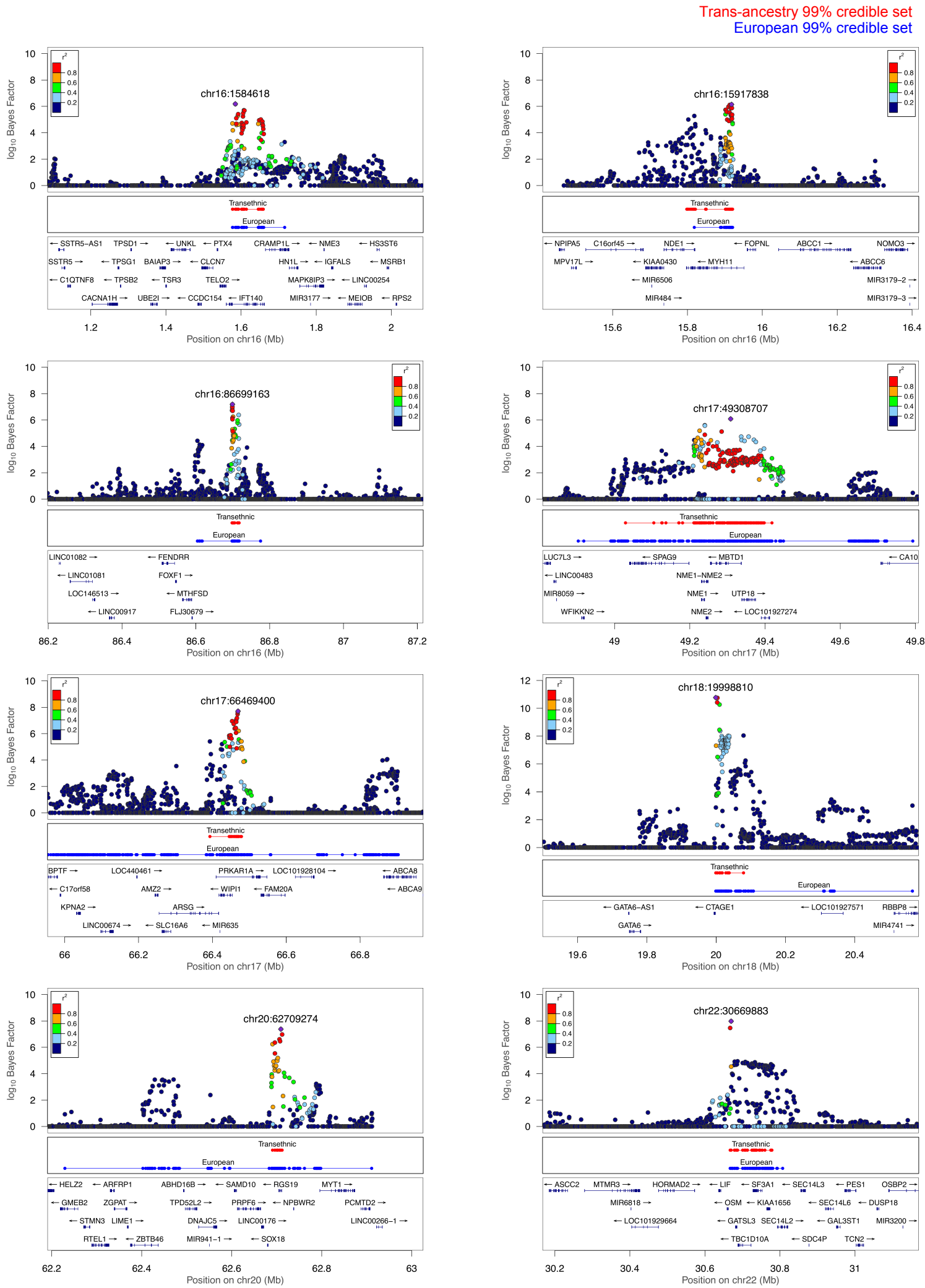
3. Novel loci in the Trans-ancestry meta-analysis



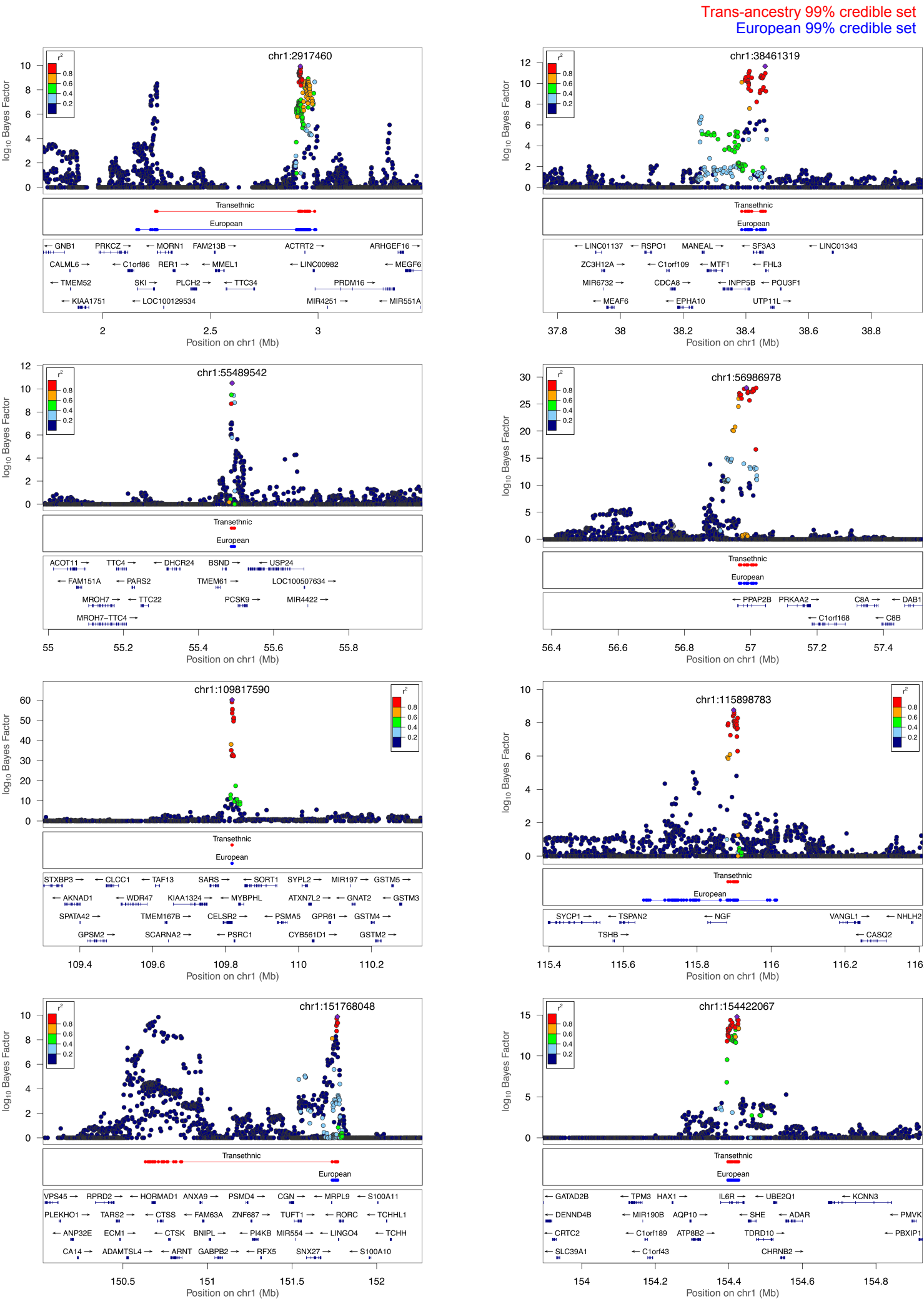
3. Novel loci in the Trans-ancestry meta-analysis



3. Novel loci in the Trans-ancestry meta-analysis

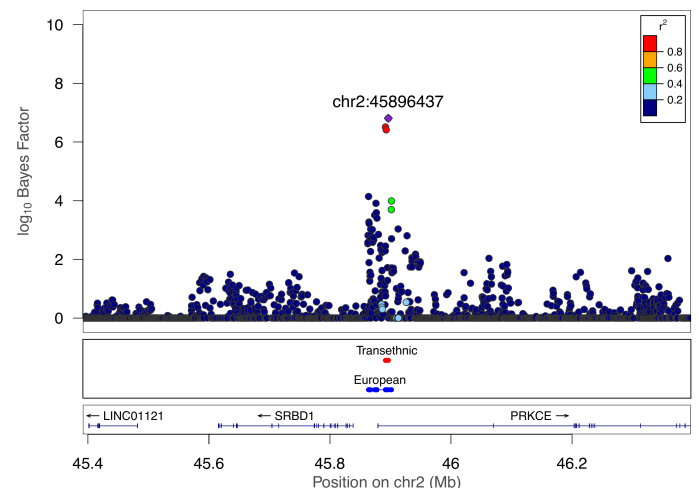
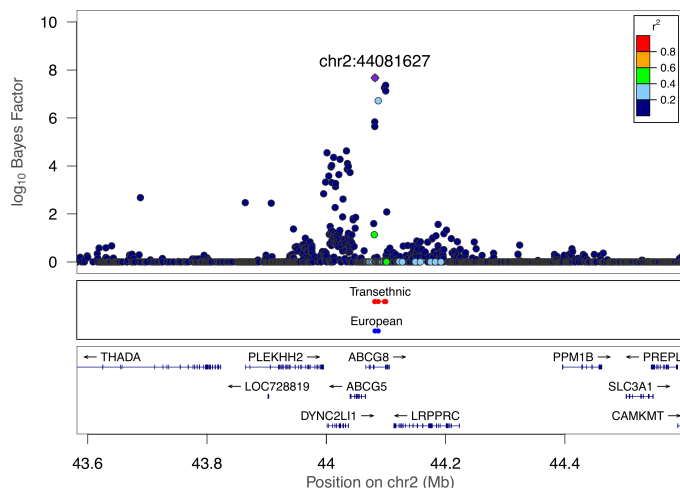
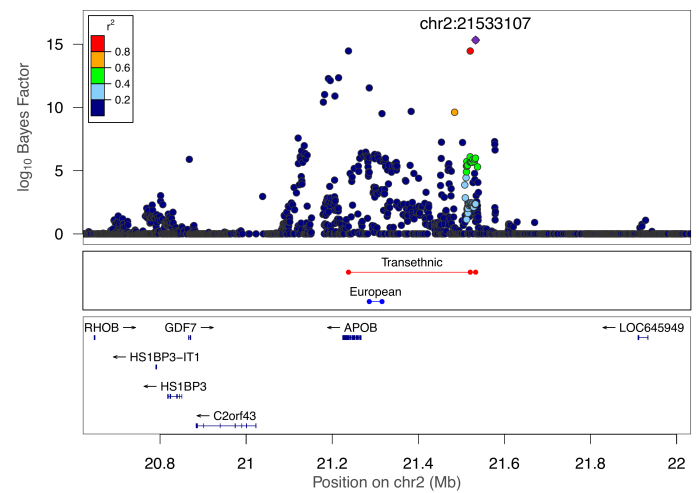
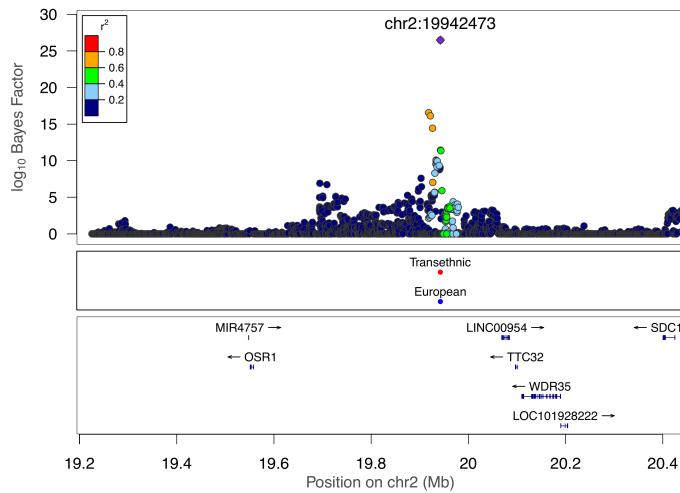
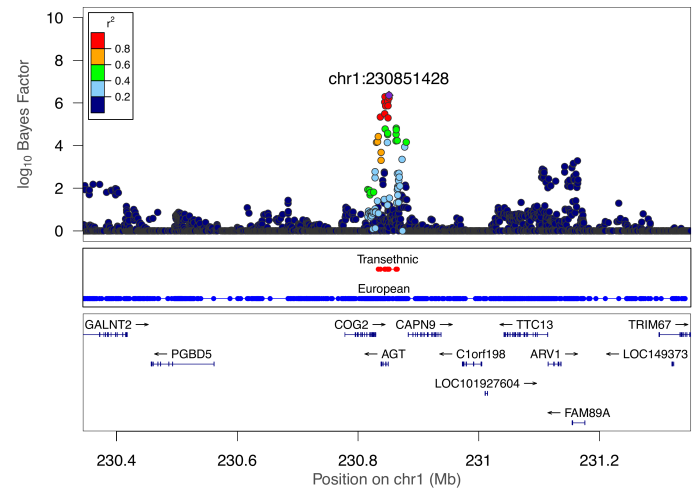
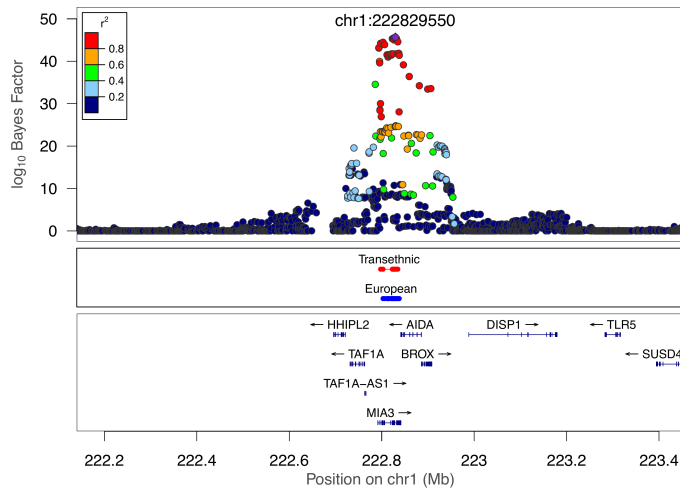
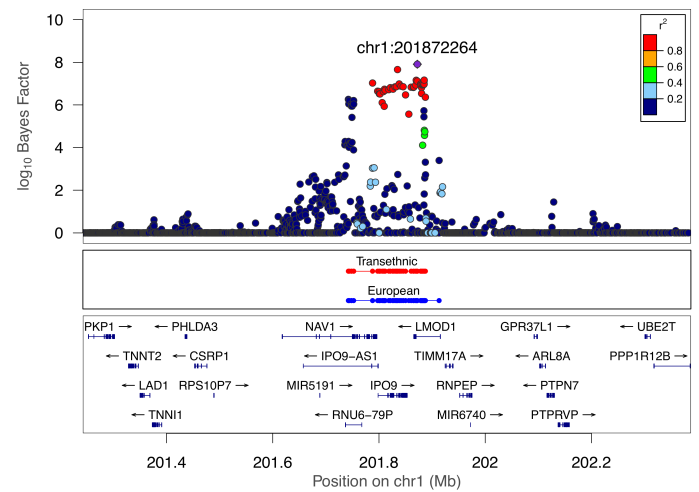
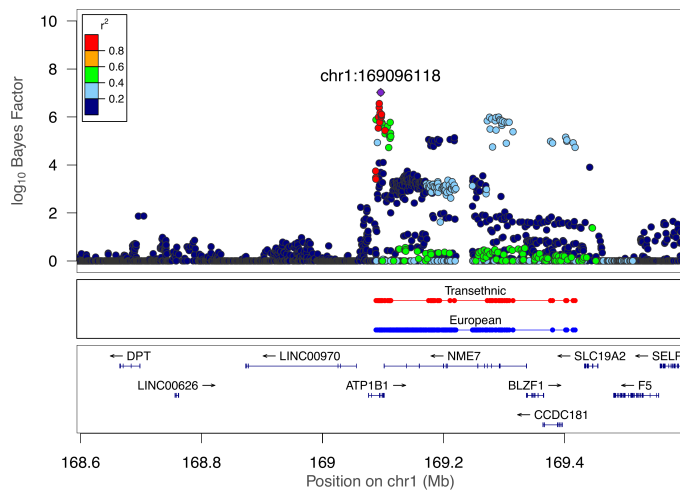


4. Previously reported loci in the Trans-ancestry meta-analysis

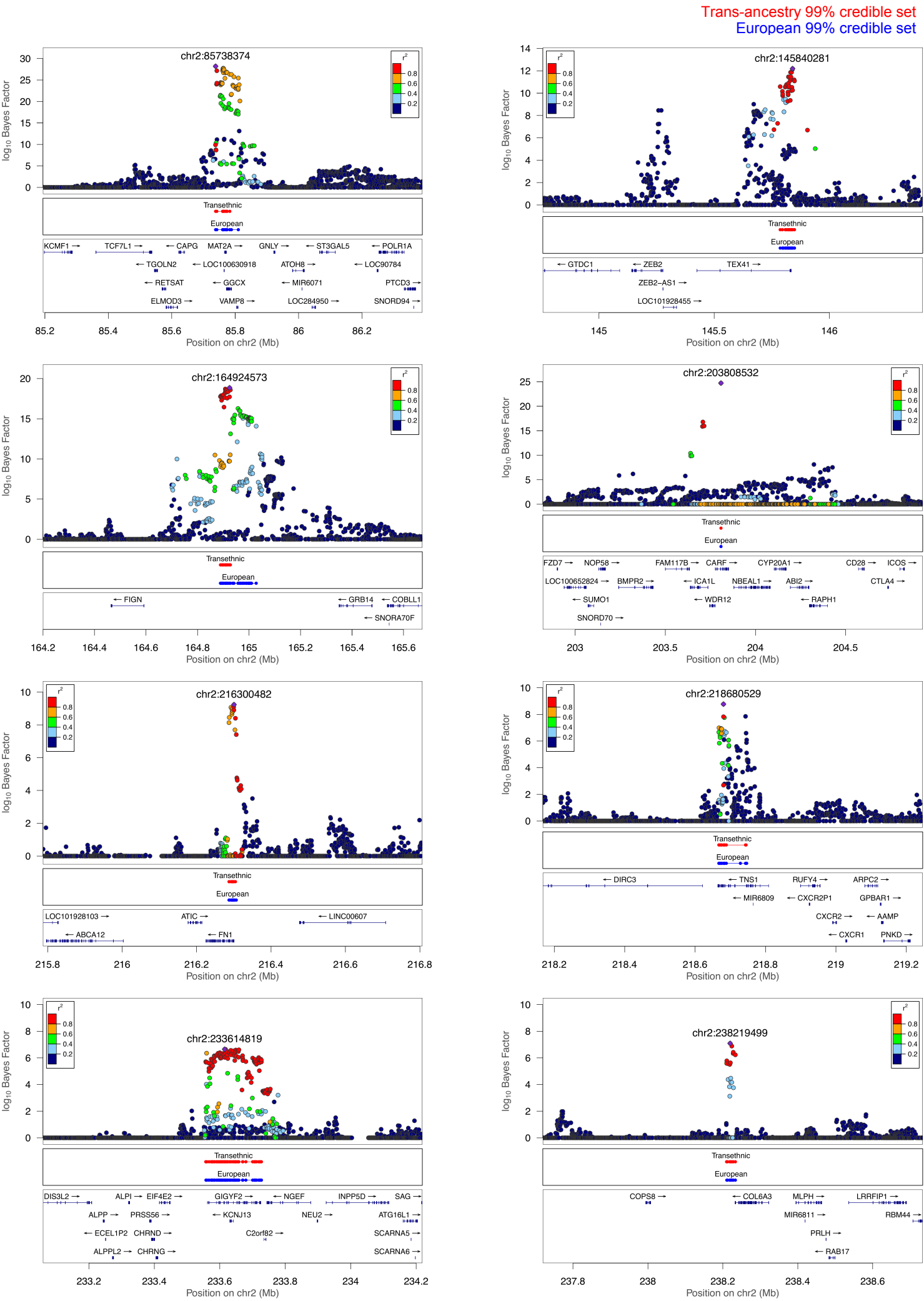


4. Previously reported loci in the Trans-ancestry meta-analysis

Trans-ancestry 99% credible set
European 99% credible set

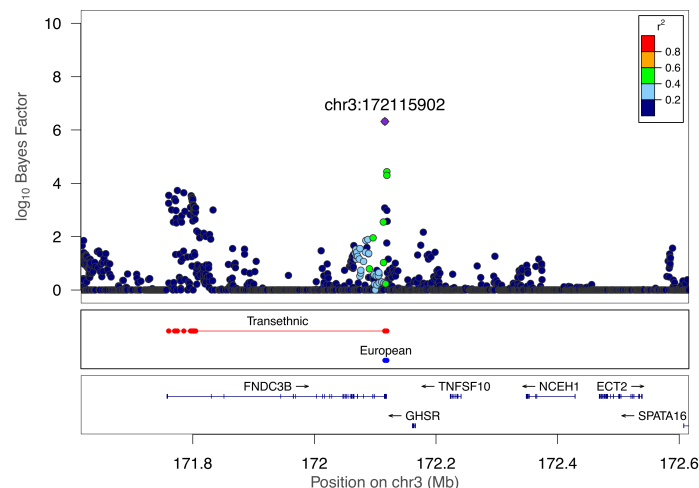
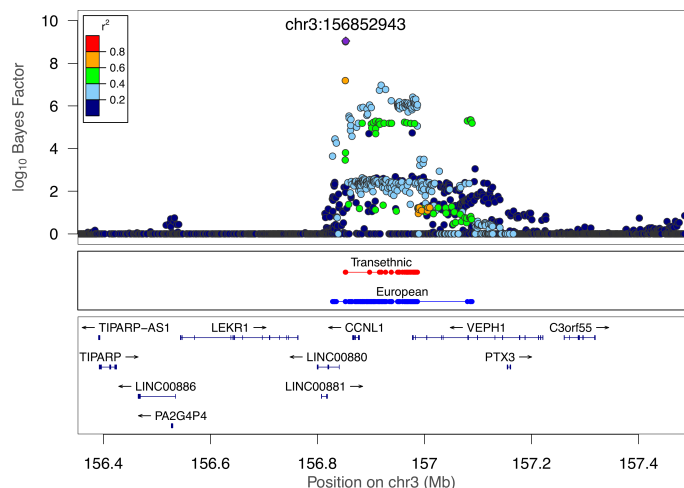
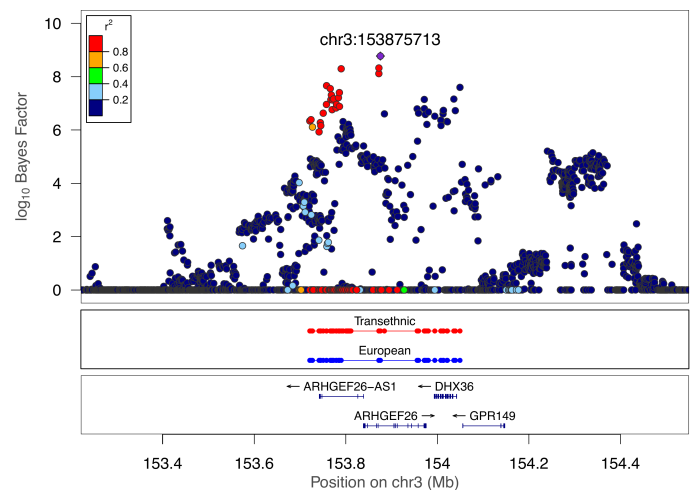
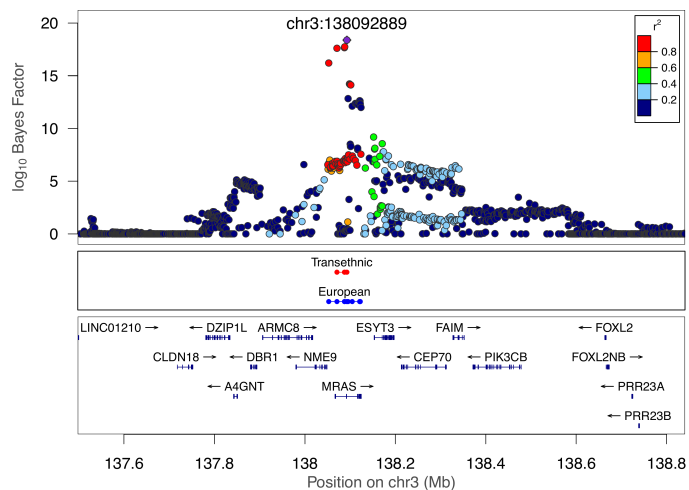
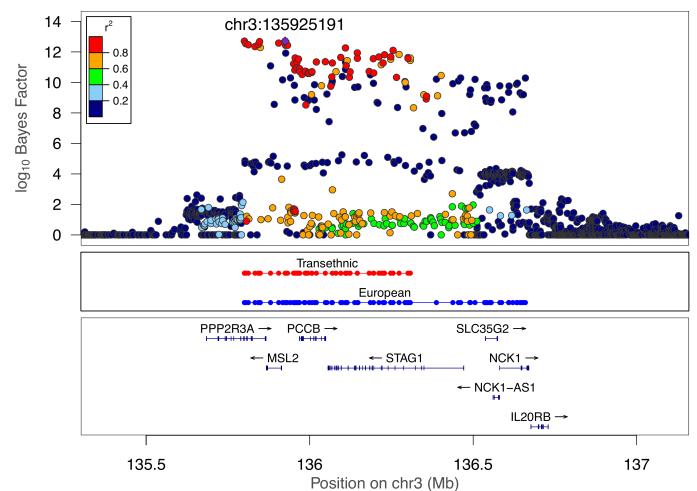
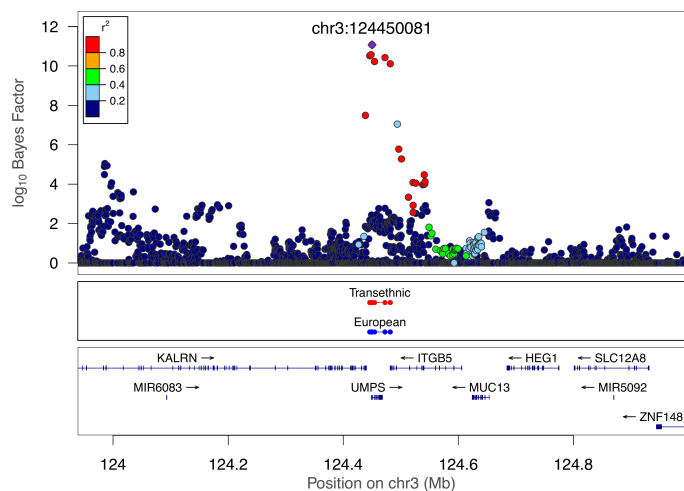
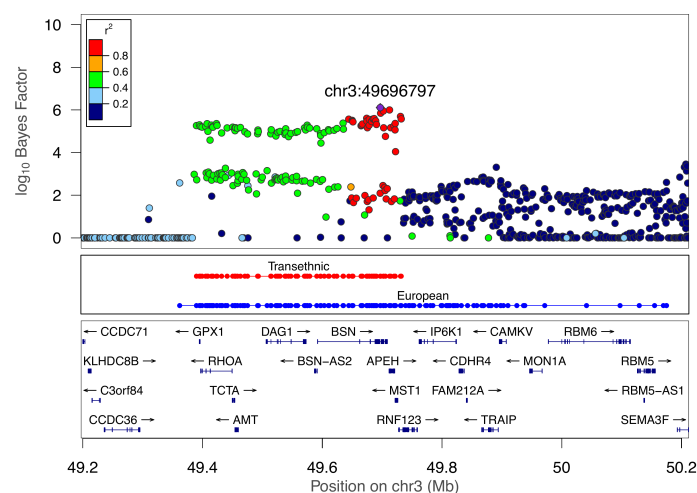
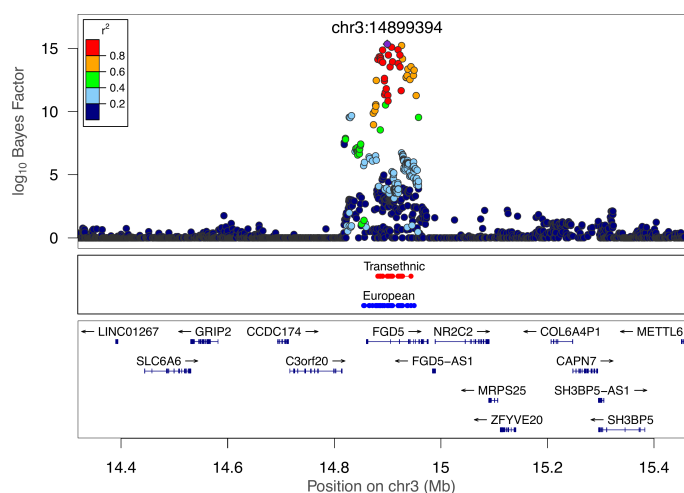


4. Previously reported loci in the Trans-ancestry meta-analysis

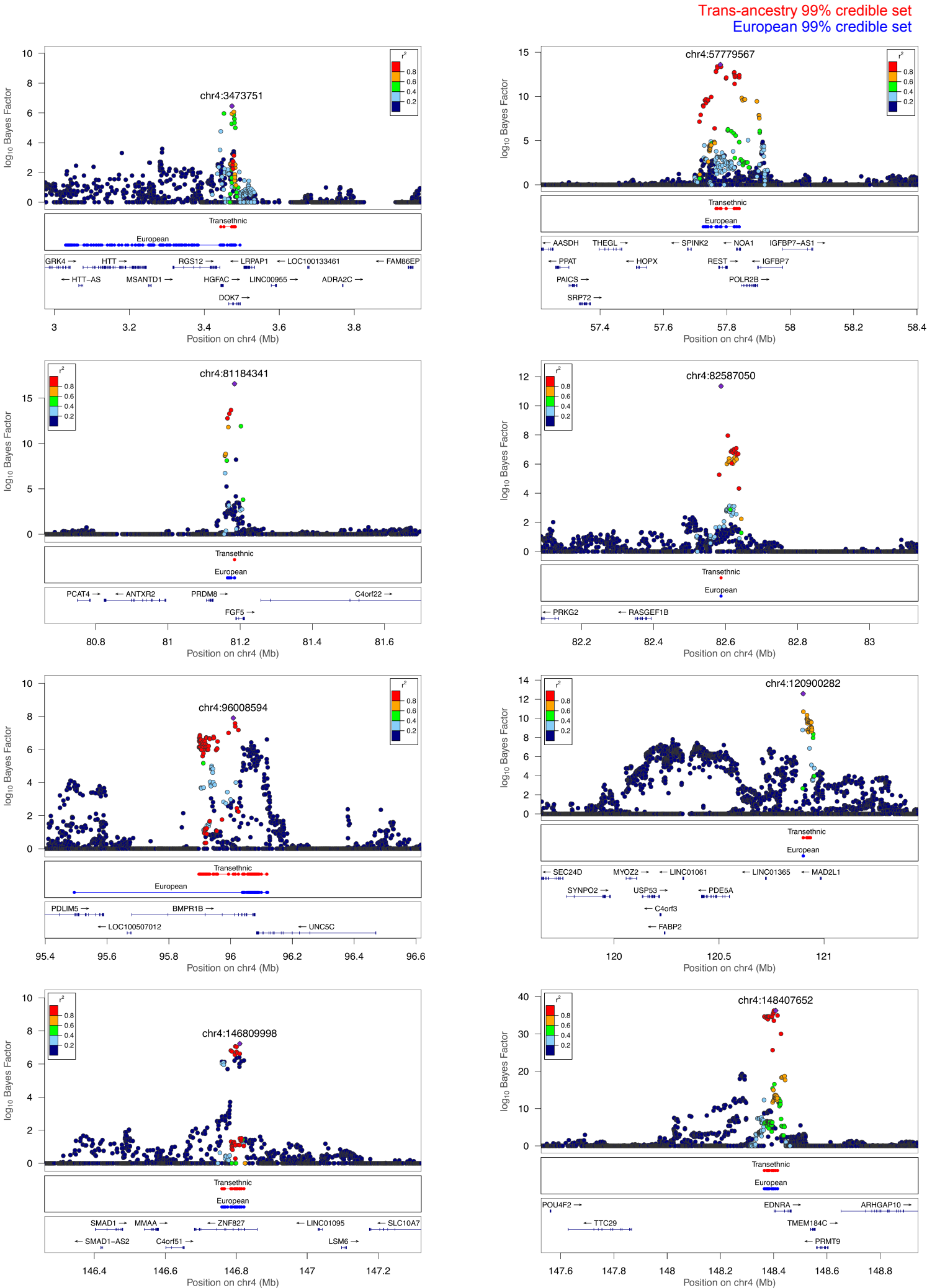


4. Previously reported loci in the Trans-ancestry meta-analysis

Trans-ancestry 99% credible set
European 99% credible set

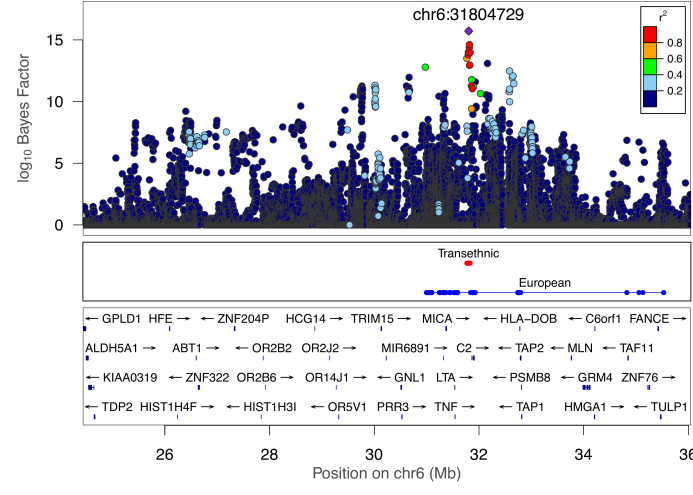
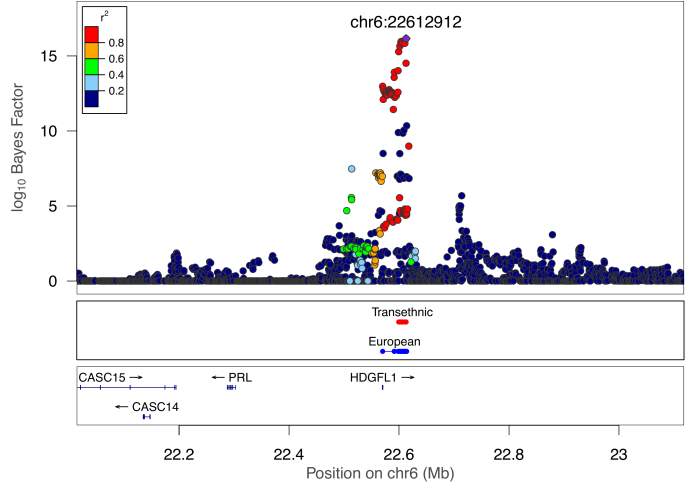
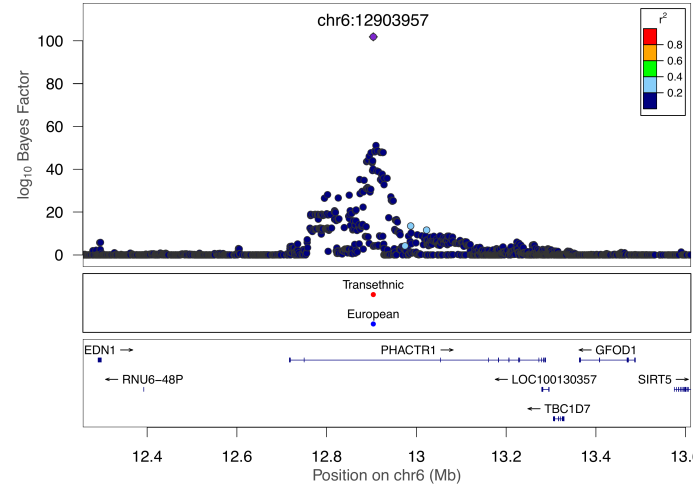
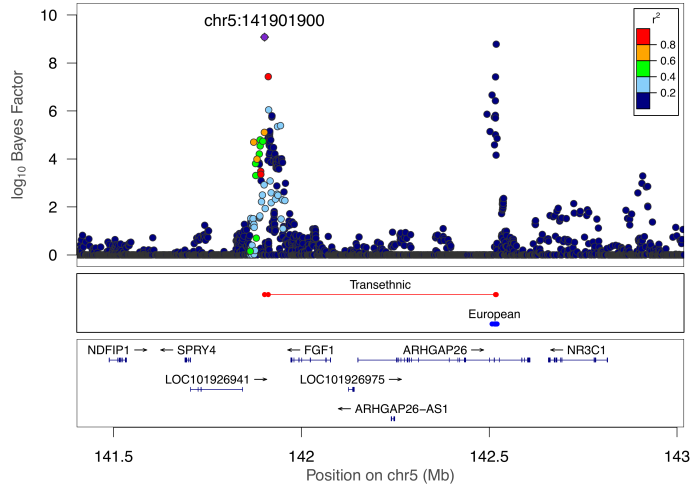
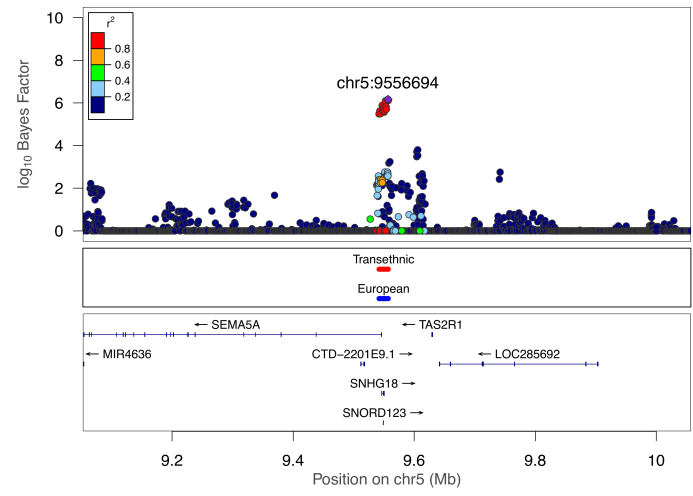
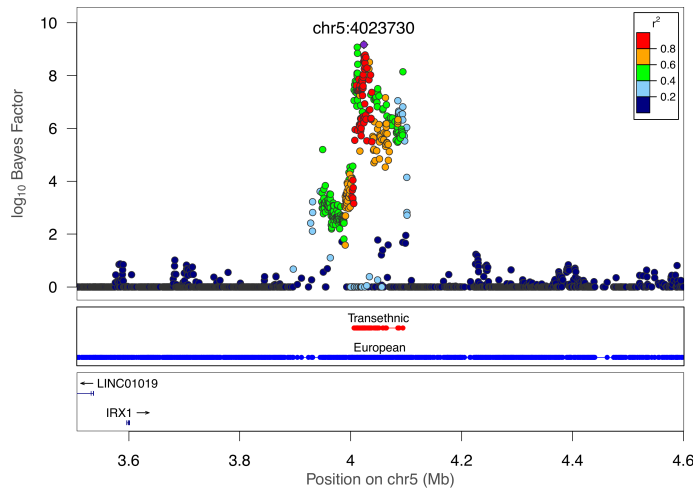
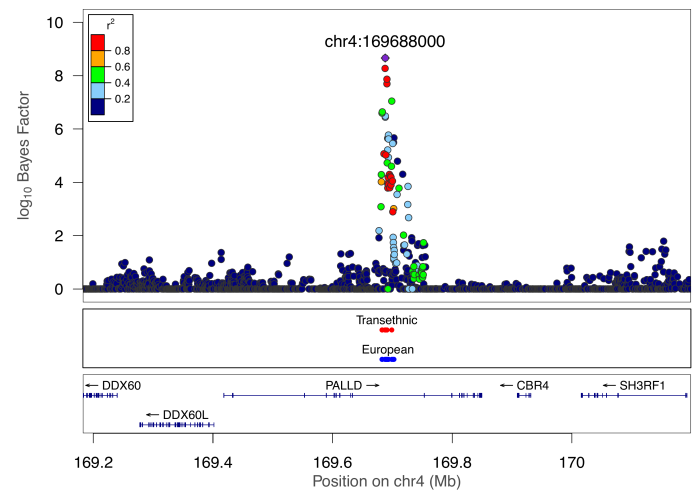
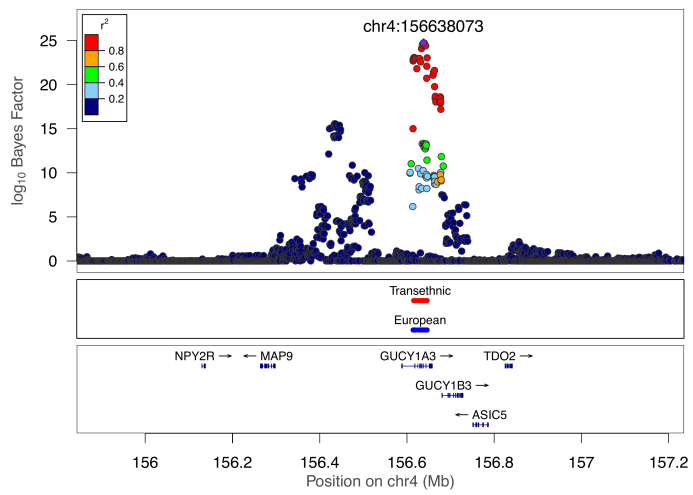


4. Previously reported loci in the Trans-ancestry meta-analysis

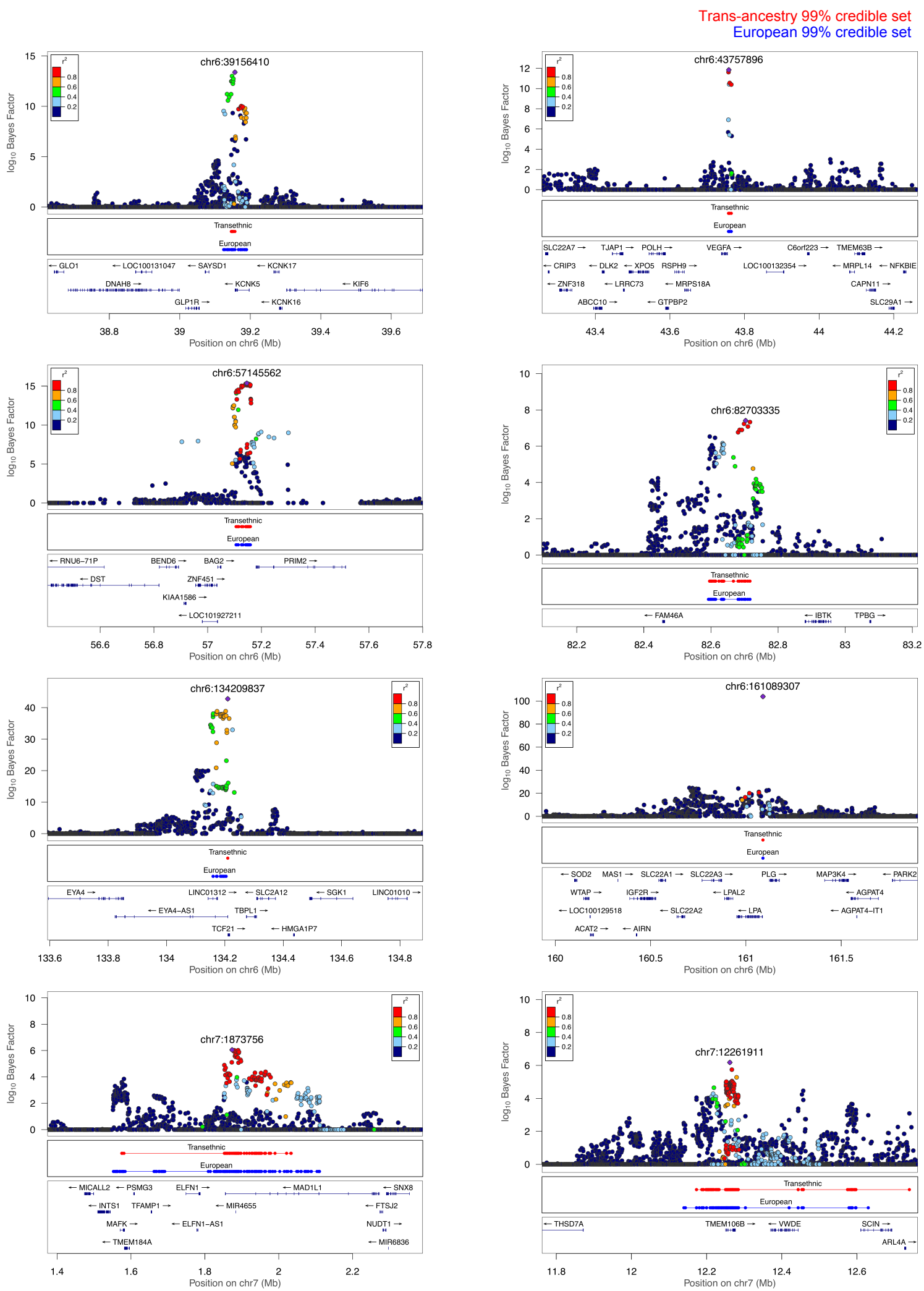


4. Previously reported loci in the Trans-ancestry meta-analysis

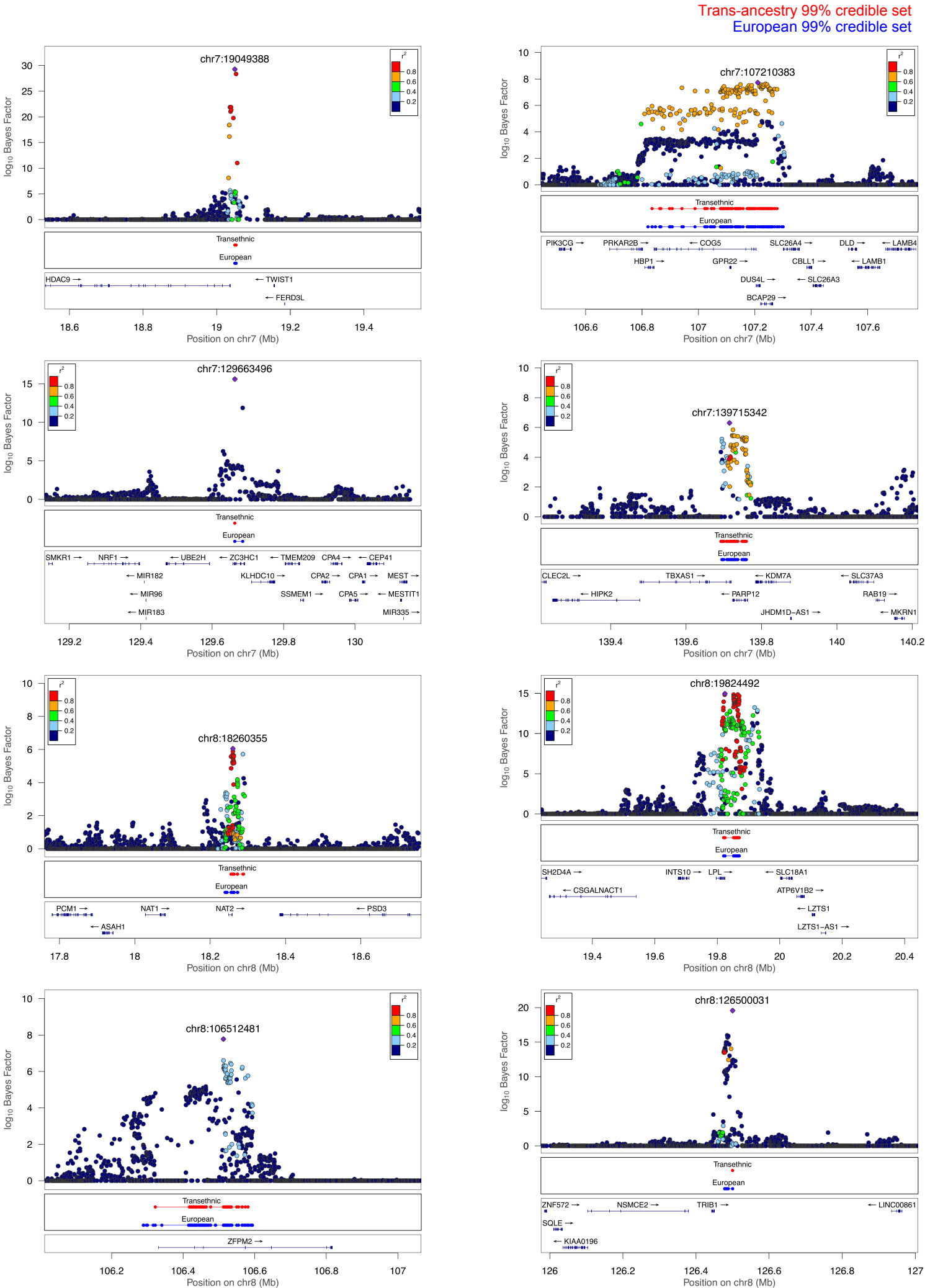
Trans-ancestry 99% credible set
European 99% credible set



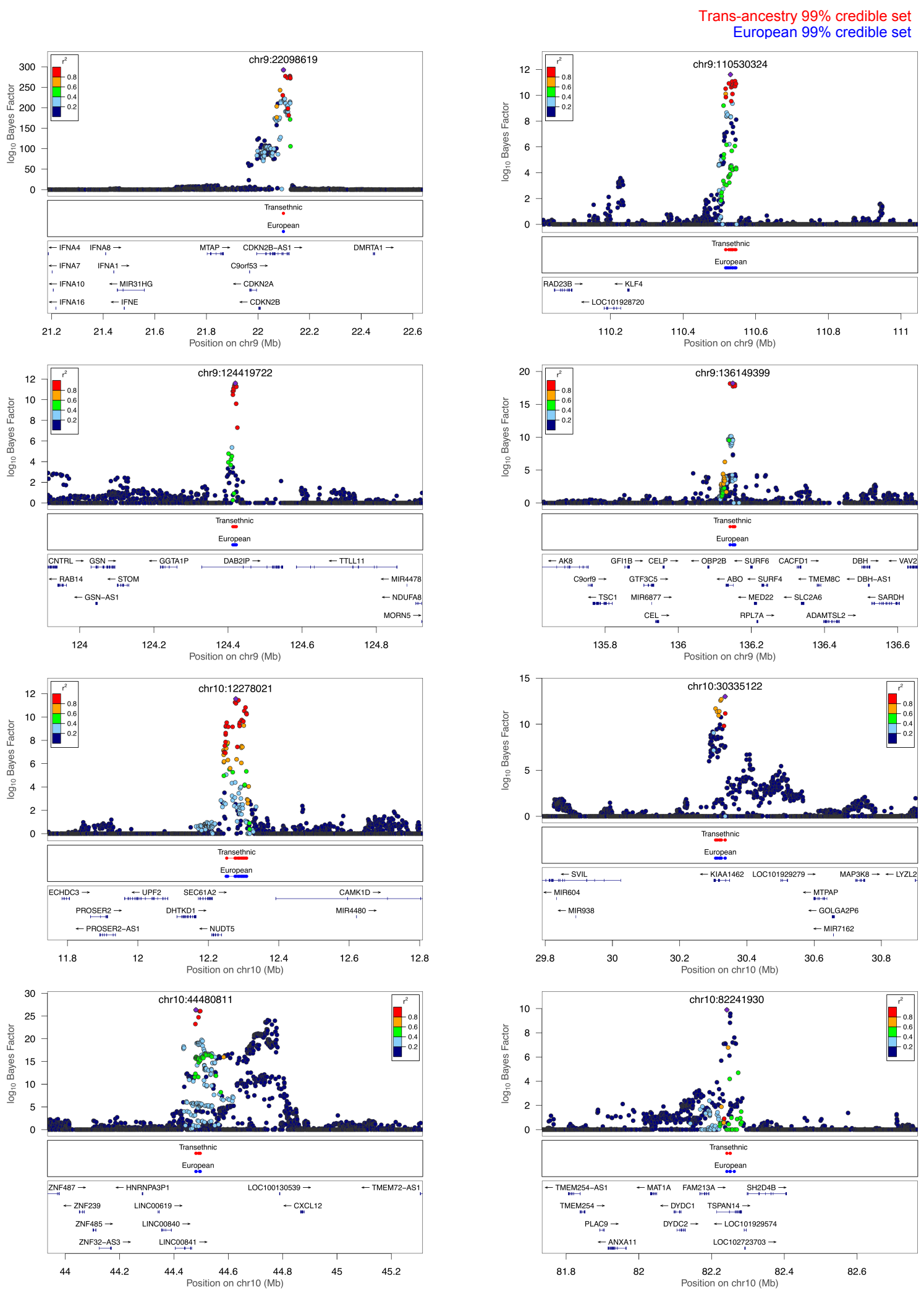
4. Previously reported loci in the Trans-ancestry meta-analysis



4. Previously reported loci in the Trans-ancestry meta-analysis

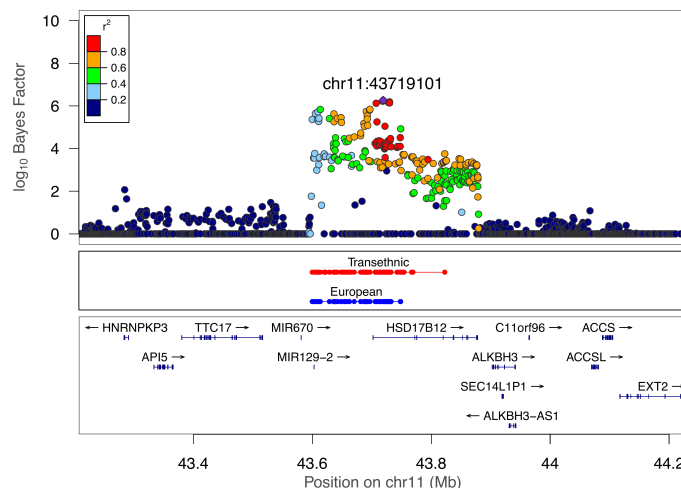
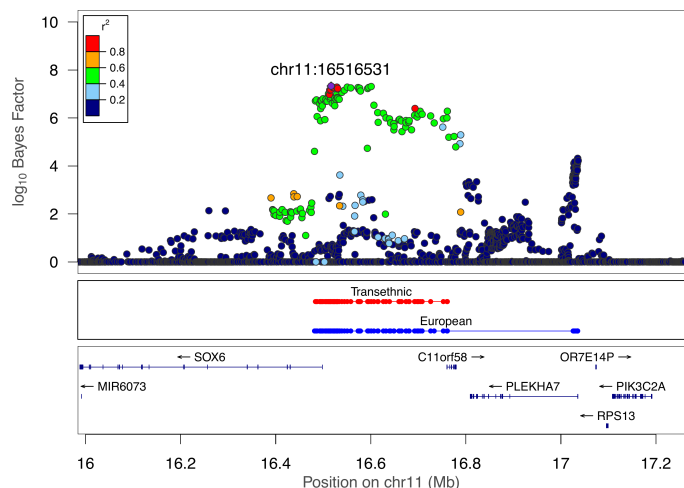
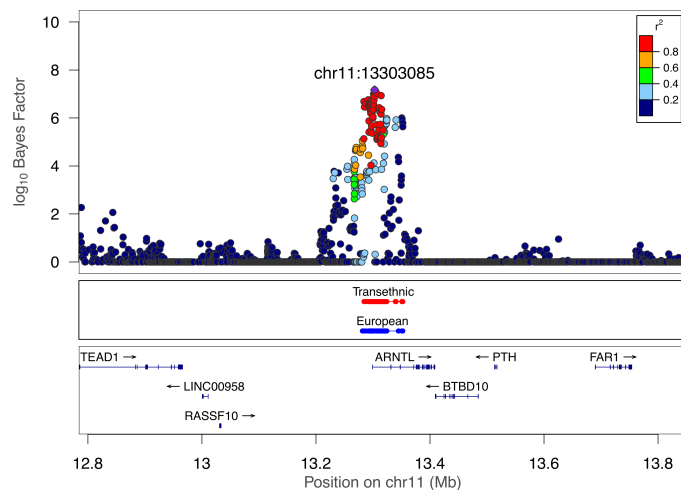
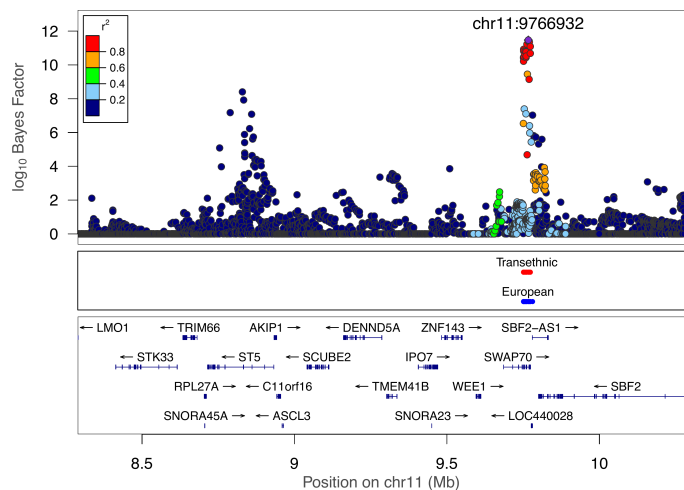
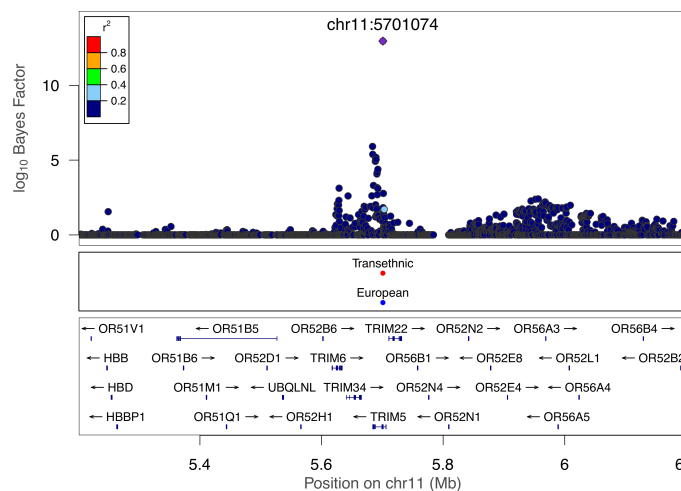
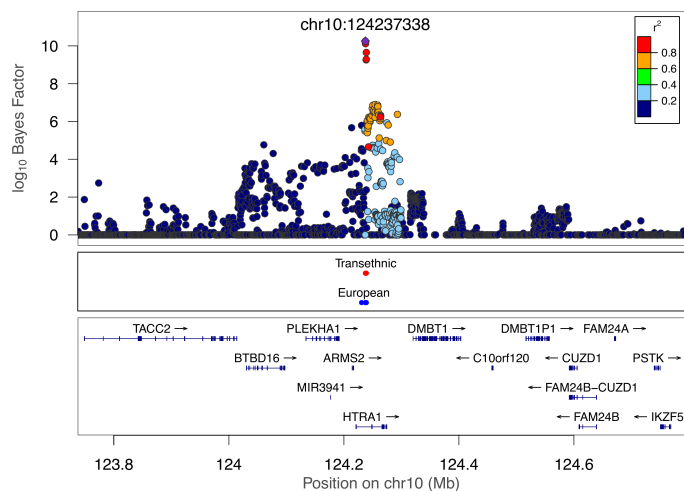
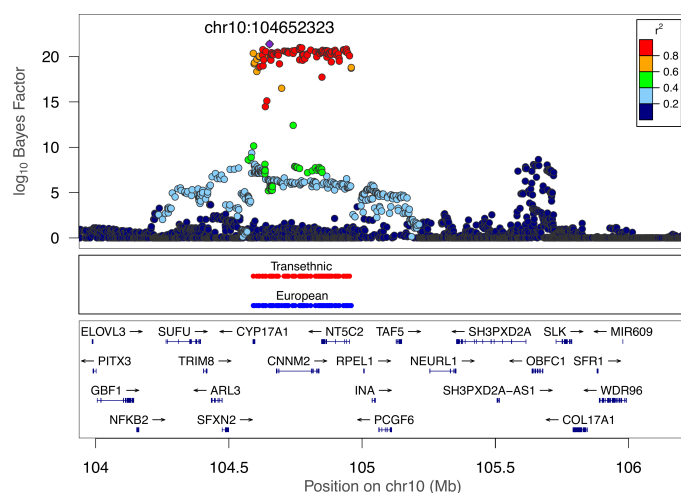
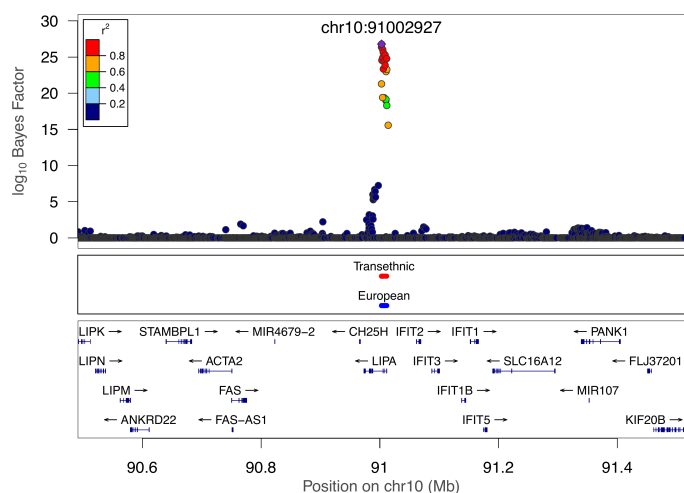


4. Previously reported loci in the Trans-ancestry meta-analysis

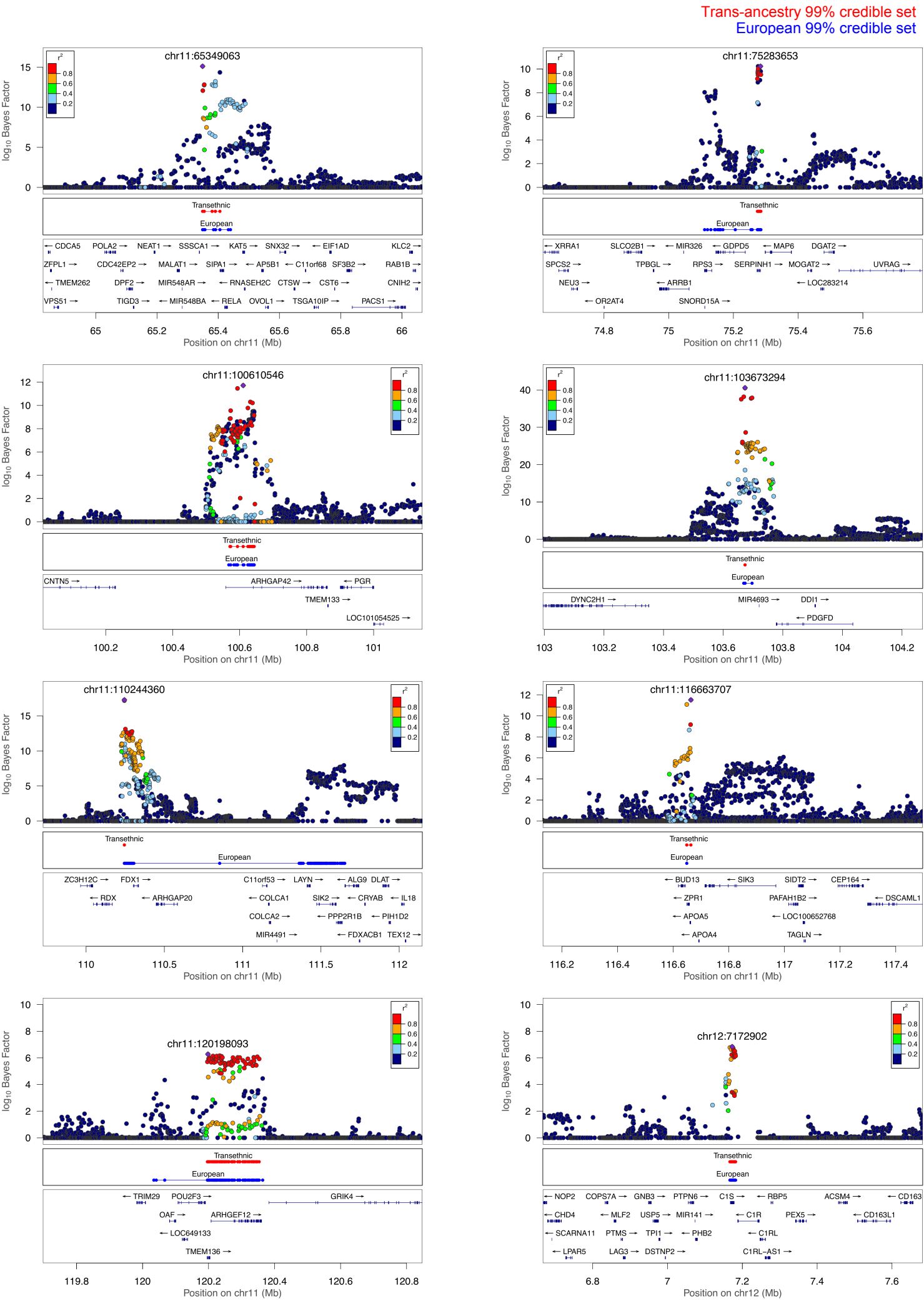


4. Previously reported loci in the Trans-ancestry meta-analysis

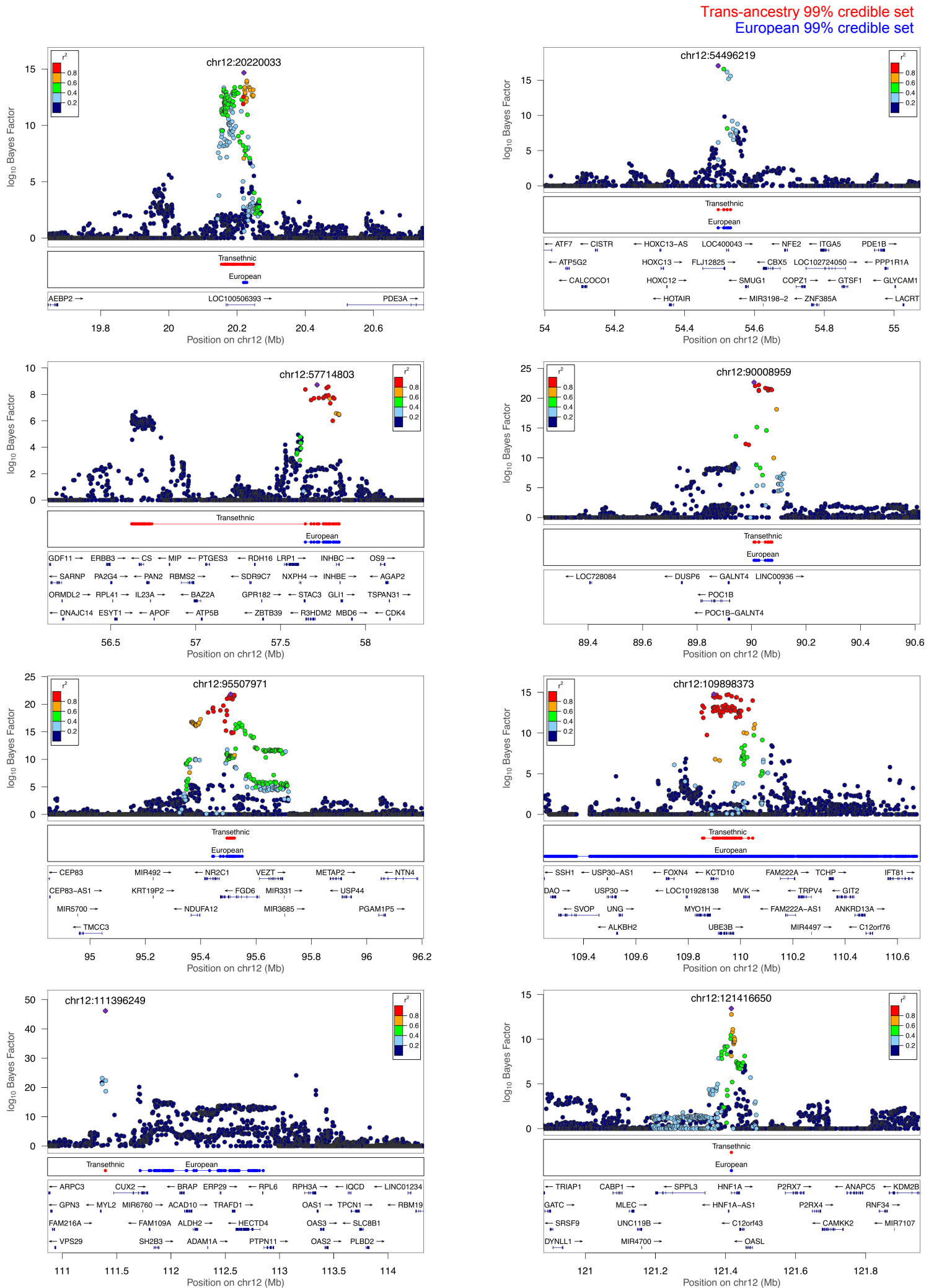
Trans-ancestry 99% credible set
European 99% credible set



4. Previously reported loci in the Trans-ancestry meta-analysis

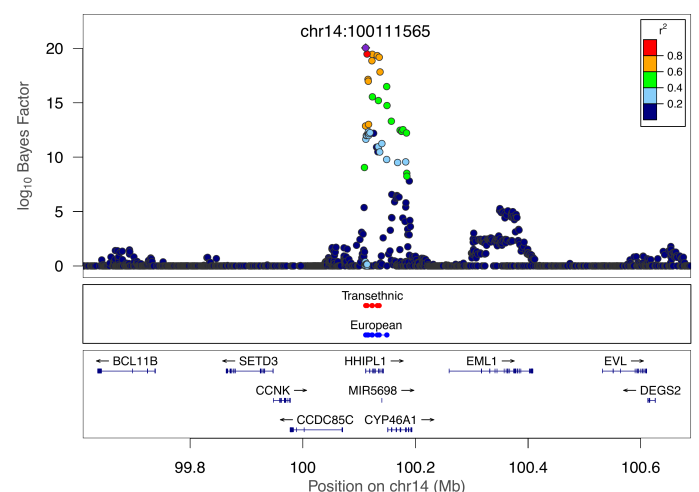
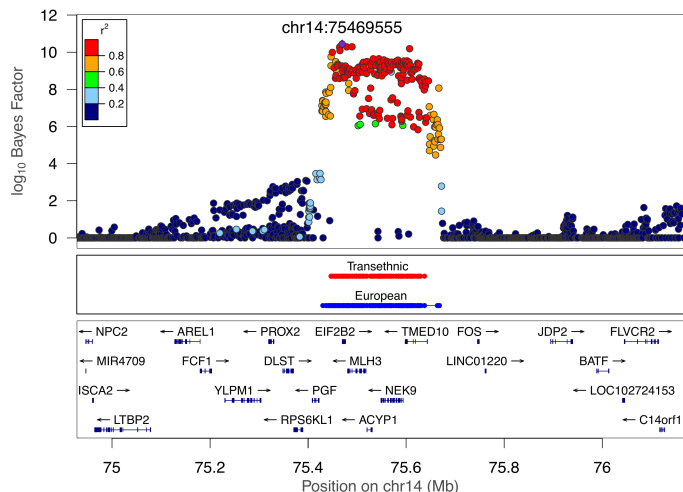
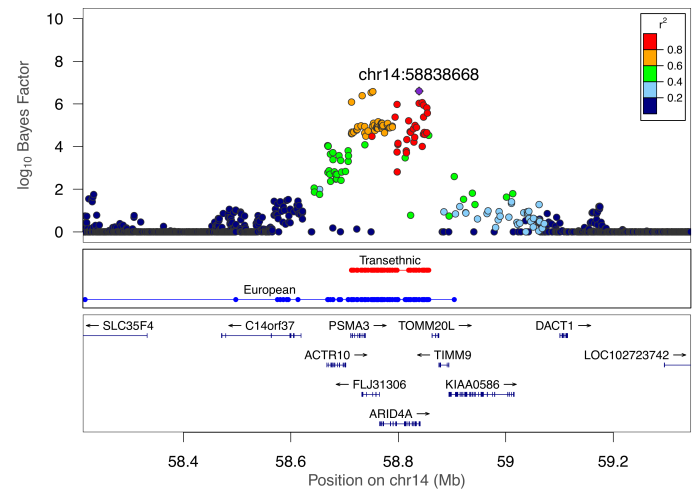
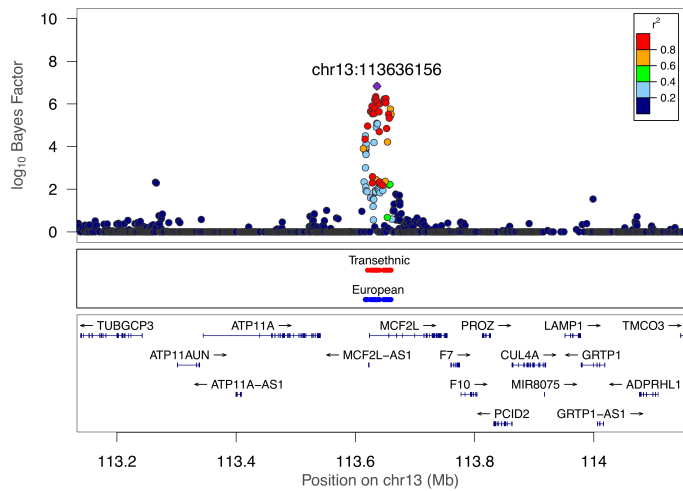
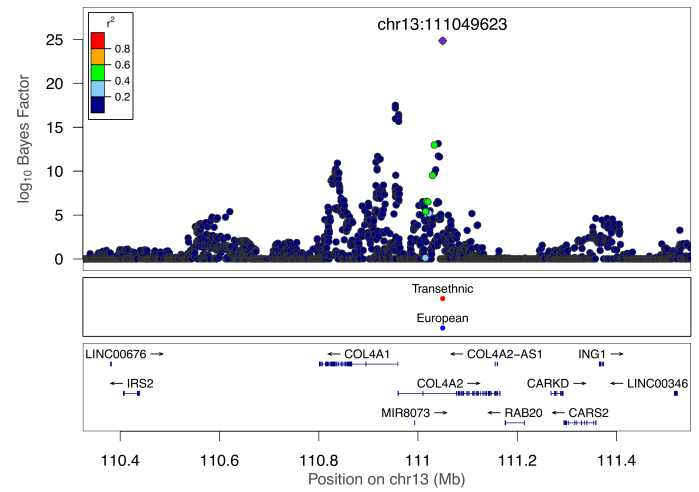
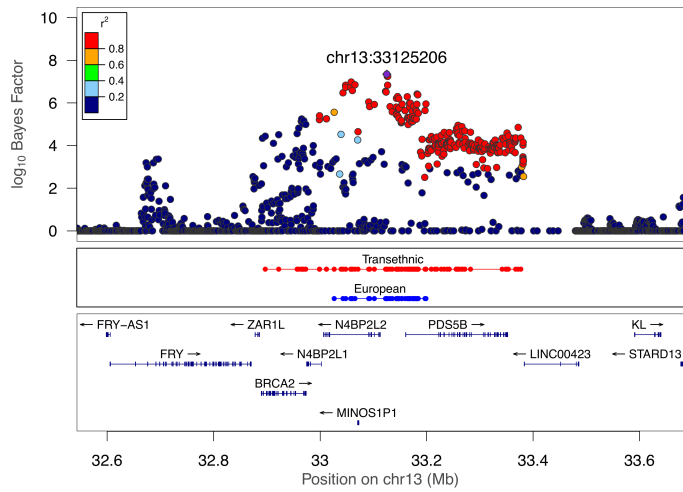
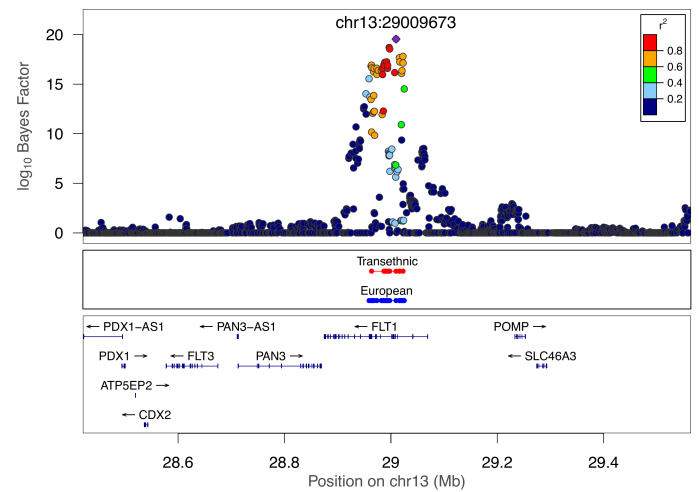
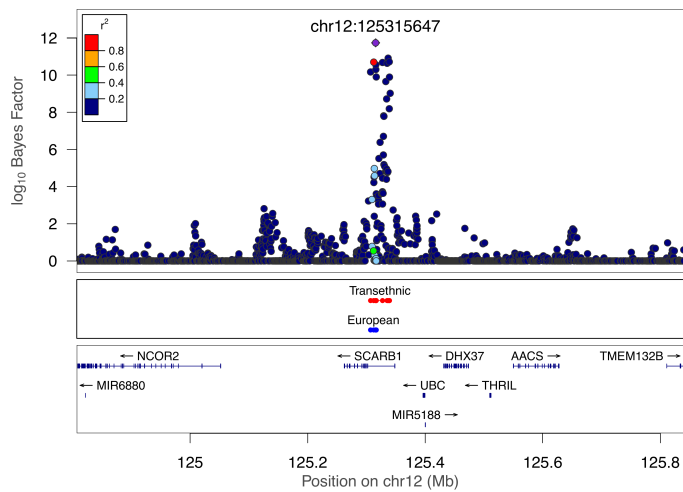


4. Previously reported loci in the Trans-ancestry meta-analysis



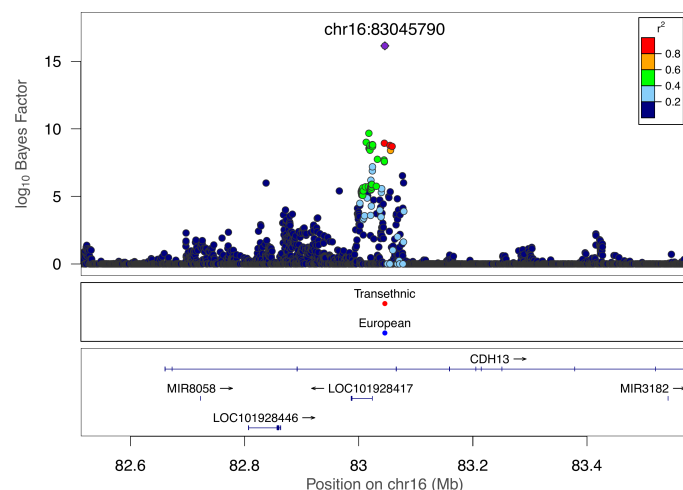
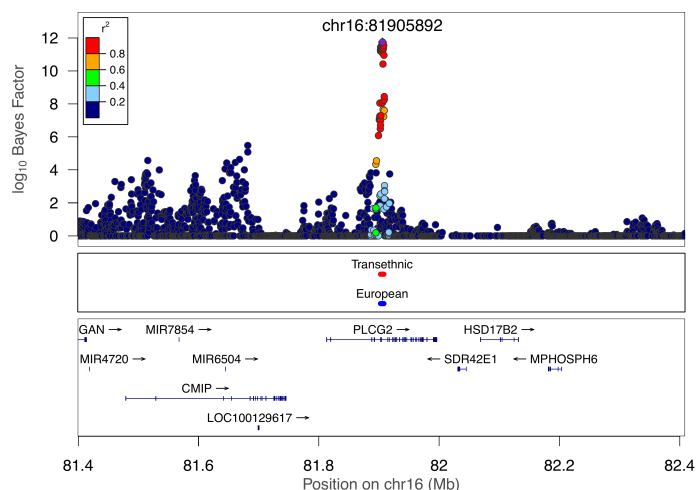
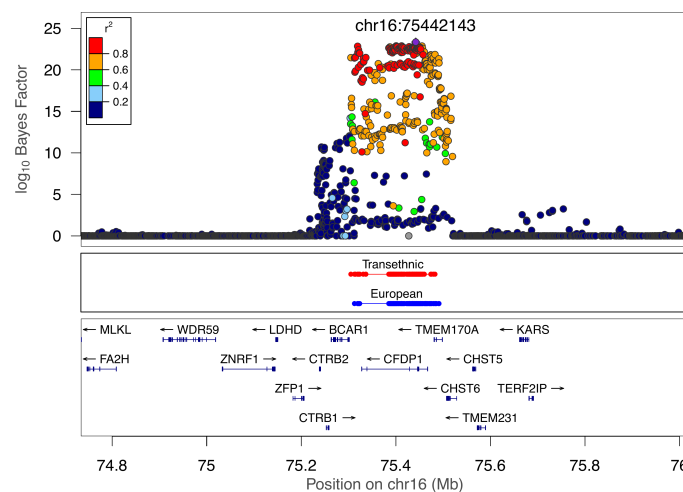
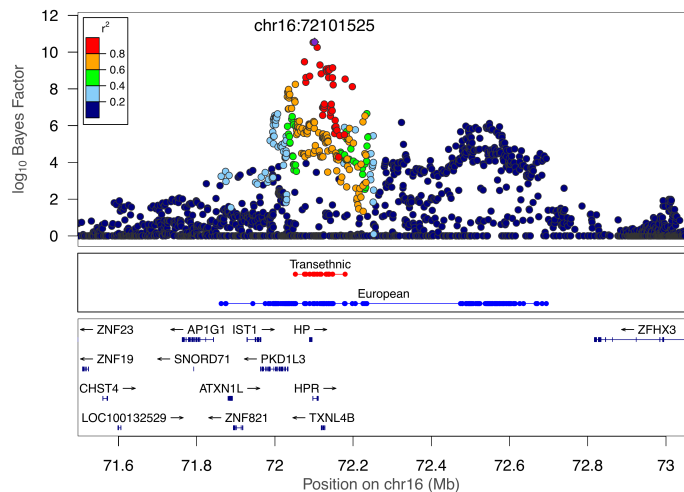
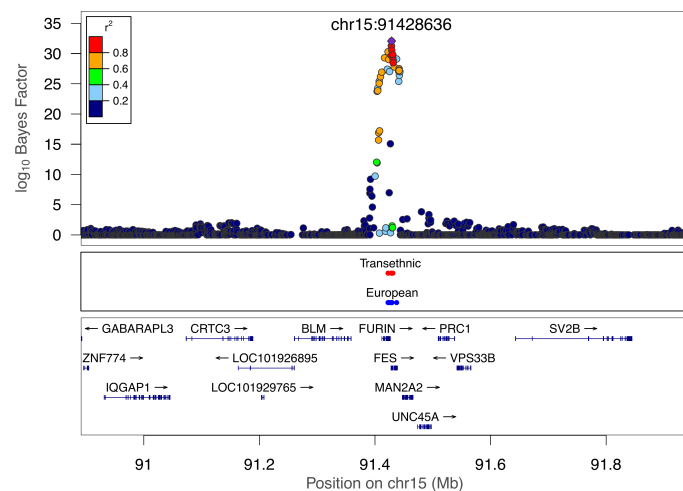
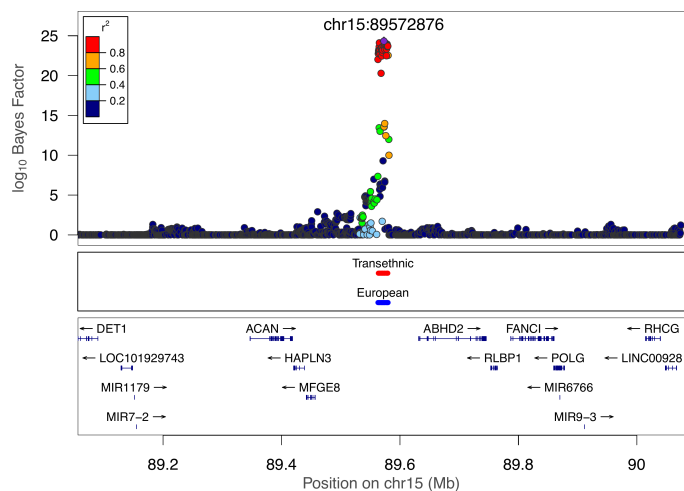
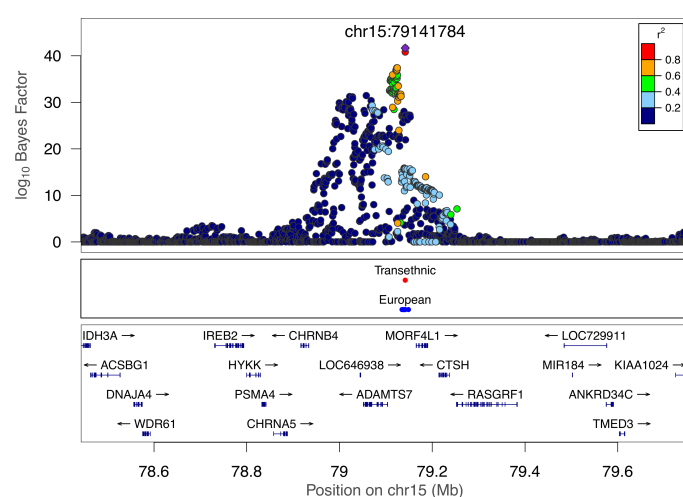
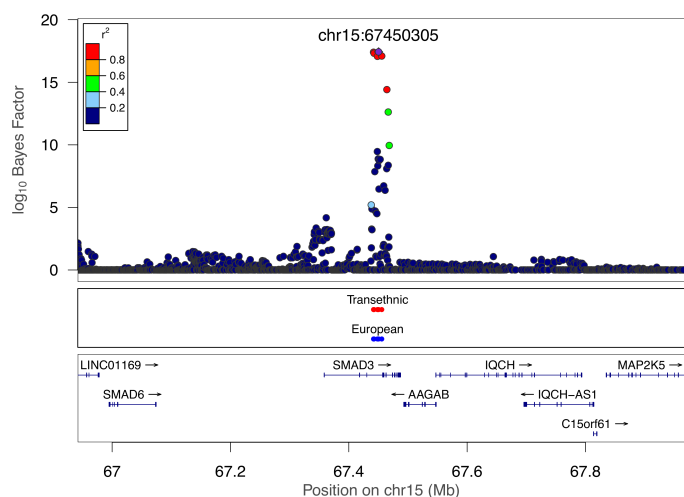
4. Previously reported loci in the Trans-ancestry meta-analysis

Trans-ancestry 99% credible set
European 99% credible set



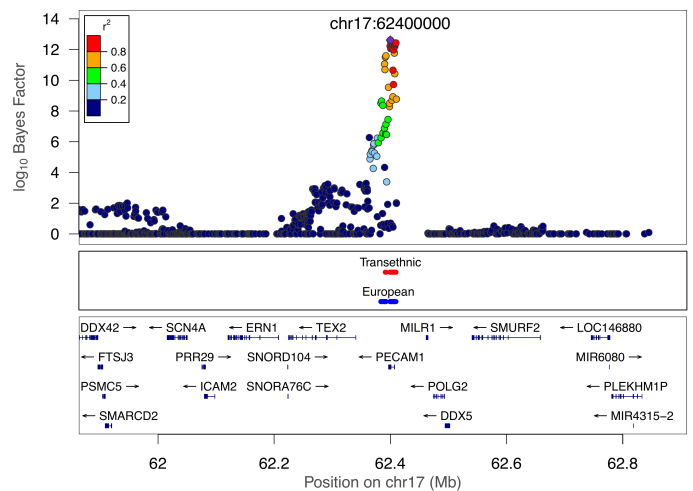
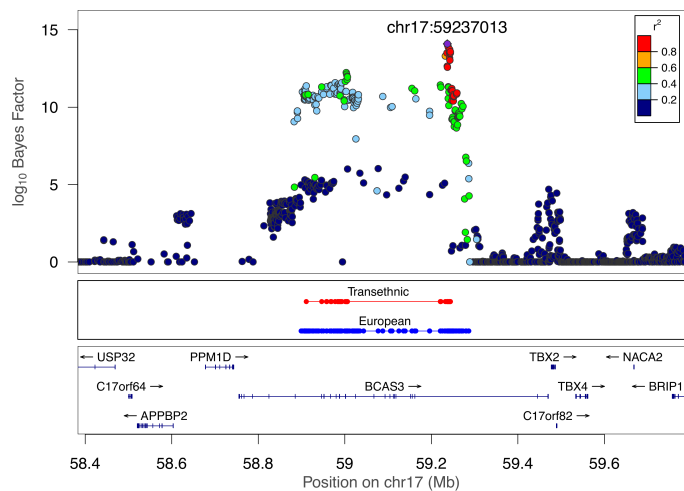
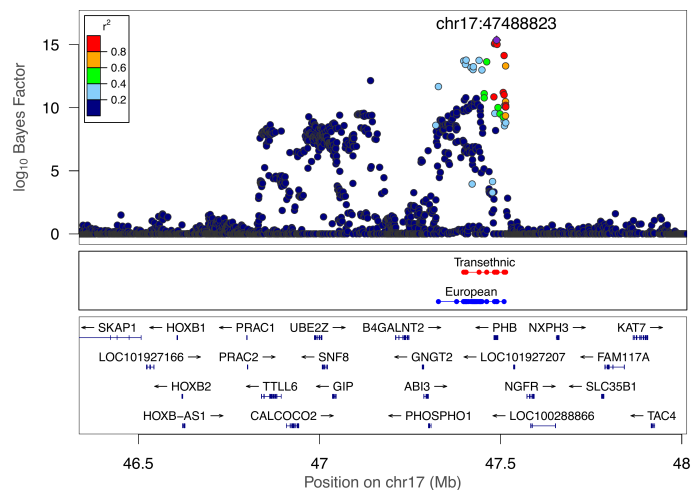
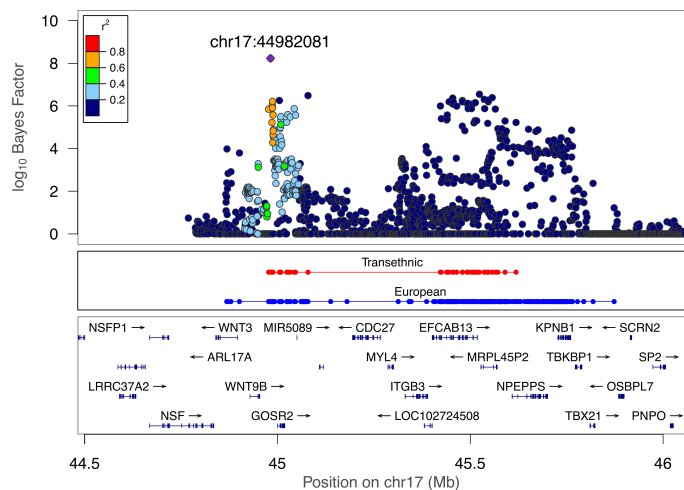
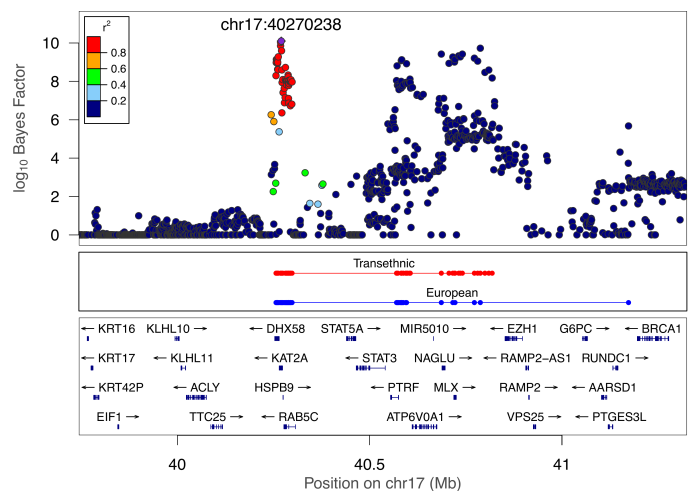
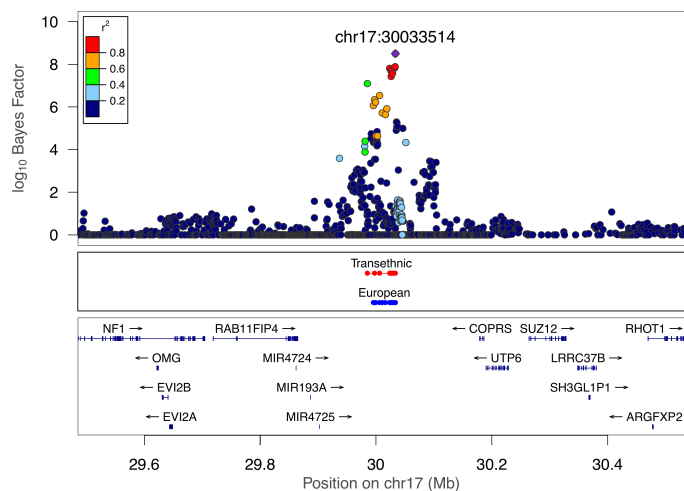
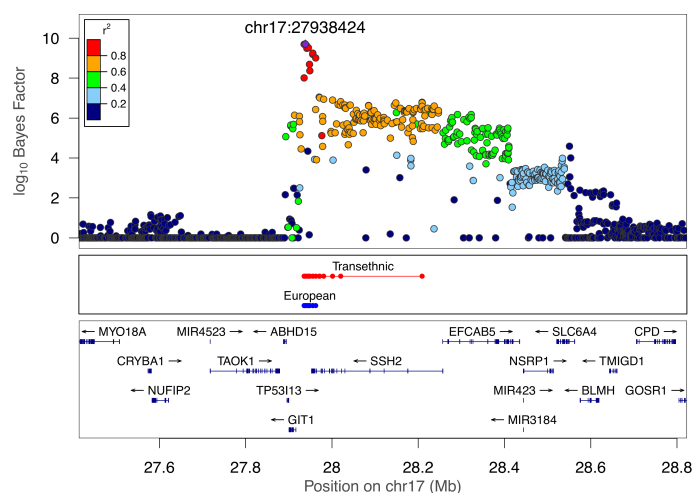
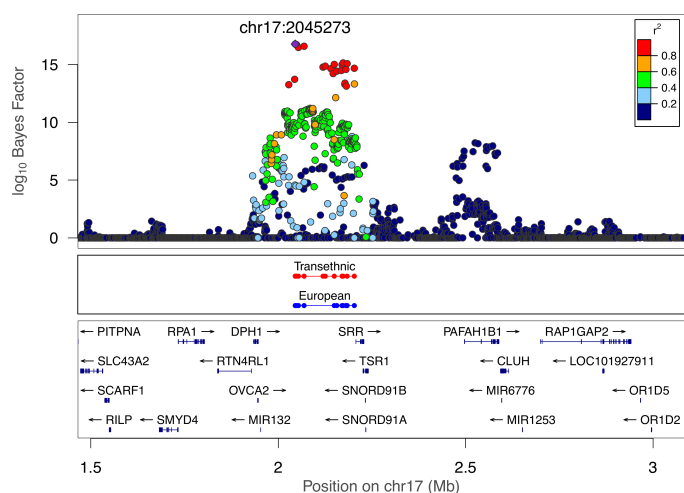
4. Previously reported loci in the Trans-ancestry meta-analysis

Trans-ancestry 99% credible set
European 99% credible set



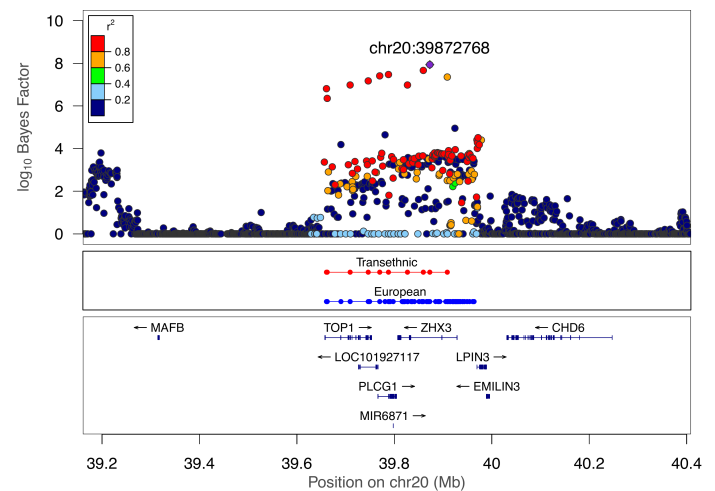
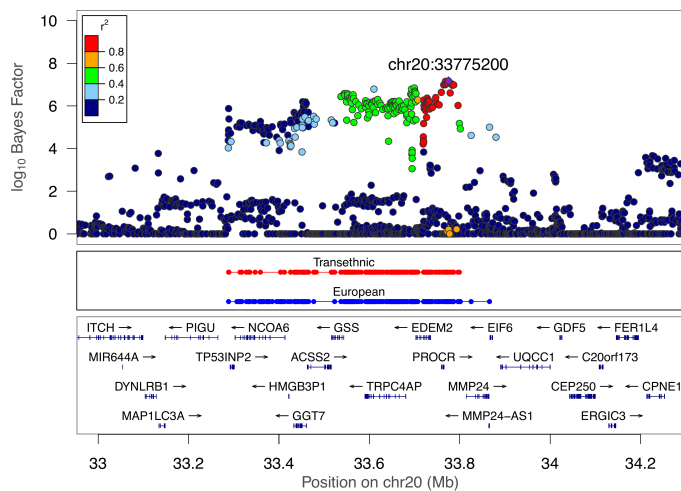
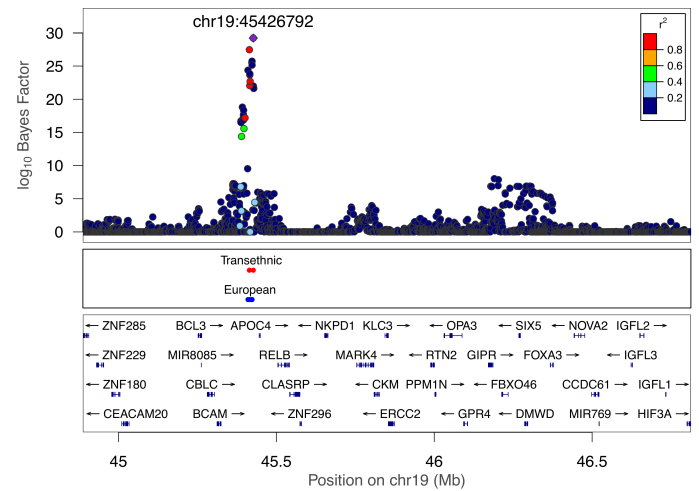
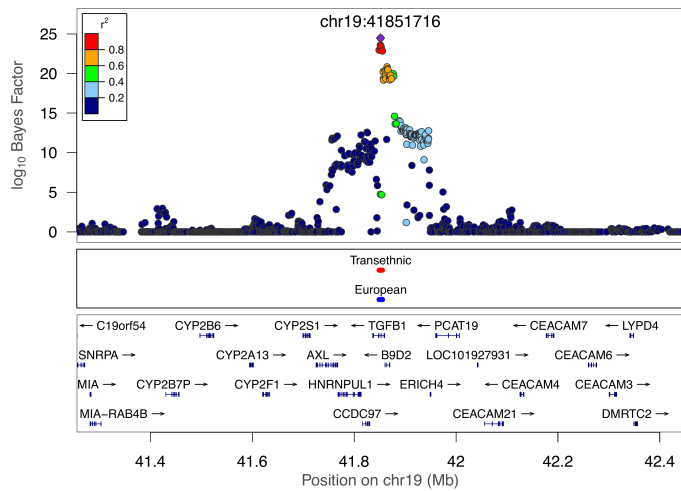
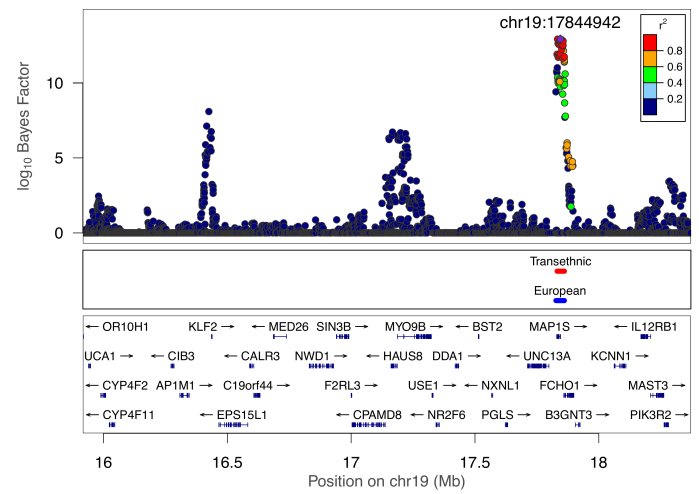
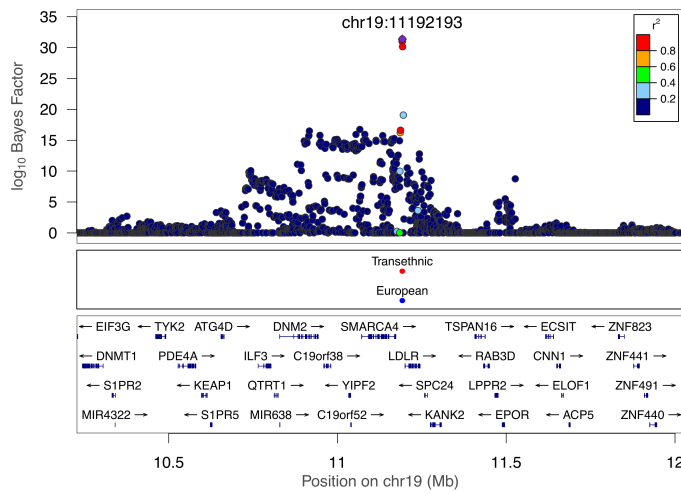
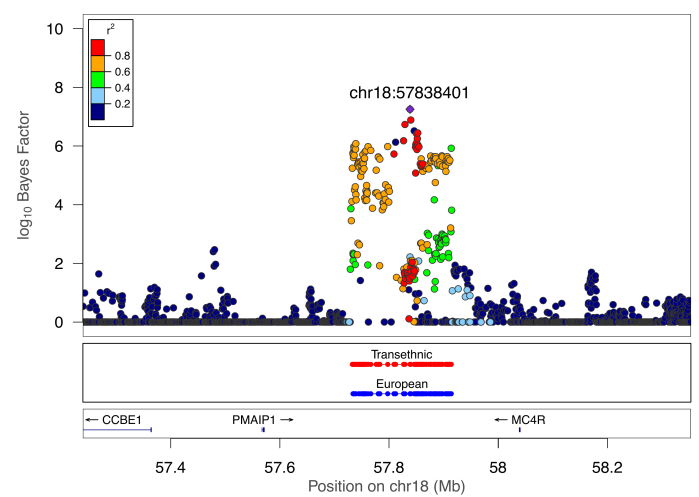
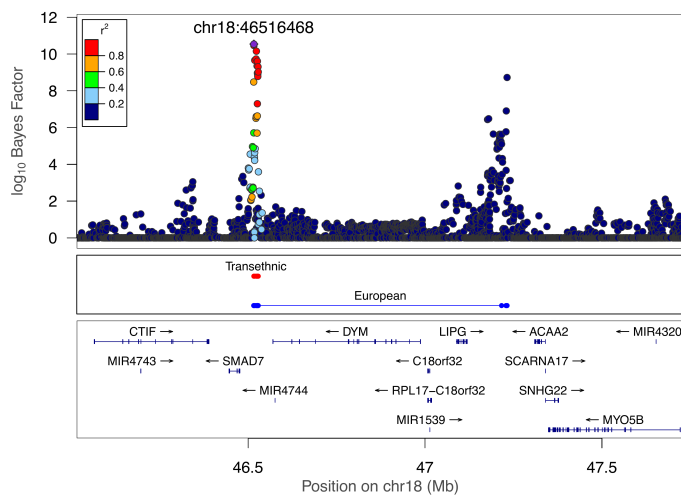
4. Previously reported loci in the Trans-ancestry meta-analysis

Trans-ancestry 99% credible set
European 99% credible set



4. Previously reported loci in the Trans-ancestry meta-analysis

Trans-ancestry 99% credible set
European 99% credible set



4. Previously reported loci in the Trans-ancestry meta-analysis

Trans-ancestry 99% credible set
European 99% credible set

