ARTICLE

Trans-ethnic Meta-analysis and Functional Annotation Illuminates the Genetic Architecture of Fasting Glucose and Insulin

Ching-Ti Liu,[1,84,*] Sridharan Raghavan,[2,3,4,84] Nisa Maruthur,[5,6,7,84] Edmond Kato Kabagambe,[8,84] Jaeyoung Hong,[1] Maggie C.Y. Ng,[9,10] Marie-France Hivert,[11,12,13] Yingchang Lu,[14,15] Ping An,[16] Amy R. Bentley,[17] Anne M. Drolet,[18] Kyle J. Gaulton,[19] Xiuqing Guo,[20] Loren L. Armstrong,[21] Marguerite R. Irvin,[22] Man Li,[7] Leonard Lipovich,[18,23] Denis V. Rybin,[1] Kent D. Taylor,[20] Charles Agyemang,[24] Nicholette D. Palmer,[9,25] Brian E. Cade,[26] Wei-Min Chen,[27] Marco Dauriz,[28] Joseph A.C. Delaney,[29] Todd L. Edwards,[8] Daniel S. Evans,[30] Michele K. Evans,[31] Leslie A. Lange,[32] Aaron Leong,[2] Jingmin Liu,[33] Yongmei Liu,[34] Uma Nayak,[27] Sanjay R. Patel,[35] Bianca C. Porneala,[2] Laura J. Rasmussen-Torvik,[36] Marieke B. Snijder,[24] Sarah C. Stallings,[83] Toshiko Tanaka,[37] Lisa R. Yanek,[38] Wei Zhao,[39] Diane M. Becker,[38,40] Lawrence F. Bielak,[39] Mary L. Biggs,[41,42] Erwin P. Bottinger,[14] Donald W. Bowden,[9,10,25] Guanjie Chen,[17] Adolfo Correa,[43] David J. Couper,[44] Dana C. Crawford,[45] Mary Cushman,[46] John D. Eicher,[47,48] Myriam Fornage,[49] Nora Franceschini,[50] Yi-Ping Fu,[51] Mark O. Goodarzi,[52] Omri Gottesman,[14] Kazuo Hara,[14,53,54] Tamara B. Harris,[55] Richard A. Jensen,[42] Andrew D. Johnson,[48] Min A. Jhun,[39] Andrew J. Karter,[56] Margaux F. Keller,[57] Abel N. Kho,[21] Jorge R. Kizer,[58,59] Ronald M. Krauss,[60] Carl D. Langefeld,[61,62] Xiaohui Li,[20]

(Author list continued on next page)

Knowledge of the genetic basis of the type 2 diabetes (T2D)-related quantitative traits fasting glucose (FG) and insulin (FI) in African ancestry (AA) individuals has been limited. In non-diabetic subjects of AA (n = 20,209) and European ancestry (EA; n = 57,292), we performed trans-ethnic (AA+EA) fine-mapping of 54 established EA FG or FI loci with detailed functional annotation, assessed their relevance in AA individuals, and sought previously undescribed loci through trans-ethnic (AA+EA) meta-analysis. We narrowed credible sets of variants driving association signals for 22/54 EA-associated loci; 18/22 credible sets overlapped with active islet-specific enhancers or transcription factor (TF) binding sites, and 21/22 contained at least one TF motif. Of the 54 EA-associated loci, 23 were shared between EA and AA. Replication with an additional 10,096 AA individuals identified two previously undescribed FI loci, chrX *FAM133A* (rs213676) and chr5 *PELO* (rs6450057). Trans-ethnic analyses with regulatory annotation illuminate the genetic architecture of glycemic traits and suggest gene regulation as a target to advance precision medicine for T2D. Our approach to utilize state-of-the-art functional annotation and implement trans-ethnic association analysis for discovery and fine-mapping offers a framework for further follow-up and characterization of GWAS signals of complex trait loci.

Introduction

The global burden of type 2 diabetes (T2D [MIM: 125853]) is borne disproportionately by populations with little genetic European ancestry (EA), especially African Americans.[1] Although environmental and behavioral factors account for a large portion of these observed race-ethnic disparities, genetic variation also contributes[2,3]

[1]Department of Biostatistics, School of Public Health, Boston University, Boston, MA 02118, USA; [2]Division of General Internal Medicine, Massachusetts General Hospital, Harvard Medical School, Boston, MA 02114, USA; [3]Department of Veterans Affairs Medical Center, Eastern Colorado Health Care System, Denver, CO 80220, USA; [4]Division of General Internal Medicine, Department of Medicine, University of Colorado School of Medicine, Denver, CO 80220, USA; [5]Division of General Internal Medicine, Johns Hopkins University School of Medicine, Baltimore, MD 21287, USA; [6]Welch Center for Prevention, Epidemiology, and Clinical Research, Johns Hopkins University, Baltimore, MD 21287, USA; [7]Department of Epidemiology, Johns Hopkins University Bloomberg School of Public Health, Baltimore, MD 21287, USA; [8]Division of Epidemiology, Department of Medicine, School of Medicine, Vanderbilt University Medical Center, Nashville, TN 37203, USA; [9]Center for Genomics and Personalized Medicine Research, Wake Forest University School of Medicine, Winston-Salem, NC 27157, USA; [10]Center for Diabetes Research, Wake Forest University School of Medicine, Winston-Salem, NC 27157, USA; [11]Department of Population Medicine, Harvard Pilgrim Health Care Institute, Harvard Medical School, Boston, MA 02215, USA; [12]Diabetes Unit, Massachusetts General Hospital, Boston, MA 02114, USA; [13]Department of Medicine, Universite de Sherbrooke, Sherbrooke, QC J1G 0A2, Canada; [14]The Charles Bronfman Institute for Personalized Medicine, The Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA; [15]The Genetics of Obesity and Related Metabolic Traits Program, The Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA; [16]Division of Statistical Genomics, Department of Genetics, School of Medicine, Washington University, St Louis, MO 63108, USA; [17]Center for Research on Genomics and Global Health, National Human Genome Research Institute, NIH, Bethesda, MD 20892, USA; [18]Center for Molecular Medicine and Genetics, School of Medicine, Wayne State University, Detroit, MI 48201, USA; [19]Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford OX3 7BN, UK; [20]Institute for Translational Genomics and Population Sciences, Department of Pediatrics, Los Angeles Biomedical Research Institute, Harbor-UCLA Medical Center, Torrance, CA 90502, USA; [21]Feinberg School of Medicine, Northwestern University, Chicago, IL 60611, USA; [22]Department of Epidemiology, School of Public Health, University of Alabama – Birmingham, Birmingham, AL 35294, USA; [23]Department of Neurology, School of Medicine, Wayne State University, Detroit, MI 48201, USA; [24]Department of Public Health, Academic Medical Center Amsterdam, Meibergdreef 15, 1105 AZ Amsterdam, the Netherlands; [25]Department of Biochemistry, Wake Forest University School of Medicine, Winston-Salem, NC 27157, USA; [26]Division of Sleep and Circadian Disorders, Brigham and

(Affiliations continued on next page)

Jingling Liang,[63] Simin Liu,[64,65] William L. Lowe, Jr.,[21] Thomas H. Mosley,[66] Kari E. North,[50] Jennifer A. Pacheco,[21] Patricia A. Peyser,[39] Alan L. Patrick,[67] Kenneth M. Rice,[41] Elizabeth Selvin,[6,7] Mario Sims,[43] Jennifer A. Smith,[39] Salman M. Tajuddin,[31] Dhananjay Vaidya,[7,38] Mary P. Wren,[25] Jie Yao,[20] Xiaofeng Zhu,[63] Julie T. Ziegler,[61,62] Joseph M. Zmuda,[68] Alan B. Zonderman,[69] Aeilko H. Zwinderman,[24] AAAG Consortium, CARe Consortium, COGENT-BP Consortium, eMERGE Consortium, MEDIA Consortium, Adebowale Adeyemo,[17] Eric Boerwinkle,[49] Luigi Ferrucci,[37] M. Geoffrey Hayes,[21] Sharon L.R. Kardia,[39] Iva Miljkovic,[68] James S. Pankow,[70] Charles N. Rotimi,[17] Michele M. Sale,[27] Lynne E. Wagenknecht,[71] Donna K. Arnett,[72] Yii-Der Ida Chen,[20] Michael A. Nalls,[73] MAGIC Consortium, Michael A. Province,[16] W.H. Linda Kao,[7,85] David S. Siscovick,[29,42,74] Bruce M. Psaty,[29,42,75,76] James G. Wilson,[77] Ruth J.F. Loos,[14,15,78] Josée Dupuis,[1,47] Stephen S. Rich,[27] Jose C. Florez,[12,79,80,81] Jerome I. Rotter,[20] Andrew P. Morris,[19,82] and James B. Meigs[2,*]

but remains understudied in persons of mostly or all genetic African ancestry (AA).[2–4] A few studies have examined the association signals of EA-associated loci with levels of fasting glucose (FG) and insulin (FI) in ethnic minorities, but on a relatively small scale.[5–7] Genome-wide association studies (GWASs) with meta-analysis in EA populations have identified more than 50 loci associated with T2D-related quantitative traits (QTs), particularly levels of fasting glucose (FG) and insulin (FI).[8] Associated SNPs at these loci are common, with modest effect sizes.[8–10] At most SNPs the causal action remains unknown, because most lie in non-coding regions of the genome. Now, these have been annotated for regulatory function.[11–14] We collected a large sample of AA

Women's Hospital, Harvard Medical School, Boston, MA 02115, USA; [27]Center for Public Health Genomics, Department of Public Health Sciences, School of Medicine, University of Virginia, Charlottesville, VA 22908, USA; [28]Division of Endocrinology, Diabetes & Metabolism, Department of Medicine, University of Verona, 37126 Verona, Italy; [29]Department of Epidemiology, University of Washington, Seattle, WA 98195, USA; [30]California Pacific Medical Center Research Institute, San Francisco, CA 94107, USA; [31]Health Disparities Research Section, Laboratory of Epidemiology and Population Sciences, National Institute on Aging, NIH, Baltimore, MD 21224, USA; [32]Department of Genetics, University of North Carolina, Chapel Hill, NC 27607, USA; [33]Fred Hutchinson Cancer Research Center, Seattle, WA 98109, USA; [34]Center for Human Genetics, Division of Public Health Sciences, Wake Forest University School of Medicine, Winston-Salem, NC 27157, USA; [35]Department of Medicine, University of Pittsburgh, Pittsburgh, PA 15213, USA; [36]Department of Preventive Medicine, Feinberg School of Medicine, Northwestern University, Chicago, IL 60611, USA; [37]Translational Gerontology Branch, National Institute of Aging at Harbor Hospital, Baltimore, MD 21225, USA; [38]GeneSTAR Research Program, Division of General Internal Medicine, Department of Medicine, Johns Hopkins University, Baltimore, MD 21287, USA; [39]Department of Epidemiology, School of Public Health, University of Michigan, Ann Arbor, MI 48109, USA; [40]Department of Health Policy and Management, Johns Hopkins University Bloomberg School of Public Health, Baltimore, MD 21287, USA; [41]Department of Biostatistics, University of Washington, Seattle, WA 98195, USA; [42]Cardiovascular Health Research Unit, Department of Medicine, School of Medicine, University of Washington, Seattle, WA 98195, USA; [43]Department of Medicine, University of Mississippi Medical Center, Jackson, MS 39216, USA; [44]Collaborative Studies Coordinating Center, Department of Biostatistics, Gillings School of Global Public Health, University of North Carolina, Chapel Hill, NC 27514, USA; [45]Department of Epidemiology and Biostatistics, Institute for Computational Biology, Case Western Reserve University, Cleveland, OH 44106, USA; [46]Department of Medicine and Pathology, University of Vermont, College of Medicine, Burlington, VT 05405, USA; [47]National Heart, Lung, and Blood Institute's Framingham Heart Study, Framingham, MA 01702, USA; [48]Population Sciences Branch, Division of Intramural Research, National Heart, Lung, and Blood Institute, NIH, Framingham, MA 01702, USA; [49]Institute of Molecular Medicine and Human Genetics Center, University of Texas Health Science Center at Houston, Houston, TX 77030, USA; [50]Department of Epidemiology, University of North Carolina, Chapel Hill, NC 27514, USA; [51]Cardiovascular Epidemiology and Human Genomics Branch, National Heart, Lung, and Blood Institute, NIH, Framingham, MA 01702, USA; [52]Division of Endocrinology, Diabetes & Metabolism, Cedars-Sinai Medical Center, Los Angeles, CA 90048, USA; [53]Department of Diabetes and Metabolic Diseases, Graduate School of Medicine, The University of Tokyo, Tokyo 113-8655, Japan; [54]Department of Diabetes, Endocrinology, and Metabolism, Tokyo Medical University, Tokyo 163-0023, Japan; [55]Laboratory of Epidemiology and Population Sciences, NIH, Bethesda, MD 20892, USA; [56]Division of Research, Kaiser Permanente, Northern California Region, Oakland, CA 94612, USA; [57]Department of Genetics and Pharmacogenomics, Merck Research Laboratories, 33 Avenue Louis Pasteur, Boston, MA 02115, USA; [58]Department of Medicine, Albert Einstein College of Medicine, Montefiore Medical Center, Bronx, NY 10461, USA; [59]Department of Epidemiology and Population Health, Albert Einstein College of Medicine, Bronx, NY 10461, USA; [60]Children's Hospital Oakland Research Institute, Oakland, CA 94609, USA; [61]Center for Public Health Genomics, Wake Forest University School of Medicine, Winston-Salem, NC 27157, USA; [62]Department of Biostatistical Sciences, Wake Forest University School of Medicine, Winston-Salem, NC 27157, USA; [63]Department of Epidemiology and Biostatistics, Case Western Reserve University, Cleveland, OH 44106, USA; [64]Department of Epidemiology, Brown University, Providence, RI 02912, USA; [65]Department of Medicine, Brown University, Providence, RI 02903, USA; [66]Division of Geriatrics/Gerontology, Department of Medicine, University of Mississippi Medical Center, Jackson, MS 39216, USA; [67]Tobago Health Studies Office, Scarborough, Tobago, Trinidad and Tobago; [68]Department of Epidemiology, University of Pittsburgh, Pittsburgh, PA 15213, USA; [69]Behavioral Epidemiology Section, Laboratory of Epidemiology & Population Science, Intramural Research Program, National Institute on Aging, NIH, Baltimore, MD 21224, US; [70]Division of Epidemiology and Community Health, School of Public Health, University of Minnesota, Minneapolis, MN 55455, USA; [71]Division of Public Health Sciences, Wake Forest University School of Medicine, Winston-Salem, NC 27157, USA; [72]University of Kentucky College of Public Health, Lexington, KY 40563, USA; [73]Laboratory of Neurogenetics, National Institute on Aging, NIH, Bethesda, MD 20892, USA; [74]The New York Academy of Medicine, New York, NY 10029, USA; [75]Department of Health Services, University of Washington, Seattle, WA 98195, USA; [76]Group Health Research Institute, Group Health Cooperative, Seattle, WA 98101, USA; [77]Department of Physiology and Biophysics, University of Mississippi Medical Center, Jackson, MS 39216, USA; [78]The Mindich Child Health and Development Institute, The Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA; [79]Center for Human Genetic Research, Massachusetts General Hospital, Boston, MA 02114, USA; [80]Programs in Metabolism and Medical & Population Genetics, Broad Institute, Cambridge, MA 02142, USA; [81]Department of Medicine, Harvard Medical School, Boston, MA 02115, USA; [82]Institute of Translational Medicine, Department of Biostatistics, University of Liverpool, Liverpool L69 3BX, UK; [83]Vanderbilt Institute for Clinical and Translational Research, Vanderbilt University Medical Center, Nashville, TN 37203, USA
[84]These authors contributed equally to this work
[85]Deceased
*Correspondence: ctliu@bu.edu (C.-T.L.), jmeigs@partners.org (J.B.M.)
http://dx.doi.org/10.1016/j.ajhg.2016.05.006.

individuals for genetic study and, taking advantage of differences in linkage disequilibrium (LD) patterns across EA and AA, used a trans-ethnic analytic approach to improve mapping resolution[15] and narrow the number of potential causal SNPs at associated loci.[15,16] We then characterized predicted SNP function with detailed annotation information from diverse sources. We hypothesized that a trans-ethnic approach would identify SNPs with high likelihood of having regulatory, causal function, with results illuminating mechanisms underlying glycemic regulation in African Americans as well as whites of European ancestry.

We created the African American Glucose and Insulin Genetic Epidemiology (AAGILE) Consortium, with up to 20,209 AA individuals from 16 cohorts, to conduct a fixed effects meta-analysis of association summary statistics at 3.3 million (HapMap2) SNPs for levels of FG and body mass index (BMI)-adjusted FI. We then combined meta-analysis results from AAGILE with those from the EA Meta-Analyses of Glucose and Insulin-related traits Consortium (MAGIC, n = 57,292)[10] with three aims in mind: (1) conduct trans-ethnic fine-mapping of 54 T2D QT loci (36 FG, 16 FI, 2 associated with both FG and FI) identified from EA and combine fine-mapping with annotation resources including RegulomeDB, ENCyclOpedia of DNA Elements (ENCODE), Islet Regulome, and Functional ANnotation of The mammalian genOMe Consortium (FANTOM);[11–14] (2) assess the biologic relevance (allelic heterogeneity, transferability, population genetic selection, and consistency of association with T2D or insulin resistance traits) of the 54 EA FG and FI loci in AA individuals; and (3) identify additional FG and FI variants by combining association results from AAGILE and MAGIC using Meta-Analysis of TRans-ethnic Association Studies (MANTRA)[15] followed by de novo or in silico replication in additional AA samples (n up to 10,096) for 62 potential additional SNPs that met pre-specified significance levels from the trans-ethnic meta-analysis. The study design is illustrated in Figure S1 and characteristics of each participating cohort are described in Table S1.

Material and Methods

Research Participants

A total of 20,209 (for FG) and 17,871 (for FI) non-diabetic men and women of African ancestry (AA) from 16 cohorts participated in stage 1 (Table S1). Additionally, up to 10,096 (for FG) and 6,669 (for FI) non-diabetic individuals from 14 cohorts were included in a stage 2 replication analyses. Participants were excluded from this study if they had a diagnosis of T2D by a physician, were on any diabetes treatment, or had a FG concentration equal to or greater than 7 mmol/L. HbA1c levels were not used as diagnostic criteria. FG and FI GWAS data for 57,292 (FG) and 52,328 (FI) EA individuals were obtained from MAGIC.[10] Each participating study has obtained institutional review board approval and all subjects provided written informed consent.

Genetic Variants

Genotyping was conducted in each cohort using commercially available genome-wide SNP arrays with quality control criteria for variants before imputation listed in Table S1. In all stage 1 discovery analyses, imputation was performed to infer ungenotyped variants and fill in missing genotypes. We used phased haplotype data from the CEU and YRI HapMap phase 2 samples for the majority of contributing studies, using MACH[17] or IMPUTE2.[18] Variants with lower imputation quality scores (MACH $r^2 < 0.30$ or IMPUTE2 information score < 0.40) were excluded from further analyses. Approximately 3.3 million directly genotyped or imputed SNPs, including ~78,000 from the X chromosome, passed the quality control filters and were evaluated for association.

Traits and Covariates

In all cohorts, fasting blood samples were obtained from participants after an overnight (≥8 hr) fast. Detailed descriptions of study-specific FG and FI measurements are given in Table S1. Analyses of untransformed levels of FG and natural logarithm-transformed FI were adjusted for age, age squared, sex, and principal components (PC) for ancestry. In addition, we adjusted FI levels for BMI to reduce confounding by obesity.[10] SNP-trait associations were tested using additive genetic models. Additional cohort-specific covariates (like study center for cohorts with multiple sites or relatedness for studies containing families) were included at the discretion of each cohort (Table S1).

Overview of Study Design and Analysis Strategy

The overall study design is shown in Figure S1. We first performed fixed-effect meta-analyses of FG and FI in AA samples. We then conducted trans-ethnic meta-analyses by combining the fixed-effect meta-analysis results from AAGILE with MAGIC. We fine-mapped FG and FI loci previously identified in EA by constructing 99% credible sets.[15] Second, we used results from the fixed-effect meta-analysis in AA to assess whether FG and FI loci identified in EA populations have genetic concordance or biological relevance in AA. Third, we carried forward 62 SNPs (not previously described to be associated with FG or FI in persons of EA) based on low fixed-effect meta-analysis p values in AA or high log-Bayes factor (log(BF)) in the combined AA and EA trans-ethnic analysis for follow-up in additional non-diabetic samples of AA to identify additional FG and FI association signals. Specifically, the threshold for SNP promotion to replication was a fixed-effect meta-analysis $p < 10^{-6}$ in AA samples, or $p < 10^{-5}$ with log(BF) > 5 in MANTRA. Meta-analyses were performed at two different sites and summary statistics were crosschecked to ensure consistency of results.

Meta-analysis of Samples from AAGILE and MAGIC Consortia

Each participating study from the AAGILE consortium performed a cohort-specific association analysis under an additive genetic model to test the genetic association of each genetic variant with FG and FI. The cohort-specific genome-wide association results were corrected with genomic control, unless $\lambda_{GC} < 1$, before meta-analyzing the GWAS results.[17] We then conducted a fixed-effects meta-analysis with inverse-variance weighting implemented in the program METAL to aggregate the cohort-specific association results.[19]

To leverage differential patterns of LD between common variants in EA and AA populations, we meta-analyzed GWAS results from AA in AAGILE and previously published EA results from MAGIC[10] using the Meta-ANalysis of Trans-ethnic Association Studies (MANTRA) software.[15,20] The results from MANTRA were used to fine-map the 54 loci (36 FG, 16 FI, 2 associated with both FG and FI) previously identified in EA samples and to prioritize variants for discovery of previously undescribed variants associated with FG and FI in AA samples.

Construction of 99% Credible Sets

To improve fine-mapping resolution, we first constructed 99% credible sets for previously reported FG and FI loci identified in EA samples (36 FG, 16 FI, 2 associated with both FG and FI).[8–10] We identified the genomic region 250 kb upstream and 250 kb downstream of the lead SNP from the EA meta-analysis and defined BF_k, obtained from MANTRA analysis, as the Bayes factor for SNP k. We calculated the posterior probability that SNP k is functional or tagging an unobserved causal SNP by $BF_k/\sum_i BF_i$, where i indexes SNPs in the locus of interest. The 99% credible sets are the collection of the minimum number of variants providing a cumulative posterior probability greater or equal to 0.99 for representing the causal variant at a given locus.[21]

Annotation of Credible Set SNPs

We focused our annotation analysis on loci that showed a 99% credible set reduction of at least 20% in the length of the genomic interval spanned by the variants in the credible set or in the number of variants included in the credible set (13 FG, 8 FI loci, and 1 locus for both FG and FI). When examining the distribution of credible set reduction across all 54 FG and FI loci, we noted that there appeared to be a natural break point between 20% and 12%, so we selected 20% reduction as a threshold defining loci with substantial reduction. At these loci, we classified SNPs in the credible set into two groups: one group with SNPs included in the 99% credible set from the trans-ethnic fine-mapping with MANTRA and a second group that included only SNPs that were included in the 99% credible set using EA samples (MAGIC) but were excluded from the trans-ethnic 99% credible set. For brevity, we henceforth refer to these categories of SNPs as the "narrowed sets" and "excluded sets," respectively.

To compare functional annotation of the narrowed versus excluded credible sets, we annotated SNPs from the narrowed and excluded sets separately using the HaploReg tool that searches for dbSNP annotation (synonymous substitution, non-synonymous substitution, lying within an intron, 5′ UTR, 3′ UTR, or lncRNA, conservation in mammals, or having unknown position or function), genomic position, distance to the nearest named known protein-coding gene, eQTL data, and transcription factor motif, transcription factor binding site (TFBS), DNASe hypersensitivity (DHS), and histone marks associated with promoters and enhancers derived from the ENCODE and Roadmap Epigenomics consortia.[12] We dichotomized the data from HaploReg for each SNP in the narrowed and excluded sets based on whether there was evidence of each specific annotation. To compare the narrow and excluded sets of SNPs for each trait, we performed a Fisher's exact test to assess differences in proportions of SNPs in each set with a specific annotation characteristic.

We visually examined overlap between trait association and regulatory annotation by plotting association statistics and regulatory data together. For each of the 22 loci with a reduction in the size of at least 20% (either length of or the count of SNPs included in) of the 99% credible set after trans-ethnic analysis, we used RegulomeDB to generate a single numeric score summarizing the strength of regulatory data associated with each SNP in the locus (within 250 kb on either side of the index SNP for that locus from MAGIC).[11] In brief, RegulomeDB uses data from ENCODE and other published literature to annotate SNPs based on overlap with TFBS, TF motifs, DHS, eQTLs, and promoter histone marks and creates a score for each SNP ranging from 1 to 7, with 1 corresponding to the strongest degree of regulatory annotation, 6 corresponding to the weakest degree of regulatory annotation, and 7 representing no data available. We then used statistical software R to make regional association plots as described above using the log(BF) for each SNP as the association statistic, with the color of each plotted SNP corresponding to its RegulomeDB score and the size of each plotted SNP corresponding to LD with the MAGIC index SNP in the YRI population. Finally, to visually examine overlap between the 99% credible regions at these 22 loci and regulatory data derived from pancreatic islets, we used the Islet Regulome Browser to generate plots with the same coordinates as represented in the regional plots and aligned the schematic representation of islet-derived regulatory annotation from the Islet Regulome Browser to the regional association plot.[13]

We performed additional annotation analyses for the narrowed set (from the trans-ethnic meta-analysis). The starting and ending chromosome position of the 99% credible region from the trans-ethnic analysis for each of the 22 loci with substantially reduced 99% credible sets were entered into the Islet Regulome Browser, and we cataloged the presence/absence of binding sites for five transcription factors (FOXA2, MAFB, NKX2.2, NKX6.1, PDX1) or the insulator protein CTCF, and histone marks associated with promoters, active enhancers, and inactive enhancers.[11,13] In addition to examining overlap between credible set SNPs and regulatory annotation, we also manually annotated the credible set intervals for the 22 loci with substantially reduced 99% credible sets. The genomic interval for each narrowed trans-ethnic credible set was examined in the UCSC Genome Browser, and we cataloged RNA expression, DHS sites, TFBS, and promoter and enhancer histone marks in cell types relevant to FG and FI, namely liver, pancreatic, adipose, and muscle.

Genomic Annotation Enrichment Analysis

We tested for enrichment of chromatin state marks and TFBS using all variants in the trans-ethnic meta-analysis credible sets. We pooled chromatin states for promoter (TssA, TssFlnk) and enhancer (EnhA, EnhWk) elements for 93 cell types (after excluding cancer lines) from the Roadmap Epigenomics Consortium[22] and used TF binding data for 165 proteins from ENCODE[23] and the Islet Regulome.[13] For a given annotation, we calculated the cumulative posterior probability of annotated variants at each locus and then averaged these values across all loci. We then generated a null distribution for this procedure by randomly shuffling the probabilities among variants at each locus, recalculating the average probability for annotated variants, and repeating this procedure 1,000,000 times. We estimated fold-enrichment for each annotation by dividing the observed value by the permuted value averaged across all permutations. We then calculated a p value for the enrichment as the proportion of permutations for which the resulting value was greater than or equal to that observed. We applied a corrected significance threshold of 0.00019 (0.05/257 annotations).

Interrogation of Transferability across Populations

We investigated the transferability to AA of EA FG- and FI-associated SNPs and loci. To evaluate SNP transferability, we tested in AA the association of index FG and FI SNPs from EA individuals in MAGIC[8–10] (i.e., EA FG or FI SNPs with $p < 5 \times 10^{-8}$). We defined SNP transferability as an EA index SNP showing in AA an association that was statistically significant ($p < 0.05$) and consistent in direction of effect as in the EA meta-analysis. To evaluate locus transferability, given differences in local LD structure across populations, we also interrogated the flanking ±250 kb regions of the index SNP in AA to search for any SNPs with a smaller association p than the EA index SNP. For locus transferability, we used a Bonferroni corrected p to determine the significance for each locus by adjusting for the effective number of independent tests within that locus, using the Li and Ji approach.[24]

Conditional Analysis of Signals with Significant SNP Associations in AA

We performed approximate conditional association analyses at loci with significant QT associations in AAGILE in order to test whether the associated AAGILE SNP was the same association signal as the MAGIC SNP. We used genome-wide complex trait analysis (GCTA)[25] for this analysis, because it allows approximate conditional analyses in results from meta-analysis without the need for individual cohort data to conduct the tests. GCTA approximated the variance-covariance matrix of genotype using estimated allele frequency from the meta-analysis results and LD between SNPs from a reference sample. We calculated the association of the AA best SNP conditional on the EA index SNP within the same locus in AA samples.

Concordance Analysis across Ancestry Groups

For this analysis, we considered SNPs that passed QC and had a MAF > 1% in both EA- and AA-specific meta-analyses. We further excluded (1) EA-associated FG or FI loci, defined as those lying 500 kb upstream or downstream of previously reported SNPs for each loci as described,[16] and (2) AT/GC SNPs to avoid the potential bias introduced by the strand misalignment between EA- and AA-specific meta-analysis. We then classified all the remaining SNPs into categories based on the association p value in the MAGIC EA samples: $p \leq 0.01$, $0.01 < p \leq 0.5$, and $0.5 < p \leq 1$. For the SNPs within each category, we then selected a set of independent SNPs by identifying the most significant SNP, omitting the SNPs within 500 kb region apart from the most significant SNP, and then repeating this process until there were no more SNPs left. We then determined the direction of effect for the EA trait-raising allele between EA and AA samples and we calculated the proportion of these selected SNPs that share the same direction of effect. To determine the significance of the excess in concordance (with 50% expected), we then conducted one-sided binomial tests.

Population Differentiation and Natural Selection at QT Loci

We applied several approaches to evaluate population differentiation and natural selection at index FG or FI SNPs, using the trait-raising allele in EA as the risk allele. First, we compared the risk-allele frequencies in EA versus AA by calculating the absolute value difference between the risk-allele frequency in EA and the risk-allele frequency in AA for each index SNP. Second, we used Wright's fixation index[26] (F_{st}) to measure the degree of the population differentiation due to genetic drift and reflected by the divergent allele frequencies. A value of F_{st} lying in range of 0 to 0.05 indicates little genetic differentiation; a value 0.05 to 0.15, moderate differentiation; and a value greater than 0.15, greater population differentiation.[27,28] We calculated F_{st} using risk-allele frequencies obtained from AAGILE cohorts for AAs and from MAGIC[8–10] for EAs. We also calculated the F_{st} using CEU and YRI HapMap2 data from two samples of equal size to ensure that the imbalance in sample size of our EA and AA samples does not obscure the population differentiation at any locus. Third, we used Haplotter[29] to calculate the integrated haplotype score (iHS) in HapMap2 data to measure the amount of extended haplotype homozygosity and hence the evidence of recent positive selection at the index SNP. Fourth, we compared the effect alleles from the index SNP of each locus against the human genome to determine whether it was the major or minor allele using the UCSC Genome Browser GRCh37/hg19 version, produced by the Genome Reference Consortium in 2009.[30]

Associations of QT Loci with T2D and Insulin Resistance-Related Traits in AA Individuals

As described below, we assembled new African ancestry cohort data sources for triglycerides and initiated collaborations with new and existing consortia for body mass index (BMI), waist-to-hip ratio adjusted for BMI (WHRadjBMI), systolic and diastolic blood pressure, hypertension (HTN), and low-density and high-density lipoprotein cholesterol (LDL-C and HDL-C) to investigate the association of T2D QT SNPs with T2D and insulin resistance-related traits (BMI, WHRadjBMI, HTN, LDL-C, HDL-C, and triglycerides) in AA individuals. Specifically, we investigated the association of 25 SNPs, including 24 most associated SNPs (14 FG, 9 FI, and 1 SNP, rs780094, associated with both FG and FI) residing in the regions demonstrating locus transferability in AA and 1 previously undescribed FI SNP (rs6450057). We also tested the association of the FI SNP on chromosome X, rs213676, with lipid traits (the only traits available for chromosome X). HapMap2-imputed GWAS meta-analysis summary statistics for 25 SNPs were obtained from the Meta-analysis of T2D in African Americans (MEDIA) consortium for T2D[31] (n up to 23,818), the African Ancestry Anthropometry Genetics Consortium (AAAGC) for BMI[32] (n up to 39,141) and WHRadjBMI[32,33] (n up to 19,049), and the Continental Origins and Genetic Epidemiology Network (COGENT) for HTN and blood pressure[34] (n up to 29,828). HapMap-imputed GWAS meta-analysis summary statistics for 26 SNPs were obtained from the Candidate gene Association Resource (CARe) consortium for LDL-C and HDL-C[35] (n up to 8,090) and the Electronic Medical Records and Genomics Network (eMERGE) for triglycerides[36] (n up to 2,838). SNPs with $p < 0.05$ and same direction of effect (i.e., FG/FI-increasing alleles associated with T2D or higher levels of the quantitative insulin-resistance traits, except for HDL-C where FG/FI-increasing alleles were expected to be associated with lower HDL-C) were considered significant.

Discovery and Replication of Previously Undescribed FG and FI SNPs in AA

For identification of additional FG- and FI-associated loci in AA, we took a two-stage (discovery followed by replication) approach (Figure S1). In the discovery stage, we used fixed-effect meta-analysis results in AA and the trans-ethnic meta-analysis (MANTRA) results as described in the Meta-analysis of Samples from AAGILE and MAGIC Consortia section. We identified 62 variants, not

previously reported to be associated with FG or FI in any ancestry, classified into three tiers of decreasing restrictiveness based on low fixed-effect meta-analysis p values in AA or high trans-ethnic meta-analysis (MANTRA) log(BF) in the combined AA and EA results, for follow-up in 10,096 additional AA samples from 16 additional, independent cohorts (Table S1). Identified variants were classified into three tiers to take forward to the replication stage. Tier 1 was variants with a fixed-effect meta-analysis $p < 10^{-6}$ in AA samples, or $p < 10^{-5}$ with log(BF) > 5 in MANTRA; tier 2 was variants with log(BF) > 4 *and* fixed-effect $p < 10^{-5}$; and tier 3 was variants with log(BF) > 4 *or* fixed-effect $p < 10^{-5}$ (Figure S1).

The 16 additional independent replication cohorts are listed in Table S1. For replication, we sought either in silico look-ups of the 62 SNPs that met criteria for one of the three tiers from the discovery stage in the cohorts with extant genotyping array data or conducted de novo genotyping for SNPs in tier 1 in additional cohorts with DNA and trait levels. Each participating replication cohort implemented the same model used for discovery analyses to evaluate associations between SNPs and traits. First we compared discovery and replication results for significance and direction of effect. Then we meta-analyzed the discovery and replication stage results to obtain a combined, fixed effect inverse variance estimate for each of the 62 SNPs. Genome-wide statistical significance was set at $p \leq 5 \times 10^{-8}$, and associations were considered to be previously undescribed if they were not in LD ($r^2 < 0.3$ or not within 500 kb of a previously reported glycemia-associated SNP).

Finally, we performed a trans-ethnic meta-analysis in MANTRA combining fixed effects estimates from the AAGILE discovery and replication combined meta-analysis with published EA results from MAGIC to get a trans-ethnic total effect size using all available data for 62 SNPs. We considered an association to have reached genome-wide significance if the $p \leq 5 \times 10^{-8}$ in fixed-effect meta-analysis or log(BF) from MANTRA was greater than 6, and we considered the association to be previously undescribed if the variants were not in LD ($r^2 < 0.3$) or not within 500 kb of a SNP previously reported to be associated with FG or FI.

Results

Trans-ethnic Fine-Mapping and Annotation of Glycemic QT Loci Established in EA Populations

To fine-map 54 loci previously associated with FG or FI in EA,[8–10] we constructed 99% credible sets, the smallest set of SNPs that accounts for 99% of the posterior probability of containing the causal variant at the locus, using meta-analysis results only from MAGIC EA samples and trans-ethnic meta-analysis results from both MAGIC EA samples and AAGILE AA samples (Table S2). Reflecting increased sample size and differences in LD structure between ancestry groups, trans-ethnic meta-analyses yielded more than 20% reduction in either the number of SNPs or the genomic interval spanned by the SNPs in credible sets for 22/54 loci (13 FG, 8 FI, and 1 associated with both FG and FI; Table 1 and Figures S2 and S3) while we also observed some loci with substantially enlarged credible sets. For 4 of these 22 loci (*GCK* [MIM: 138079] and *ADCY5* [MIM: 600293] for FG, *PPP1R3B* [MIM: 610541] for FI, and *GCKR* [MIM: 600842] for both) with a >20% reduction in the credible set, the credible set included a single SNP. We observed the greatest reduction (95%) in the number of SNPs in the credible set at the *FOXA2* locus, where the genomic width of the credible set was also greatly reduced (from ~46 kb to ~4 kb; 92% reduction), and the extent of LD surrounding the index SNP was less in AA than EA (Figure 1). The narrowed trans-ethnic CSs contained previously described functional variants at several loci, including the coding SNP rs1801282 (GenBank: NC_000003.12; g.12351626C>G [p.Pro12Ala]) in *PPARG* (MIM: 601487)[37] and rs7903146 in *TCF7L2* (MIM: 602228), which has been shown to overlap an islet enhancer and modify enhancer activity.[38] In contrast, the coding SNP rs1260326 at the *GCKR* locus, presumed to be causal based on prior studies,[39,40] was excluded from the 99% credible set. With the exception of non-synonymous variants at *DPYSL5* (MIM: 608383), *COBLL1* (MIM: 610318)-*GRB14* (MIM: 601524), and *PPARG* (Table 2), the greatly reduced credible sets mapped predominantly to non-coding sequences.

At 22 loci with a >20% reduction in the credible set, we compared functional annotations of SNPs in the trans-ethnically generated credible sets ("narrowed" set) to those of SNPs excluded from the EA-only credible sets ("excluded" set) using HaploReg annotation.[12] For the narrowed set of nine FI loci, the SNP annotation from dbSNP indicated only a reduction in intronic SNPs and an increase in unknown function SNPs compared to the excluded set (Table S3). However, regulatory annotation data showed that the narrowed set of nine FI loci was enriched for enhancer-associated chromatin marks and eQTLs when compared to the excluded set (76.3% versus 66.0%, $p = 0.004$ for enhancer marks; 39.4% versus 28.5%, $p = 0.002$ for eQTLs [Table S3]). At the 14 FG loci, we observed an enrichment of SNPs in 3′ UTR of genes, but no enrichment of regulatory annotations in the narrow set compared to the excluded set (Table S4). Table S5 provides annotation information from HaploReg for each SNP in these 22 derived credible sets, and the Supplemental Note and Table S6 show results from more extensive annotation from publicly available regulatory data. As an example, manual annotation of the *FOXA2* locus shows that the top SNP lies just upstream of an lncRNA, LINC00261, with evidence for expression in liver and pancreas and overlapping numerous TFBSs in liver cell lines (Supplemental Note). This and prior evidence that this lncRNA can regulate *FOXA2* expression[41] implicate the lncRNA as a possible causal transcript at this locus.

To more specifically examine whether variants at the 22 loci with greatly narrowed credible sets were enriched for individual TFBS or cell-type-specific chromatin marks, we employed a permutation test. We observed significant enrichment ($p < 1.9 \times 10^{-4}$) for FG loci at binding sites for MAFB ($p = 2 \times 10^{-6}$), NKX2-2 ($p = 2 \times 10^{-6}$), FOXA2 ($p = 1.6 \times 10^{-5}$), and PDX1 ($p = 1.0 \times 10^{-4}$) as well as for chromatin marks in pancreatic islets ($p = 1.2 \times 10^{-4}$) (Figure 2A, Table S7). Among FI loci, we observed nominally significant enrichment for chromatin marks in

Table 1. 22 EA-Associated Type 2 Diabetes Quantitative Traits Loci with Substantially Reduced 99% Credible Sets Based on Trans-ethnic Fine Mapping

		99% Credible Set: European Ancestry Only		99% Credible Set: Trans-ethnic		99% Credible Set Reduction	
Locus[a]	Chr	# SNPs[b]	Range[c] (bp)	# SNPs[b]	Range[c] (bp)	% SNPs[d]	% Range[e]
Fasting Glucose-Associated Loci							
FOXA2	20	40	46,365	2	3,872	95.0	91.6
GCK	7	7	25,107	1	1	85.7	100.0
KL	13	696	496,262	147	492,550	78.9	0.7
ADCY5	3	4	31,042	1	1	75.0	100.0
GCKR	2	2	11,663	1	1	50.0	100.0
PROX1	1	11	18,286	6	13,550	45.5	25.9
DPYSL5	2	87	294,065	50	269,667	42.5	–8.3
IGF2BP2	3	64	317,522	38	355,236	40.6	–11.9
CDKN2B-ANRIL	9	7	5,914	5	4,515	28.6	23.7
ADRA2A	10	33	68,716	26	68,110	21.2	0.9
TCF7L2	10	5	36,312	4	15,268	20.0	58.0
FADS1	11	20	58,394	16	57,823	20.0	1.0
DGKB-TMEM195	7	11	143,605	10	2,182	9.1	98.5
CRY2	11	10	57,088	11	14,850	–10.0	74.0
Fasting Insulin-Associated Loci							
ARL15	5	319	498,585	22	33,535	93.1	93.3
PPP1R3B	8	8	9,510	1	1	87.5	100.0
COBLL1-GRB14	2	14	51,528	3	11,540	78.6	77.6
IRS1	2	43	137,640	13	68,951	69.8	49.9
GCKR	2	3	11,663	1	1	66.7	100.0
FAM13A	4	43	243,374	21	243,374	51.2	0.0
ANKRD55-MAP3K1	5	417	497,027	218	487,103	47.7	2.0
UHRF1BP1	6	13	217,136	9	136,609	30.8	37.1
PPARG	3	14	60,448	11	56,618	21.4	6.3

The 22 EA-associated T2D QT loci include 13 fasting glucose, 8 fasting insulin, and 1 associated with both. Substantial reduction is defined as greater than 20% reduction in its genomic length or the number of SNPs.
[a]For ease of comparison to previous studies, the loci are named based on the historically identified nearest protein-coding gene or genes to the index SNP in European ancestry.
[b]# SNPs is the number of SNPs included in the 99% credible set.
[c]Range is defined as the maximum genomic distance based on hg18 among the SNPs included in the 99% credible set.
[d]%SNPs is (the number of SNPs in the EA-based 99% credible set − the number of SNPs in the trans-ethnic analysis-based 99% credible set)/the number of SNPs in the EA-based 99% credible set.
[e]%Range is (the range of the EA-based 99% credible set − the range of the trans-ethnic analysis-based 99% credible set)/the range of the EA-based 99% credible set.

adipose cells ($p = 0.048$) and for several TFBSs such as MAFK ($p = 0.0038$) and RXRA ($p = 0.0071$) (Figure 2A, Table S7).

We incorporated information from RegulomeDB and the Islet Regulome Browser[11,13] to better visualize the relationship between trait association and regulatory annotation at the 22 loci with >20% reduction in the 99% credible sets. Of the 14 FG loci, seven (*CRY2* [MIM: 603732], *DPYSL5*, *FADS1* [MIM: 606148], *FOXA2*, *GCKR*, *IGF2BP2* [MIM: 608289], and *KL* [MIM: 604824]) contained a SNP with a RegulomeDB score ≤ 3, consistent with moderate evidence for regulatory function (Figure S4). At four of these seven loci (*CRY2*, *FADS1*, *FOXA2*, and *GCKR*), the same SNP with strong regulatory annotation also had genome-wide significant evidence of association in the trans-ethnic meta-analysis ($\log(BF) > 6$) (Figures 2B, 2C, and S4). Similarly, of the nine FI loci with reduced credible set size, six of the credible regions (*ANKRD55* [MIM: 615189]-*MAP3K1* [MIM: 600982], *ARL15*, *FAM13A* [MIM: 613299], *GCKR*, *PPARG*, and *UHRF1BP*) contained a SNP with a RegulomeDB score ≤ 3, and at three of these loci (*GCKR*, *PPARG*, and *UHRF1BP1*), the SNP with strong

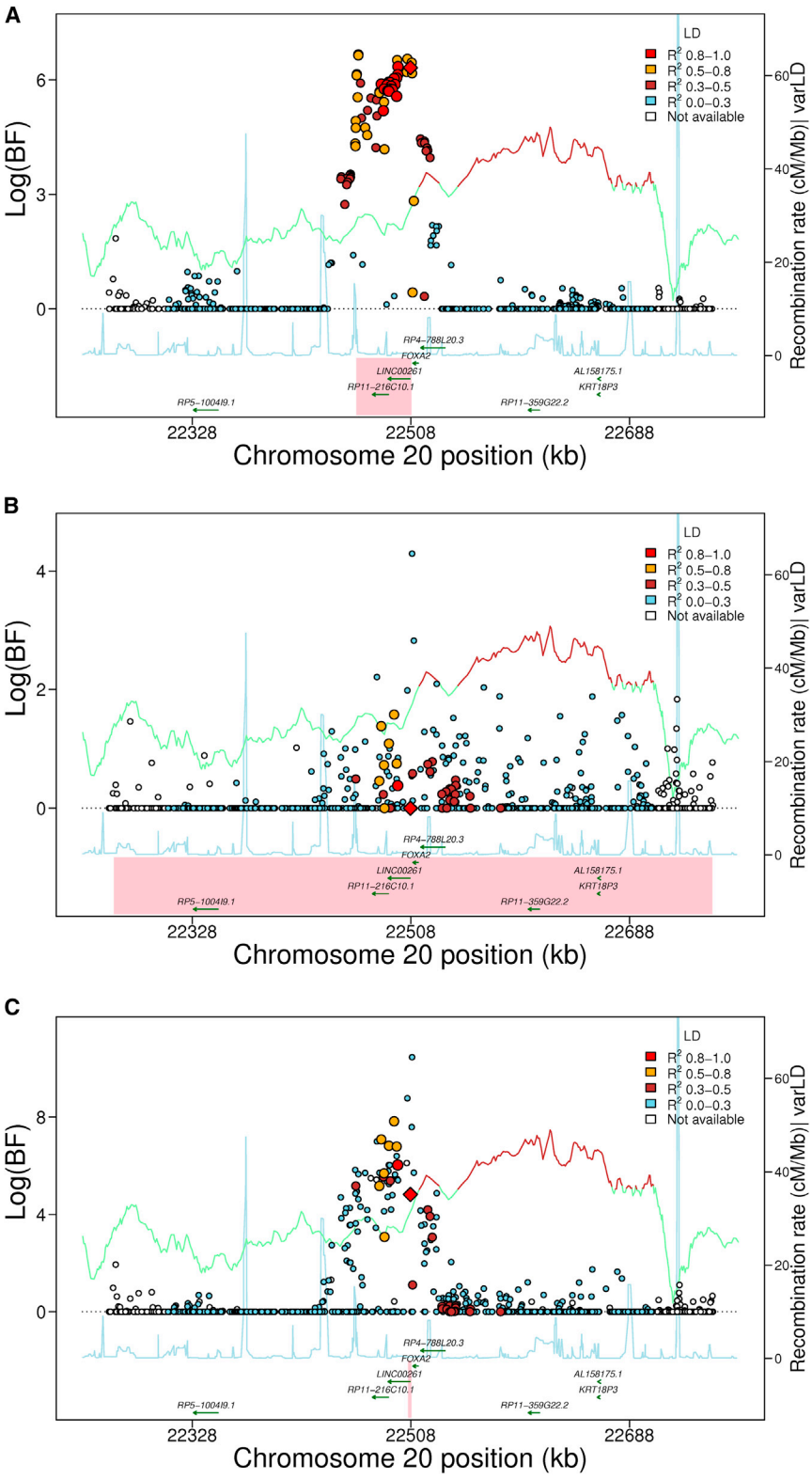


Figure 1. Trans-ethnic Analysis of Glycemic Quantitative Loci Provides Narrowed Intervals Spanned by the 99% Credible Set

Data are 500 kb regional association plots for fasting glucose at *FOXA2*, centered at the index SNP identified from European ancestry (EA) samples. The x axis denotes genomic position and the y axis denotes the log (BF), recombination rate, and varLD information[61] (a measure to quantify LD variation differences comparing populations). The red diamond data point represents the index SNP within the region previously reported from EA samples. The color of each data point indicates its LD value (r^2) with the index SNP based on HapMap2 (YRI for AA results and CEU for EA results): white, r^2 not available; blue, $r^2 = 0.0–0.2$; brown, $r^2 = 0.2–0.5$; orange, $r^2 = 0.5–0.8$; red, $r^2 = 0.8–1.0$. The blue line represents the recombination rate. The green line shows the varLD score at each SNP and is highlighted with dark brown if the varLD score is >95th percentile of the genome-wide varLD score, comparing LD information between YRI and CEU HapMap2 samples.[61] The interval spanned by the 99% credible set is highlighted in pink.

(A) Association results for fasting glucose in the *FOXA2* region in EA individuals. The 99% credible set contains 40 SNPs that span an interval of 46,365 bp.

(B) Association results for fasting glucose in the *FOXA2* region in AA individuals. The association signal is weaker than in EA samples, leading to a wider interval spanned by the 99% credible set.

(C) Association results for fasting glucose in the *FOXA2* region in both EA and AA individuals. The 99% credible set contains 2 SNPs and spans an interval of 3,872 bp, a 95% reduction in the number of SNPs and a 91.6% reduction in the length of the credible set interval.

regulatory annotation also had genome-wide significant evidence of association in the trans-ethnic analysis (Figure S4). Then, from overlay of Islet Regulome Browser data, we found that 8 of 14 substantially narrowed credible sets for FG and 3 of 9 for FI had either a TFBS or an active islet-specific enhancer within the narrowed credible region (Figures 2B, 2C, and S4). For example, the narrowed credible region at the *FOXA2* locus, which also overlaps an lncRNA as noted above, falls within an active C3 enhancer cluster and contains binding sites for both NKX2-2 and FOXA2, raising the possibility that the causal genetic mechanism at this locus involves regulation of *FOXA2*, the lncRNA at the locus, or both (Figure 2B). The narrowed credible region at the *CRY2* locus also overlaps a C3 enhancer cluster in islets that contains an NKX2-2 TFBS (Figure 2C). Furthermore, at three FG loci (*GCK*, *ADCY5* [MIM: 600293], and *GCKR*) and both FI loci (*GCKR* and *PPP1R3B* [MIM: 610541]) whose 99% credible set was narrowed to a single variant, the remaining credible set SNP

Table 2. Genomic Annotation Characteristics at 22 EA-Associated Type 2 Diabetes Quantitative Traits Loci with Substantially Reduced 99% Credible Sets Based on Trans-ethnic Fine Mapping

		dbSNP					ENCODE[23]				Islet Regulome Browser[13,a]
Locus[b]	# SNPs[c]	Syn	Non-Syn	Intronic	eQTL	TF Motif	TFBS	Promoter	Enhancer	DHS	PIACT[d]
Fasting Glucose-Associated Loci											
FOXA2	2	0	0	1	0	2	2	2	2	2	AT
GCK	1	0	0	0	1	1	0	0	1	1	PT
KL	147	1	0	53	24	122	19	18	94	36	PIACT
ADCY5	1	0	0	1	0	0	0	1	1	1	T
GCKR	1	0	0	1	1	1	1	0	1	0	I
PROX1	5	0	0	0	6	5	1	1	4	1	PAT
DPYSL5	50	1	1	37	23	44	7	9	42	24	PIACT
IGF2BP2	38	0	0	30	32	34	8	8	35	15	PIACT
CDKN2B-ANRIL	5	0	0	0	3	5	0	0	4	2	AT
ADRA2	26	0	0	0	1	22	0	1	12	6	IT
TCF7L2	4	0	0	3	0	2	0	1	3	2	AT
FADS1	16	0	0	12	16	15	7	6	16	10	PAT
DGKB-TMEM195	10	0	0	0	2	8	0	3	10	0	A
CRY2	2	0	0	1	2	2	0	1	2	0	PACT
Fasting Insulin-Associated Loci											
ARL15	22	0	0	22	18	18	3	3	17	5	IT
PPP1R3B	1	0	0	1	1	1	0	0	1	0	–
COBLL1-GRB14	3	0	1	0	3	3	1	0	1	0	T
IRS1	13	0	0	0	13	11	1	0	7	2	C
GCKR	1	0	0	1	1	1	1	0	1	0	I
FAM13A	21	0	0	21	18	15	3	8	20	6	PIACT
ANKRD55-MAP3K1	200	0	0	0	63	161	20	17	147	67	IAT
UHRF1BP1	9	0	0	8	9	8	2	0	4	5	PAT
PPARG	11	0	1	10	11	8	2	4	11	5	–

Abbreviations are as follows: Syn, synonymous SNP; non-syn, non-synonymous SNP; eQTL, expression quantitative trait loci; TF motif, transcription factor motif; TFBS, transcription factor binding site; DHS, DNase I-hypersensitive sites.
[a]The information was obtained on December 1, 2014.
[b]For ease of comparison to previous studies, the loci (13 fasting glucose, 8 fasting insulin, and 1 more for both) are named based on the historically identified nearest protein-coding gene or genes to the index SNP in EA.
[c]#SNPs: the number of SNPs in trans-ethnic analysis-based 99% credible set.
[d]PIACT: P, I, A, C, and T represent promoter, inactive enhancer, active enhancer, CTCF insulator, and transcription factor binding site (TFBS), respectively.

overlapped with regulatory annotation in either ENCODE, Roadmap, or the Islet Regulome (Table 2). Based on data derived from the Islet Regulome Browser, all 14 FG and 7 of 9 FI loci had evidence of regulatory function in pancreatic islets within the credible set (Table 2). Nearly all loci with narrowed credible sets contained at least one transcription factor (TF) motif within the 99% credible set (13/14 FG and 9/9 FI) and had at least one variant associated with *cis*-eQTL data (11/14 FG and all 9 FI). In addition, all 14 FG and 9 FI loci contained some regulatory evidence; in contrast, only three of the loci (*DPYSL5*, *COBLL1-GRB14*, and *PPARG*) contained a nonsynonymous variant in their credible sets (Table 2).

Relevance of European T2D QT Loci for African Americans

To evaluate the relevance in AA individuals of genetic determinants of FG and FI identified from EA studies, we examined SNP and locus association transferability, allele frequency differences, and patterns of association between FG/FI SNPs and glycemia-related traits. We tested SNP transferability (defined as whether the index EA SNP was associated with the same trait in AA, with the same direction of effect, and $p < 0.05$) at 54 EA-associated FG or FI loci. Of 36 EA FG index SNPs, 11 SNPs reached SNP transferability criteria (binomial $p = 9.87 \times 10^{-8}$ for observing 11/36 meeting SNP transferability criteria) (Table 3). Of

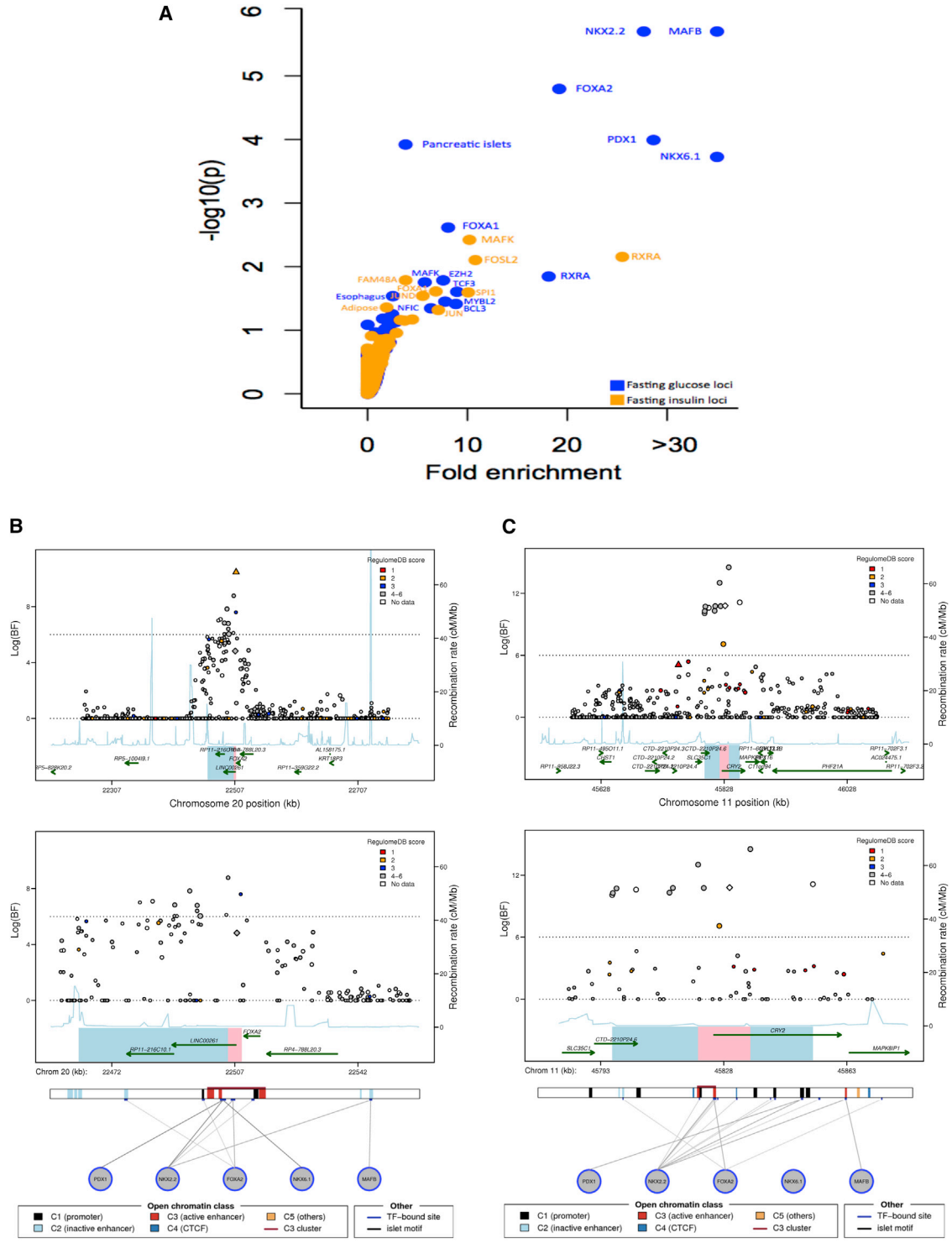


Figure 2. Trans-ethnic Fine-Mapping of Glycemic Quantitative Trait Loci Highlights Overlap between Trait-Associated SNPs and Predicted Regulatory Function

(A) Analysis for enrichment of posterior probabilities in SNPs overlapping transcription factor binding sites (TFBSs) and cell-type-specific enhancer and promoter marks at 22 (13 FG, 8 FI, and 1 both FG and FI) substantially narrowed 99% credible sets. x axis shows fold-enrichment above null, the y axis shows $-\log_{10}(P)$ for enrichment, and FI and FG are indicated by yellow and blue points, respectively. TFBSs and cell types with enhancer or promoter marks with p value for enrichment below 0.01 are labeled.

(B) Regional association plots for fasting glucose after trans-ethnic analysis demonstrating overlap between regulatory annotation and narrowed credible regions at the *FOXA2* locus.

(C) Regional association plots for fasting glucose after trans-ethnic analysis demonstrating overlap between regulatory annotation and narrowed credible regions at the *CRY2* locus.

(legend continued on next page)

18 EA index FI SNPs, 2 met SNP transferability criteria (Table 3). We also found excess concordance in direction of effect of the SNPs comparing EA with AA samples, irrespective of the strength of association: of 36 EA FG index SNPs, 28 SNPs shared the same direction of effect in the AA sample (binomial $p = 5.97 \times 10^{-4}$); of 18 EA FI index SNPs, 14 SNPs shared the same direction of effect in the AA sample (binomial $p = 0.015$) (Figure S5). For both traits, SNPs that met transferability criteria tended to have larger effect sizes than those not meeting the criteria, and the magnitudes of effect in EA were similar to those in AA samples (Figure S5). There was genome-wide excess of directional concordance of SNP effects comparing EA with AA (binomial $p = 0.021$ for FG and binomial $p = 0.016$ for FI) when considering SNPs independent of previously reported T2D QT associations and with $p \leq 0.01$ in EA (Table S8). This evidence supports the hypothesis that trans-ethnic meta-analysis in large samples could reveal additional SNPs associated with glycemic traits.

We also evaluated locus transferability (defined as whether any SNP within ±250 kb of the index EA SNP was associated with the same trait in AA at a Bonferroni-corrected $p < 0.05$) at the 54 EA-associated FG or FI loci. Loci were transferable from the EA to AA individuals (adjusting for the effective number of SNPs tested in each region) for 15/38 FG loci and for 10/18 FI loci (Table 4, Figure S2). At six FG loci (*GCK*, *ADCY5*, *GCKR*, *CRY2*, *PPP1R3B*, and *MTNR1B* [MIM: 600804]) and two FI loci (*PDGFC* [MIM: 608452] and *GCKR*), the index SNPs from the EA sample and the most significantly associated SNPs in the AA sample were either in LD ($r^2 \geq 0.20$) in YRI or were the same SNP (Table 4). In contrast, for the remaining loci, the index SNP from the EA sample and the most associated SNP in the AA sample were not in LD ($r^2 < 0.20$ in AA sample). For these loci with low LD between EA index SNP and the most associated AA SNP, we found a change in effect size > 10% for the AA SNP after conditioning on the index SNP in EA at only three FG loci (*SLC30A8* [MIM: 611145], *PPP1R3B*, and *GCK*) (Table S9). These results show ancestrally derived allelic heterogeneity giving more than one variant signal at these FG loci.

Allele Frequency Differences and Selection

FG-raising (38 SNPs) and FI-raising (18 SNPs) allele frequencies for EA index SNPs differed widely comparing AA with EA populations (absolute allele frequency differences ranged from 0.007 to 0.825 for FG index SNPs and 0.017 to 0.540 for FI index SNPs) (Table S10, Figure S5). We estimated Wright's fixation index (F_{ST})[26] to demonstrate whether selection pressure has resulted in widely different allele frequencies at any SNPs in AA versus EA populations. F_{ST} estimates were consistent with moderate to substantial population differentiation for a minority of FG and FI SNPs: $F_{ST} > 0.15$ at four FG loci (*ADRA2A* [MIM: 104210], *PCSK1* [MIM: 162150], *OR4S1*, and *ARAP1* [MIM: 606646]) and at one FI locus (*UHRF1BP1*) in the AAGILE and MAGIC data. There was also evidence of recent positive selection (absolute value of iHS > 2) for one FG locus (*FOXA2*) and three FI loci (*UHRF1BP1*, *HIP1* [MIM: 601767], and *MAP3K19*) in the EA, yet no evidence of recent selection in AA (Table S10).

Associations between FG/FI SNPs and Insulin Resistance-Related Metabolic Traits

Many FG/FI-associated loci identified in EA samples were also associated with T2D and other insulin resistance-related traits.[8,9] We investigated these associations in AA individuals at 25 SNPs, including 24 AA SNPs (14 FG, 9 FI, and 1 associated with both FG and FI) at loci that showed transferability, plus the previously undescribed FI SNP (rs6450057) described in the next section. A second FI SNP (also described in the next section), rs213676, was not interrogated for association with metabolic traits due to scant chromosome X data. Table S11 summarizes the association results of these 25 FG/FI SNPs with T2D, body mass index (BMI), BMI-adjusted waist-to-hip ratio (WHR), blood pressure, hypertension (HTN), and lipid levels in AA from several consortia.[32–36,42] 14 of 25 (56%) FG/FI SNPs were associated with T2D or an insulin resistance-related trait ($p < 0.05$ with an effect in the expected direction [e.g., FG-raising SNP associated with increased risk of T2D]). Six SNPs (24%; 4 [*ADCY5*, *RREB1* (MIM: 602209), *MTNR1B* (MIM: 600804), and *FOXA2*] of 15 FG SNPs and 2 [*COBLL1-GRB14* and *ARL15*] of 11 FI SNPs) were associated with higher odds of T2D in AA samples (n up to 23,818).[31] 40% (6/15) of the FG-raising alleles and 45% (5/11) of the FI-raising alleles were associated with insulin resistance-related traits in AA samples. The FI SNP rs6717858 (*COBLL1-GRB14*) was associated with three traits (T2D, BMI-adjusted WHR, and HDL-C). rs17811863 (*PDGFC*), also an FI SNP, was associated with both systolic and diastolic blood pressure.

Previously Undescribed Glycemic Quantitative Trait Loci

The strategy we used for glycemic loci discovery in AAGILE is shown in Figure S1. Results of the discovery analysis are shown in Figure S6. AAGILE GWAS results were combined with MAGIC EA GWAS results[10] in a two-stage meta-analysis approach. In the discovery stage, results from 16 studies (n = 20,209) in the AAGILE AA GWAS fixed effects meta-analysis were combined with results from 29 studies

(B and C) The index SNP in European ancestry (MAGIC) is represented by a diamond; the best SNP in African ancestry (AAGILE) is represented by a triangle. SNPs are colored according to the score assigned in RegulomeDB[11] with lower score corresponding to stronger level of evidence supporting regulatory function; data from the Islet Regulome Browser[13] for the genomic interval is shown below regional association plots. The interval spanned by the 99% credible set using EA data only is represented by combining blue and pink regions; interval spanned by the narrowed 99% credible set after trans-ethnic analysis is shown in pink.

Table 3. SNP Transferability in AA Individuals at 13 EA-Associated Type 2 Diabetes Quantitative Trait SNPs

				EA Association[a]			AA Association[a]			
Locus[b]	Index SNP[c]	Chr	Alleles[d]	EAF[e]	Effect[f]	SE	EAF[e]	Effect[f]	SE	p
Fasting Glucose-Associated Loci										
MTNR1B	rs10830963	11	G/C	0.30	0.067	0.003	0.08	0.089	0.012	9.29×10^{-15}
G6PC2	rs560887	2	C/T	0.70	0.075	0.003	0.93	0.059	0.013	2.67×10^{-06}
ADCY5	rs11708067	3	A/G	0.78	0.027	0.003	0.85	0.036	0.008	6.27×10^{-06}
GCKR	rs780094	2	C/T	0.62	0.029	0.003	0.82	0.032	0.008	2.03×10^{-05}
GCK	rs4607517	7	A/G	0.16	0.062	0.004	0.11	0.041	0.010	6.84×10^{-05}
GLIS3	rs7034200	9	A/C	0.49	0.018	0.003	0.63	0.019	0.006	1.82×10^{-03}
KL	rs576674	13	G/A	0.15	0.017	0.003	0.60	0.018	0.006	2.06×10^{-03}
SLC30A8	rs13266634	8	C/T	0.68	0.027	0.004	0.90	0.024	0.010	1.82×10^{-02}
MADD	rs7944584	11	A/T	0.75	0.021	0.003	0.95	0.029	0.014	3.37×10^{-02}
DGKB-TMEM195	rs2191349	7	T/G	0.52	0.030	0.003	0.60	0.013	0.006	3.70×10^{-02}
GRB10	rs6943153	7	T/C	0.34	0.015	0.002	0.68	0.012	0.006	4.81×10^{-02}
Fasting Insulin-Associated Loci[g]										
COBLL1-GRB14	rs7607980	2	T/C	0.87	0.027	0.004	0.84	0.042	0.008	2.85×10^{-07}
GCKR	rs780094	2	C/T	0.62	0.032	0.004	0.82	0.025	0.008	1.92×10^{-03}

SNP transferability is defined as the association $p < 0.05$ in AA and sharing the same trait-raising allele between EA and AA. 13 EA-identified T2D QT SNPs in 12 loci, including 10 fasting glucose loci, 1 fasting insulin locus, and 1 locus associated with both.
[a]EA Association results refer to the association results using samples of European ancestry in previous publications;[8–10] AA Association results refer to the association results using AAGILE samples of African ancestry assembled in this study.
[b]For ease of comparison to previous studies, the loci are named based on the historically identified nearest protein-coding gene or genes to the index SNP in European ancestry.
[c]Index SNPs are the most significant SNPs previously reported in MAGIC publications. All of these SNPs reach genome-wide significant level ($p < 5 \times 10^{-8}$) in the original study.
[d]EA trait-raising allele/other allele.
[e]Frequency of EA trait-raising allele.
[f]Effect of EA trait-raising allele ([mmol/L] for fasting glucose and [pmol/L] for fasting insulin per trait-raising allele).
[g]The association with BMI-adjusted fasting insulin.

(n = 57,292) from the MAGIC EA GWAS[10] for trans-ethnic meta-analysis using MANTRA. A total of 62 SNPs met prespecified multi-tiered criteria for stage 2 follow-up, with 12 SNPs in tier 1, 10 in tier 2, and 40 in tier 3 (Figure S1 and Table S12). Follow-up in the second stage, with up to 10,096 additional AA samples from 14 studies (Table S1), yielded two previously undescribed SNPs in loci associated with FI that exceeded GWAS significance thresholds (Table 5). We found no previously unknown FG loci.

In the fixed effects meta-analysis of AA samples, we identified a previously undescribed SNP (rs213676) on chromosome X near *FAM133A* associated with FI ($p = 2.4 \times 10^{-8}$) (Figure 3A). This FI SNP was not included in the trans-ethnic meta-analysis because MAGIC[10] did not report chromosome X results. Although rs213676 is in a region without known genes, this region might be of regulatory significance because it is known to harbor a TFBS in pancreatic islets[11,13] (Figure 3A).

The other previously undescribed FI SNP, rs6450057 on chromosome 5 (Figure 3B), resides near four putative lncRNA genes and the *PELO* (MIM: 605757) (or *ITGA1* [MIM: 192968]) gene. In trans-ethnic analyses, this locus achieved genome-wide significance (log(BF) = 7.1) and trans-ethnic fine-mapping reduced the credible set at this locus from 229 SNPs to just 3 SNPs. Interestingly, the rs6450057 T allele is associated with higher FI ($p = 3.1 \times 10^{-6}$) in AA samples but with lower FI ($p = 9.2 \times 10^{-5}$) in EA samples in ancestry-specific fixed-effect meta-analyses (Table 5). The discordant direction of effect at this locus was observed across nearly all SNPs at the locus, regardless of LD with rs6450057 (Figure 3C). However, the direction of effects did not show a clear pattern in the association analysis after conditioning on rs6450057, implying that the signal at this locus was driven by rs6450057 (Figure S7). As with many of the glycemic QT loci described above, the 99% credible set at this locus did not include coding variants but did overlap an active C3 enhancer in pancreatic islets[11,13] (Figure 3D).

Discussion

We assembled a large sample of AA individuals, combined resulting data with published data from EA individuals, and used trans-ethnic fine mapping to narrow the genomic interval containing putative causal SNPs for 22 of 54

previously identified FI and FG loci. We demonstrated that many of the genetic variants associated with FG and FI are predicted to have regulatory function, with few having predicted protein-coding function. The results show that although a substantial portion of the genetic architecture underlying these T2D-associated traits is shared across EA and AA populations, allelic heterogeneity suggests that there are also genetic variants unique to AA populations. Finally, we identified two previously undescribed FI loci, bringing to 56 the number of FG- and FI-associated loci in humans.

Fine mapping combined with regulatory annotation provides a plausible functional explanation for the many T2D-associated GWAS loci that reside in non-protein-coding regions of the genome.[43–45] Previous GWAS findings from MAGIC show complete overlap of loci associated with HOMA-B (a measure of beta cell function[46]) and FG,[9] so it would be expected that fine-mapping of FG loci might identify regulatory function in islets. On the other hand, FI is typically considered a marker of insulin resistance.[47] However, insulin resistance does not account for all of the variability in FI,[48,49] and fasting hyperinsulinemia itself, due to hypersecretion of insulin by beta cells, might be causal in the pathogenesis of T2D.[50–52] Our finding that some FI loci had predicted regulatory function in islets is supportive of this evidence. At many loci, for instance at *FOXA2* and *PPP1R3B*, the narrowed credible sets from trans-ethnic analysis coupled with genomic annotation focused attention on lncRNA transcripts rather than the nearest protein-coding gene, by convention generally assumed to be the putative causal transcript. At the *FOXA2* locus, the trans-ethnic credible set combined with genomic annotation highlights regulatory functionality in glycemia-related tissues—enhancer marks and TFBS in pancreas and liver—as well as an lncRNA that might affect *FOXA2* expression, raising two possible causal regulatory mechanisms for altered FG.[12,13,41] Awareness of the regulatory nature of some genetic determinants of FG and FI provides insight into novel approaches for the regulation of glucose homeostasis. In particular, regulatory targets might be amenable to post-genomic manipulation (e.g., by genome editing, use of antisense oligonucleotides, or enzyme hijacking) as suggested in other areas.[53,54] For instance, by knowing that polymorphisms in aldehyde dehydrogenase 2 (ALDH2) enzyme are associated with poor alcohol metabolism in some Asian populations, the enzyme hijacking technique has been used to upregulate a related, but naturally unimportant, enzyme (ALDH3A1), thereby improving alcohol metabolism and reducing cancer risk in mice.[53] Such techniques could in the future be extended to T2D prevention and control if accessible regulatory pathways are elucidated.

At the *GCKR* locus, trans-ethnic fine mapping provided added information to the prior knowledge of this locus identified from studies in EA populations. The 99% credible set constructed using trans-ethnic analysis results at *GCKR* contained only one non-coding SNP, rs780094 (GenBank: NC_000002.12; g.27518370T>C), the most strongly FG-associated SNP in both EA and AA. However, prior fine-mapping[39] in EA and functional studies[40] have implicated rs1260326, a nonsynonymous variant (GenBank: NC_000002.12; g.27508073T>C [p.Leu446Pro]), as a likely causal SNP at this locus. This missense variant was excluded from the narrowed credible set. This could imply that the lead non-coding SNP rs780094, which has strong evidence as residing in a TFBS, is also a causal variant at the locus (Table S5). On the basis of statistical evidence, we were unable to distinguish the association of these two SNPs in EA samples due to high LD ($r^2 = 0.93$). However, in AA the evidence of association with FG was several orders of magnitude stronger for the non-coding SNP, rs780094 ($p = 2.2 \times 10^{-5}$), than the coding SNP, rs1260326 ($p = 0.03$) and their LD is weaker ($r^2 = 0.47$). Both SNPs may play a role at the *GCKR* locus; a causal variant tagged by rs780094 might be common to both ancestries, resulting in the narrowed trans-ethnic credible set observed here, while the nonsynonymous variant rs1260326 might have greater functional impact in EA than in AA individuals. Alternatively, the actual causal SNP could be in LD with both of these SNPs, and more dense imputation or deep sequencing might reveal additional SNPs carried on haplotypes with these SNPs. Since crystal structural analysis of the GCKR protein has not identified the 446 residue as critical for binding of regulating molecules (fructose 1-phosphate and fructose 6-phosphate), genetic heterogeneity involving both coding and regulatory functional variation at the locus remains a plausible hypothesis.[55]

Analyses of the relevance of glycemic QT loci in AA suggest that genetic determinants of human glucose regulation are more similar than different across human populations. We observed an excess of consistency in direction of effect of FG and FI SNPs comparing AA with EA, regardless of statistical significance of SNPs in AA, and a substantial portion (50% for FG and 56% for FI) of EA index SNP or loci were transferable to AA individuals. As in previous studies in EA individuals,[8,9] most of the transferable T2D EA loci were also associated with T2D or insulin resistance-related traits in AA individuals, demonstrating common genetic pathways underlying glycemic QTs and other metabolic traits. We also found that several FG-raising and FI-raising alleles were at least nominally associated with lower odds of T2D or "better" metabolic trait profiles in the AA samples. Many of the loci previously observed to have this discordant pattern of associations across traits in EA, including *GCKR*,[9,39,40,56,57] *MADD* (MIM: 603584),[9] *PDGFC*,[10] and *FOXA2*,[10] had a similar pattern in our AA sample, demonstrating that the complexity of the genetic architecture of these traits is shared across populations.

By combining AA with EA information, including chromosome X variants in AA, we identified two previously undescribed FI SNPs near *FAM133A* and *PELO*, increasing the total number of human FG/FI-associated loci from 54 to

Table 4. Locus Transferability in African Individuals for 24 EA-Identified Type 2 Diabetes Quantitative Traits Associations

	SNP Information				Best SNP Association						LD[a] in YRI and CEU			
Locus[b]	Index SNP in EA[c]	Best SNP in AA[d]	Best SNP Alleles[e]	EAF[f]	$Effect_{uc}$[g]	SE	p	adj-p[h]	$Effect_c$[i]	% Change in Effect[j]	R^2_{YRI}	D'_{YRI}	R^2_{CEU}	D'_{CEU}
Fasting Glucose-Associated Loci														
MTNR1B	rs10830963	rs10830963	G/C	0.079	0.089	0.012	9.3×10^{-15}	3.5×10^{-3}	NA	NA	1.00	1.00	1.00	1.00
GCK	rs4607517	rs1799884	T/C	0.174	0.047	0.007	2.0×10^{-10}	8.6×10^{-3}	0.038	19.3	0.47	1.00	1.00	1.00
G6PC2	rs560887	rs830193	C/T	0.816	0.037	0.008	1.2×10^{-6}	2.9×10^{-3}	0.035	5.2	0.00	0.08	0.07	0.58
FOXA2	rs6048205	rs1203907	T/C	0.505	0.027	0.006	3.7×10^{-6}	3.3×10^{-3}	0.025	7.7	0.18	0.92	0.66	1.00
RREB1	rs17762454	rs557074	G/T	0.455	0.027	0.006	4.2×10^{-6}	4.1×10^{-3}	0.025	4.4	0.01	0.48	0.02	0.51
SLC30A8	rs13266634	rs10505311	G/T	0.837	0.036	0.008	5.6×10^{-6}	3.7×10^{-3}	0.025	30.8	0.19	1.00	NA	NA
ADCY5	rs11708067	rs11708067	A/G	0.846	0.036	0.008	6.3×10^{-6}	4.5×10^{-3}	NA	NA	1.00	1.00	1.00	1.00
GCKR	rs780094	rs780094	C/T	0.819	0.032	0.008	2.0×10^{-5}	1.4×10^{-2}	NA	NA	1.00	1.00	1.00	1.00
CRY2	rs11605924	rs11038651	T/C	0.828	0.033	0.008	3.8×10^{-5}	5.3×10^{-3}	0.030	9.4	0.55	0.85	0.12	0.44
MADD	rs7944584	rs1052373	T/C	0.481	0.022	0.006	1.4×10^{-4}	1.1×10^{-2}	0.022	0.0	NA	NA	0.15	1.00
PPP1R3B	rs4841132	rs7004769	A/G	0.228	0.024	0.007	2.7×10^{-4}	2.2×10^{-3}	0.018	26.1	0.44	1.00	0.38	1.00
PROX1	rs340874	rs2282387	C/G	0.519	0.021	0.006	4.6×10^{-4}	3.4×10^{-3}	0.021	–1.2	0.01	0.27	0.00	0.09
IKBKAP	rs16913693	rs7038936	C/T	0.669	0.021	0.006	5.3×10^{-4}	3.9×10^{-3}	NA[k]	NA	0.04	0.29	0.01	0.39
ADRA2A	rs10885122	rs12569523	A/T	0.368	0.021	0.006	5.4×10^{-4}	3.6×10^{-3}	0.020	1.9	0.01	0.11	0.00	0.25
PCSK1	rs13179048	rs7722200	T/C	0.773	0.023	0.007	8.2×10^{-4}	7.0×10^{-3}	0.023	–0.9	0.04	1.00	0.66	0.83
Fasting Insulin-Associated Loci														
COBLL1-GRB14	rs7607980	rs6717858	T/C	0.281	0.036	0.007	8.6×10^{-8}	5.3×10^{-3}	0.034	6.4	0.06	1.00	0.22	1.00
ARL15	rs4865796	rs6876198	C/T	0.295	0.031	0.007	2.0×10^{-6}	3.8×10^{-3}	0.030	4.0	0.11	0.89	0.09	0.87
PPP1R3B	rs4841132	rs9949	G/A	0.231	0.029	0.007	6.9×10^{-5}	2.2×10^{-3}	0.027	6.7	0.11	0.51	0.01	0.23
IRS1	rs2943634	rs4413154	G/A	0.041	0.062	0.016	1.2×10^{-4}	4.9×10^{-3}	0.062	1.3	0.01	1.00	0.08	1.00
ANKRD55-MAP3K1	rs459193	rs7700714	A/G	0.436	0.023	0.006	1.8×10^{-4}	2.3×10^{-3}	0.024	–3.6	0.00	0.03	0.04	0.29
FAM13A	rs3822072	rs17799176	C/G	0.934	0.043	0.012	4.3×10^{-4}	5.8×10^{-3}	0.043	0.2	0.00	0.22	0.02	0.43
HIP1	rs1167800	rs11465341	C/T	0.035	0.060	0.017	4.5×10^{-4}	8.1×10^{-3}	0.060	0.0	0.00	1.00	NA	NA
MAP3K19	rs1530559	rs13405563	T/C	0.924	0.043	0.013	5.5×10^{-4}	8.8×10^{-3}	0.045	–3.6	0.00	0.79	0.04	1.00
PDGFC	rs4691380	rs17811863	G/A	0.333	0.022	0.006	6.2×10^{-4}	4.7×10^{-3}	0.020	8.9	0.21	0.49	0.77	0.93
GCKR	rs780094	rs780094	C/T	0.817	0.025	0.008	1.9×10^{-3}	1.4×10^{-2}	NA	NA	1.00	1.00	1.00	1.00

(legend on next page)

[a]LD information between the EA index SNP and the best SNP in AA.
[b]24 EA-identified T2D QT (14 glucose, 9 fasting insulin, and 1 associated with both) loci with locus-wide significant, $p < 0.05$/effective number of independent SNPs within each locus, associations. For ease of comparison to previous studies, the loci are named based on the historically identified nearest protein-coding gene or genes to the index SNP in European ancestry.
[c]Index SNPs are the most significant SNPs previously reported in MAGIC publications. All of these SNPs reach genome-wide significant level ($p < 5 \times 10^{-8}$) in the original study.
[d]The best SNP in AA is the most significant SNP within ± 250 kb region of the EA index SNP.
[e]AA trait-raising allele/other allele.
[f]Frequency of AA trait-raising allele (effect allele) for the best SNP.
[g]Effect of AA trait-raising allele ([mmol/L] for fasting glucose and [pmol/L] for fasting insulin per trait-raising allele). This is unconditional association of best SNP in AA samples.
[h]Indicates the p value is adjusted for locus-wise multiple comparison and calculated as min(1, P *the number of effective SNPs). The information for the number of effective SNPs is available in Table S8.
[i]$Effect_{uc}$ is the association of the best SNP unconditional on the EA index SNP in AA samples and $Effect_c$ is the association of AA best SNP conditional on EA index SNP in AA samples.
[j]Percent change in association effect for the best SNP in the AA samples. It is calculated as the (unconditional association effect − conditional association effect)/(unconditional association effect). More detailed information for conditional and unconditional association results are available in Table S8.
[k]The association results for EA index SNP is not available in AA.

56. The lead SNP at the *PELO* locus, rs6450057, had discordant effects in the AA and EA samples: the FI-raising allele in AA lowered FI in EA. Discordant effects of common variants on complex human traits have been observed in other traits such as breast cancer[58] and serum protein levels[59] but not in T2D QT. This may be due to the lack of trans-ethnic study of T2D QT, the use of fixed-effects approaches that do not adequately account for heterogeneity in LD or allelic frequencies across populations, gene-environment interactions with differential exposures across ancestral populations, or gene-gene interactions with substantially different allele frequencies for the interacting variants across ancestries. Alternatively, EA and AA groups may have different causal variants that are in moderate LD with the lead SNP, thus boosting the association signal in trans-ethnic meta-analyses that take account of the heterogeneity in allelic effects between ancestry groups. Haplotype analysis in both EA and AA samples did not clearly elucidate differential SNP contributions to FI levels at the *PELO* locus. Although our analyses are unable to distinguish between possible mechanisms driving the discordant effects of rs6450057 in EA and AA populations, evaluation of the locus, with high-density imputation and/or whole-genome sequencing data from individuals of EA and AA as well as laboratory-based examination of the clustered enhancer in pancreatic islets at this locus, could help explain the observed divergent associations.

A major strength of this investigation was the large sample size of AA, a hitherto under-studied ancestral group with a heavy burden of hyperglycemia, insulin resistance, and T2D. The large sample of AA individuals allowed genome-wide trans-ethnic discovery, fine-mapping, and (combined with new annotation resources) detailed prediction of regulatory function. Yet, despite the large sample size, our study still had modest power to detect loci with small effects. With sample size of 30,305 (discovery + follow up) in AA, we had only 15% power to detect an EA-identified median FG effect size of 0.0196 mmol/L with SD of 0.5 mmol/L, for variants with MAF of 0.3. In addition, the reduced size of LD blocks in AA populations decreased the probability that a tagging SNP resided on the same haplotype as a causal variant, further limiting power for discovery. Also, reduced LD haplotype and corresponding greater genetic diversity made imputation for AA samples more challenging. The HapMap2 panel, used as the imputation panel for majority of contributing studies in this analysis, did not provide complete coverage of common and low-frequency genetic variation in AA samples; the latest reference panels from 1000 Genomes Phase 3 should provide better coverage although whole-genome sequencing and laboratory analysis of genetic function are going to be necessary to unequivocally determine causal variants that may have similar or different effects across ancestral groups. As for other trans-ethnic efforts,[16,60] we have reported the improvement in fine-mapping resolution in terms of the reduction in the number of SNPs or the length of the genomic interval to which they

Table 5. Two Previously Undescribed Loci Associated with Fasting Insulin in African Ancestry Individuals

SNP[a]	Chr	Position (build 36)	Nearest Gene	Alleles[b]	African Ancestry Association[c]				European Ancestry Association				Trans-ethnic Association	
					N_{AA}[d]	EAF[e]	Effect[f]	p	N_{EA}[d]	EAF[e]	Effect[f]	p	N_{Total}[d]	log(BF)
rs213676	X	93358147	*FAM133A*	C/G	14,043	0.98	0.147	2.37×10^{-8}	0[g]	NA	NA	NA	14,043	6.26
rs6450057	5	51683121	*PELO*	T/C	20,853	0.4	0.027	3.11×10^{-6}	52,372	0.37	−0.011	9.21×10^{-5}	73,225	7.1

[a]Two SNPs in previously undescribed loci are associated with fasting insulin at genome-wide significance, i.e., $p < 5 \times 10^{-8}$ in AA individuals or with logBF > 6 in trans-ethnic meta-analysis.
[b]Trait-raising allele/other allele in the sample of African ancestry.
[c]Association results from the samples of African ancestry based on the joint analysis of samples from stage 1 and stage 2.
[d]N_{AA}: the sample size of African ancestry; N_{EA}: the sample size of European ancestry; N_{total}: total sample size of EA and AA.
[e]Frequency of effect allele (i.e., AA trait-raising allele).
[f]Effect of AA trait-raising allele ([pmol/L] per trait-raising allele).
[g]Association analysis result is unavailable in MAGIC EA samples because the original MAGIC[10] effort analyzed only autosomal variants.

map. The improvement from trans-ethnic analysis varied from one locus to another as shown in Table S2. We used 20% reduction in credible set size to provide an overall assessment of fine-mapping resolution (albeit crude). However, fine-mapping resolutions depends on many factors, such as the size of LD blocks and the availability of high-quality read-depth for the reference datasets used in the imputation, and will therefore vary from one locus to another. Furthermore, the improvement in resolution offered by trans-ethnic meta-analysis relies on the extent of LD differences with the causal variant between ancestry groups and the increase in sample size. However, importantly, credible set sizes can increase after trans-ethnic meta-analysis, which most often occurs due to different underlying causal variants across ancestry groups. In this scenario, the credible set captures multiple association signals driven by each causal variant, which will therefore be larger than that observed for ethnic-specific analyses.

In conclusion, by using AA and EA trans-ethnic analysis, we narrowed the genomic interval containing likely causal variants for a large number of biologically plausible FG and FI loci and demonstrated that many FG and FI loci probably have regulatory, rather than protein-coding, function. The observed effects of genetic variants on glycemic traits might result from multiple regulatory functions residing in the same genomic region; for example, concurrent presence of an enhancer and lncRNA at the FOXA2 locus, an attribute that could lead to synergy in function. We also showed that there are probably both shared and unique genetic determinants of T2D QTs across European and African ancestral populations. We identified two previously undescribed FI loci, bringing the total number of identified FG and FI loci in humans to 56. Our finding of the predicted regulatory significance of many FG and FI loci is particularly noteworthy, given the prior uncertainty about the functional relevance of most GWAS findings for T2D and related QTs. Our study provides a framework for further follow-up of GWAS signals seen in EA and other ancestral populations. This approach using trans-ethnic meta-analysis for discovery and transferability combined with trans-ethnic fine-mapping and state-of-the-art annotation will lead the way to an understanding of the functional, and ultimately therapeutic, implications of genetic variation underlying glucose homeostasis, T2D risk, and other complex disorders.

Supplemental Data

Supplemental Data include consortium membership, sources of data, author contributions, acknowledgments, 7 figures, and 12 tables and can be found with this article online at http://dx.doi.org/10.1016/j.ajhg.2016.05.006.

Conflicts of Interest

B.M.P. serves on the DSMB of a clinical trial for the device manufacturer Zoll LifeCor and on the Steering Committee for the Yale

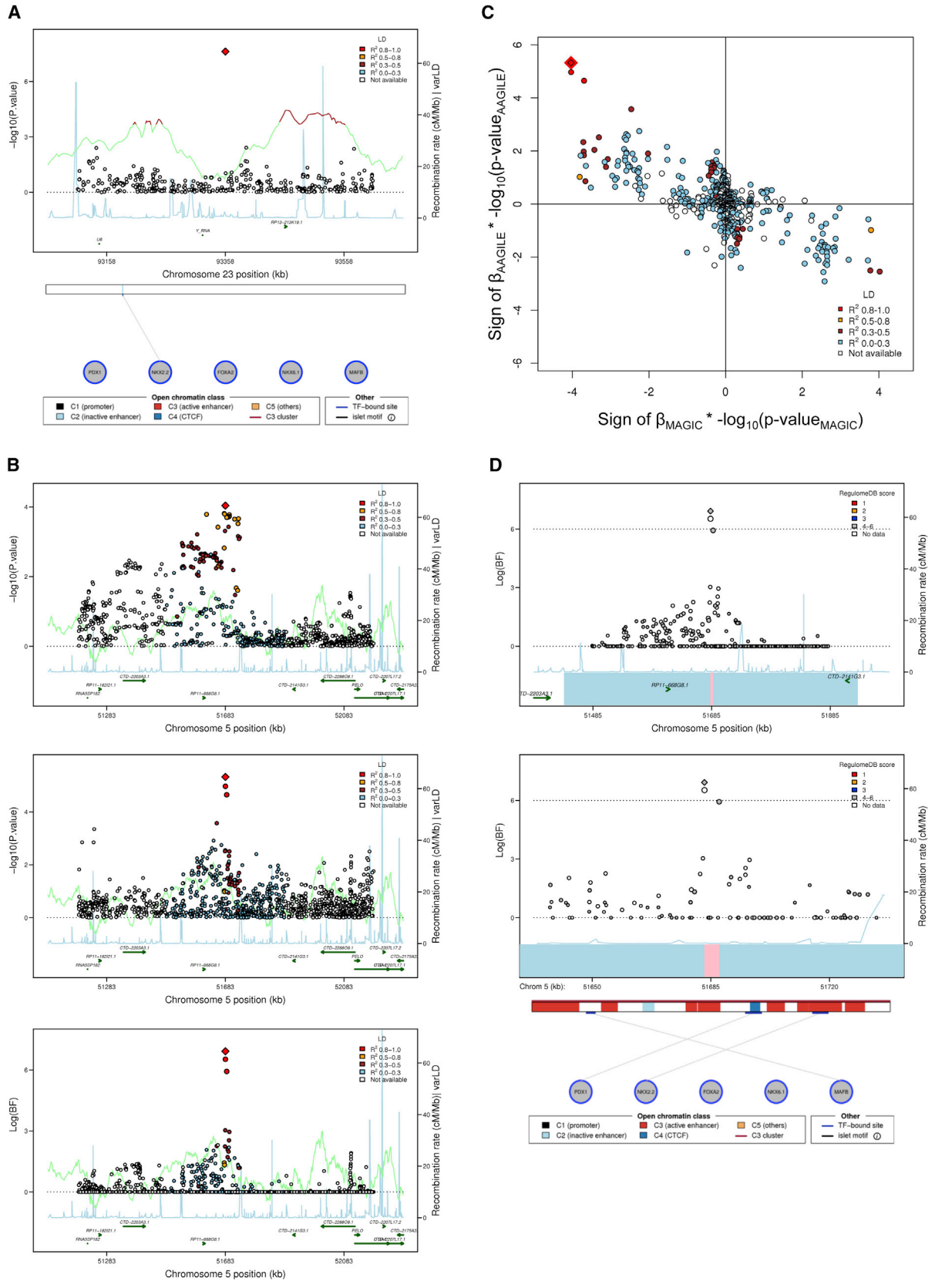


Figure 3. Regional Association Plots and Functional Annotation for Two Previously Undescribed Loci

(A) Regional association plot of a previously undescribed chromosome X locus associated with fasting insulin in AA individuals. The nearest gene is *FAM133A*. Top panel shows association signal on the y axis (−log(*P*)) and genomic position on chromosome X on the x axis. The red diamond data point represents the lead SNP (rs213676) within the region. The blue line represents the recombination

(legend continued on next page)

Open Data Access Project funded by Johnson & Johnson. D.V. is a consultant for Consumable Science, Inc. J.R.K. holds stock in Pfizer, Inc., and Gilead Sciences, Inc.

Received: February 8, 2016
Accepted: May 2, 2016
Published: June 16, 2016

Web Resources

AAGILE consortium, http://aagileandmedia.partners.org/
HaploReg, http://www.broadinstitute.org/mammals/haploreg/haploreg.php
Human Islet Regulome Browser, http://isletregulome.org/
OMIM, http://www.omim.org/
RefSeq, http://www.ncbi.nlm.nih.gov/RefSeq

References

1. Danaei, G., Finucane, M.M., Lu, Y., Singh, G.M., Cowan, M.J., Paciorek, C.J., Lin, J.K., Farzadfar, F., Khang, Y.H., Stevens, G.A., et al.; Global Burden of Metabolic Risk Factors of Chronic Diseases Collaborating Group (Blood Glucose) (2011). National, regional, and global trends in fasting plasma glucose and diabetes prevalence since 1980: systematic analysis of health examination surveys and epidemiological studies with 370 country-years and 2·7 million participants. Lancet *378*, 31–40.
2. Spanakis, E.K., and Golden, S.H. (2013). Race/ethnic difference in diabetes and diabetic complications. Curr. Diab. Rep. *13*, 814–823.
3. Cheng, C.Y., Reich, D., Haiman, C.A., Tandon, A., Patterson, N., Selvin, E., Akylbekova, E.L., Brancati, F.L., Coresh, J., Boerwinkle, E., et al. (2012). African ancestry and its correlation to type 2 diabetes in African Americans: a genetic admixture analysis in three U.S. population cohorts. PLoS ONE *7*, e32840.
4. Rosenberg, N.A., Huang, L., Jewett, E.M., Szpiech, Z.A., Jankovic, I., and Boehnke, M. (2010). Genome-wide association studies in diverse populations. Nat. Rev. Genet. *11*, 356–366.
5. Fesinmeyer, M.D., Meigs, J.B., North, K.E., Schumacher, F.R., Bůžková, P., Franceschini, N., Haessler, J., Goodloe, R., Spencer, K.L., Voruganti, V.S., et al. (2013). Genetic variants associated with fasting glucose and insulin concentrations in an ethnically diverse population: results from the Population Architecture using Genomics and Epidemiology (PAGE) study. BMC Med. Genet. *14*, 98.
6. Rasmussen-Torvik, L.J., Guo, X., Bowden, D.W., Bertoni, A.G., Sale, M.M., Yao, J., Bluemke, D.A., Goodarzi, M.O., Chen, Y.I., Vaidya, D., et al. (2012). Fasting glucose GWAS candidate region analysis across ethnic groups in the Multi-ethnic Study of Atherosclerosis (MESA). Genet. Epidemiol. *36*, 384–391.
7. Liu, C.T., Ng, M.C., Rybin, D., Adeyemo, A., Bielinski, S.J., Boerwinkle, E., Borecki, I., Cade, B., Chen, Y.D., Djousse, L., et al. (2012). Transferability and fine-mapping of glucose and insulin quantitative trait loci across populations: CARe, the Candidate Gene Association Resource. Diabetologia *55*, 2970–2984.
8. Scott, R.A., Lagou, V., Welch, R.P., Wheeler, E., Montasser, M.E., Luan, J., Mägi, R., Strawbridge, R.J., Rehnberg, E., Gustafsson, S., et al.; DIAbetes Genetics Replication and Meta-analysis (DIAGRAM) Consortium (2012). Large-scale association analyses identify new loci influencing glycemic traits and provide insight into the underlying biological pathways. Nat. Genet. *44*, 991–1005.
9. Dupuis, J., Langenberg, C., Prokopenko, I., Saxena, R., Soranzo, N., Jackson, A.U., Wheeler, E., Glazer, N.L., Bouatia-Naji, N., Gloyn, A.L., et al.; DIAGRAM Consortium; GIANT Consortium; Global BPgen Consortium; Anders Hamsten on behalf of Procardis Consortium; MAGIC investigators (2010). New genetic loci implicated in fasting glucose homeostasis and their impact on type 2 diabetes risk. Nat. Genet. *42*, 105–116.
10. Manning, A.K., Hivert, M.F., Scott, R.A., Grimsby, J.L., Bouatia-Naji, N., Chen, H., Rybin, D., Liu, C.T., Bielak, L.F., Prokopenko, I., et al.; DIAbetes Genetics Replication And Meta-analysis (DIAGRAM) Consortium; Multiple Tissue

rate. The color of each data point indicates its LD value (r^2) with the lead SNP at the locus based on HapMap2 YRI: white, r^2 not available; blue, $r^2 = 0.0–0.2$; brown, $r^2 = 0.2–0.5$; orange, $r^2 = 0.5–0.8$; red, $r^2 = 0.8–1.0$. The bottom part of the figure shows Islet Regulome Browser data[13] for the genomic interval shown in the regional association plot.
(B) Regional association plot of a previously undescribed chromosome 5 locus associated with fasting insulin in trans-ethnic meta-analysis. The nearest gene is *PELO* (also called *ITGA1*). The red diamond data point represents the lead SNP (rs6450057) within the region. The blue line represents the recombination rate. The color of each data point indicates its LD value (r^2) with the lead SNP at the locus based on HapMap2: white, r^2 not available; blue, $r^2 = 0.0–0.2$; brown, $r^2 = 0.2–0.5$; orange, $r^2 = 0.5–0.8$; red, $r^2 = 0.8–1.0$.The top two panels show association with $-\log(P)$ on the y axis and genomic position on chromosome 5 on the x axis. The top panel shows association in European ancestry (EA) individuals in MAGIC with plotted points colored according to LD with the lead SNP in HapMap CEU individuals; the middle panel shows association in African ancestry (AA) samples with plotted points colored according to LD with lead SNP in HapMap YRI individuals. The bottom panel shows the trans-ethnic association, with log(BF) on the y axis and genomic position on the x axis, and plotted points colored according to LD with lead SNP in HapMap YRI individuals.
(C) Comparison of the direction of effect on fasting insulin levels of SNPs at the previously undescribed chromosome 5 locus (*PELO*) comparing EA with AA samples. The product of the sign of the beta-coefficient for FI level and –log(*P*) for each SNP at the locus in EA samples (MAGIC) and in AA samples (AAGILE) is plotted on the x and y axes, respectively. The red diamond data point represents the lead SNP (rs6450057) within the region. The color of each data point indicates its LD value (r^2) with the lead SNP at the locus based on HapMap2 YRI: white, r^2 not available; blue, $r^2 = 0.0–0.2$; brown, $r^2 = 0.2–0.5$; orange, $r^2 = 0.5–0.8$; red, $r^2 = 0.8–1.0$. SNPs exhibit opposite effects on fasting insulin level with similar association p values across the LD spectrum at the locus.
(D) Regional association plot of a previously undescribed chromosome 5 locus, with plotted points colored according to RegulomeDB score. The diamond data point represents the lead SNP (rs6450057) within the region; the 99% credible set derived from EA data only is represented by combining the blue and pink boxes; the 99% credible set derived from trans-ethnic analysis is represented by the pink box. Top panel shows a 500 kb genomic span, the lower panel shows a 100 kb genomic span. Data from the Islet RegulomeDB, aligned in the bottom part of the panel, shows binding sites for five transcription factors and regulatory genomic marks in pancreatic islets.

Human Expression Resource (MUTHER) Consortium (2012). A genome-wide approach accounting for body mass index identifies genetic variants influencing fasting glycemic traits and insulin resistance. Nat. Genet. *44*, 659–669.
11. Boyle, A.P., Hong, E.L., Hariharan, M., Cheng, Y., Schaub, M.A., Kasowski, M., Karczewski, K.J., Park, J., Hitz, B.C., Weng, S., et al. (2012). Annotation of functional variation in personal genomes using RegulomeDB. Genome Res. *22*, 1790–1797.
12. Ward, L.D., and Kellis, M. (2012). HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. Nucleic Acids Res. *40*, D930–D934.
13. Pasquali, L., Gaulton, K.J., Rodríguez-Seguí, S.A., Mularoni, L., Miguel-Escalada, I., Akerman, I., Tena, J.J., Morán, I., Gómez-Marín, C., van de Bunt, M., et al. (2014). Pancreatic islet enhancer clusters enriched in type 2 diabetes risk-associated variants. Nat. Genet. *46*, 136–143.
14. Lizio, M., Harshbarger, J., Shimoji, H., Severin, J., Kasukawa, T., Sahin, S., Abugessaisa, I., Fukuda, S., Hori, F., Ishikawa-Kato, S., et al.; FANTOM consortium (2015). Gateways to the FANTOM5 promoter level mammalian expression atlas. Genome Biol. *16*, 22.
15. Morris, A.P. (2011). Transethnic meta-analysis of genomewide association studies. Genet. Epidemiol. *35*, 809–822.
16. Mahajan, A., Go, M.J., Zhang, W., Below, J.E., Gaulton, K.J., Ferreira, T., Horikoshi, M., Johnson, A.D., Ng, M.C., Prokopenko, I., et al.; DIAbetes Genetics Replication And Meta-analysis (DIAGRAM) Consortium; Asian Genetic Epidemiology Network Type 2 Diabetes (AGEN-T2D) Consortium; South Asian Type 2 Diabetes (SAT2D) Consortium; Mexican American Type 2 Diabetes (MAT2D) Consortium; Type 2 Diabetes Genetic Exploration by Nex-generation sequencing in muylti-Ethnic Samples (T2D-GENES) Consortium (2014). Genome-wide trans-ancestry meta-analysis provides insight into the genetic architecture of type 2 diabetes susceptibility. Nat. Genet. *46*, 234–244.
17. Devlin, B., and Roeder, K. (1999). Genomic control for association studies. Biometrics *55*, 997–1004.
18. Howie, B., Fuchsberger, C., Stephens, M., Marchini, J., and Abecasis, G.R. (2012). Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. Nat. Genet. *44*, 955–959.
19. Willer, C.J., Li, Y., and Abecasis, G.R. (2010). METAL: fast and efficient meta-analysis of genomewide association scans. Bioinformatics *26*, 2190–2191.
20. Hong, J., Lunetta, K.L., Cupples, L.A., Dupuis, J., and Liu, C.T. (2016). Evaluation of a two-stage approach in trans-ethnic meta-analysis in genome-wide association studies. Genet. Epidemiol. *40*, 284–292.
21. Maller, J.B., McVean, G., Byrnes, J., Vukcevic, D., Palin, K., Su, Z., Howson, J.M., Auton, A., Myers, S., Morris, A., et al.; Wellcome Trust Case Control Consortium (2012). Bayesian refinement of association signals for 14 loci in 3 common diseases. Nat. Genet. *44*, 1294–1301.
22. Kundaje, A., Meuleman, W., Ernst, J., Bilenky, M., Yen, A., Heravi-Moussavi, A., Kheradpour, P., Zhang, Z., Wang, J., Ziller, M.J., et al.; Roadmap Epigenomics Consortium (2015). Integrative analysis of 111 reference human epigenomes. Nature *518*, 317–330.
23. Bernstein, B.E., Birney, E., Dunham, I., Green, E.D., Gunter, C., Snyder, M., and Consortium, E.P.; ENCODE Project Consortium (2012). An integrated encyclopedia of DNA elements in the human genome. Nature *489*, 57–74.
24. Li, J., and Ji, L. (2005). Adjusting multiple testing in multilocus analyses using the eigenvalues of a correlation matrix. Heredity (Edinb) *95*, 221–227.
25. Yang, J., Ferreira, T., Morris, A.P., Medland, S.E., Madden, P.A., Heath, A.C., Martin, N.G., Montgomery, G.W., Weedon, M.N., Loos, R.J., et al.; Genetic Investigation of ANthropometric Traits (GIANT) Consortium; DIAbetes Genetics Replication And Meta-analysis (DIAGRAM) Consortium (2012). Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. Nat. Genet. *44*, 369–375, S1–S3.
26. Wright, S. (1965). The interpretation of population structure by F-statistics with special regard to systems of mating. Evolution *19*, 395–420.
27. Wright, S. (1978). Evolution and the Genetics of Population, Variability Within and Among Natural Populations (University of Chicago Press).
28. Balloux, F., and Lugon-Moulin, N. (2002). The estimation of population differentiation with microsatellite markers. Mol. Ecol. *11*, 155–165.
29. Voight, B.F., Kudaravalli, S., Wen, X., and Pritchard, J.K. (2006). A map of recent positive selection in the human genome. PLoS Biol. *4*, e72.
30. Kent, W.J., Sugnet, C.W., Furey, T.S., Roskin, K.M., Pringle, T.H., Zahler, A.M., and Haussler, D. (2002). The human genome browser at UCSC. Genome Res. *12*, 996–1006.
31. Ng, M.C., Shriner, D., Chen, B.H., Li, J., Chen, W.M., Guo, X., Liu, J., Bielinski, S.J., Yanek, L.R., Nalls, M.A., et al.; FIND Consortium; eMERGE Consortium; DIAGRAM Consortium; MuTHER Consortium; MEta-analysis of type 2 DIabetes in African Americans Consortium (2014). Meta-analysis of genome-wide association studies in African Americans provides insights into the genetic architecture of type 2 diabetes. PLoS Genet. *10*, e1004517.
32. Monda, K.L., Chen, G.K., Taylor, K.C., Palmer, C., Edwards, T.L., Lange, L.A., Ng, M.C., Adeyemo, A.A., Allison, M.A., Bielak, L.F., et al.; NABEC Consortium; UKBEC Consortium; BioBank Japan Project; AGEN Consortium (2013). A meta-analysis identifies new loci associated with body mass index in individuals of African ancestry. Nat. Genet. *45*, 690–696.
33. Liu, C.T., Monda, K.L., Taylor, K.C., Lange, L., Demerath, E.W., Palmas, W., Wojczynski, M.K., Ellis, J.C., Vitolins, M.Z., Liu, S., et al. (2013). Genome-wide association of body fat distribution in African ancestry populations suggests new loci. PLoS Genet. *9*, e1003681.
34. Franceschini, N., Fox, E., Zhang, Z., Edwards, T.L., Nalls, M.A., Sung, Y.J., Tayo, B.O., Sun, Y.V., Gottesman, O., Adeyemo, A., et al.; Asian Genetic Epidemiology Network Consortium (2013). Genome-wide association analysis of blood-pressure traits in African-ancestry individuals reveals common associated genes in African and non-African populations. Am. J. Hum. Genet. *93*, 545–554.
35. Lettre, G., Palmer, C.D., Young, T., Ejebe, K.G., Allayee, H., Benjamin, E.J., Bennett, F., Bowden, D.W., Chakravarti, A., Dreisbach, A., et al. (2011). Genome-wide association study of coronary heart disease and its risk factors in 8,090 African Americans: the NHLBI CARe Project. PLoS Genet. *7*, e1001300.
36. McCarty, C.A., Chisholm, R.L., Chute, C.G., Kullo, I.J., Jarvik, G.P., Larson, E.B., Li, R., Masys, D.R., Ritchie, M.D., Roden,

D.M., et al.; eMERGE Team (2011). The eMERGE Network: a consortium of biorepositories linked to electronic medical records data for conducting genomic studies. BMC Med. Genomics *4*, 13.
37. Gouda, H.N., Sagoo, G.S., Harding, A.H., Yates, J., Sandhu, M.S., and Higgins, J.P. (2010). The association between the peroxisome proliferator-activated receptor-gamma2 (PPARG2) Pro12Ala gene variant and type 2 diabetes mellitus: a HuGE review and meta-analysis. Am. J. Epidemiol. *171*, 645–655.
38. Gaulton, K.J., Nammo, T., Pasquali, L., Simon, J.M., Giresi, P.G., Fogarty, M.P., Panhuis, T.M., Mieczkowski, P., Secchi, A., Bosco, D., et al. (2010). A map of open chromatin in human pancreatic islets. Nat. Genet. *42*, 255–259.
39. Orho-Melander, M., Melander, O., Guiducci, C., Perez-Martinez, P., Corella, D., Roos, C., Tewhey, R., Rieder, M.J., Hall, J., Abecasis, G., et al. (2008). Common missense variant in the glucokinase regulatory protein gene is associated with increased plasma triglyceride and C-reactive protein but lower fasting glucose concentrations. Diabetes *57*, 3112–3121.
40. Beer, N.L., Tribble, N.D., McCulloch, L.J., Roos, C., Johnson, P.R., Orho-Melander, M., and Gloyn, A.L. (2009). The P446L variant in GCKR associated with fasting plasma glucose and triglyceride levels exerts its effect through increased glucokinase activity in liver. Hum. Mol. Genet. *18*, 4081–4088.
41. Jiang, W., Liu, Y., Liu, R., Zhang, K., and Zhang, Y. (2015). The lncRNA DEANR1 facilitates human endoderm differentiation by activating FOXA2 expression. Cell Rep. *11*, 137–148.
42. Ng, M.C., Saxena, R., Li, J., Palmer, N.D., Dimitrov, L., Xu, J., Rasmussen-Torvik, L.J., Zmuda, J.M., Siscovick, D.S., Patel, S.R., et al. (2013). Transferability and fine mapping of type 2 diabetes loci in African Americans: the Candidate Gene Association Resource Plus Study. Diabetes *62*, 965–976.
43. Schaub, M.A., Boyle, A.P., Kundaje, A., Batzoglou, S., and Snyder, M. (2012). Linking disease associations with regulatory information in the human genome. Genome Res. *22*, 1748–1759.
44. Maurano, M.T., Humbert, R., Rynes, E., Thurman, R.E., Haugen, E., Wang, H., Reynolds, A.P., Sandstrom, R., Qu, H., Brody, J., et al. (2012). Systematic localization of common disease-associated variation in regulatory DNA. Science *337*, 1190–1195.
45. Claussnitzer, M., Dankel, S.N., Klocke, B., Grallert, H., Glunk, V., Berulava, T., Lee, H., Oskolkov, N., Fadista, J., Ehlers, K., et al.; DIAGRAM+Consortium (2014). Leveraging cross-species transcription factor binding site patterns: from diabetes risk loci to disease mechanisms. Cell *156*, 343–358.
46. Matthews, D.R., Hosker, J.P., Rudenski, A.S., Naylor, B.A., Treacher, D.F., and Turner, R.C. (1985). Homeostasis model assessment: insulin resistance and beta-cell function from fasting plasma glucose and insulin concentrations in man. Diabetologia *28*, 412–419.
47. Laakso, M. (1993). How good a marker is insulin level for insulin resistance? Am. J. Epidemiol. *137*, 959–965.
48. Kahn, S.E., Prigeon, R.L., McCulloch, D.K., Boyko, E.J., Bergman, R.N., Schwartz, M.W., Neifing, J.L., Ward, W.K., Beard, J.C., Palmer, J.P., et al. (1993). Quantification of the relationship between insulin sensitivity and beta-cell function in human subjects. Evidence for a hyperbolic function. Diabetes *42*, 1663–1672.
49. Weyer, C., Hanson, R.L., Tataranni, P.A., Bogardus, C., and Pratley, R.E. (2000). A high fasting plasma insulin concentration predicts type 2 diabetes independent of insulin resistance: evidence for a pathogenic role of relative hyperinsulinemia. Diabetes *49*, 2094–2101.
50. Ferrannini, E., Natali, A., Bell, P., Cavallo-Perin, P., Lalic, N., and Mingrone, G.; European Group for the Study of Insulin Resistance (EGIR) (1997). Insulin resistance and hypersecretion in obesity. J. Clin. Invest. *100*, 1166–1173.
51. Corkey, B.E. (2012). Banting lecture 2011: hyperinsulinemia: cause or consequence? Diabetes *61*, 4–13.
52. Corkey, B.E. (2012). Diabetes: have we got it all wrong? Insulin hypersecretion and food additives: cause of obesity and diabetes? Diabetes Care *35*, 2432–2437.
53. Chen, C.H., Cruz, L.A., and Mochly-Rosen, D. (2015). Pharmacological recruitment of aldehyde dehydrogenase 3A1 (ALDH3A1) to assist ALDH2 in acetaldehyde and ethanol metabolism in vivo. Proc. Natl. Acad. Sci. USA *112*, 3074–3079.
54. Doudna, J.A., and Charpentier, E. (2014). Genome editing. The new frontier of genome engineering with CRISPR-Cas9. Science *346*, 1258096.
55. Pautsch, A., Stadler, N., Löhle, A., Rist, W., Berg, A., Glocker, L., Nar, H., Reinert, D., Lenter, M., Heckel, A., et al. (2013). Crystal structure of glucokinase regulatory protein. Biochemistry *52*, 3523–3531.
56. Saxena, R., Voight, B.F., Lyssenko, V., Burtt, N.P., de Bakker, P.I., Chen, H., Roix, J.J., Kathiresan, S., Hirschhorn, J.N., Daly, M.J., et al.; Diabetes Genetics Initiative of Broad Institute of Harvard and MIT, Lund University, and Novartis Institutes of BioMedical Research (2007). Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels. Science *316*, 1331–1336.
57. Vaxillaire, M., Cavalcanti-Proença, C., Dechaume, A., Tichet, J., Marre, M., Balkau, B., and Froguel, P.; DESIR Study Group (2008). The common P446L polymorphism in GCKR inversely modulates fasting glucose and triglyceride levels and reduces type 2 diabetes risk in the DESIR prospective general French population. Diabetes *57*, 2253–2257.
58. Udler, M.S., Ahmed, S., Healey, C.S., Meyer, K., Struewing, J., Maranian, M., Kwon, E.M., Zhang, J., Tyrer, J., Karlins, E., et al. (2010). Fine scale mapping of the breast cancer 16q12 locus. Hum. Mol. Genet. *19*, 2507–2515.
59. Franceschini, N., van Rooij, F.J., Prins, B.P., Feitosa, M.F., Karakas, M., Eckfeldt, J.H., Folsom, A.R., Kopp, J., Vaez, A., Andrews, J.S., et al.; LifeLines Cohort Study (2012). Discovery and fine mapping of serum protein loci through transethnic meta-analysis. Am. J. Hum. Genet. *91*, 744–753.
60. Liu, C.T., Buchkovich, M.L., Winkler, T.W., Heid, I.M., Borecki, I.B., Fox, C.S., Mohlke, K.L., North, K.E., Adrienne Cupples, L., Consortium, A.A.A.G., et al.; African Ancestry Anthropometry Genetics Consortium; GIANT Consortium (2014). Multiethnic fine-mapping of 14 central adiposity loci. Hum. Mol. Genet. *23*, 4738–4744.
61. Ong, R.T., and Teo, Y.Y. (2010). varLD: a program for quantifying variation in linkage disequilibrium patterns between populations. Bioinformatics *26*, 1269–1270.

The American Journal of Human Genetics, Volume *99*

Supplemental Data

Trans-ethnic Meta-analysis and Functional Annotation Illuminates the Genetic Architecture of Fasting Glucose and Insulin

Ching-Ti Liu, Sridharan Raghavan, Nisa Maruthur, Edmond Kato Kabagambe, Jaeyoung Hong, Maggie C.Y. Ng, Marie-France Hivert, Yingchang Lu, Ping An, Amy R. Bentley, Anne M. Drolet, Kyle J. Gaulton, Xiuqing Guo, Loren L. Armstrong, Marguerite R. Irvin, Man Li, Leonard Lipovich, Denis V. Rybin, Kent D. Taylor, Charles Agyemang, Nicholette D. Palmer, Brian E. Cade, Wei-Min Chen, Marco Dauriz, Joseph A.C. Delaney, Todd L. Edwards, Daniel S. Evans, Michele K. Evans, Leslie A. Lange, Aaron Leong, Jingmin Liu, Yongmei Liu, Uma Nayak, Sanjay R. Patel, Bianca C. Porneala, Laura J. Rasmussen-Torvik, Marieke B. Snijder, Sarah C. Stallings, Toshiko Tanaka, Lisa R. Yanek, Wei Zhao, Diane M. Becker, Lawrence F. Bielak, Mary L. Biggs, Erwin P. Bottinger, Donald W. Bowden, Guanjie Chen, Adolfo Correa, David J. Couper, Dana C. Crawford, Mary Cushman, John D. Eicher, Myriam Fornage, Nora Franceschini, Yi-Ping Fu, Mark O. Goodarzi, Omri Gottesman, Kazuo Hara, Tamara B. Harris, Richard A. Jensen, Andrew D. Johnson, Min A. Jhun, Andrew J. Karter, Margaux F. Keller, Abel N. Kho, Jorge R. Kizer, Ronald M. Krauss, Carl D. Langefeld, Xiaohui Li, Jingling Liang, Simin Liu, William L. Lowe, Jr.,, Thomas H. Mosley, Kari E. North, Jennifer A. Pacheco, Patricia A. Peyser, Alan L. Patrick, Kenneth M. Rice, Elizabeth Selvin, Mario Sims, Jennifer A. Smith, Salman M. Tajuddin, Dhananjay Vaidya, Mary P. Wren, Jie Yao, Xiaofeng Zhu, Julie T. Ziegler, Joseph M. Zmuda, Alan B. Zonderman, Aeilko H. Zwinderman, AAAG Consortium, CARe Consortium, COGENT-BP Consortium, eMERGE Consortium, MEDIA Consortium, Adebowale Adeyemo, Eric Boerwinkle, Luigi Ferrucci, M. Geoffrey Hayes, Sharon L.R. Kardia, Iva Miljkovic, James S. Pankow, Charles N. Rotimi, Michele M. Sale, Lynne E. Wagenknecht, Donna K. Arnett, Yii-Der Ida Chen, Michael A. Nalls, MAGIC Consortium, Michael A. Province, W.H. Linda Kao, David S. Siscovick, Bruce M. Psaty, James G. Wilson, Ruth J.F. Loos, Josée Dupuis, Stephen S. Rich, Jose C. Florez, Jerome I. Rotter, Andrew P. Morris, and James B. Meigs

Supplemental Materials

Table of Brief Contents

Table of contents

Figure S1 Schematic study diagram

Figure S2 Venn diagram of trans-ethnic analysis and transferability results

Figure S3 Trans-ethnic fine-mapping of 22 loci (13 FG, 8 FI, and 1 both FG and FI) with greater than 20% reduction in the 99% credible set.

FG loci

Figure S3A. *FOXA2* locus
Figure S3B. *GCK* locus
Figure S3C. *CRY2* locus
Figure S3D. *KL* locus
Figure S3E. *ADCY5* locus
Figure S3F. *GCKR* locus
Figure S3G. *PROX1* locus
Figure S3H. *DPYSL5* locus
Figure S3I. *IGF2BP2* locus
Figure S3J. *CDKN2B* locus
Figure S3K. *ADRA2A* locus
Figure S3L. *TCF7L2* locus
Figure S3M. *FADS1* locus
Figure S3N. *DGKB-TMEM195* locus

FI loci

Figure S3O. *ARL15* locus
Figure S3P. *PPP1R3B* locus
Figure S3Q. *COBLL1-GRB14* locus
Figure S3R. *IRS1* locus
Figure S3S. *GCKR* locus
Figure S3T. *FAM13A* locus
Figure S3U. *ANKRD55-MAP3K1* locus
Figure S3V. *UHRF1BP1* locus
Figure S3W. *PPARG* locus

Figure S4 Plots of regional association and RegulomeDB and Islet Regulome Browser information at 14 FG and 9 FI loci with substantially narrowed credible sets after trans-ethnic analysis

FG loci

Figure S4A. *TCF7L2* locus
Figure S4B. *ADRA2* locus
Figure S4C. *DGKB-TMEM195* locus
Figure S4D. *FADS1* locus
Figure S4E. *PROX1* locus
Figure S4F. *GCK* locus
Figure S4G. *ADCY5* locus
Figure S4H. *GCKR* locus
Figure S4I. *CDKN2B* locus
Figure S4J. *FOXA2* locus
Figure S4K. *CRY2* locus
Figure S4L. *DPYSL5* locus
Figure S4M. *IGF2BP2* locus
Figure S4N. *KL* locus

FI loci

Figure 4SO. *ARL15* locus
Figure 4SP. *COBLL1-GRB14* locus
Figure 4SQ. *IRS1* locus
Figure 4SR. *UHRF1BP1* locus
Figure 4SS. *FAM13A* locus
Figure 4ST. *PPARG* locus
Figure 4SU. *GCKR* locus
Figure 4SV. *PPP1R3B* locus
Figure 4SW. *ANKRD55-MAP3K1* locus

Figure S5 Concordance of effect size and Comparison of EA trait-raising allele Frequency in EA and AA

Figures S5A Effect size comparison for EA FG SNPs
Figures S5B Effect size comparison for EA FI SNPs
Figures S5C Allele frequency comparison for EA FG SNPs
Figures S5D Allele frequency comparison for EA FI SNPs

Figure S6 Genome-wide association plots and quantile-quantile (QQ) plots for FG and FI

Figures S6A Miami plots of association with FG
Figures S6B the QQ plots for FG
Figures S6C Miami plots of association with FI
Figures S6D the QQ plots for FI
Figures S6E genome-wide association plots of trans-ethnic meta-analysis results for FG
Figures S6F genome-wide association plots of trans-ethnic meta-analysis results for FI

Figure S7 Conditional analysis at *PELO*/rs6450057

Figures S7A the comparison for unconditional association results with HapMap 2 CEU LD information
Figures S7B the comparison for conditional association results with HapMap 2 CEU LD information
Figures S7C the comparison for unconditional association results with HapMap 2 YRI LD information
Figures S7D the comparison for conditional association results with HapMap 2 YRI LD information

Supplemental Figure Legends

Supplemental Tables and Legends

Supplemental References

Participating consortia and investigators

The data used in the current report were derived from published results from The Meta-Analyses of Glucose and Insulin-related traits Consortium (**MAGIC**)[1] and from unpublished meta-analyses from the African American Glucose and Insulin genetic Epidemiology (**AAGILE**) consortium. Only non-diabetic individuals were included in the trans-ethnic meta-analysis that combined data from MAGIC and AAGILE consortium. All data on participants of European ancestry (EA) were obtained from MAGIC (n = 51,750) while all data on participants of African ancestry (AA) were obtained from the AAGILE Consortium (n = 20,209). Cohorts that contributed data to the MAGIC consortium and names of the steering committee members are listed on the consortium website: http://www.magicinvestigators.org/.

The AAGILE consortium includes AA individuals from 16 cohorts. Key characteristics for each discovery and replication study sample are shown in **Table S1**. Investigators from AAGILE consortium are listed in the author list while those from the MAGIC consortium are shown below.

MAGIC investigator [1]

Manning AK, Hivert MF, Scott RA, Grimsby JL, Bouatia-Naji N, Chen H, Rybin D, Liu CT, Bielak LF, Prokopenko I, Amin N, Barnes D, Cadby G, Hottenga JJ, Ingelsson E, Jackson AU, Johnson T, Kanoni S, Ladenvall C, Lagou V, Lahti J, Lecoeur C, Liu Y, Martinez-Larrad MT, Montasser ME, Navarro P, Perry JR, Rasmussen-Torvik LJ, Salo P, Sattar N, Shungin D, Strawbridge RJ, Tanaka T, van Duijn CM, An P, de Andrade M, Andrews JS, Aspelund T, Atalay M, Aulchenko Y, Balkau B, Bandinelli S, Beckmann JS, Beilby JP, Bellis C, Bergman RN, Blangero J, Boban M, Boehnke M, Boerwinkle E, Bonnycastle LL, Boomsma DI, Borecki IB, Böttcher Y, Bouchard C, Brunner E, Budimir D, Campbell H, Carlson O, Chines PS, Clarke R, Collins FS, Corbatón-Anchuelo A, Couper D, de Faire U, Dedoussis GV, Deloukas P, Dimitriou M, Egan JM, Eiriksdottir G, Erdos MR, Eriksson JG, Eury E, Ferrucci L, Ford I, Forouhi NG, Fox CS, Franzosi MG, Franks PW, Frayling TM, Froguel P, Galan P, de Geus E, Gigante B, Glazer NL, Goel A, Groop L, Gudnason V, Hallmans G, Hamsten A, Hansson O, Harris TB, Hayward C, Heath S, Hercberg S, Hicks AA, Hingorani A, Hofman A, Hui J, Hung J, Jarvelin MR, Jhun MA, Johnson PC, Jukema JW, Jula A, Kao WH, Kaprio J, Kardia SL, Keinanen-Kiukaanniemi S, Kivimaki M, Kolcic I, Kovacs P, Kumari M, Kuusisto J, Kyvik KO, Laakso M, Lakka T, Lannfelt L, Lathrop GM, Launer LJ, Leander K, Li G, Lind L, Lindstrom J, Lobbens S, Loos RJ, Luan J, Lyssenko V, Mägi R, Magnusson PK, Marmot M, Meneton P, Mohlke KL, Mooser V, Morken MA, Miljkovic I, Narisu N, O'Connell J, Ong KK, Oostra BA, Palmer LJ, Palotie A, Pankow JS, Peden JF, Pedersen NL, Pehlic M, Peltonen L, Penninx B, Pericic M, Perola M, Perusse L, Peyser PA, Polasek O, Pramstaller PP, Province MA, Räikkönen K, Rauramaa R, Rehnberg E, Rice K, Rotter JI, Rudan I, Ruokonen A, Saaristo T, Sabater-Lleal M, Salomaa V, Savage DB, Saxena R, Schwarz P, Seedorf U, Sennblad B, Serrano-Rios M, Shuldiner AR, Sijbrands EJ, Siscovick DS, Smit JH, Small KS, Smith NL, Smith AV, Stančáková A, Stirrups K, Stumvoll M, Sun YV, Swift AJ, Tönjes A, Tuomilehto J, Trompet S, Uitterlinden AG, Uusitupa M, Vikström M, Vitart V, Vohl MC, Voight BF, Vollenweider P, Waeber G, Waterworth DM, Watkins H, Wheeler E, Widen E, Wild SH, Willems SM, Willemsen G, Wilson JF, Witteman JC, Wright AF, Yaghootkar H, Zelenika D, Zemunik T, Zgaga L; DIAbetes Genetics Replication And Meta-analysis (DIAGRAM) Consortium; Multiple Tissue Human Expression Resource (MUTHER) Consortium, Wareham NJ, McCarthy MI, Barroso I, Watanabe RM, Florez JC, Dupuis J, Meigs JB, Langenberg C.

Sources of data for pleiotropy studies

Consortia that contributed results for associations between fasting glucose or fasting insulin SNPs and insulin-related traits (i.e., hypertension, systolic and diastolic blood pressure, triglycerides, high density lipoprotein cholesterol, low density lipoprotein cholesterol, body mass index and waist-to-hip ratio-adjusted for BMI) are shown below:

1. **Continental Origins and Genetic Epidemiology Network (COGENT) consortium[2]**

   The COGENT consortium provided association results for hypertension and for systolic and diastolic blood pressure.

2. **Electronic Medical Records and Genomics Network (eMERGE)[3]**

   The eMERGE consortium provided triglyceride data from BioVU at Vanderbilt University Medical Center (https://victr.vanderbilt.edu/pub/biovu/) and triglyceride and blood pressure (hypertension and systolic and diastolic blood pressure) data from Mt. Sinai School of Medicine.

3. **The National, Heart, Lung and Blood Institute's Candidate gene Association Resource (CARe)[4]**

   The CARe consortium contributed association results for high density lipoprotein cholesterol and low density lipoprotein cholesterol.

4. **MEta-analysis of type 2 DIabetes in African Americans (MEDIA) Consortium[5]**

   The MEDIA consortium provided association results for type 2 diabetes.

5. **African Ancestry Anthropometry Genetics (AAAG) Consortium[6,7]**

   The AAAG consortium provided association results for body mass index and waist-hip-ratios.

Author contributions

The contributions of authors are summarized below:

Assembling and steering the consortium: AA, APM, BMP, CNR, CTL, DJC, DKA, DS, DWB, EB, EPB, IM, JBM, JCF, JD, JGW, JIR, JSP, LEW, LF, MAN, MAP, MMS, RJFL, SSR, YIC

Genotyping and data imputation: AA, EPB, GC, JAS, JIR, JY, KDT, LJRT, MAP, MF, MMS, MOG, RJFL, SLRK, TT, XG, YIC, YL

Phenotyping of the study participants: ANK, DS, EPB, JAP, JIR, JK, JMZ, MAP, MLB, PAP, RJFL, RK, SLRK, YIC, YL

Statistical analyses: AA, ABZ, AC, ADJ, AHZ, AJK, ALP, AMD, ARB, BEC, BP, CA, CDL, CTL, DC, DV, DS, DSE, DVR, DWB, EKK, EPB, ES, GC, JACD, JD, JDE, JH, JL, JL, JTZ, JY, KEN, KJG, KMR, LAL, LFB, LL, LLA, LRY, MAJ, MAP, MBS, MC, MCYN, MFK, MGH, MKE, ML, MPW, MRI, MS, NA, NF, NM, OG, PA, PAP, RAJ, RK, SL, SR, SRP, SS, ST, TBH, THM, TLE, TT, UN, WLL, WMC, WMC, WZ, XG, XL, XZ, YL, YL, YPF

SNP annotation: AMD, KJG, LL, JH, SR, KH

Drafting of the manuscript: AMD, APM, ASL, CTL, DS, DMB, DV, EKK, EPB, ES, JBM, JCF, JD, JGW, JH, JIR, KDT, KH, KJG, LL, LRY, MCYN, MD, MFH, MGH, ML, MOG, MMS, NM, PA, RJFL, RK, SR, SSR, WHLK, WMC, XG, XL, YIC, YL

Enhancing the manuscript for intellectual content: All authors

Approval of the final version: All authors

SNP annotation

Following trans-ethnic meta-analysis of data from **MAGIC** (EA participants) and the **AAGILE** consortium (AA participants), we identified 14 fasting glucose (FG) and 9 fasting insulin (FI) loci (**Table S5**) in which the number of SNPs in a locus or the size of the genomic region likely to harbor the causative SNP was reduced by at least 20%. For each of the SNPs in the narrowed region, referred to as the 99% credible set, we used HaploReg V2 to annotate their biological relevance. HaploReg provides useful annotation information for the SNP of interest as well as those within a user-specified LD. The fully operational web version is available at http://www.broadinstitute.org/mammals/haploreg/haploreg.php. HaploReg reports any evidence for regulatory chromatin marks, DNAse I hypersensitivity sites (DHSI), transcription factor binding sites (TFBS), transcription factor binding motifs, or expression quantitative trait loci (eQTL) overlapping with each SNP of interest and thus gives mechanistic insights into how non-coding SNPs may lead to a given disease condition. Ward and Kellis provide more detailed information on HaploReg and its usage.[8]

After HaploReg we performed additional annotation manually using annotation resources in the public domain (i.e., RegulomeDB, ENCODE, Islet Regulome and FANTOM) to further characterize potential regulatory functions of the variants in the credible sets. Use of HaploReg together with manual annotation revealed whether a given SNP lies in a location with a histone mark suggestive of regulatory activity, a DHSI, TFBS, or if the SNP is in a gene expressed in a diabetes-relevant tissue, e.g., the liver or pancreas. In the case of histone marks we also evaluated whether the histone mark corresponds to an enhancer or promoter. Finally, we catalogued other traits with significant associations reported within the credible set from the NHGRI GWAS Catalog. Below we provide detailed information on the potential biological relevance of SNPs in each of the 14 FG and 9 FI loci that we annotated; the index SNP in EA indicated in parentheses.

1. **FOXA2 (rs6048205)**
   a. Credible set interval in hg18: chr20:22505099-22508971
   b. Credible set interval in hg19: chr20:22557099-22560971
   c. The region contains the 5' end of TCONS_00028636 (lincRNA)
      i. Displays especially high expression in liver cell lines
      ii. Is accompanied by moderate to high transcription levels (RNA-seq) only in HepG2 cell lines
      iii. The EST BG655894 indicates expression in pancreatic Islets
   d. There are five broad and weak H3K4Me1 peaks in the region
      i. Especially prominent in Embryonic stem cells
   e. There are 51 TFBSs in the region, all but 2 are expressed in liver cell lines
   f. The lead SNP is 20bp telomeric (p arm) to a DHSI (19/125) which is found in pancreatic islet cell lines
   g. The lead SNP sits inside 10 TFBSs
      i. They are all found in liver cell lines
      ii. Includes FOXA1 and FOXA2
   h. The RNA-seq information from the TFBSs in the region as well as the expression profile of the lincRNA indicate that the locus plays a role in the liver. The lead SNPs position in several binding sites may lead to a disruption in the binding of TFBSs.

2. **GCK (rs4607517)**

This credible set contained a single SNP; annotation overlap for the SNP can be found in Supplemental Table 5.

3. CRY2 (rs11605924)

a. Credible set interval in hg18: chr11:45820718-45835568
b. Credible set interval in hg19: chr11:45864142-45878992
c. Contains 5' end of CRY2
d. There is one NHGRI SNP
    i. rs11605924 – fasting glucose traits
e. There are 7 moderate to weak HK4Me1 peaks in the region
f. There are 3 weak and 1 strong H3K4Me3 peaks in the region
g. There are 2 weak and 2 strong H3K27Ac peaks in the region
h. The lead SNP is 1kbp upstream of a weak H3K4Me1 peak
    i. DHSI (27) found in liver cell lines
    ii. 9 TFBSs
i. 5.6kbp upstream of weak H3K4Me1 and H3K27Ac peaks
    i. DHSI (57) found in muscle and pancreatic islets
    ii. 3TFBSs
j. 1kbp downstream of a weak H3K4Me1 peak
    i. DHSI (3) and (32)
k. 3.8kbp downstream of dual H3K4me1, H3K4Me3, and H3K27Ac peaks
    i. DHSI (111) and (125)
    ii. 101 TFBSs
l. 6.5kbp telomeric (p arm) to the SNP is a strong H3K4Me3 peak
m. 8kbp telomeric to the SNP are weak H3K4Me1, H3K27Ac, H3M4Me3 peaks
    i. DHSI (125)
    ii. 49 TFBSs
n. The lead SNP lies in the promoter region of CRY2, which explains its proximity to several H3K4Me1 peaks. Because CRY2 impacts metabolism, a polymorphism in the gene's promoters may alter the function.

4. KL (rs576674)

a. Credible set interval in hg18: chr13:32209005-32701555
b. Credible set interval in hg19: chr13:33311005-33803555
c. Four genes in the region
    i. 3' end of PDS5B
    ii. KL
    iii. STARD13
    iv. TCONS_00021632 (lincRNA)
d. Four NHGRI GWAS SNPs
    i. rs2555603 – BMI
    ii. (not related) rs2555603 – aneurysm
    iii. (not related) rs642899 –behavioral disinhibition
    iv. (not related) rs990324 – total ventricular volume
e. There are 29 H3K4Me1 peaks throughout the region
f. There are two H3K4Me3 peaks (weak)
g. 15 H3K27Ac peaks throughout the region
h. The lead SNP lies within a FOXA2 binding site found in liver cell lines
    i. This is clustered near three other binding sites, all found in liver cell lines

i. 10.5kbp telomeric (q arm) to a moderate H3K4Me1 peak and weak H3K4Me3 and H3K27Ac peaks
    i. DHSI (66) found in muscle and pancreatic cells
    ii. 6 TFBSs
j. 4.6kbp upstream of a DHSI (26) in pancreatic islets
    i. 12 TFBSs
k. The lead SNP is in an intergenic region, and may serve as a bidirectional, distant regulator to both TCONS_00021632 and KL. The SNP may impact the binding of FOXA2 to its TFBS, and exert its effects primarily in the liver.

5. ADCY5 (rs11708067)

This credible set contained a single SNP; annotation overlap for the SNP can be found in Supplemental Table 5.

6. GCKR (rs780094)

This credible set contained a single SNP; annotation overlap for the SNP can be found in Supplemental Table 5.

7. PROX1 (rs340874)

a. Credible set interval in hg18: chr1: 212212012-212230298
b. Credible set interval in hg19: chr1:214145389-214163675
c. Contains the 5'UTRs of PROX-AS1 and PROX1
d. Exonic to the mRNA AK096113, overlaps PROX-AS1. Found in liver cell lines
e. The 5'UTR of PROX-AS1 overlaps the mRNA AK096113, which comes from human liver cells
f. There are 2 NHGRI SNPs in the region
    i. rs2075423 – Type 2 Diabetes
    ii. rs340874 – Fasting glucose traits
g. There are 4 H3K4Me1 peaks throughout the region
h. There is one weak and one strong H3K4Me3 peaks
i. There are two H3K27Ac peaks
j. The 5'UTR of PROX1 is overlaps with a level of moderate transcription according to RNA-seq, especially strong in HepG2 cell lines (2.2kbp telomeric to (p arm) lead SNP)
    i. There are also moderate H3K4Me1 and H3K27Ac peaks and a strong H3K4Me3 peak
    ii. There is a DHSI (116/125) with 36 TFBSs (31 present in liver cells)
        1. TFBSs include FOXA1 and FOXA2
k. 12.5kbp downstream, there is a large H3K4Me1 peak, with a DHSI (25/125) found in HSMM cells and one TFBS
l. 8.2kbp downstream, there are strong H3K4Me1 dual peaks and small H3K37Ac peaks. There is a DHSI (27/125) and 9 TFBSs
m. The lead SNP lies within an area of mild H3K4Me1 expression and several different DHSI and 41 TFBSs (31 expressed in liver cells)
    i. The SNP is within two TFBSs: EZH2 and CTBP2
n. The histone modifications are consistent with the location of the SNP within the promoter region of PROX1 and PROX1-AS1. This area may play an important role in the liver, as indicated by the tissue specificity of TFBSs, RNA-seq, and human mRNA.

8. **DPYSL5 (rs1371614)**
    a. Credible set interval in hg18: chr2: 26957635-27251700
    b. Credible set interval in hg19: chr2:27104131-27398196
    c. Contains 12 different genes
        i. DPYSL5
        ii. MAPRE3
        iii. TMEM214
        iv. OST4
        v. KHK
        vi. EMILIN1
        vii. AGBL5
        viii. CGREF1
        ix. ABHD1
        x. PREB
        xi. C2orf53
        xii. TCF23
    d. The 3'UTR of DYPSL5 contains high transcription levels, according to RNA-seq
    e. Two NHGRI SNPs
        i. rs1371614 – Fasting glucose traits
        ii. (not related) rs7588926 – Response to cytadine analogues
    f. There are 23 H3K4Me1 peaks throughout the region
    g. There are 12 H3K4Me3 peaks throughout the region
    h. There are 14 H3K27Ac peaks throughout the region
    i. The lead SNP lies with a DHSI (3) and 2 TFBSs, FOXA1 and FOXA2, both found in liver cell lines
    j. 4kbp downstream of a small H3K4Me1 peak
        i. DHSI (26) found in liver, muscle, and pancreatic cell lines
        ii. 23 TFBSs, all found in liver cell lines
    k. 13kbp upstream of an area of weak H3K4Me1 modification
        i. DHSI (113) and 13 TFBSs
    l. This is a larger transethnic region, with explains the large amount of genes with associated histone modifications. The SNP may impact the binding of FOXA1 and FOXA2, specifically in liver cells, which would affect liver metabolism.

9. **IGF2BP2 (rs7651090)**
    a. Credible set interval in hg18: chr3: 186750043-187067565
    b. Credible set interval in hg19: chr3:185267349-185584871
    c. Five genes in the region
        i. LIPH
        ii. SENP2
        iii. IGF2BP2
            1. Human mRNA BC021290 indicates expression in pancreas
        iv. C3Orf65
        v. TCONS_00006340 (lincRNA)
    d. Five NHGRI GWAS SNPs
        i. (not related) rs720390 – height
        ii. rs1374910 – Type 2 Diabetes
        iii. rs6769511 – Type 2 Diabetes

iv. rs1470579 – Type 2 Diabetes
v. rs4402960 – Type 2 Diabetes

e. There are 24 H3K4Me1 peaks throughout the region
f. There are 3 H3K4Me3 peaks throughout the region
g. There are 15 H3K27Ac peaks throughout the region
h. The lead SNP is intronic to IGF2BP2
i. 1.3kbp downstream of large H3K4Me1 and H3K27Ac peaks
    i. DHSI (93) and 28 TFBSs
j. 12kbp downstream of a DHSI (87)
    i. 29 TFBSs
k. 4.7kbp upstream of a moderate H3K4Me1 peak
    i. DHSI (12)
l. 13kbp upstream of an area of moderate H3K4Me1
    i. DHSI (60) found in muscle and pancreatic cells
    ii. 5 TFBSs
m. The large Type 2 Diabetes SNP cloud indicates the importance of the region in that disease. The lead SNP is intronic to IGF2BP2, and 27 kbp from the nearest exon. Its role may be with the nearby putative promoter, located 1.3 kbp away. Further validation could show if this promoter is associated with IGF2BP2, or the downstream lincRNA TCONS_00006340.

10. CDKN2B (rs10811661)

a. Credible set interval in hg18: chr9: 22118180-22124094
b. Credible set interval in hg19: chr9:22128180-22134094
c. No genes within the region
d. 4 NHGRI GWAS SNPs
    i. rs7020996 – Type 2 diabetes
    ii. rs2383208 – Type 2 Diabetes
    iii. rs10965250 – Type 2 Diabetes
    iv. rs10811661 – Type 2 Diabetes
e. There are three moderate, broad H3M4Me1 peak, especially prominent in blood vessel cells
    i. 4kbp, 4.6kbp, and 5.4kbp telomeric (p arm) to the lead SNP
f. 4.5kbp telomeric to the SNP, there is a DHSI (36/125) which is in HSMM and pancreatic islet cells. There are three TFBSs nearby, including FOXA2
g. The entire region is intergenic and contains a large Type 2 Diabetes SNP cloud. This indicates the importance of the enhancer region, which overlaps a DHSI site with tissue specificity in diabetes-relevant tissues.

11. ADRA2A (rs10885122)

a. Credible set interval in hg18: chr10: 112960941-113029657
b. Credible set interval in hg19: chr10:112970951-113039667
c. No genes in region
d. Only histone modification is a very small H3K4Me1 peak between 113,005,278-113,007,774 only in HSMM cells
e. There is a DHSI (8/125) in HSMM and PanIsletD.
f. There are 9 TFBSs in the region, and 7 are present in liver cell lines
g. **When using lift over, the lead SNP lies just outside the credible region** (113042093)

h. This trans-ethnic region contains little biological information, being both intergenic and having little regulatory information. However, the information that is available shows tissue specificity to diabetes-relevant tissues, and may impact a genomic element not currently reported.

12. TCF7L2 (rs7903146)

a. Credible set interval in hg18: chr10: 114742493-114778805
b. Credible set interval in hg19: chr10:114752503-114788815
c. The entire region is intronic to TCF7L2
    i. Human mRNA FJ010174 indicates expression in pancreatic, hepatic, renal, muscle and adipose cells
d. There are five NHGRI GWAS SNPs
    i. rs12243326 – 2 hour glucose challenge
    ii. (not related) rs7904519 – Breast cancer
    iii. rs7903146 – Type 2 diabetes
    iv. rs4506565 – fasting glucose traits/Type 2 Diabetes
    v. rs7901695 – Type 2 diabetes
e. There are 5 H3K4Me1 peaks (two strong, three weak)
    i. The weak peaks are located 2.3kbp, 3.8kbp, and 5kbp downstream from the lead SNP
f. 8.8kbp downstream, there is a DHSI (8/125) preset in liver cell lines with 21 TFBSs, all present in liver cell lines
    i. This coincides with a very strong peak of conservation
g. The large Type 2 Diabetes SNP cloud, along with the tissue specificity information from the human mRNA, confirms the region's importance in diabetes. The SNP may play a role in the promoter regions and impact the transcription of TCF7L2, which is important in maintaining blood glucose levels. (http://www.ncbi.nlm.nih.gov/gene/6934)

13. FADS1 (rs174550)

a. Credible set interval in hg18: chr11: 61307932-61366326
b. Credible set interval in hg19: chr11:61551356-61609750
c. There are five genes in the region
    i. MYRF
    ii. FEN1
        1. EST CA868349 indicating expression in pancreas
    iii. TMEM258
    iv. FADS1
    v. FADS2
        1. EST BP237803 indicating expression in liver
d. There are 20 NHGRI GWAS SNPs
    i. rs174541 – metabolite levels
    ii. rs4246215 – platelet counts, phospholipid levels
    iii. rs174538 – blood metabolite levels
    iv. rs102275 – blood cholesterol/metabolite levels, metabolic syndrome
    v. (not related) rs174537 – colorectal cancer
    vi. rs174556 – blood metabolite levels
    vii. rs174555 – blood fatty acid levels

viii. rs174550 – fasting glucose related traits
ix. rs174548 – blood metabolite levels
x. rs174547 – metabolic traits
xi. rs174546 – metabolic syndrome, cholesterol
xii. rs174578 – blood metabolite levels
xiii. rs2727271 – blood metabolite levels
xiv. rs2727270 – fatty acid levels
xv. (not related) rs174583 – response to statins
xvi. (not related) rs174577 – P wave duration
xvii. rs174574 – phospholipid levels
xviii. rs1535 – metabolic syndromw
xix. (not related) rs174570 – glycated hemoglobin levels
xx. rs968567 – blood metabolite levels

e. Seven H3K27Ac peaks
   i. Two are associated with H3K4Me1 peaks
f. There are three H3KeMe3 peaks, all associated with H3K27Ac peaks
g. There are eleven H3K4Me1 peaks, seven are associated with H3K27Ac peaks
h. The lead SNP is intronic to FADS1
i. The lead SNP is inside of a H3K4Me1 peak
j. 12kbp upstream are strong H3K4Me3 and H3K27Ac peaks
   i. associated with several DHSI (55,117,124,125,14). The 14/125 DHSI is in both hepatocytes and pancreatic cells. There are over 150 associated TFBSs
   ii. On either side of this region are H3K4Me1 peaks
k. 3.9kbp upstream of H3K4Me1 and K3K17Ac peaks
   i. DHSI (12) found in pancreatic islet cells
   ii. 42 TFBSs
l. 11kbp telomeric (q arm) to H3K4Me3 and H3K27Ac peaks
   i. DHSI (125) and (116)
   ii. 87 TFBSs
m. This region contains an expansive SNP cloud, which strengthens the case for the importance of the region in Type 2 Diabetes. Furthermore, several of the genes in the region are associated with metabolism and cholesterol, including FADS1 and FADS2. The location of the lead SNP within a promoter region may impact the transcription of these elements.

14. DGKB-TMEM195 (rs2191349)

a. Credible set interval in hg18: chr7: 14888532-15032137
b. Credible set interval in hg19: chr7:14922007-15065612
c. Contains the 5’ end of DGKB
d. Contains 3 NHGRI SNPs
   i. rs10244051 – metabolic traits
   ii. rs2191349 – fasting glucose related traits
   iii. rs6947830 – metabolic syndrome
e. There are three small H3K4Me1 peaks
   i. 8.7kbp telomeric (p arm) there is a DHSI (104) with HSMM, HepG2, and PanIslets cell lines. There are 16 TFBSs
f. This transethnic region is mostly intergenic, with few biological elements. However, the presence of a SNP cloud indicates the importance of this region to Type 2 diabetes. The

SNP cloud is clustered around a small H3K4Me1 peak, which may be involved in distant regulation.

15. ARL15 (rs4865796)

a. Credible set interval in hg18: chr5: 53059217-53557802
b. Credible set interval in hg19: chr5:53023460-53522045
c. There are two genes in the region
    i. TCONS_l2_00022897 (lincRNA)
    ii. 3' end of ARL15
        1. EST AV660016 indicating expression in liver
d. There are 5 NHGRI SNPs in the region
    i. rs4311394 – adiponectin levels
    ii. rs6450176 – adiponectin levels/cholesterol
    iii. (not related) rs273218 – migraine
    iv. rs702634 – Type 2 diabetes
    v. (not related) rs255758 – rheumatoid arthritis
e. There are 14 H3K4Me1 peaks
        i. There is one small peak located 1.5kbp downstream of the SNP
        ii. 7kbp upstream there is a small peak, associated with a DHSI (42) present in HepG2 Cells and 7 TFBSs
f. There are 3 H3K17Ac peaks, each corresponding to a H3K4Me1 peak
g. There are numerous DHSI and TFBS throughout the region, as well as low-levels of transcription according to RNA-seq
h. The lead SNP is located near a promoter region intronic to ARL15. While this may regulate the gene, which is shown to have tissue specificity to the liver and play a role in glucose levels via adiponectin (PMID: 20011104), it may also impact regulation of the downstream lncRNA TCONS_l2_00022897.

16. PPP1R3B (rs4841132)

This credible set contained a single SNP (rs1461729); annotation overlap for the SNP can be found in Supplemental Table 5.

17. COBLL1-GRB14 (rs7607980)

a. Credible set interval in hg18: chr2: 165214970-165266498
b. Credible set interval in hg19: chr2:165506724-165558252
c. There are two genes in the region
    i. 3' end of COBLL1
        1. EST CB270545 indicates expression in adipose tissue
    ii. TCONS_00004484 (lncRNA)
        1. Increased expression in liver and adrenal tissue.
d. Five NHGRI GWAS SNPs
        i. rs10195252: Triglycerides
        ii. rs13389219: Waist hip ratio
        iii. (not related) rs6717858: sexual dimorphism is anthropometric traits

iv. rs12328675: cholesterol
v. rs7607980: fasting insulin traits

e. Between 165,536,302-165,558,065 (coinciding with COBLL1), there is moderate to high transcription according to RNA-seq
   i. This is especially prominent in HepG2 cells
f. 1.3kbp downstream, there is a DHSI (5/125) present in HepG2 cells, with 25 TFBSs (17 present in liver cells)
g. 3.1kbp upstream, there is a DHSI (32) present in HSMM, PANC-1, and HepG2, associated with 7 TFBSs (all in liver cells)
h. There is a moderate H3K4Me1 peak and small H3K27Ac peak in the region
i. This trans-ethnic region shows several instances of tissue specificity in liver tissue, including information from TFBSs, DHSI, and lncRNA expression profiles. The lead SNP is exonic to COBLL1, which plays a role in cholesterol levels (PubMed: 17903299). This variant may impact the function of this gene.

18. IRS1 (rs2943634)

a. Credible set interval in hg18: chr2: 226735108-226872748
b. Credible set interval in hg19: chr2:227026864-227164504
c. There are three genes in the region
   i. TCONS_l2_00015614 (lncRNA) – especially prominent in thyroid
   ii. TCONS_00003502 (lncRNA)
   iii. TCONS_00004599 (lncRNA)
d. There are seven NHGRI GWAS SNPs
   i. rs2972146 – triglycerides/HDL cholesterol
   ii. rs2943641 – Type 2 Diabetes
   iii. rs2943650 – adiposity
   iv. rs1515110 – adiponectin levels
   v. rs2972146 – triglycerides/HDL cholesterol
   vi. (not related) rs2943636 - sexual dimorphism is anthropometric traits
   vii. rs2943634 – fasting insulin traits
e. There are five H3K4Me1 peaks
   i. 12.7kbp centromeric (q arm) – especially prominent in HSMM cell lines
   ii. 10kbp telomeric – especially prominent in HSMM cell lines
f. There are three H3K27Ac peaks, each coinciding with a H3K4Me1 peak
   i. 12.7kbp centromeric (q arm) – especially prominent in HSMM cell lines
   ii. 10kbp telomeric – especially prominent in HSMM cell lines
g. 12.7kbp centromeric is a DHSI (21) with 9 TFBS (all in liver cells)
h. 10kbp telomeric are two DHSIs with 27 TFBSs
i. This region contains only non-coding SNPs, highlighting their importance in metabolic disorders. The SNP cloud resides in an intergenic area of the region, with relatively few genomic features. However, the genomic features that are present display tissue specificity to Type 2 Diabetes relevant tissues. Further validation could uncover if this area serves as a distant regulator to another genomic area involved in Type 2 Diabetes.

19. GCKR (rs780094)

This credible set contained a single SNP; annotation overlap for the SNP can be found in Supplemental Table 5.

20. ANKRD55-MAP3K1 (rs459193)

a. Credible set interval in hg18: chr5: 55595454-56092481
b. Credible set interval in hg19: chr5:55559697-56056724
c. There are five genes in the region
    i. TCONS_00009667 (lncRNA)
    ii. TCONS_00010339 (lncRNA) – especially prominent in kidney and adrenal
    iii. TCONS_00009669 (lncRNA) – especially prominent in kidney and adrenal
    iv. TCONS_00010343 (lncRNA) – especially prominent in kidney and adrenal
        1. 3' end located less than 1kbp from lead SNP
    v. TCONS_00010346 (lncRNA)
d. There are 11 NHGRI GWAS SNPs
    i. rs9686661 – triglycerides
    ii. (not related) rs11743303 - sexual dimorphism is anthropometric traits
    iii. rs6867983 – waist circumference/triglycerides
    iv. rs30360 – fasting insulin/insulin resistance
    v. (not related) rs456867 – urate levels
    vi. (not related) rs1020388 – Celiac's disease
    vii. (not related) rs889312 – breast cancer
    viii. (not related) rs16886181 – breast cancer
    ix. (not related) rs16886165 – breast cancer
    x. (not related) rs16886034 – breast cancer
    xi. (not related) rs16886113 – breast cancer
e. There are approximately 28 H3K4Me1 peaks
f. The region has 1 H3K4Me3 peak
g. There are seven H3K27Ac peaks, each coinciding with a H3K4Me1 peak
h. 6kbp centromeric (q arm) is an area of high conservation and transcription according to RNA-seq. While there is no gene, histone mod, or TFBS corresponding to this region, there are 41 ESTs
i. 9kb centromeric there is a H3K4Me1 peak. There is a DHSI (41) with 10 TFBSs
    i. The DHSI is present in HSMMtube and PanIsletD cell lines
j. 3.1kbp centromeric is a DHSI (6) present in HepG2 cells. There are 6 TFBSs, all present in liver cell lines
k. 5.5 kbp telomeric is a H3K4Me1 peak. There is a DHSI (48) with 11 TFBSs.
    i. The DHSI is present in HSMM and HepG2 cell lines
l. 15 kbp telomeric is a broad H3K4Me1 peak. This encompasses several DHSI and TFBSs
m. This region contains numerous lncRNA and regulatory elements, further emphasizing their importance. The tissue specificity of these elements in Type 2 Diabetes tissues, such as liver and muscle tissue, indicates that these features may be playing a role in the metabolic disorder. Further validation could be done to identify the specific targets of these areas.

21. FAM13A (rs3822072)

a. Credible set interval in hg18: chr4: 89840095-90083469
b. Credible set interval in hg19: chr4:89621072-89864446
c. Contains two genes
    i. 3' end of HERC3
    ii. FAM13A
        1. The entire region of the gene shows transcription levels particularly prominent in HepG2 cell lines
d. There are two NHGRI GWAS SNPs
    i. (not related) rs2609255 – lung disease

ii. rs3822072 – HDL cholesterol
e. The region has 15 H3K4Me1 peaks
f. There are 2 H3K4Me3 peaks (each coinciding with an Me3 peak)
g. There are 6 H3K27Ac peaks (each coinciding with an Me3 peak)
h. The lead SNP is intronic to FAM13A, is within a DHSI (4) and a TFBS (GATA3)
i. 3.3kbp upstream there are peaks for each form of histone modification.
   i. DHSI (119) and approximately 70 TFBSs
j. 9kbp downstream is a DHSI (17) with 11 TFBSs, all present in liver cells
k. The lead SNP in this region resides in the promoter region of the short isoform FAM13A, and may impact its rate of transcription. The tissue-specificity of the RNA-seq and TFBSs to the liver indicates where this effect may occur.

22. UHRF1BP1 (rs4646949)

a. Credible set interval in hg18: chr6: 34872900-35090036
b. Credible set interval in hg19: chr6:34764922-34982058
c. Contains three genes
   i. 3' end of UHRF1BP1
   ii. TAF11
   iii. 5' end of ANKS1A
d. There are 6 NHGRI GWAS SNPs
   i. (not related) rs3734266 – lupus
   ii. (not related) rs2140418 – alcoholism
   iii. (not related) rs1535001 – lupus
   iv. (not related) rs847845 – lung cancer
   v. (not related) rs12205331 – CAD
   vi. rs4646949 – fasting insulin traits
e. There are 18 H3K4Me1 peaks
f. There is one H3K4Me3 peak
g. There are eight H3K27Ac peaks
h. The lead SNP is inside an area of high transcription according to RNA-seq and in between the short isoform of UHRF1BP1 and TAF11, but intronic to the long isoform of UHRF1BP1
i. 1kbp telomeric (p arm) is a small H3K4Me1 peak
   i. DHSI (1) with 6TFBSs
j. 12kbp telomeric are strong H3K4Me1 and H3K27Ac peaks
   i. DHSI (33) In HSMM and PanIsletD cell lines
   ii. 33 TFBSs, with 32 in blood tissue
k. 10kbp centromeric are strong H3K4Me3 and H3K27Ac peaks, and a moderate H3K3Me1 peak
   i. There are several DHSIs (35,20,125,110)
   ii. Over 100 TFBSs
l. The lead SNP resides near the 3' end of two different genes, which explains its high rate of transcription. However, the small H3K4Me1 peak where the lead SNP resides may be a distant regulator of the downstream target ANKS1A, which has been previously identified in an obesity and Type 2 diabetes study (PMCID: PMC3364960).

23. PPARG (rs17036328)

a. Credible set interval in hg18: chr3: 12311507-12371955
b. Credible set interval in hg19: chr3:12336507-12396955
c. The entire region is intronic to PPARG
d. There are 2 NHGRI GWAS SNPs

    i. (not related) rs11128603 - PAI-1 levels
    ii. rs1801282 – Fasting glucose and Type 2 Diabetes
e. The region has only one strong H3K4Me1 peak, and six other weak peaks
f. There are 3 weak H3K27Ac peaks, each coinciding with a Me1 peak
g. The lead SNP is 3kbp downstream of the strong H3K4Me1 peak and a weak Me3 peak
    i. There is a DHSI (67) found in muscle and pancreatic cell lines
    ii. There are 17 TFBSs
h. 1.7kbp Upstream of a DHSI (51) found in muscle and pancreatic cell lines. There are 27 TFBSs
i. 4.4 kbp upstream of a DHSI (40) found in HepG2 and HSMM cell lines
    i. There are 24 TFBSs
j. While the region is intronic to the long isoform of PPARG, the lead SNP is 2.7kbp upstream of the 5' end of the short isoform. The nearby promotor signals, DHSI, and TFBSs may be involved in the regulation of this isoform. Genetic variations impacting PPARG may increase the risk for Type 2 Diabetes (PubMed: 15797964, PubMed 15592662, PubMed: 12882888).

Acknowledgements

ARIC (DJC, EB, ES, JSP, KEN, ML, NM, THM): We thank the staff and participants of the ARIC Study for their important contributions. The Atherosclerosis Risk in Communities Study is carried out as a collaborative study supported by National Heart, Lung, and Blood Institute contracts HHSN268201100005C, HHSN268201100006C, HHSN268201100007C, HHSN268201100008C, HHSN268201100009C, HHSN268201100010C, HHSN268201100011C, and HHSN268201100012C, and grants R01HL087641, R01HL59367 and R01HL086694; National Human Genome Research Institute contract U01HG004402; and National Institutes of Health contract HHSN268200625226C. KEN is supported by grants R01DK089256 and 13GRNT16490017. Infrastructure was partly supported by Grant Number UL1RR025005, a component of the National Institutes of Health and NIH Roadmap for Medical Research. Dr. Selvin was supported by NIH/NIDDK grants K24DK106414 and 2R01DK089174.

BioMe (EPB, KH, RJFL, YL): The Mount Sinai BioMe Biobank Program is supported by the Andrea and Charles Bronfman Philanthropies.

BLSA (LF, TT): The BLSA was supported in part by the Intramural Research Program of the National Institutes of Health, National Institute on Aging.

CARDIA (DSS, EKK, LJRT, MF): The Coronary Artery Risk Development in Young Adults (CARDIA) Study is funded by contracts from the National Heart, Lung, and Blood Institute to the University of Alabama at Birmingham (HHSN268201300025C and HHSN268201300026C), Northwestern University (HHSN268201300027C), University of Minnesota (HHSN268201300028C), Kaiser Foundation Research Institute (HHSN268201300029C), and The Johns Hopkins University School of Medicine (HHSN268200900041C). CARDIA also receives partial support from the National Institute on Aging Intramural Research Program. The National Human Genome Research Institute supported genotyping of CARDIA participants through grants U01-HG-004729, U01-HG-004446, and U01-HG-004424.

CFS (BEC, SRP): This research was conducted in part using data and resources from the Cleveland Family Study (HL46380, HL113338), Case Western Reserve University (M01 RR00080) and the Broad Institute and supported by the National Heart Lung Blood Institute and the National Institutes of Health. Dr. Cade is supported by HL07901.

CHS (BMP, DSS, JIR, JK, JZ, KR, MOG, MLB, RAJ, YIC): This research was supported by contracts HHSN268201200036C, HHSN268200800007C, HL087562, HL103612, HL105756, HL120393, N01HC55222, N01HC85079, N01HC85080, N01HC85081, N01HC85082, N01HC85083, N01HC85086, and grant U01HL080295 from the National Heart, Lung, and Blood Institute (NHLBI), with additional contribution from the National Institute of Neurological Disorders and Stroke (NINDS). Additional support was provided by R01AG023629 from the National Institute on Aging (NIA). A full list of principal CHS investigators and institutions can be found at CHS-NHLBI.org. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

COGENT (JL, NF, XZ): COGENT investigators are supported by NIH grants 5R21HL123677-02 (NF), HG003054(XZ), and 5T32HL007567(XZ).

Family Heart Study (PA, MAP): This research was supported by NIH grants RO1-HL-087700 and RO1-HL-088215 from NHLBI.

Framingham Heart Study (AL, BP, CTL, DR, JBM, JCF, JD, JH, SR): The authors acknowledge the Framingham Heart Study participants and staff. This work was supported by NIH grants R01DK078616 (AL, BP, CTL, JBM, JCF, JD, SR) and K24DK080140 (AL, BP, SR, JBM). JCF is supported by an MGH Research Scholars Award. AL is supported by a Canadian Diabetes Association postdoctoral research fellowship

Framingham OMNI (ADJ, JDE, YF): This research was conducted in part using data and resources from the Framingham Heart Study of the National Heart Lung and Blood Institute of the National Institutes of Health and Boston University School of Medicine. This work was partially supported by the National Heart, Lung and Blood Institute's Framingham Heart Study (Contract No. N01-HC-25195). A portion of this research utilized the Linux Cluster for Genetic Analysis (LinGA-II) funded by the Robert Dawson Evans Endowment of the Department of Medicine at Boston University School of Medicine and Boston Medical Center. Genotyping of the Framingham Heart Study OMNI cohort was supported by National Heart, Lung and Blood Institute Division of Intramural Research program funds.

GeneSTAR (DMB, DV, LRY): GeneSTAR was supported by NIH grants through the National Heart, Lung, and Blood Institute (HL58625-01A1, HL59684, HL071025-01A1, U01HL72518, and HL087698) and the National Institute of Nursing Research (NR0224103) and by M01-RR000052 to the Johns Hopkins General Clinical Research Center.

GENOA (WZ, LFB, MAJ, JAS, PAP, SLRK): Support for GENOA was provided by the National Heart, Lung and Blood Institute (HL119443, HL054464, HL054457, HL054481, and HL087660) of the National Institutes of Health. Genotyping was performed at the Mayo Clinic by Stephan T. Turner, MD, Mariza de Andrade PhD, Julie Cunningham, PhD. We thank Eric Boerwinkle, PhD and Megan L. Grove from the Human Genetics Center and Institute of Molecular Medicine and Division of Epidemiology, University of Texas Health Science Center, Houston, Texas, USA for their help with genotyping. We would also like to thank the families that participated in the GENOA study.

HANDLS (ABZ, MAN, MFK, MKE, ST): This research was supported by the Intramural Research Program of the NIH, National Institute on Aging and the National Center on Minority Health and Health Disparities (project # Z01-AG000513 and human subjects protocol # 09-AG-N248). Data analyses for the HANDLS study utilized the high-performance computational capabilities of the Biowulf Linux cluster at the National Institutes of Health, Bethesda, Md. (http://biowulf.nih.gov).

Health ABC (DSE, MAN, TBH, YL): This research was supported by NIA contracts N01AG62101, N01AG62103, and N01AG62106. The genome-wide association study was funded by NIA grant 1R01AG032098- 01A1 to Wake Forest University Health Sciences, and genotyping services were provided by the Center for Inherited Disease Research (CIDR). CIDR

is fully funded through a federal contract from the National Institutes of Health to The Johns Hopkins University, contract number HHSN268200782096C. This study utilized the high-performance computational capabilities of the Biowulf Linux cluster at the National Institutes of Health, Bethesda, Md. (http://biowulf.nih.gov). This research was supported in part by the Intramural Research Program of the NIH, National Institute on Aging.

HELIUS (AHZ, CA, MBS): The HELIUS study is conducted by the Academic Medical Center Amsterdam and the Public Health Service of Amsterdam; both organisations provided core support for HELIUS. The HELIUS study is also funded by the Dutch Heart Foundation, the Netherlands Organization for Health Research and Development (ZonMw), and the European Union (FP-7).

HUFS/AADM (AA, ARB, CNR, GC): The contents of this paper are solely the responsibility of the authors and do not necessarily represent the official view of the National Institutes of Health. The study was supported by National Institutes of Health grants S06GM008016-320107 to C. Rotimi and S06GM008016-380111 to A. Adeyemo, both from the NIGMS/MBRS/SCORE Program. Participant enrollment for HUFS was carried out at the Howard University General Clinical Research Center, which is supported by grant 2M01RR010284 from the National Center for Research Resources, a component of the National Institutes of Health. This research was supported in part by the Intramural Research Program of the National Human Genome Research Institute in the Center for Research in Genomics and Global Health (CRGGH—Z01HG200362). CRGGH is also supported by National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK), Center for Information Technology, and the Office of the Director at the National Institutes of Health. Support for participant recruitment and initial genetic studies of the AADM study was provided by NIH grant No. 3T37TW00041-03S2 from the Office of Research on Minority Health. Support for the Africa America Diabetes Mellitus (AADM) study is also provided by the National Institute on Minority Health and Health Disparities, NIDDK and NHGRI.

HyperGEN (DKA, MRI): HyperGEN: Genetics of Left Ventricular Hypertrophy, ancillary to the Family Blood Pressure Program, http://clinicaltrials.gov/ct/show/NCT00005267. Funding sources included National Heart, Lung, and Blood Institute grant R01 HL55673 and cooperative agreements (U01) with the National Heart, Lung, and Blood Institute: HL54471, HL54515 (UT); HL54472, HL54496

IRAS/IRASFS (AJK, CDL, JIR, JTZ, KT, LEW, MG, MPW, NDP, SSR, XG, YDIC): Computing resources for analysis of IRAS and IRASFS data were provided, in part, by the Wake Forest School of Medicine Center for Public Health Genomics. This study and investigators were supported by the following grants/contracts from the National Institutes of Health: DK081350 (Palmer), HL047887 (IRAS), HL047889 (IRAS), HL047890 (IRAS), HL47902 (IRAS), HL060944 (IRASFS), HL061019 (IRASFS), HL060919 (IRASFS).

JHS (AC, JGW, LAL, MS): The Jackson Heart Study is supported by contracts HHSN268201300046C, HHSN268201300047C, HHSN268201300048C, HHSN268201300049C, HHSN268201300050C from the National Heart, Lung, and Blood

Institute and the National Institute on Minority Health and Health Disparities. The authors thank the participants and data collection staff of the Jackson Heart Study.

MEDIA (CNR, DWB, MCYN, SL): We thank all the study participants for their valuable contributions to the parent studies (ARIC, CARDIA, CFS, CHS, FamHS, FIND, GeneSTAR, GENOA, HANDLS, Health ABC, HUFS, JHS, MESA, MESA Family, SIGNET-REGARDS, WFSM and WHI) of the discovery stage of the MEDIA consortium. We thank the contributions of investigators and staff of the parent studies for data collection, genotyping, data analysis and sharing of association results to the MEDIA consortium. The work at the coordinating center (Wake Forest) of the MEDIA consortium was supported by NIH grant R01 DK066358 (DWB).

MESA and MESA Family (JIR, JSP, JY, KT, MG, SSR, XG, YDIC): This research was supported by a research grant and contract NIDDK DK 078616 from the NIH on Common Genetic Variants and Quantitative Diabetic Traits (MAGIC). MESA, MESA Family, and the MESA SHARe project are conducted and supported by the National Heart, Lung, and Blood Institute (NHLBI) in collaboration with MESA investigators. Support is provided by grants and contracts N01-HC-95159, N01-HC-95160, N01-HC-95161, N01-HC-95162, N01-HC-95163, N01-HC-95164, N01-HC-95165, N01-HC-95166, N01-HC-95167, N01-HC-95168, N01-HC-95169 and by grants UL1-TR-000040 and UL1-RR-025005 from NCRR. Funding for MESA SHARe genotyping was provided by NHLBI Contract N02-HL-6-4278. The provision of genotyping data was supported in part by the National Center for Advancing Translational Sciences, CTSI grant UL1TR000124, and the National Institute of Diabetes and Digestive and Kidney Disease Diabetes Research Center (DRC) grant DK063491 to the Southern California Diabetes Endocrinology Research Center.

NUgene (ANK, LLA, JAP, MGH): This work was funded in part by U01HG004609 and U01HG006388 as part of the eMERGE Network. The Northwestern University Enterprise Data Warehouse was funded in part by a grant from the National Center for Research Resources, UL1RR025741.

PARC (JIR, KT, MG, MOG, RMK, XG, XL, YDIC): This research was supported by the National Institutes of Health: grant U19HL069757 from the National Heart, Lung, and Blood Institute; grants R01DK79888 and P30DK063491 from the National Institute of Diabetes and Digestive and Kidney Diseases; and grant UL1TR000124 from the National Center for Advancing Translational Sciences.

REGARDS (MC, WMC): The authors thank the other investigators, the staff, and the participants of the REGARDS study for their valuable contributions. A full list of participating REGARDS investigators and institutions can be found at http://www.regardsstudy.org. This study was supported by a cooperative agreement U01 NS041588 from the National Institute of Neurological Disorders and Stroke (NINDS).

SIGNET (MC, MS, WMC): The Sea Islands Genetics Network (SIGNET) was supported by R01 DK084350 (MM Sale), and consists of data from the REasons for Geographic And Racial Differences in Stroke (REGARDS) cohort, (U01 NS041588; G Howard), Project SuGAR (Sea

Islands Genetic African American Registry) (W.M. Keck Foundation; WT Garvey), a South Carolina Center of Biomedical Research Excellence (COBRE) in Oral Health Project P20 RR017696 (PI: Kirkwood; Sub-award: JK Fernandes), and the Systemic Lupus Erythematosus in Gullah Health (SLEIGH) study (PI: GS Gilkeson; K23 AR052364, DL Kamen; UL1 RR029882, KT Brady), DK084350.

Tobago Health Study (ALP, IM, JMZ): Tobago Health Study was supported, in part, by funding or in-kind services from the Division of Health and Social Services and Tobago House of Assembly, and by grants from the National Institute of Arthritis and Musculoskeletal and Skin Diseases (R01AR049747) and the National Institute of Diabetes and Digestive and Kidney Diseases (K01-DK083029 and R01-DK097084).

Vanderbilt University BioVU and eMERGE (DCC, EKK, SCS, TLE): The dataset used in the analyses described were obtained from Vanderbilt University Medical Center's BioVU which is supported by institutional funding and by the Vanderbilt CTSA grant ULTR000445 from NCATS/NIH. Genome-wide genotyping was funded by NIH grants RC2GM092618 from NIGMS/OD and U01HG004603 from NHGRI/NIGMS. The eMERGE Network was initiated and funded by NHGRI through the following grants: U01HG006378 (Vanderbilt University); and U01HG006385 (Vanderbilt University serving as the Coordinating Center). TLE was also supported by 5R21HL121429.

WHI (JL): We thank the WHI investigators, staff, and study participants for their outstanding dedication and commitment.

Miscellaneous (APM): Andrew P Morris is a Wellcome Trust Senior Fellow in Basic Biomedical Science under award WT098017

Conflict of Interest and Disclaimer

Conflicts of Interest

BMP serves on the DSMB of a clinical trial for the device manufacturer Zoll LifeCor and on the Steering Committee for the Yale Open Data Access Project funded by Johnson & Johnson. DV is a consultant for Consumable Science, Inc. JRK holds the ownership of stock in Pfizer, Inc, and Gilead Sciences, Inc.

Disclaimer

The views expressed in this manuscript are those of the authors and do not necessarily represent the views of the National Heart, Lung, and Blood Institute; the National Institutes of Health; or the U.S. Department of Health and Human Services.

Supplemental Figures

Table of Contents for Supplemental Figures and Legends

Figure S4X. *FOXA2* locus
Figure S4Y. *CRY2* locus
Figure S4Z. *DPYSL5* locus
Figure S4AA. *IGF2BP2* locus
Figure S4BB. *KL* locus

FI loci

Figure 4SX. *ARL15* locus
Figure 4SY. *COBLL1-GRB14* locus
Figure 4SZ. *IRS1* locus
Figure 4SAA. *UHRF1BP1* locus
Figure 4SBB. *FAM13A* locus
Figure 4SCC. *PPARG* locus
Figure 4SDD. *GCKR* locus
Figure 4SEE. *PPP1R3B* locus
Figure 4SFF. *ANKRD55-MAP3K1* locus

Figure S5 Concordance of effect size and Comparison of EA trait-raising allele Frequency in EA and AA

Figures S5A Effect size comparison for EA FG SNPs
Figures S5B Effect size comparison for EA FI SNPs
Figures S5C Allele frequency comparison for EA FG SNPs
Figures S5D Allele frequency comparison for EA FI SNPs

Figure S6 Genome-wide association plots and quantile-quantile (QQ) plots for FG and FI

Figures S6A Miami plots of association with FG
Figures S6B the QQ plots for FG
Figures S6C Miami plots of association with FI
Figures S6D the QQ plots for FI
Figures S6E genome-wide association plots of trans-ethnic meta-analysis results for FG
Figures S6F genome-wide association plots of trans-ethnic meta-analysis results for FI

Figure S7 Conditional analysis at *PELO*/rs6450057

Figures S7A the comparison for unconditional association results with HapMap 2 CEU LD information
Figures S7B the comparison for conditional association results with HapMap 2 CEU LD information
Figures S7C the comparison for unconditional association results with HapMap 2 YRI LD information
Figures S7D the comparison for conditional association results with HapMap 2 YRI LD information

Supplemental Figure Legends

Figure S1.

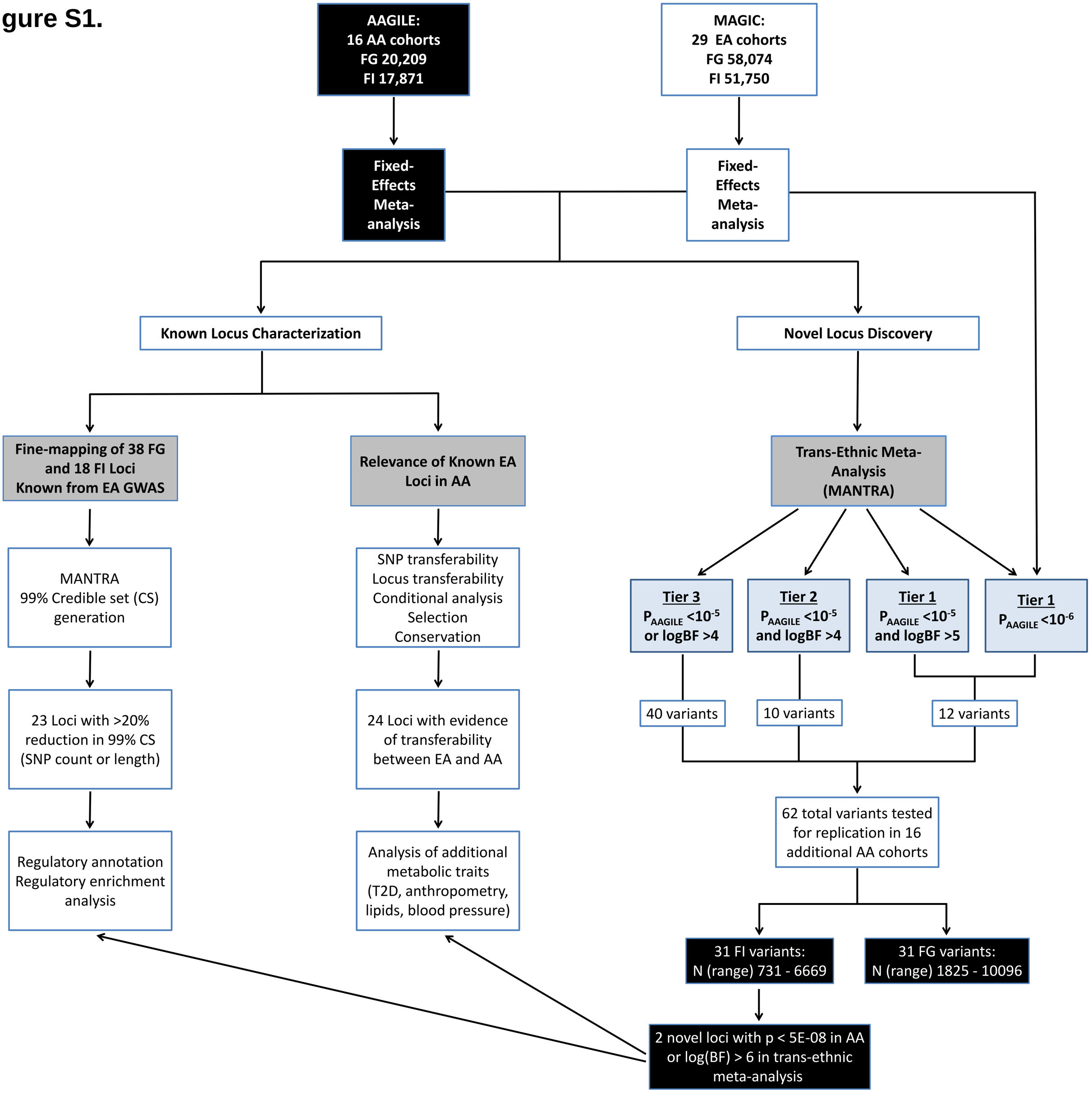

AAGILE:
16 AA cohorts
FG 20,209
FI 17,871
MAGIC:
29 EA cohorts
FG 58,074
FI 51,750
Fixed-Effects Meta-analysis
Fixed-Effects Meta-analysis
Known Locus Characterization
Novel Locus Discovery
Fine-mapping of 38 FG and 18 FI Loci Known from EA GWAS
Relevance of Known EA Loci in AA
Trans-Ethnic Meta-Analysis (MANTRA)
MANTRA
99% Credible set (CS) generation
SNP transferability
Locus transferability
Conditional analysis
Selection
Conservation
Tier 3
P_AAGILE <10^-5 or logBF >4
Tier 2
P_AAGILE <10^-5 and logBF >4
Tier 1
P_AAGILE <10^-5 and logBF >5
Tier 1
P_AAGILE <10^-6
23 Loci with >20% reduction in 99% CS (SNP count or length)
24 Loci with evidence of transferability between EA and AA
40 variants
10 variants
12 variants
Regulatory annotation
Regulatory enrichment analysis
Analysis of additional metabolic traits (T2D, anthropometry, lipids, blood pressure)
62 total variants tested for replication in 16 additional AA cohorts
31 FI variants:
N (range) 731 - 6669
31 FG variants:
N (range) 1825 - 10096
2 novel loci with p < 5E-08 in AA or log(BF) > 6 in trans-ethnic meta-analysis

Figure S2.

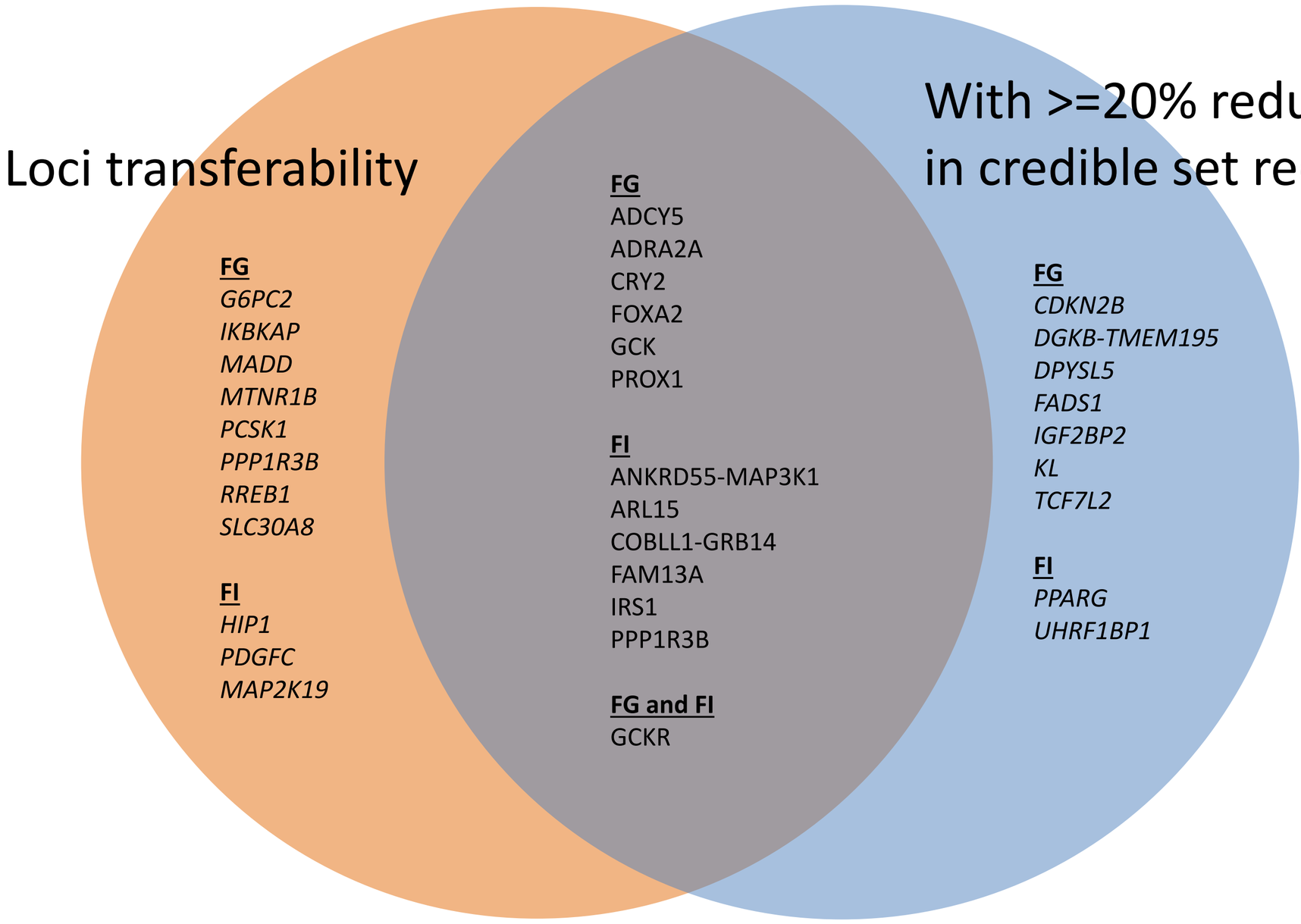
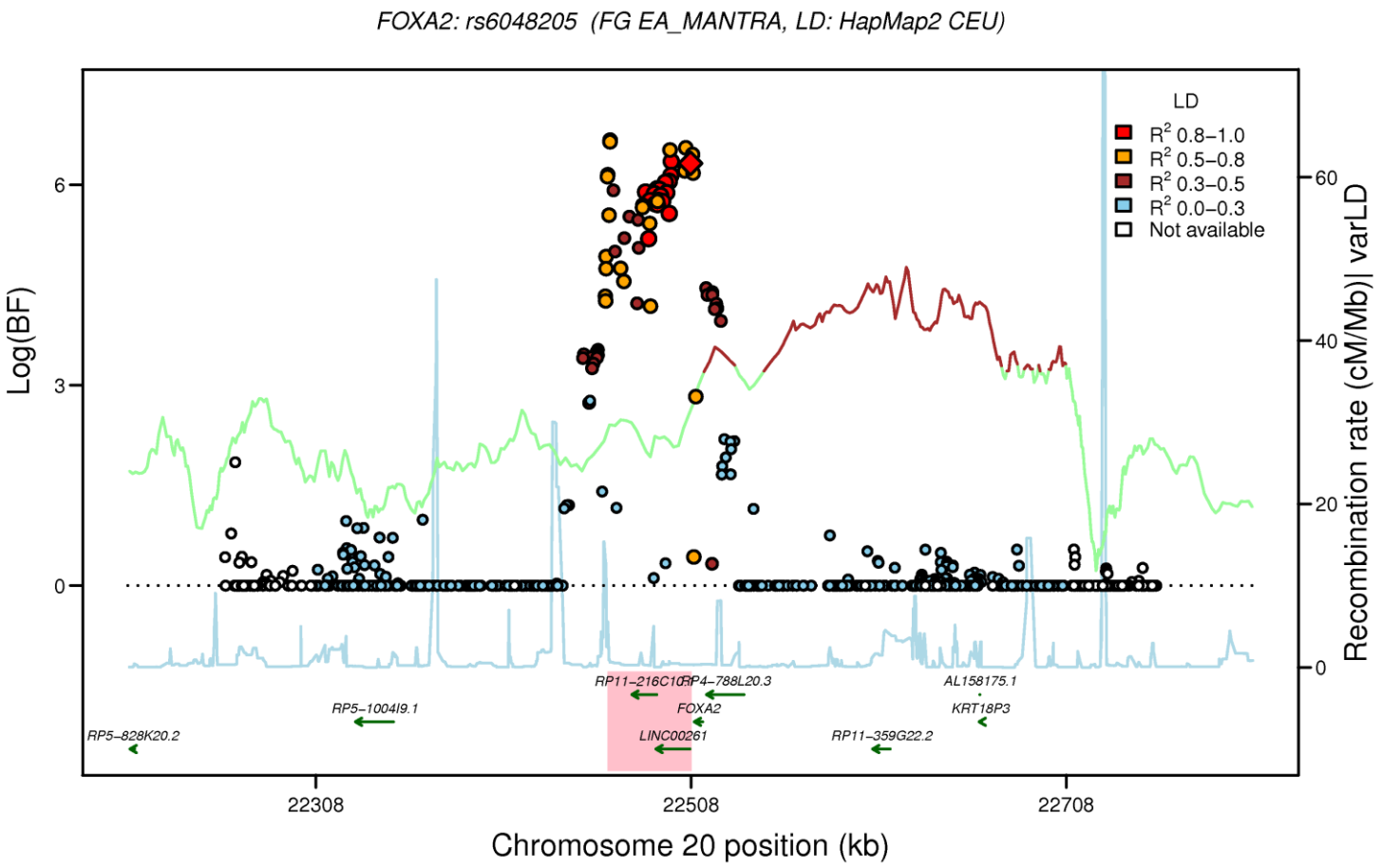
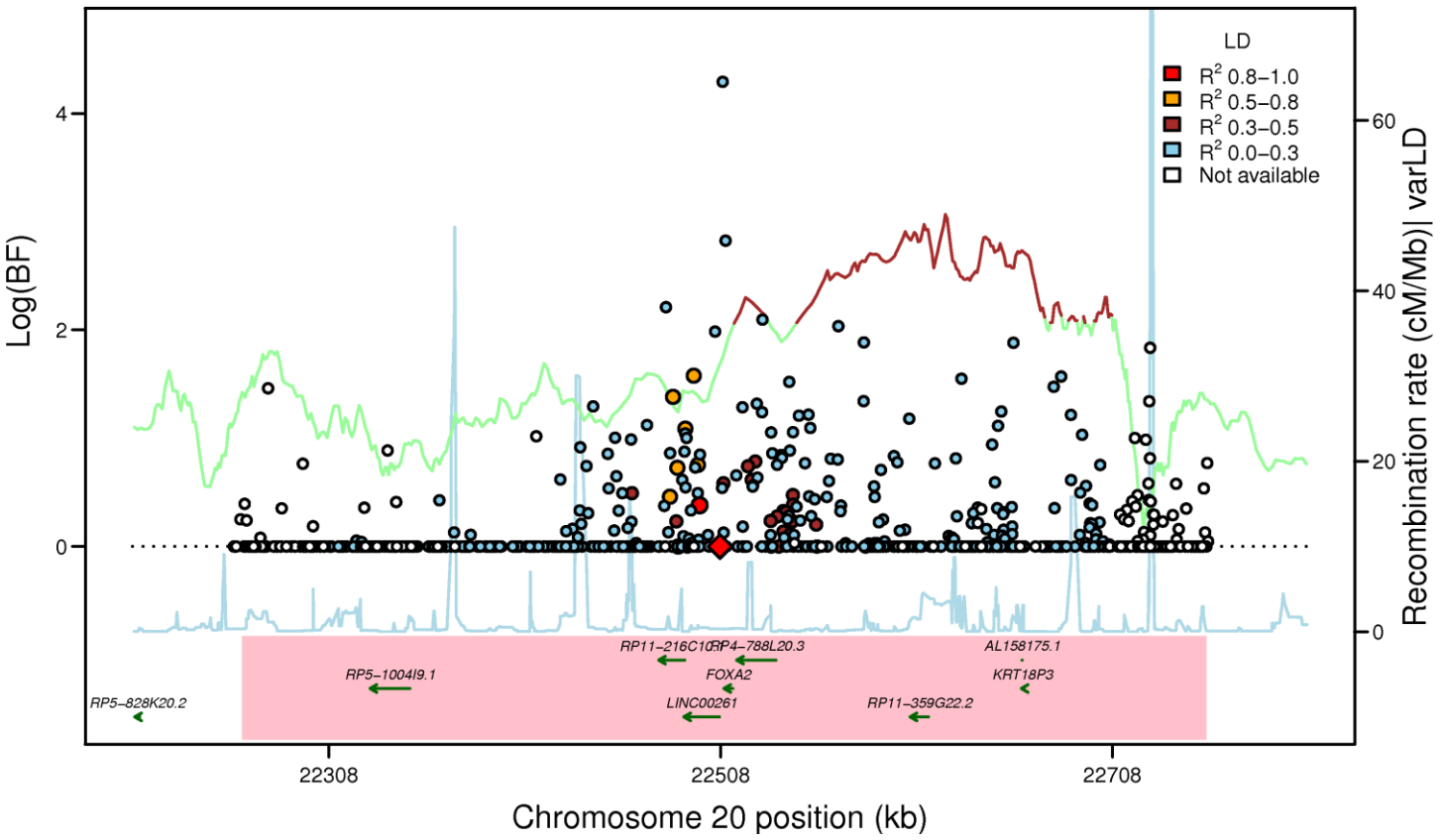

Loci transferability
With >=20% reduction in credible set reduction
FG
G6PC2
IKBKAP
MADD
MTNR1B
PCSK1
PPP1R3B
RREB1
SLC30A8
FI
HIP1
PDGFC
MAP2K19
FG
ADCY5
ADRA2A
CRY2
FOXA2
GCK
PROX1
FI
ANKRD55-MAP3K1
ARL15
COBLL1-GRB14
FAM13A
IRS1
PPP1R3B
FG and FI
GCKR
FG
CDKN2B
DGKB-TMEM195
DPYSL5
FADS1
IGF2BP2
KL
TCF7L2
FI
PPARG
UHRF1BP1

Figure S3A


FOXA2: rs6048205 (FG EA_MANTRA, LD: HapMap2 CEU)
LD
R2 0.8–1.0
R2 0.5–0.8
R2 0.3–0.5
R2 0.0–0.3
Not available
Log(BF)
Recombination rate (cM/Mb)| varLD
RP11–216C10.3
RP4–788L20.3
AL158175.1
RP5–1004I9.1
FOXA2
KRT18P3
RP5–828K20.2
LINC00261
RP11–359G22.2
Chromosome 20 position (kb)


FOXA2: rs6048205 (FG AA_MANTRA, LD: HapMap2 YRI)
LD
R2 0.8–1.0
R2 0.5–0.8
R2 0.3–0.5
R2 0.0–0.3
Not available
Log(BF)
Recombination rate (cM/Mb)| varLD
RP11–216C10.3
RP4–788L20.3
AL158175.1
RP5–1004I9.1
FOXA2
KRT18P3
RP5–828K20.2
LINC00261
RP11–359G22.2
Chromosome 20 position (kb)

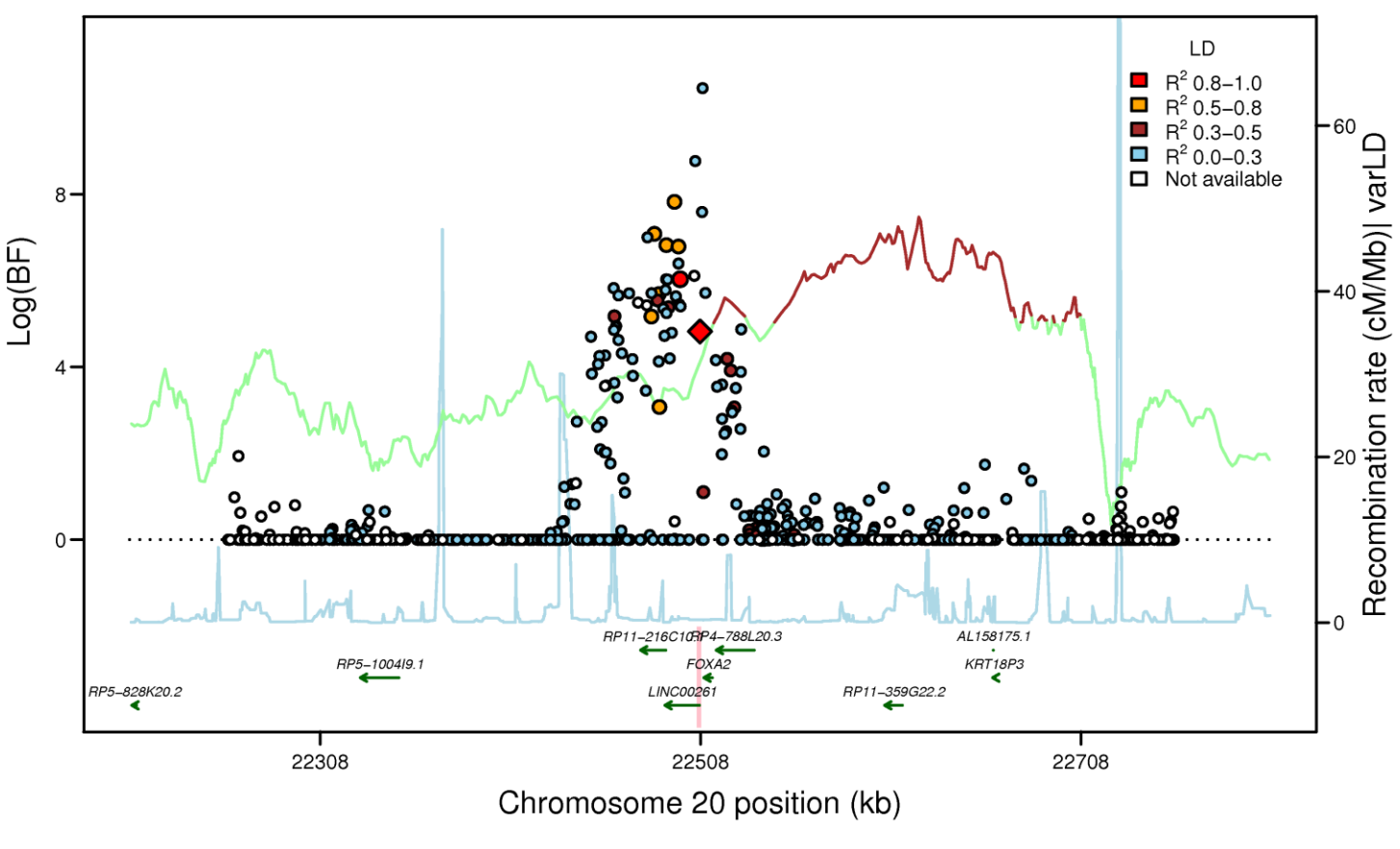
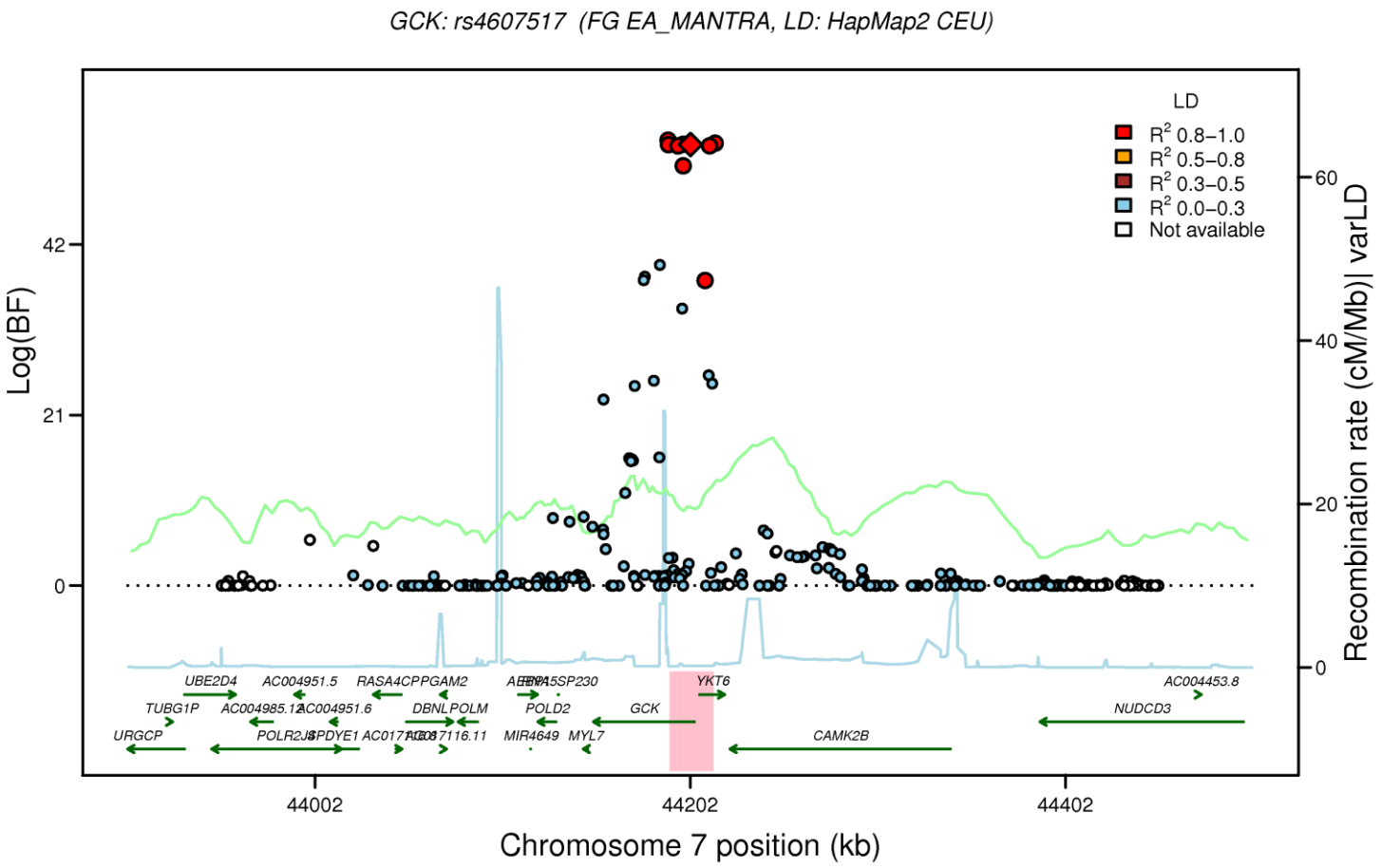
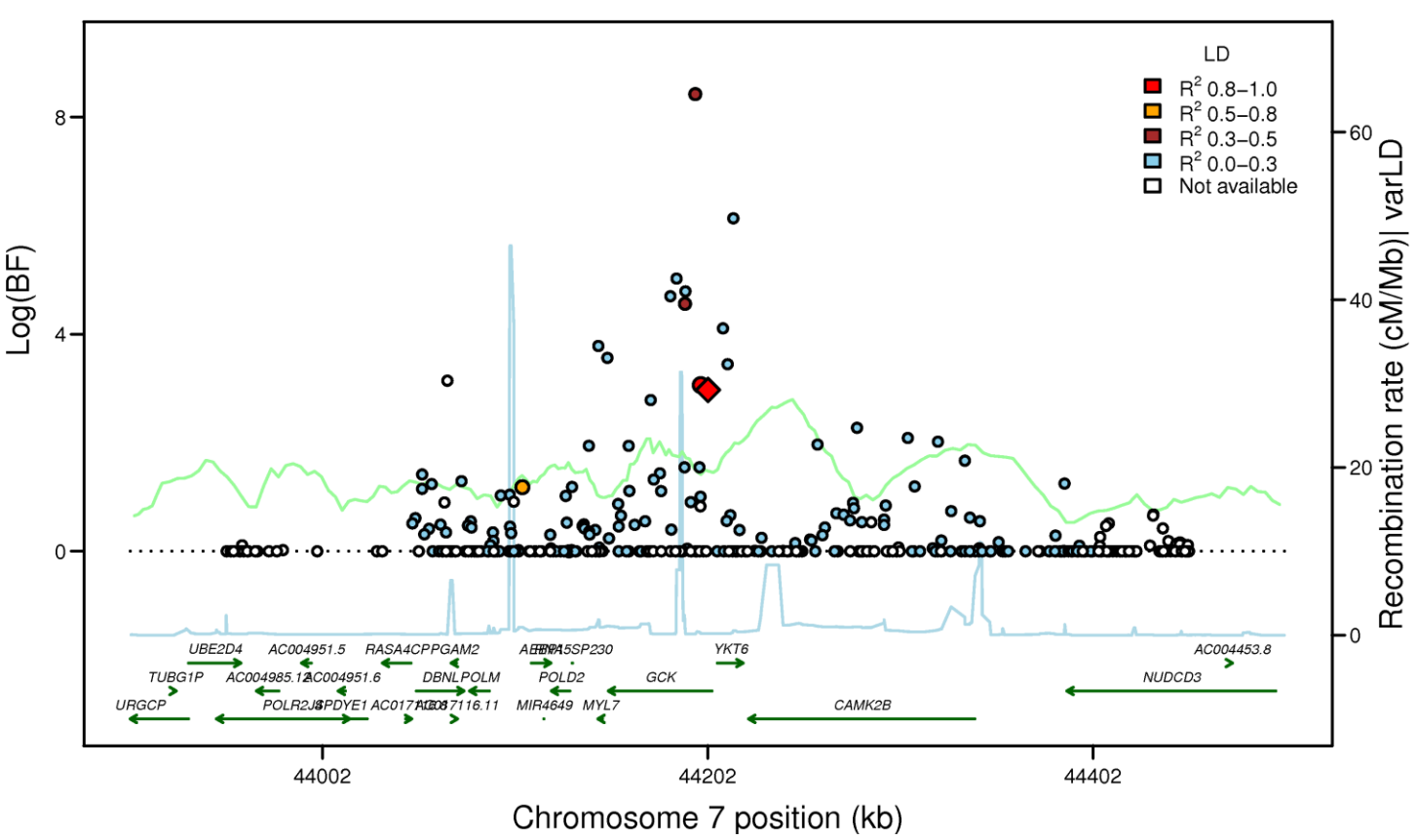

FOXA2: rs6048205 (FG TE_MANTRA, LD: HapMap2 YRI)
LD
R2 0.8–1.0
R2 0.5–0.8
R2 0.3–0.5
R2 0.0–0.3
Not available
Log(BF)
Recombination rate (cM/Mb)| varLD
RP11–216C10.3
RP4–788L20.3
AL158175.1
RP5–1004I9.1
FOXA2
KRT18P3
RP5–828K20.2
LINC00261
RP11–359G22.2
Chromosome 20 position (kb)

Figure S3B


GCK: rs4607517 (FG EA_MANTRA, LD: HapMap2 CEU)
LD
R2 0.8–1.0
R2 0.5–0.8
R2 0.3–0.5
R2 0.0–0.3
Not available
Log(BF)
42
21
0
Recombination rate (cM/Mb)| varLD
60
40
20
0
UBE2D4
AC004951.5
RASA4CP
PGAM2
YKT6
AC004453.8
TUBG1P
AC004985.12
AC004951.6
DBNL
POLM
POLD2
GCK
NUDCD3
URGCP
POLR2J3
SPDYE1
MIR4649
MYL7
CAMK2B
44002
44202
44402
Chromosome 7 position (kb)


GCK: rs4607517 (FG AA_MANTRA, LD: HapMap2 YRI)
LD
R2 0.8–1.0
R2 0.5–0.8
R2 0.3–0.5
R2 0.0–0.3
Not available
Log(BF)
8
4
0
Recombination rate (cM/Mb)| varLD
60
40
20
0
UBE2D4
AC004951.5
RASA4CP
PGAM2
YKT6
AC004453.8
TUBG1P
AC004985.12
AC004951.6
DBNL
POLM
POLD2
GCK
NUDCD3
URGCP
POLR2J3
SPDYE1
MIR4649
MYL7
CAMK2B
44002
44202
44402
Chromosome 7 position (kb)

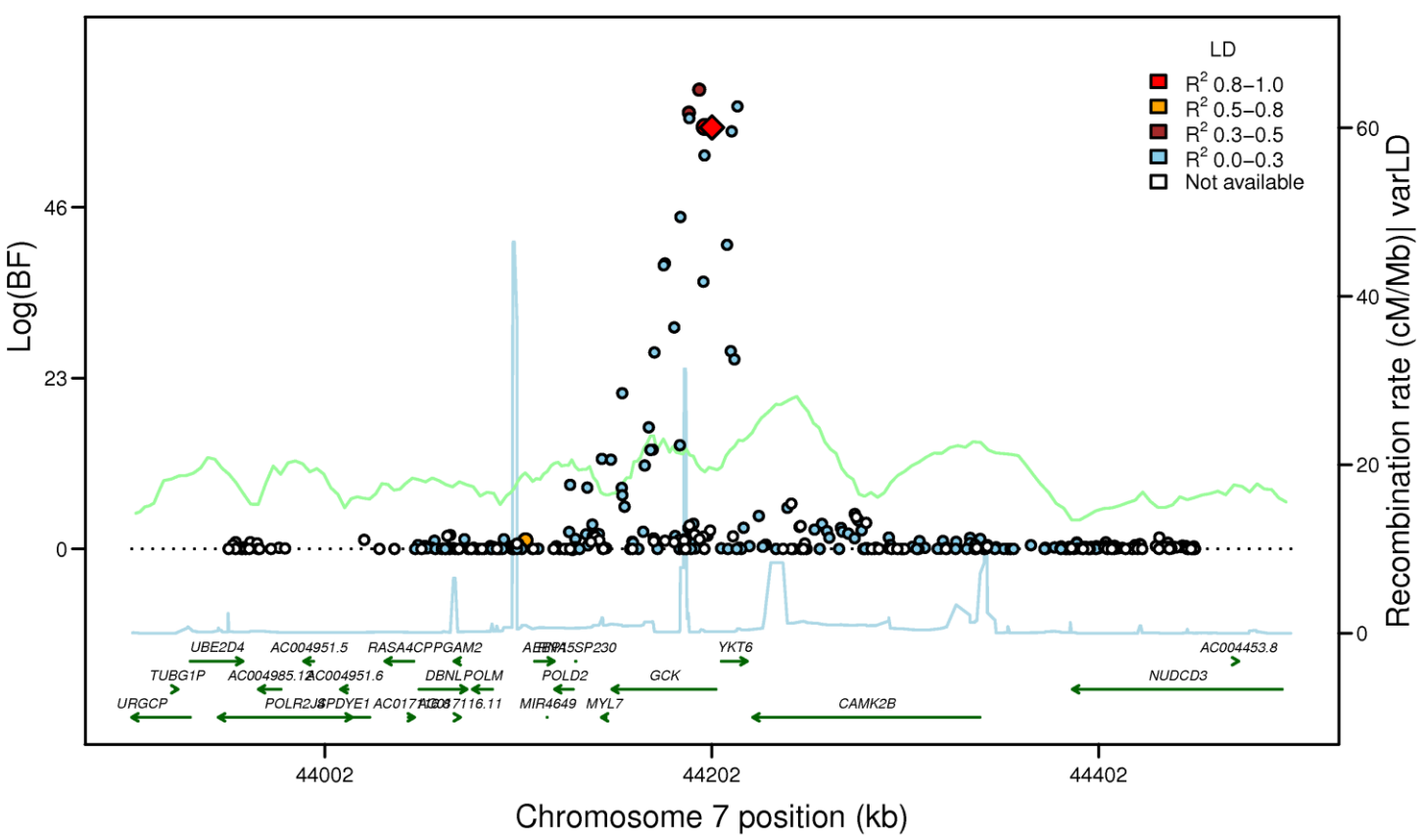
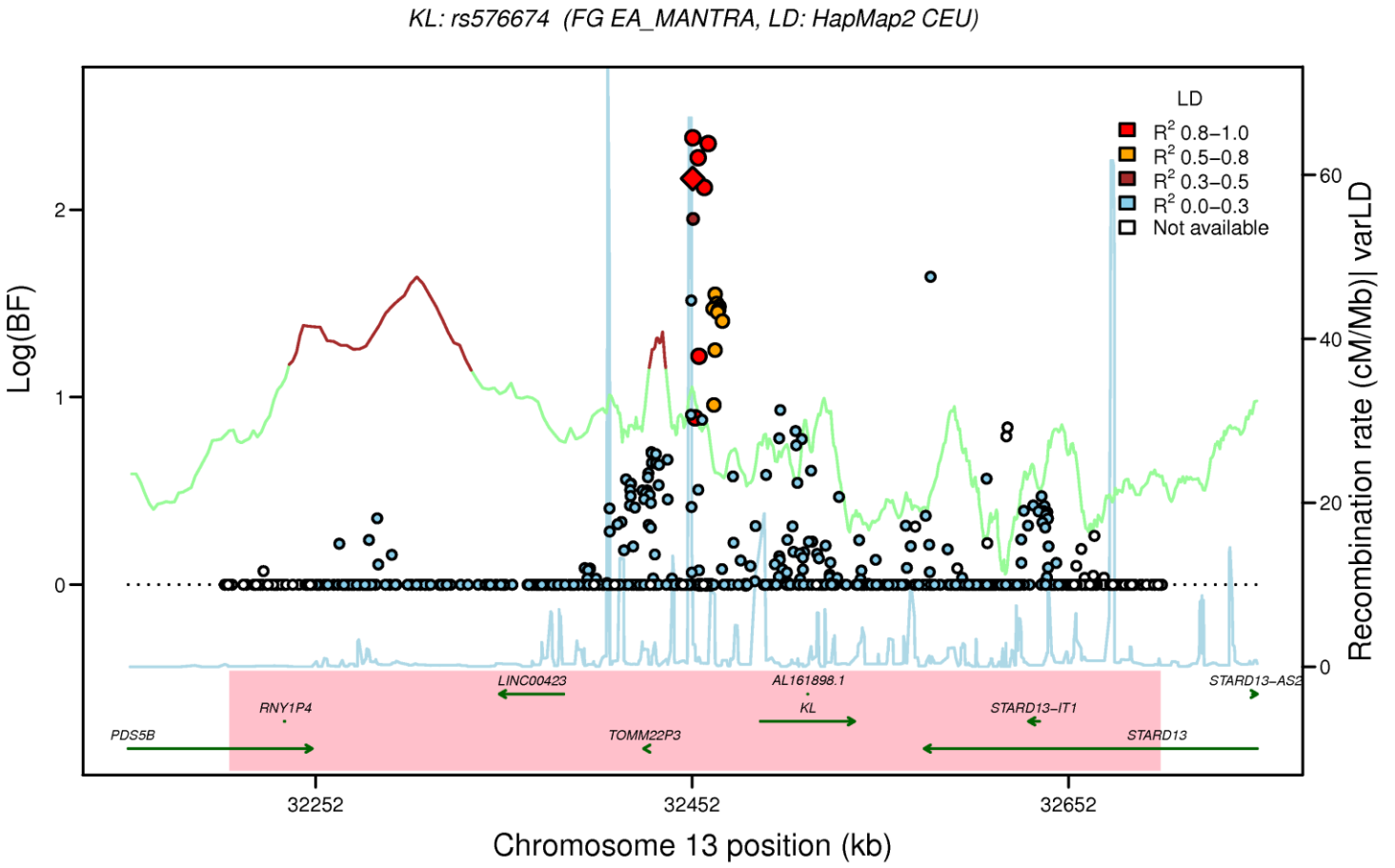
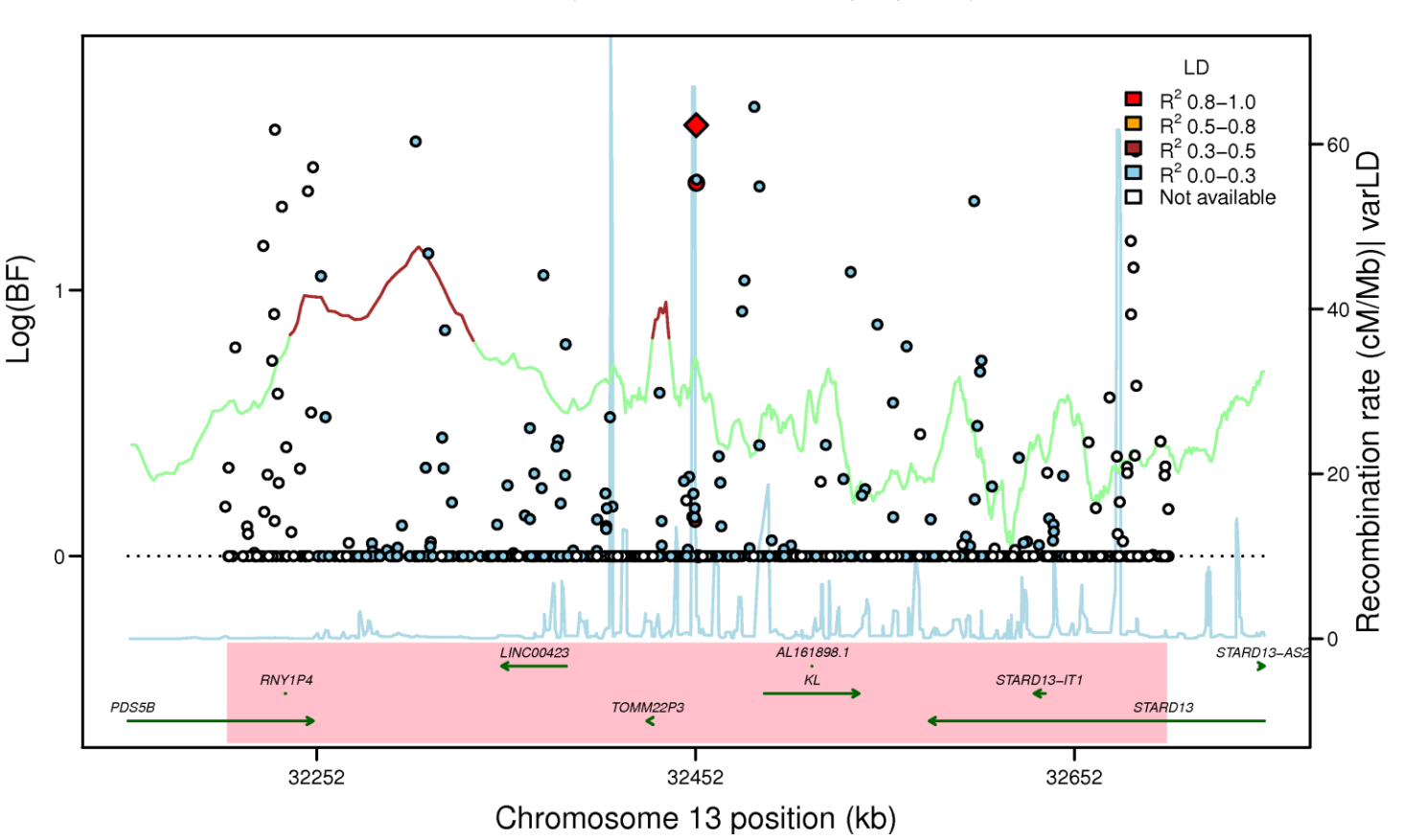

GCK: rs4607517 (FG TE_MANTRA, LD: HapMap2 YRI)
LD
R2 0.8–1.0
R2 0.5–0.8
R2 0.3–0.5
R2 0.0–0.3
Not available
Log(BF)
46
23
0
Recombination rate (cM/Mb)| varLD
60
40
20
0
UBE2D4
AC004951.5
RASA4CP
PGAM2
YKT6
AC004453.8
TUBG1P
AC004985.12
AC004951.6
DBNL
POLM
POLD2
GCK
NUDCD3
URGCP
POLR2J3
SPDYE1
MIR4649
MYL7
CAMK2B
44002
44202
44402
Chromosome 7 position (kb)

Figure S3C


KL: rs576674 (FG EA_MANTRA, LD: HapMap2 CEU)
LD
R2 0.8–1.0
R2 0.5–0.8
R2 0.3–0.5
R2 0.0–0.3
Not available
Log(BF)
Recombination rate (cM/Mb)| varLD
LINC00423
AL161898.1
STARD13–AS2
RNY1P4
KL
STARD13–IT1
PDS5B
TOMM22P3
STARD13
32252
32452
32652
Chromosome 13 position (kb)


KL: rs576674 (FG AA_MANTRA, LD: HapMap2 YRI)
LD
R2 0.8–1.0
R2 0.5–0.8
R2 0.3–0.5
R2 0.0–0.3
Not available
Log(BF)
Recombination rate (cM/Mb)| varLD
LINC00423
AL161898.1
STARD13–AS2
RNY1P4
KL
STARD13–IT1
PDS5B
TOMM22P3
STARD13
32252
32452
32652
Chromosome 13 position (kb)

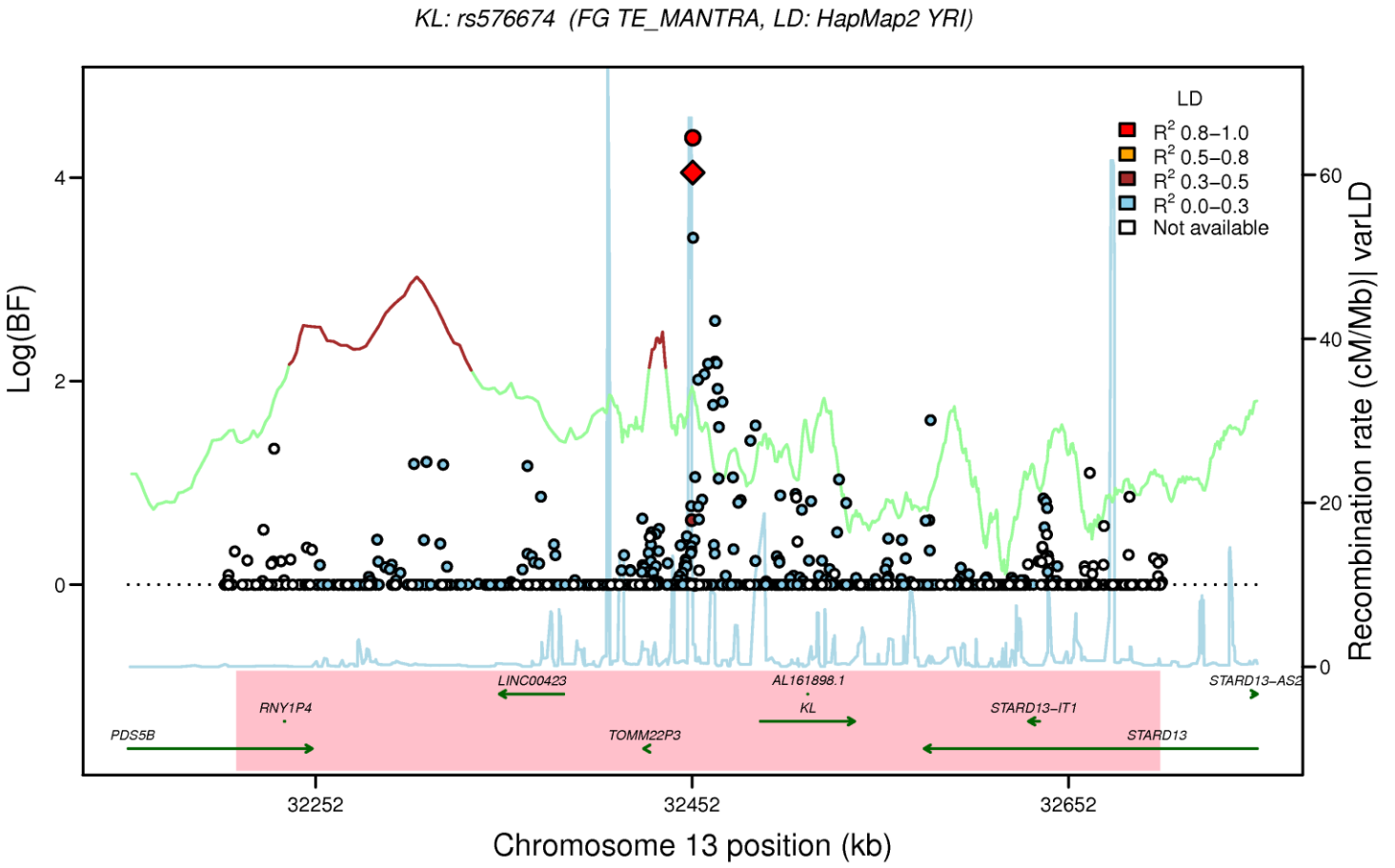
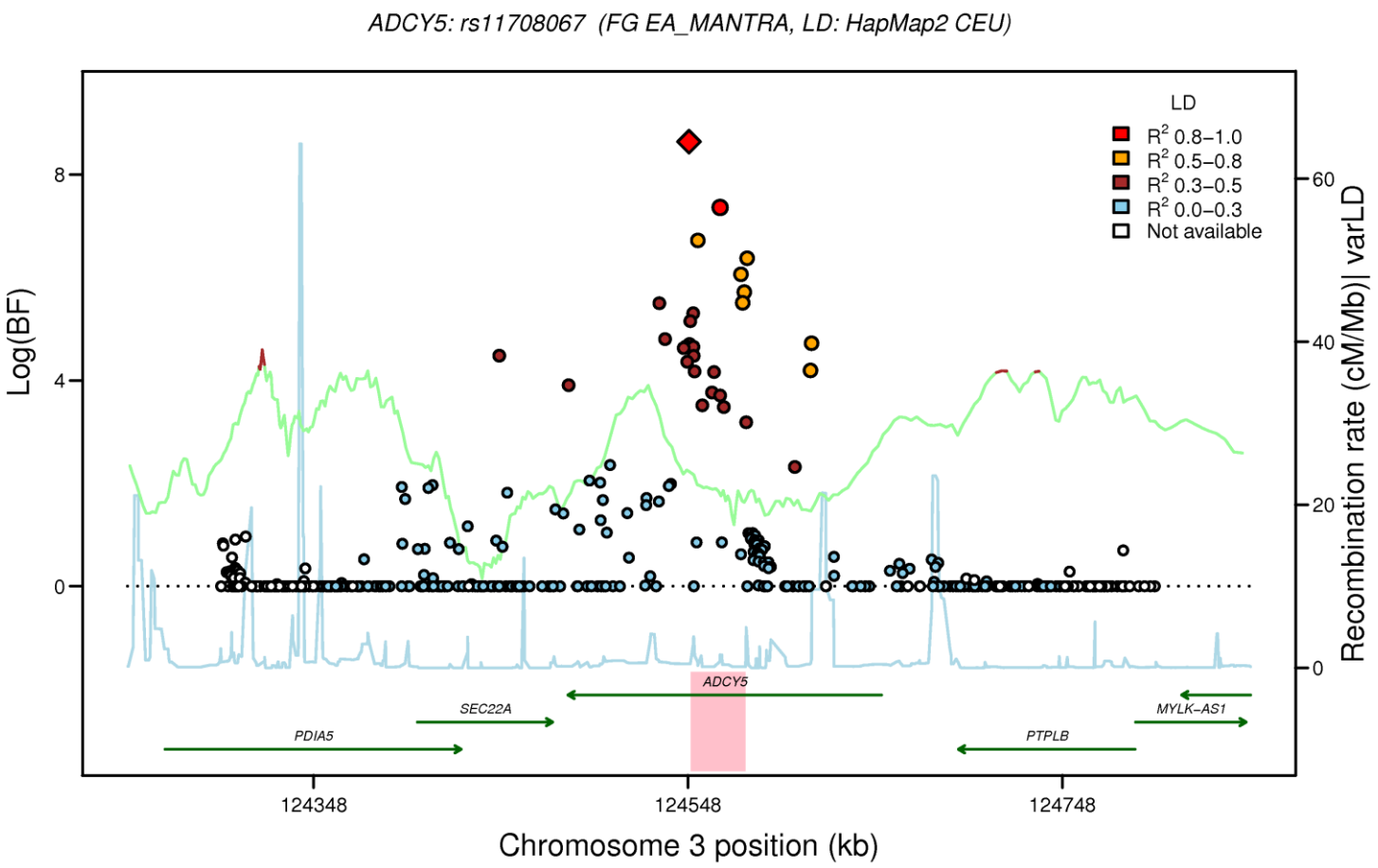
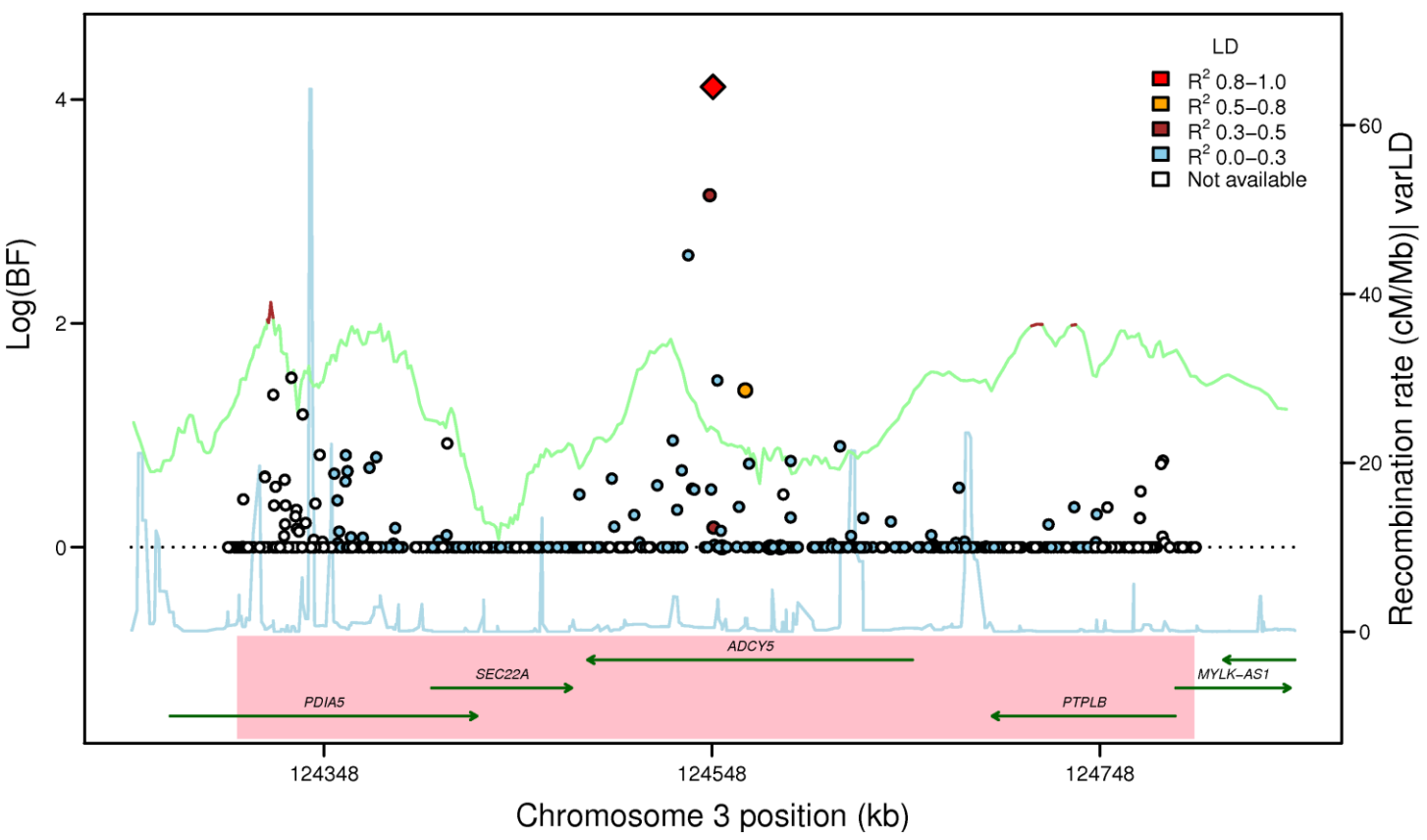

KL: rs576674 (FG TE_MANTRA, LD: HapMap2 YRI)
LD
R2 0.8–1.0
R2 0.5–0.8
R2 0.3–0.5
R2 0.0–0.3
Not available
Log(BF)
Recombination rate (cM/Mb)| varLD
LINC00423
AL161898.1
STARD13–AS2
RNY1P4
KL
STARD13–IT1
PDS5B
TOMM22P3
STARD13
32252
32452
32652
Chromosome 13 position (kb)

Figure S3D


ADCY5: rs11708067 (FG EA_MANTRA, LD: HapMap2 CEU)
LD
R2 0.8–1.0
R2 0.5–0.8
R2 0.3–0.5
R2 0.0–0.3
Not available
Log(BF)
Recombination rate (cM/Mb)| varLD
ADCY5
SEC22A
PDIA5
PTPLB
MYLK–AS1
124348
124548
124748
Chromosome 3 position (kb)


ADCY5: rs11708067 (FG AA_MANTRA, LD: HapMap2 YRI)
LD
R2 0.8–1.0
R2 0.5–0.8
R2 0.3–0.5
R2 0.0–0.3
Not available
Log(BF)
Recombination rate (cM/Mb)| varLD
ADCY5
SEC22A
PDIA5
PTPLB
MYLK–AS1
124348
124548
124748
Chromosome 3 position (kb)

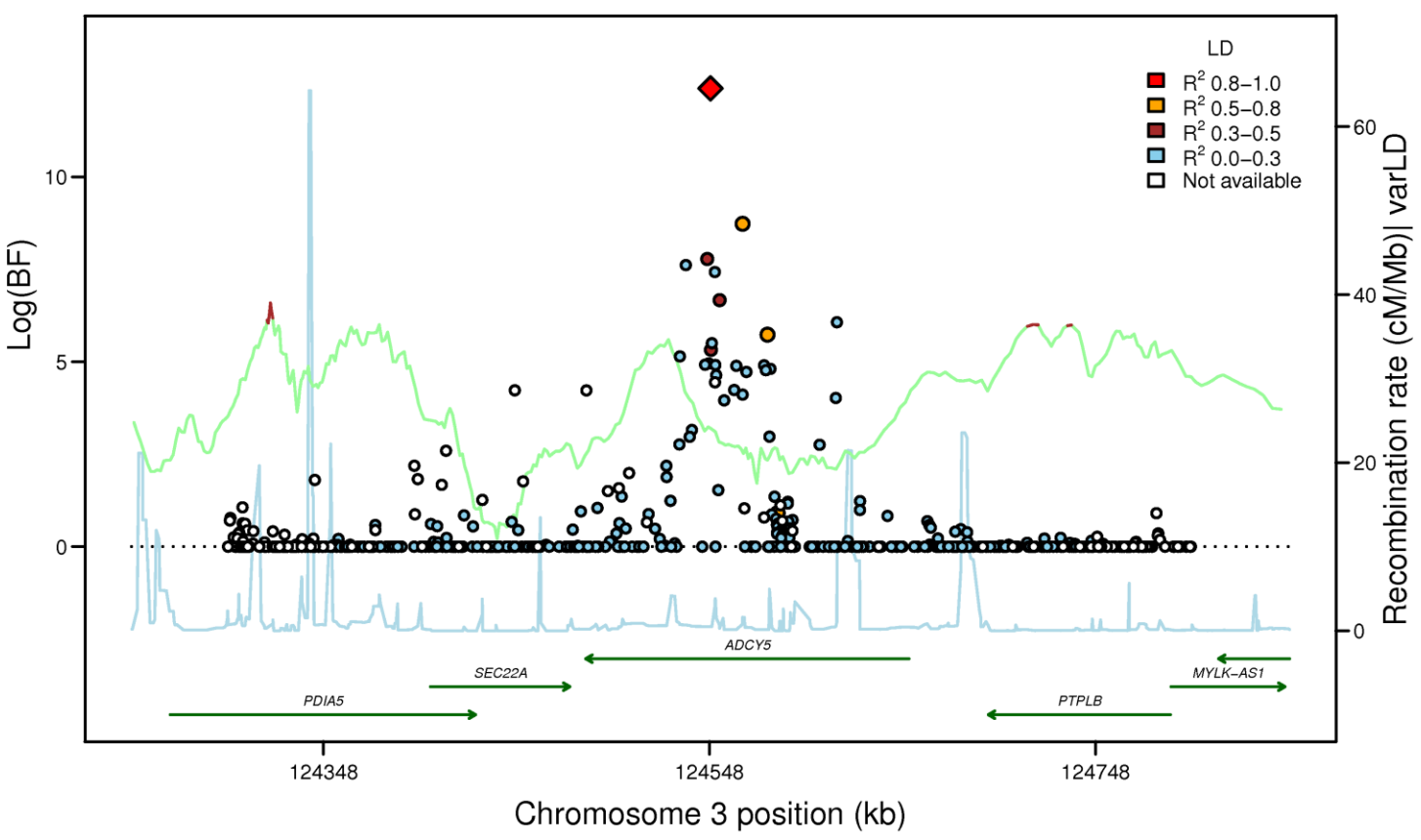
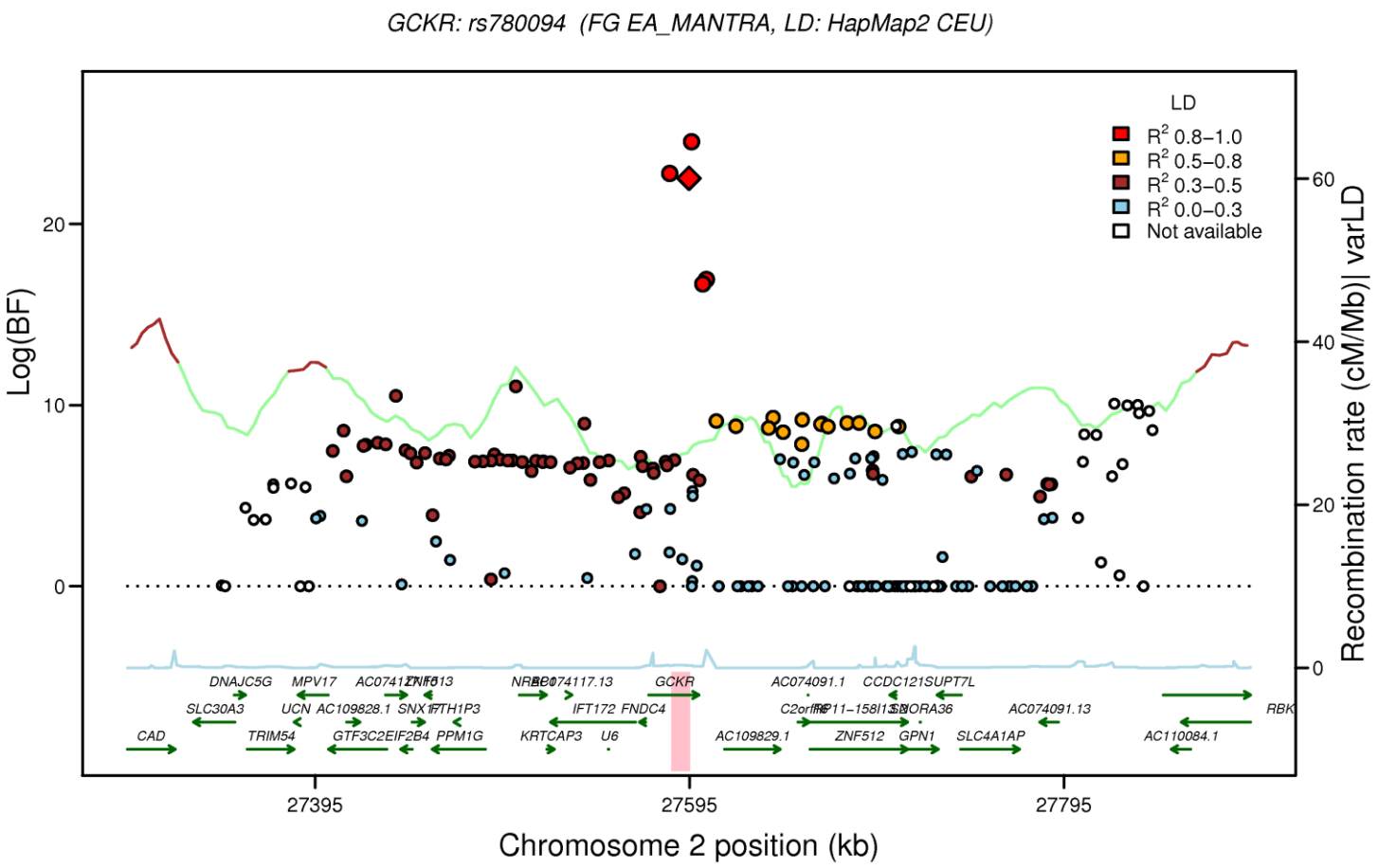
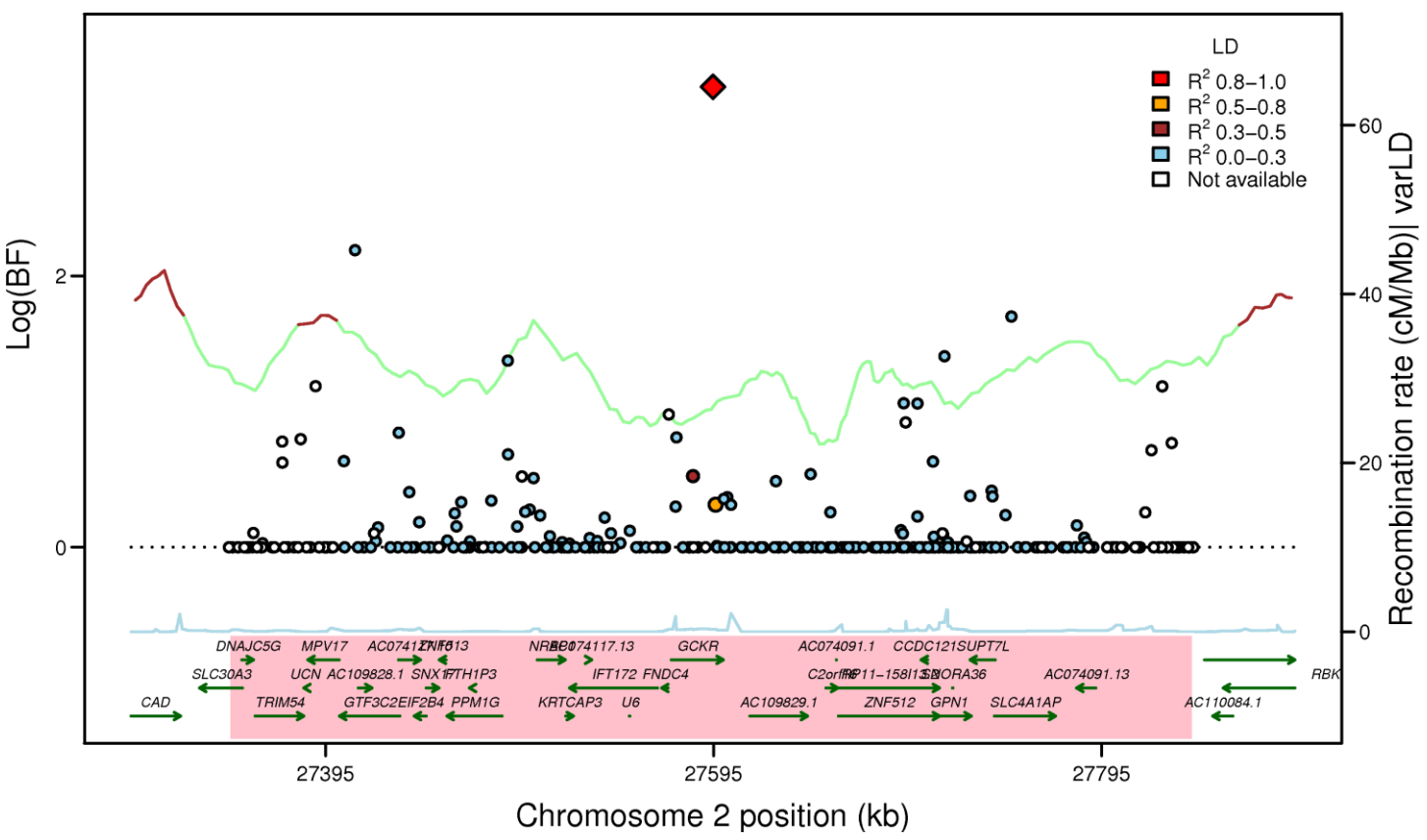

ADCY5: rs11708067 (FG TE_MANTRA, LD: HapMap2 YRI)
LD
R2 0.8–1.0
R2 0.5–0.8
R2 0.3–0.5
R2 0.0–0.3
Not available
Log(BF)
Recombination rate (cM/Mb)| varLD
ADCY5
SEC22A
PDIA5
PTPLB
MYLK–AS1
124348
124548
124748
Chromosome 3 position (kb)

Figure S3E


GCKR: rs780094 (FG EA_MANTRA, LD: HapMap2 CEU)
LD
R2 0.8–1.0
R2 0.5–0.8
R2 0.3–0.5
R2 0.0–0.3
Not available
Log(BF)
Recombination rate (cM/Mb)| varLD
Chromosome 2 position (kb)
27395
27595
27795
DNAJC5G MPV17 AC074117.13 GCKR AC074091.1 CCDC121SUPT7L
SLC30A3 UCN AC109828.1 IFT172 FNDC4 AC074091.13
CAD TRIM54 GTF3C2EIF2B4 PPM1G KRTCAP3 U6 AC109829.1 ZNF512 GPN1 SLC4A1AP AC110084.1


GCKR: rs780094 (FG AA_MANTRA, LD: HapMap2 YRI)
LD
R2 0.8–1.0
R2 0.5–0.8
R2 0.3–0.5
R2 0.0–0.3
Not available
Log(BF)
Recombination rate (cM/Mb)| varLD
Chromosome 2 position (kb)
27395
27595
27795
DNAJC5G MPV17 AC074117.13 GCKR AC074091.1 CCDC121SUPT7L
SLC30A3 UCN AC109828.1 IFT172 FNDC4 AC074091.13
CAD TRIM54 GTF3C2EIF2B4 PPM1G KRTCAP3 U6 AC109829.1 ZNF512 GPN1 SLC4A1AP AC110084.1

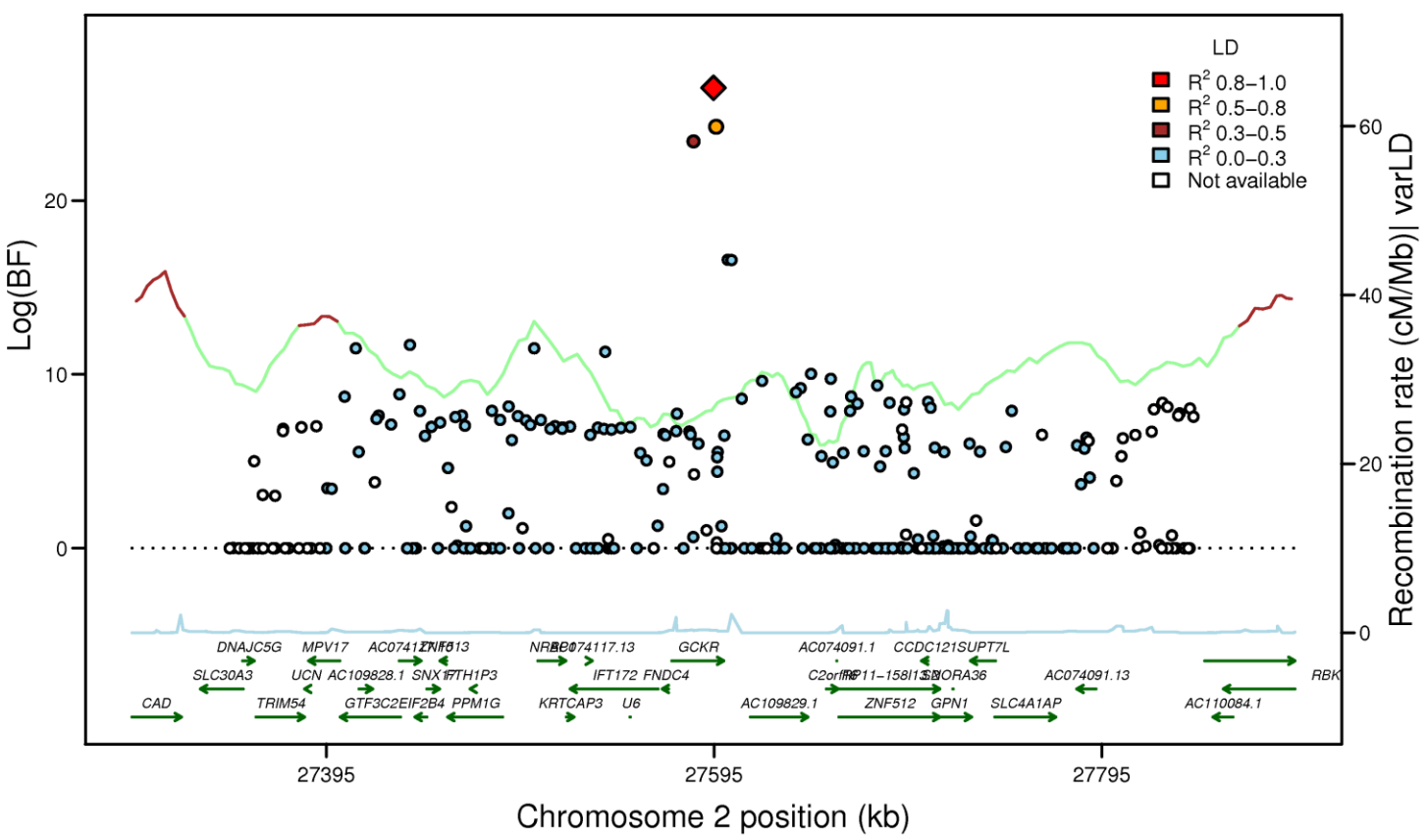
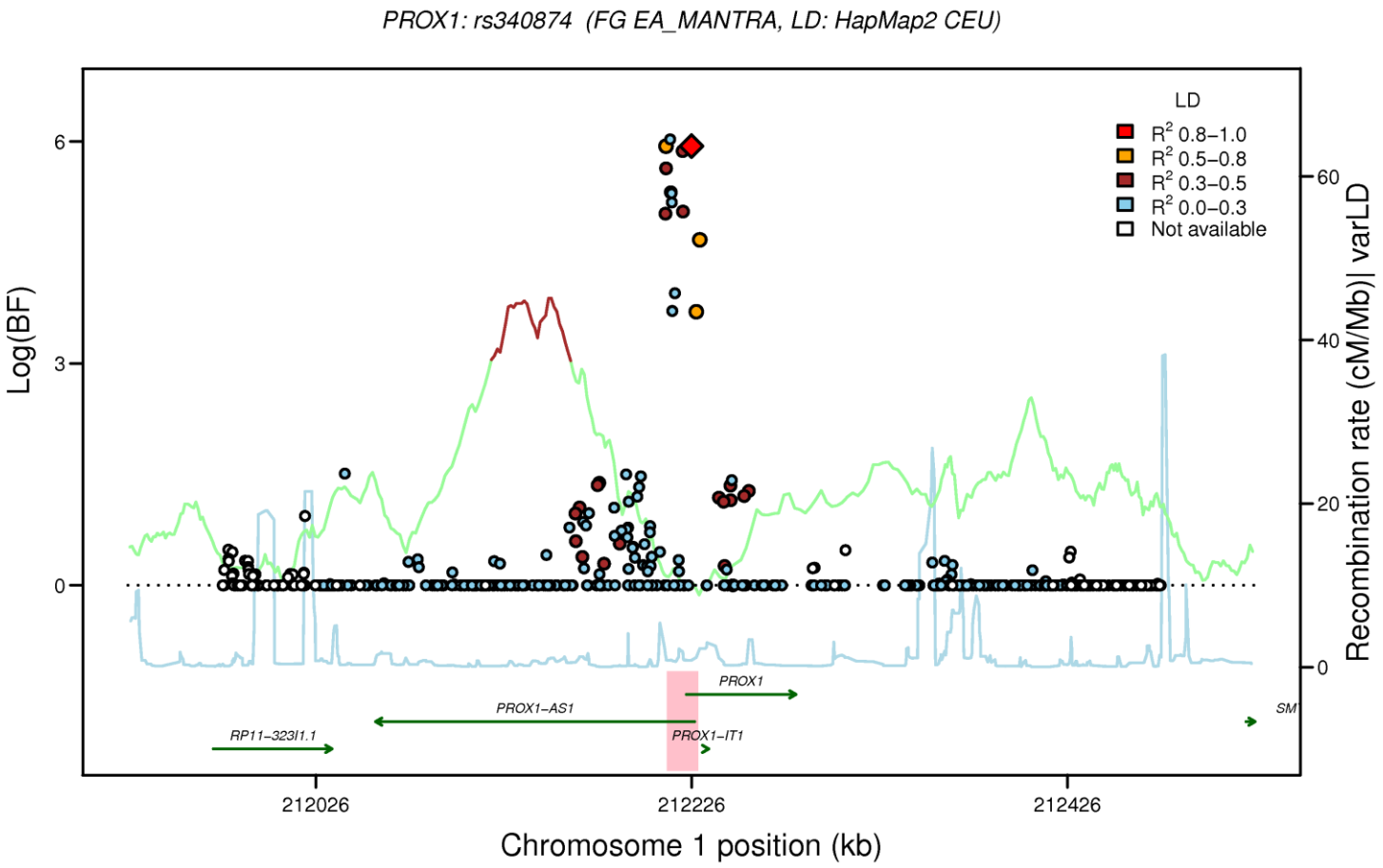
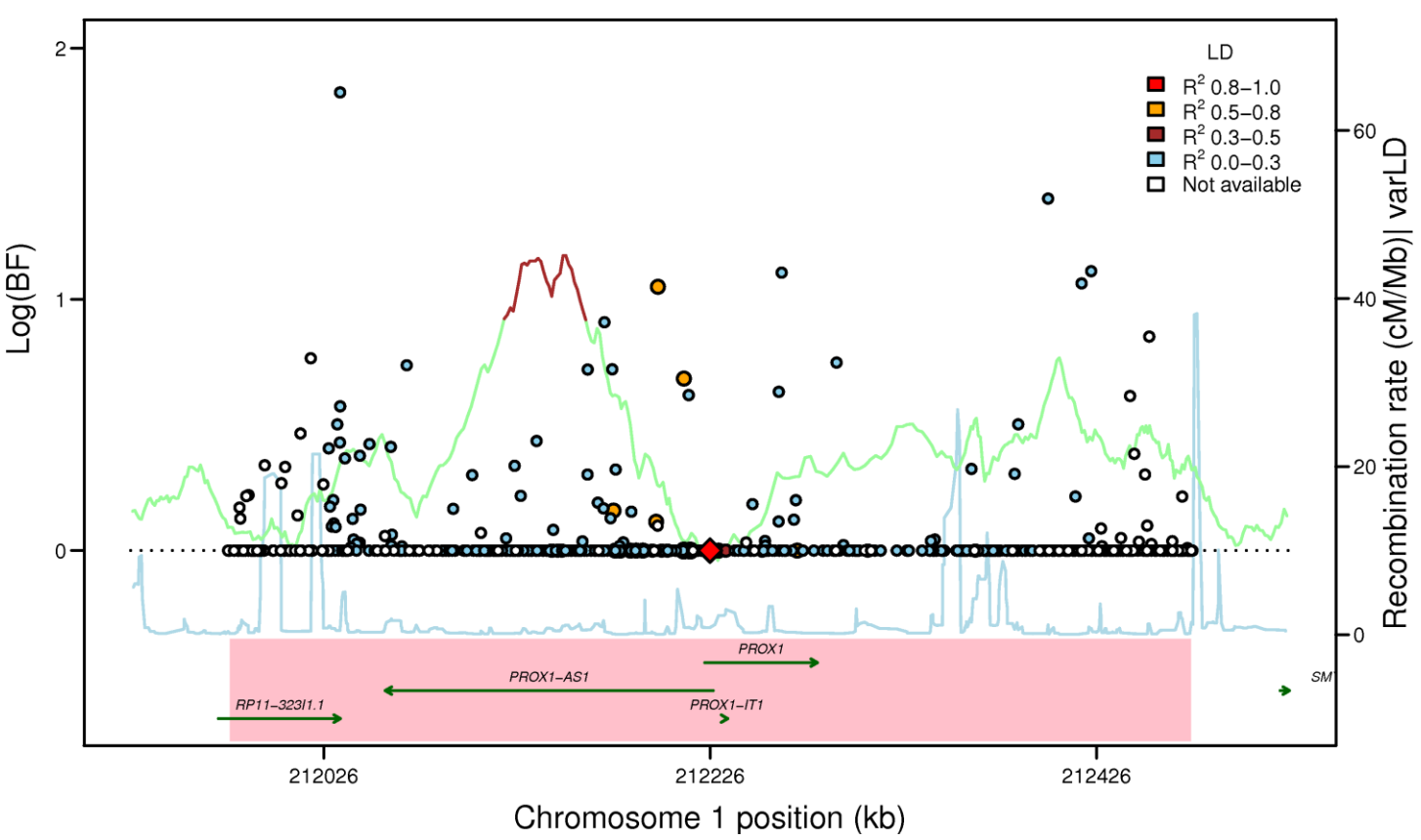

GCKR: rs780094 (FG TE_MANTRA, LD: HapMap2 YRI)
LD
R2 0.8–1.0
R2 0.5–0.8
R2 0.3–0.5
R2 0.0–0.3
Not available
Log(BF)
Recombination rate (cM/Mb)| varLD
Chromosome 2 position (kb)
27395
27595
27795
DNAJC5G MPV17 AC074117.13 GCKR AC074091.1 CCDC121SUPT7L
SLC30A3 UCN AC109828.1 IFT172 FNDC4 AC074091.13
CAD TRIM54 GTF3C2EIF2B4 PPM1G KRTCAP3 U6 AC109829.1 ZNF512 GPN1 SLC4A1AP AC110084.1

Figure S3F


PROX1: rs340874 (FG EA_MANTRA, LD: HapMap2 CEU)
LD
R2 0.8–1.0
R2 0.5–0.8
R2 0.3–0.5
R2 0.0–0.3
Not available
Log(BF)
Recombination rate (cM/Mb)| varLD
PROX1
PROX1–AS1
RP11–323I1.1
PROX1–IT1
212026
212226
212426
Chromosome 1 position (kb)


PROX1: rs340874 (FG AA_MANTRA, LD: HapMap2 YRI)
LD
R2 0.8–1.0
R2 0.5–0.8
R2 0.3–0.5
R2 0.0–0.3
Not available
Log(BF)
Recombination rate (cM/Mb)| varLD
PROX1
PROX1–AS1
RP11–323I1.1
PROX1–IT1
212026
212226
212426
Chromosome 1 position (kb)

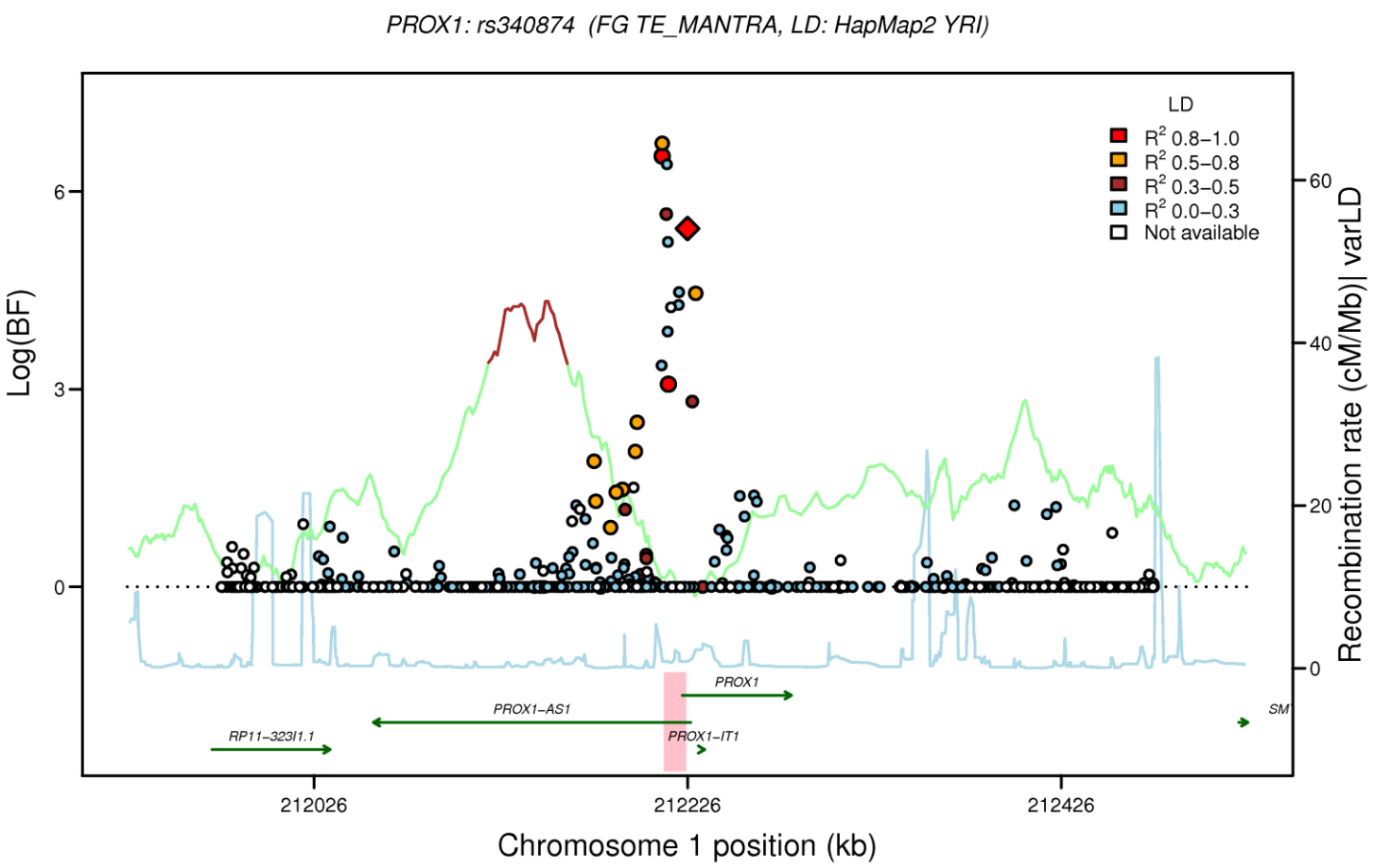
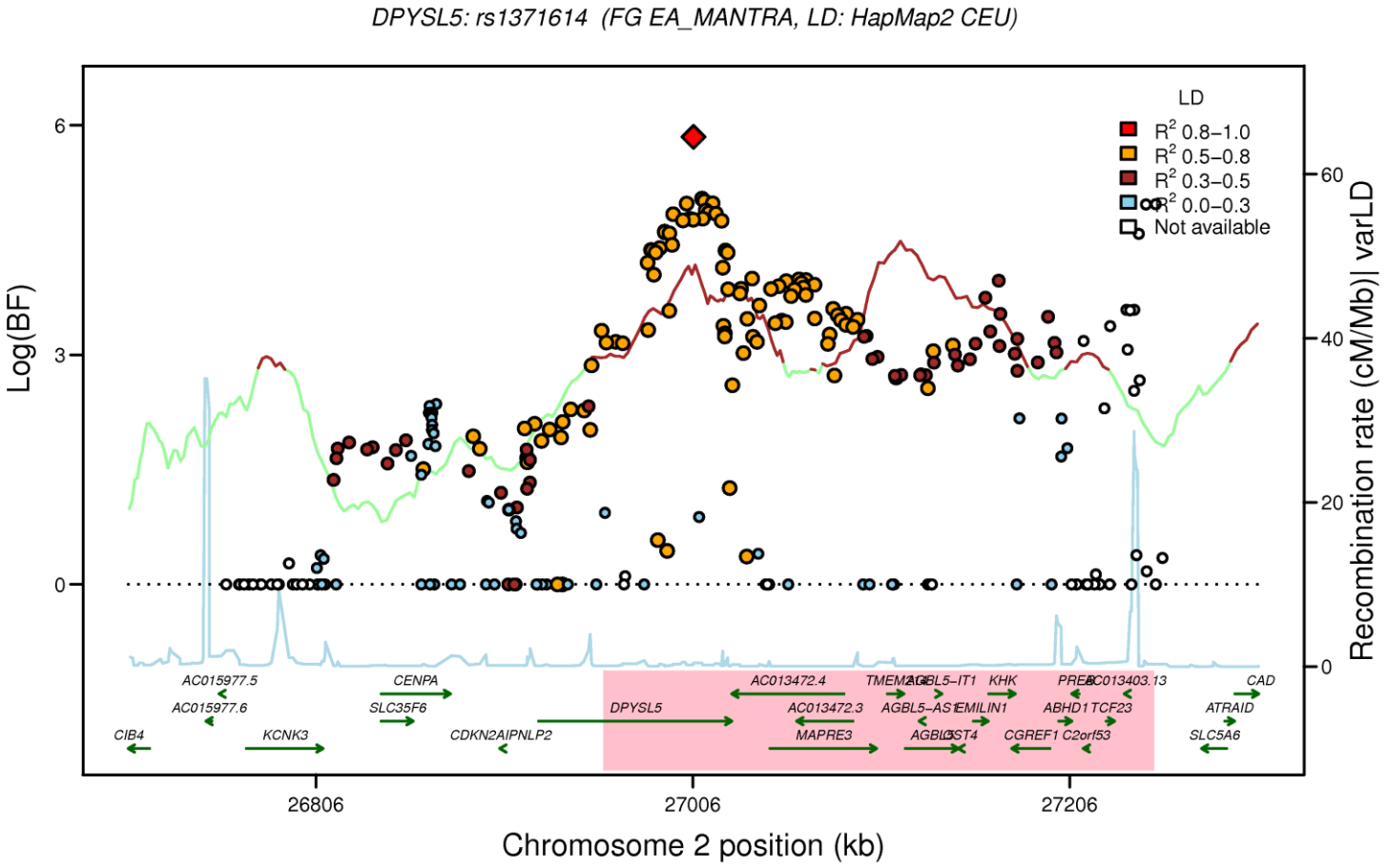
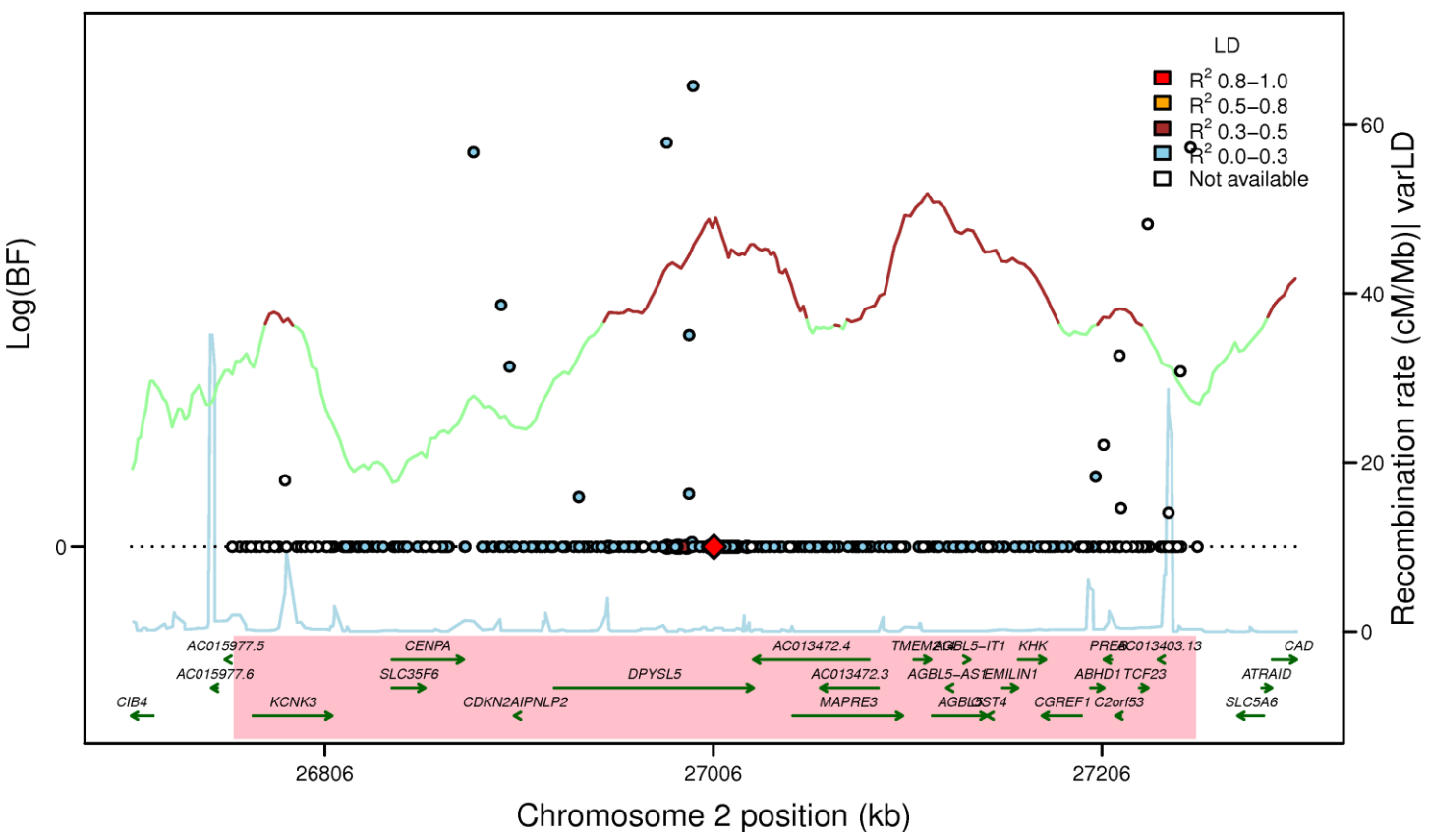

PROX1: rs340874 (FG TE_MANTRA, LD: HapMap2 YRI)
LD
R2 0.8–1.0
R2 0.5–0.8
R2 0.3–0.5
R2 0.0–0.3
Not available
Log(BF)
Recombination rate (cM/Mb)| varLD
PROX1
PROX1–AS1
RP11–323I1.1
PROX1–IT1
212026
212226
212426
Chromosome 1 position (kb)

Figure S3G


DPYSL5: rs1371614 (FG EA_MANTRA, LD: HapMap2 CEU)
LD
R2 0.8–1.0
R2 0.5–0.8
R2 0.3–0.5
R2 0.0–0.3
Not available
Log(BF)
Recombination rate (cM/Mb)| varLD
Chromosome 2 position (kb)
26806
27006
27206
AC015977.5
AC015977.6
CIB4
CENPA
SLC35F6
KCNK3
CDKN2AIPNLP2
DPYSL5
AC013472.4
AC013472.3
MAPRE3
KHK
EMILIN1
ABHD1
TCF23
C2orf53
ATRAID
SLC5A6
CAD


DPYSL5: rs1371614 (FG AA_MANTRA, LD: HapMap2 YRI)
LD
R2 0.8–1.0
R2 0.5–0.8
R2 0.3–0.5
R2 0.0–0.3
Not available
Log(BF)
Recombination rate (cM/Mb)| varLD
Chromosome 2 position (kb)
26806
27006
27206
AC015977.5
AC015977.6
CIB4
CENPA
SLC35F6
KCNK3
CDKN2AIPNLP2
DPYSL5
AC013472.4
AC013472.3
MAPRE3
KHK
EMILIN1
ABHD1
TCF23
C2orf53
ATRAID
SLC5A6
CAD

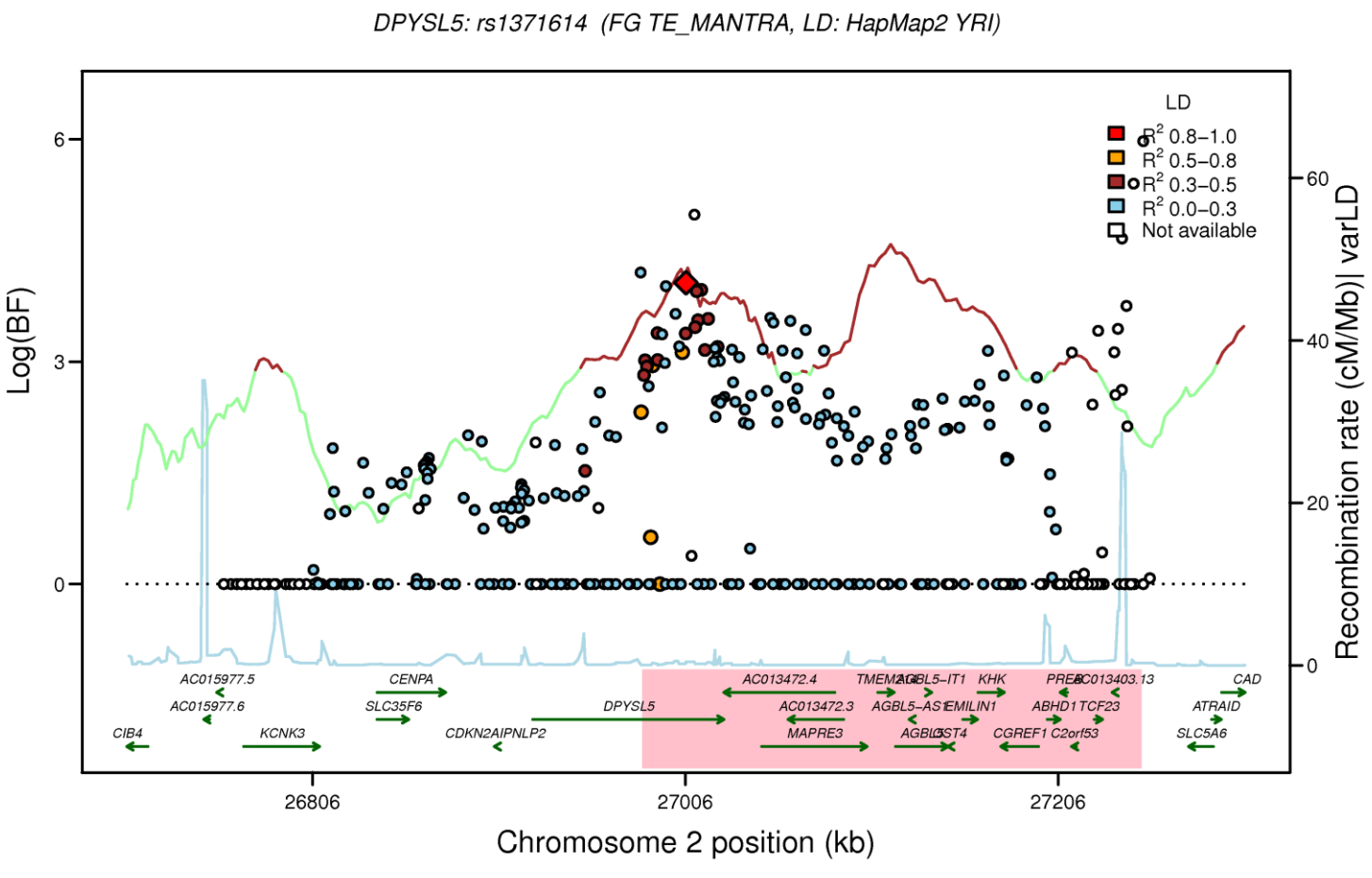
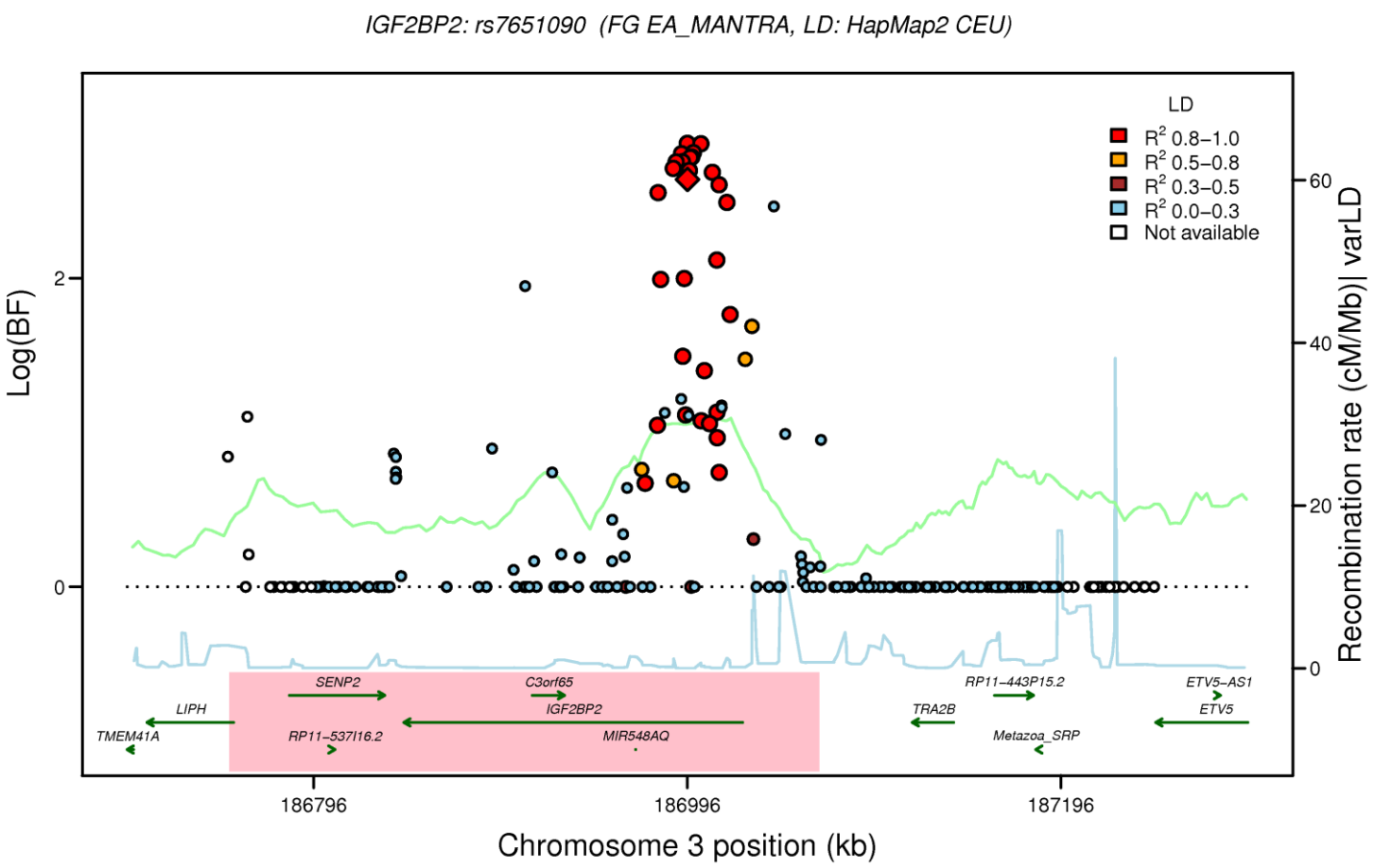
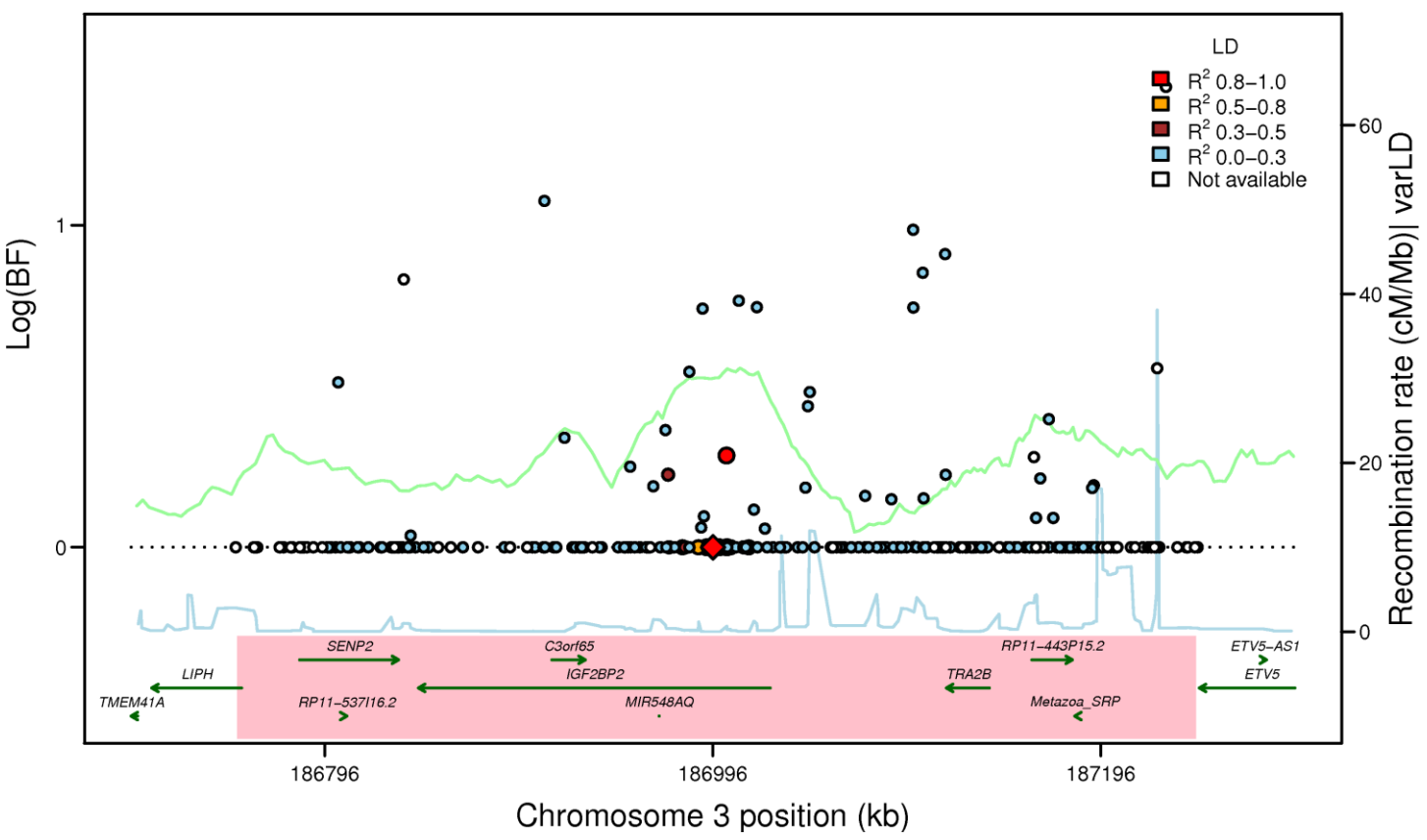

DPYSL5: rs1371614 (FG TE_MANTRA, LD: HapMap2 YRI)
LD
R2 0.8–1.0
R2 0.5–0.8
R2 0.3–0.5
R2 0.0–0.3
Not available
Log(BF)
Recombination rate (cM/Mb)| varLD
Chromosome 2 position (kb)
26806
27006
27206
AC015977.5
AC015977.6
CIB4
CENPA
SLC35F6
KCNK3
CDKN2AIPNLP2
DPYSL5
AC013472.4
AC013472.3
MAPRE3
KHK
EMILIN1
ABHD1
TCF23
C2orf53
ATRAID
SLC5A6
CAD

Figure S3H


IGF2BP2: rs7651090 (FG EA_MANTRA, LD: HapMap2 CEU)
LD
R2 0.8–1.0
R2 0.5–0.8
R2 0.3–0.5
R2 0.0–0.3
Not available
Log(BF)
Recombination rate (cM/Mb)| varLD
SENP2
C3orf65
RP11–443P15.2
ETV5–AS1
LIPH
IGF2BP2
TRA2B
ETV5
TMEM41A
RP11–537I16.2
MIR548AQ
Metazoa_SRP
186796
186996
187196
Chromosome 3 position (kb)


IGF2BP2: rs7651090 (FG AA_MANTRA, LD: HapMap2 YRI)
LD
R2 0.8–1.0
R2 0.5–0.8
R2 0.3–0.5
R2 0.0–0.3
Not available
Log(BF)
Recombination rate (cM/Mb)| varLD
SENP2
C3orf65
RP11–443P15.2
ETV5–AS1
LIPH
IGF2BP2
TRA2B
ETV5
TMEM41A
RP11–537I16.2
MIR548AQ
Metazoa_SRP
186796
186996
187196
Chromosome 3 position (kb)

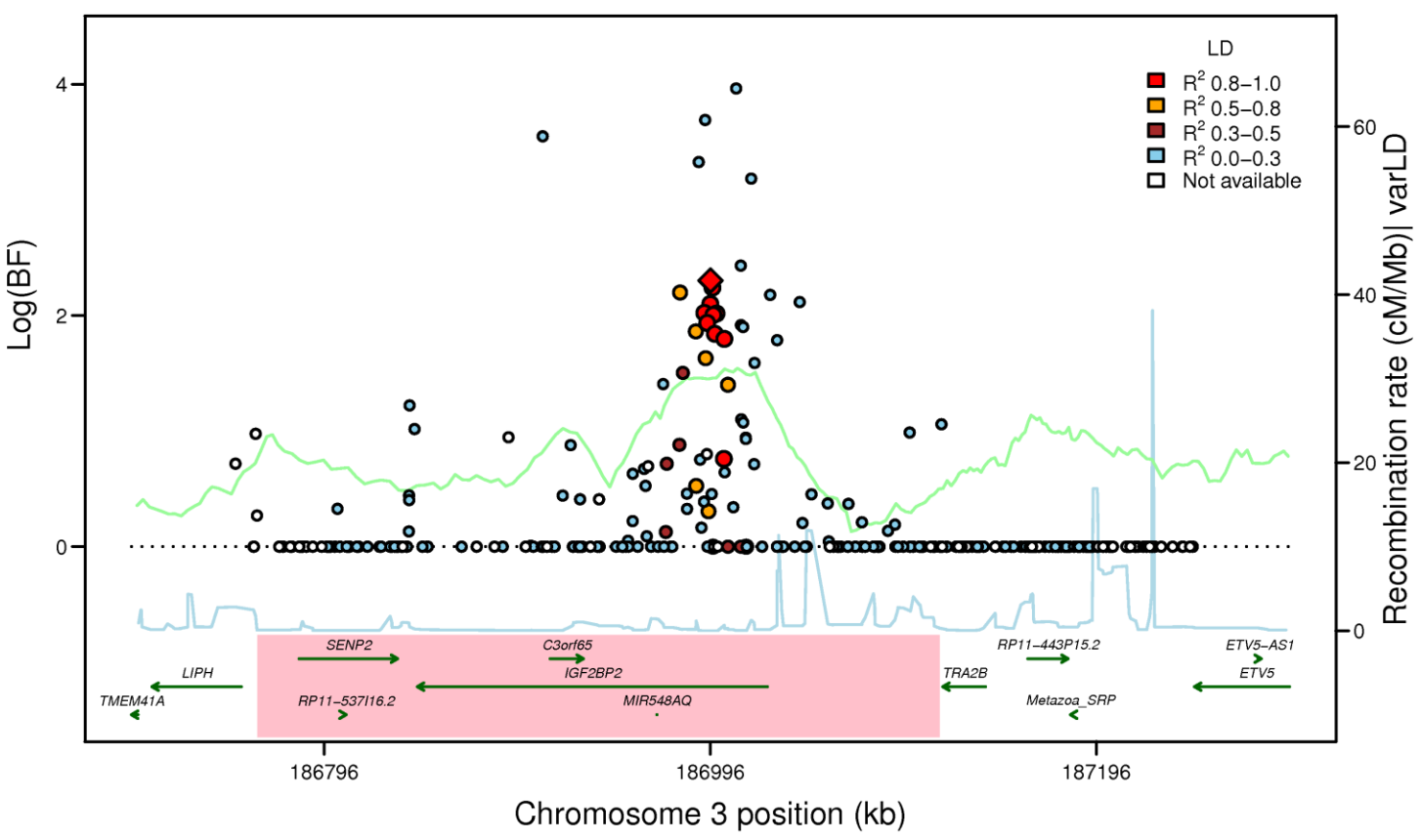
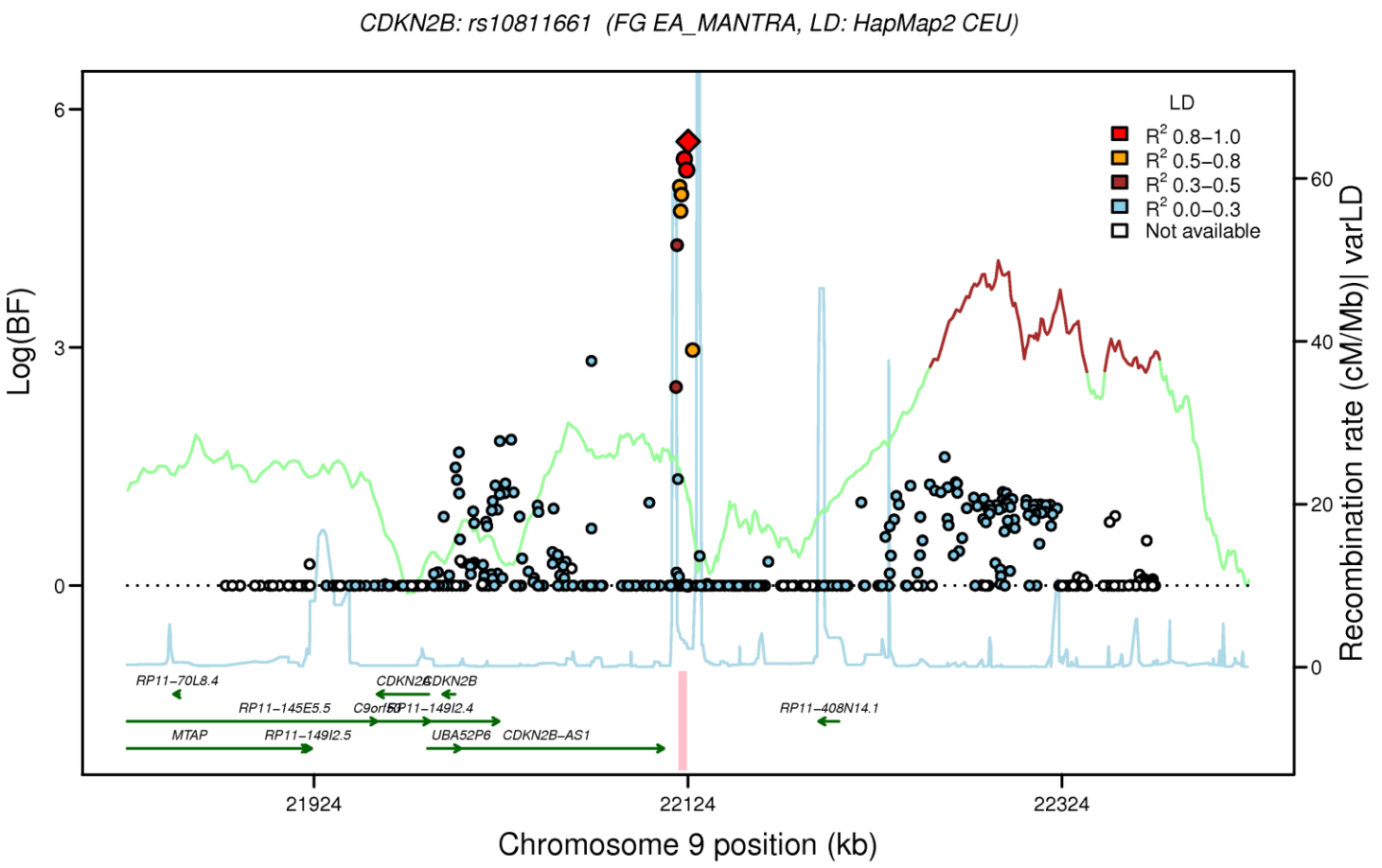
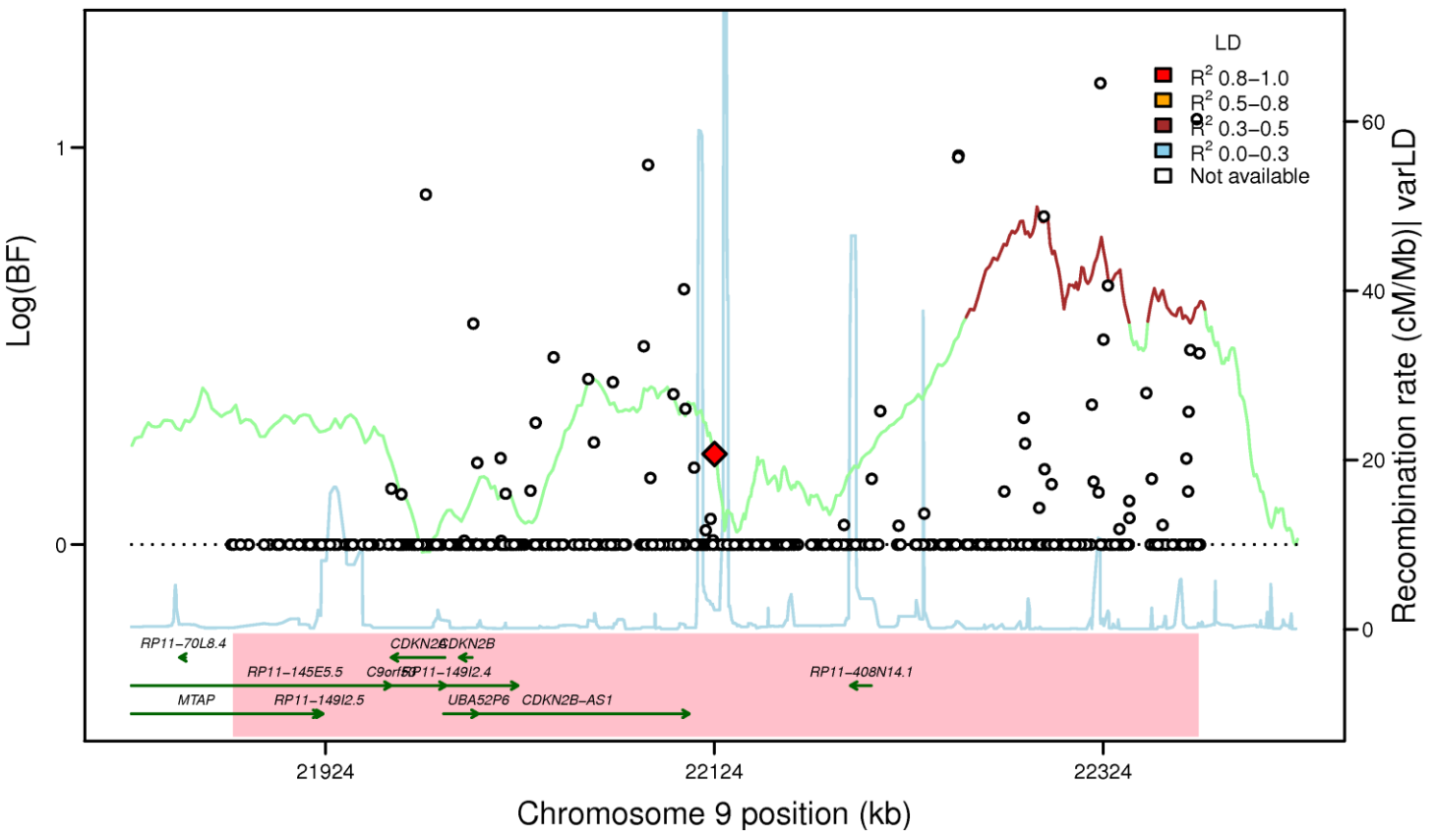

IGF2BP2: rs7651090 (FG TE_MANTRA, LD: HapMap2 YRI)
LD
R2 0.8–1.0
R2 0.5–0.8
R2 0.3–0.5
R2 0.0–0.3
Not available
Log(BF)
Recombination rate (cM/Mb)| varLD
SENP2
C3orf65
RP11–443P15.2
ETV5–AS1
LIPH
IGF2BP2
TRA2B
ETV5
TMEM41A
RP11–537I16.2
MIR548AQ
Metazoa_SRP
186796
186996
187196
Chromosome 3 position (kb)

Figure S3I


CDKN2B: rs10811661 (FG EA_MANTRA, LD: HapMap2 CEU)
LD
R2 0.8–1.0
R2 0.5–0.8
R2 0.3–0.5
R2 0.0–0.3
Not available
Log(BF)
Recombination rate (cM/Mb)| varLD
RP11–70L8.4
CDKN2A
CDKN2B
RP11–145E5.5
C9orf53
RP11–149I2.4
RP11–408N14.1
MTAP
RP11–149I2.5
UBA52P6
CDKN2B–AS1
21924
22124
22324
Chromosome 9 position (kb)


CDKN2B: rs10811661 (FG AA_MANTRA, LD: HapMap2 YRI)
LD
R2 0.8–1.0
R2 0.5–0.8
R2 0.3–0.5
R2 0.0–0.3
Not available
Log(BF)
Recombination rate (cM/Mb)| varLD
RP11–70L8.4
CDKN2A
CDKN2B
RP11–145E5.5
C9orf53
RP11–149I2.4
RP11–408N14.1
MTAP
RP11–149I2.5
UBA52P6
CDKN2B–AS1
21924
22124
22324
Chromosome 9 position (kb)

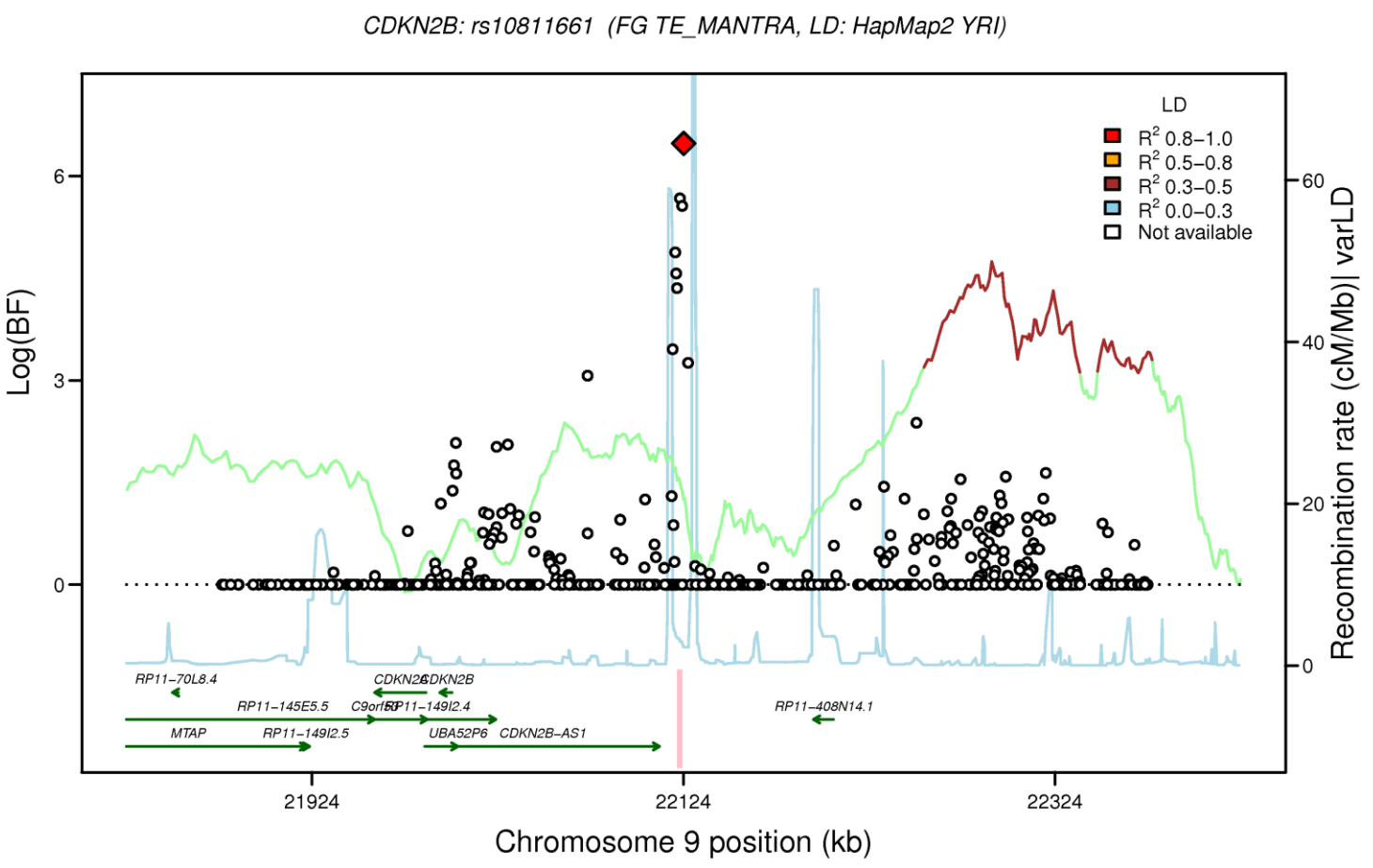
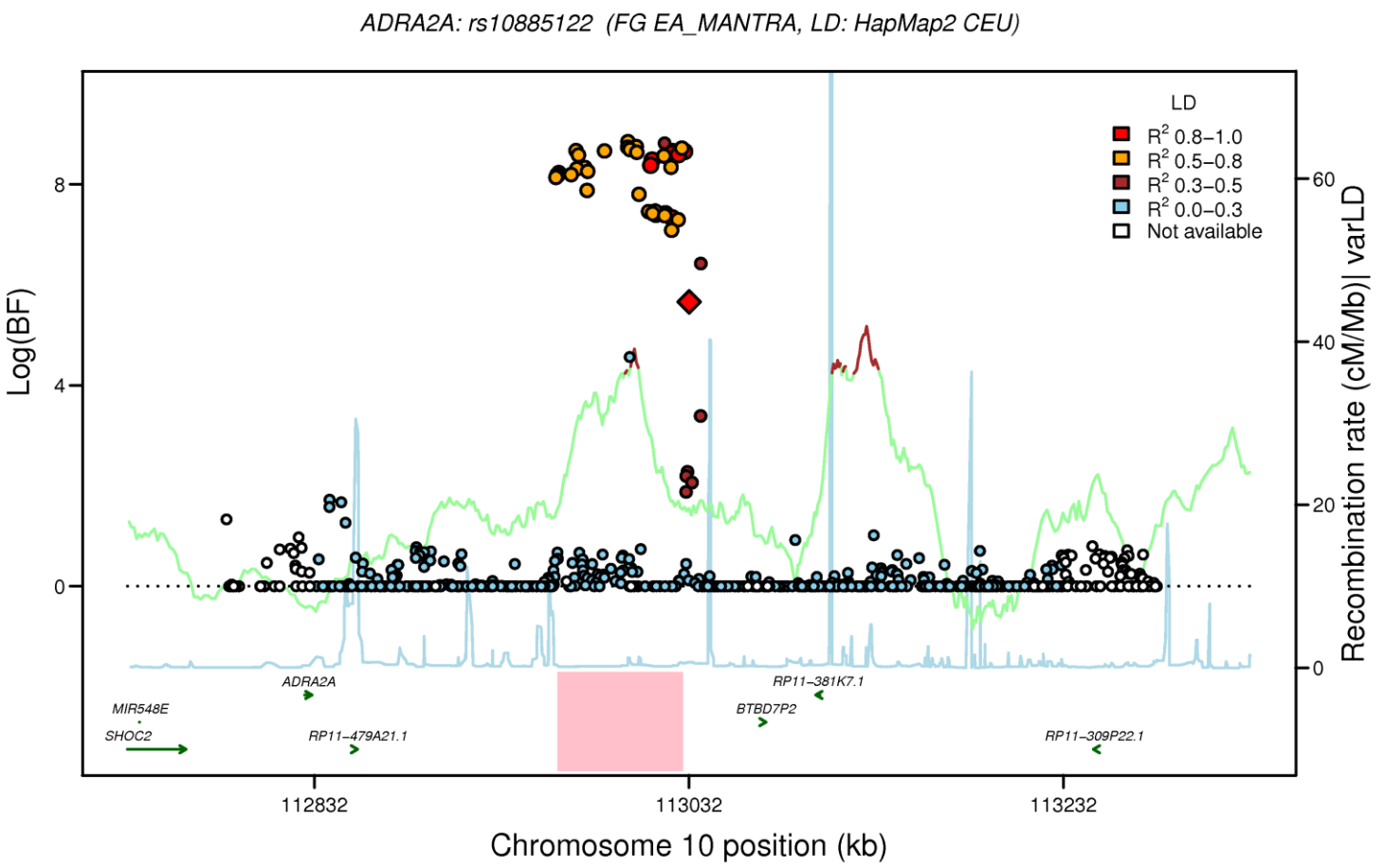
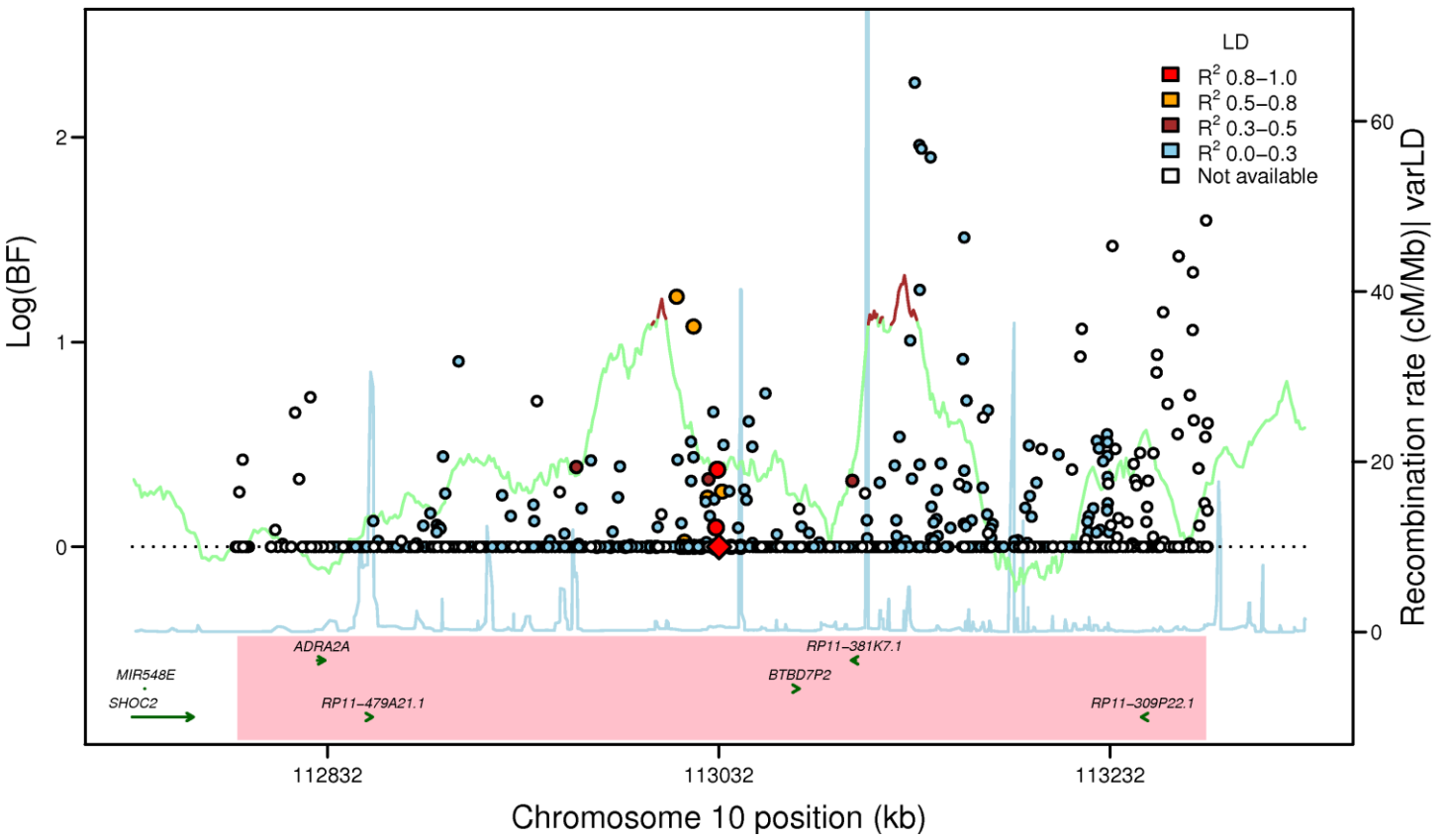

CDKN2B: rs10811661 (FG TE_MANTRA, LD: HapMap2 YRI)
LD
R2 0.8–1.0
R2 0.5–0.8
R2 0.3–0.5
R2 0.0–0.3
Not available
Log(BF)
Recombination rate (cM/Mb)| varLD
RP11–70L8.4
CDKN2A
CDKN2B
RP11–145E5.5
C9orf53
RP11–149I2.4
RP11–408N14.1
MTAP
RP11–149I2.5
UBA52P6
CDKN2B–AS1
21924
22124
22324
Chromosome 9 position (kb)

Figure S3J


ADRA2A: rs10885122 (FG EA_MANTRA, LD: HapMap2 CEU)
LD
R2 0.8–1.0
R2 0.5–0.8
R2 0.3–0.5
R2 0.0–0.3
Not available
Log(BF)
Recombination rate (cM/Mb)| varLD
ADRA2A
RP11–381K7.1
MIR548E
BTBD7P2
SHOC2
RP11–479A21.1
RP11–309P22.1
112832
113032
113232
Chromosome 10 position (kb)


ADRA2A: rs10885122 (FG AA_MANTRA, LD: HapMap2 YRI)
LD
R2 0.8–1.0
R2 0.5–0.8
R2 0.3–0.5
R2 0.0–0.3
Not available
Log(BF)
Recombination rate
ADRA2A
RP11–381K7.1
MIR548E
BTBD7P2
SHOC2
RP11–479A21.1
RP11–309P22.1
112832
113032
113232
Chromosome 10 position (kb)

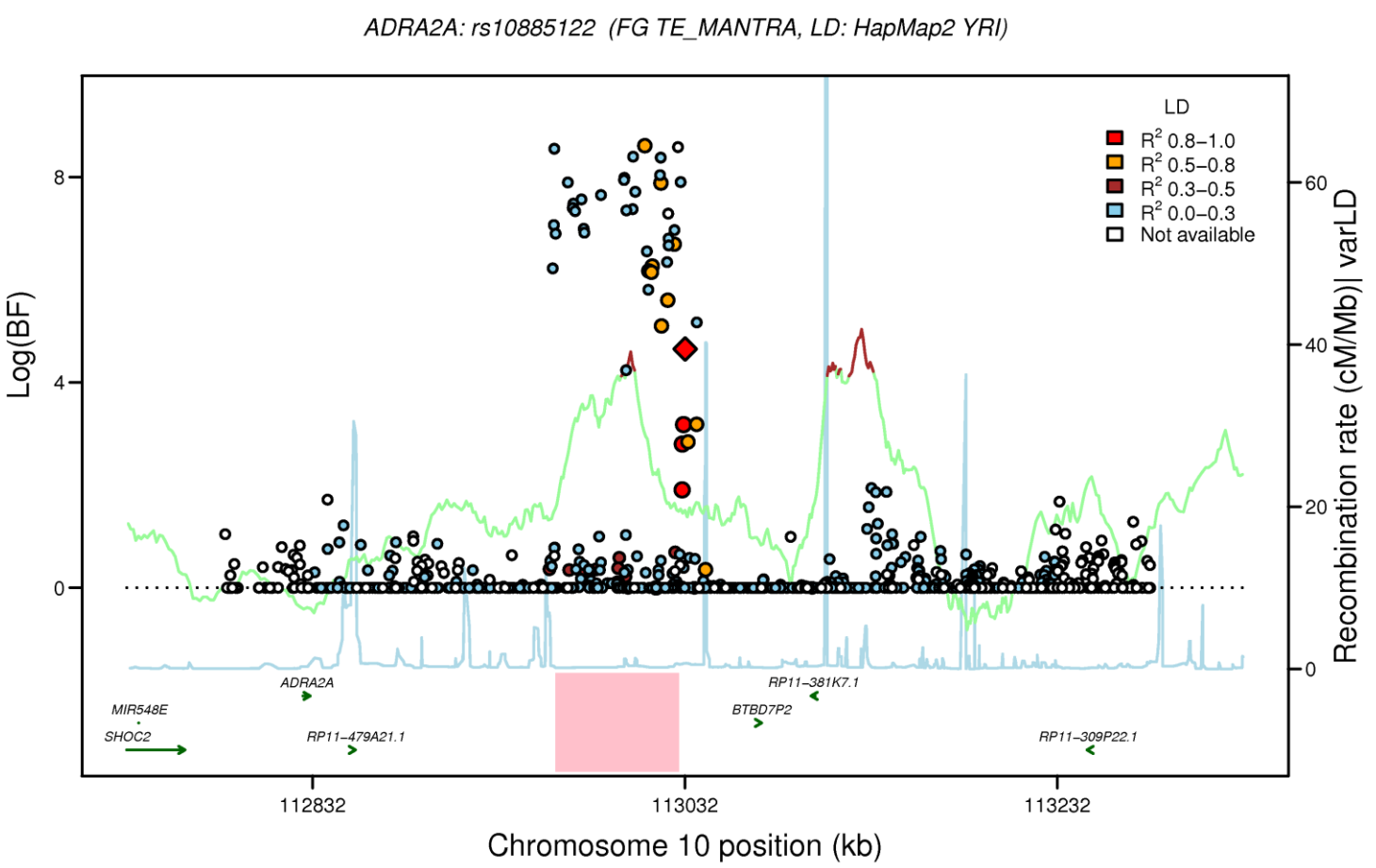
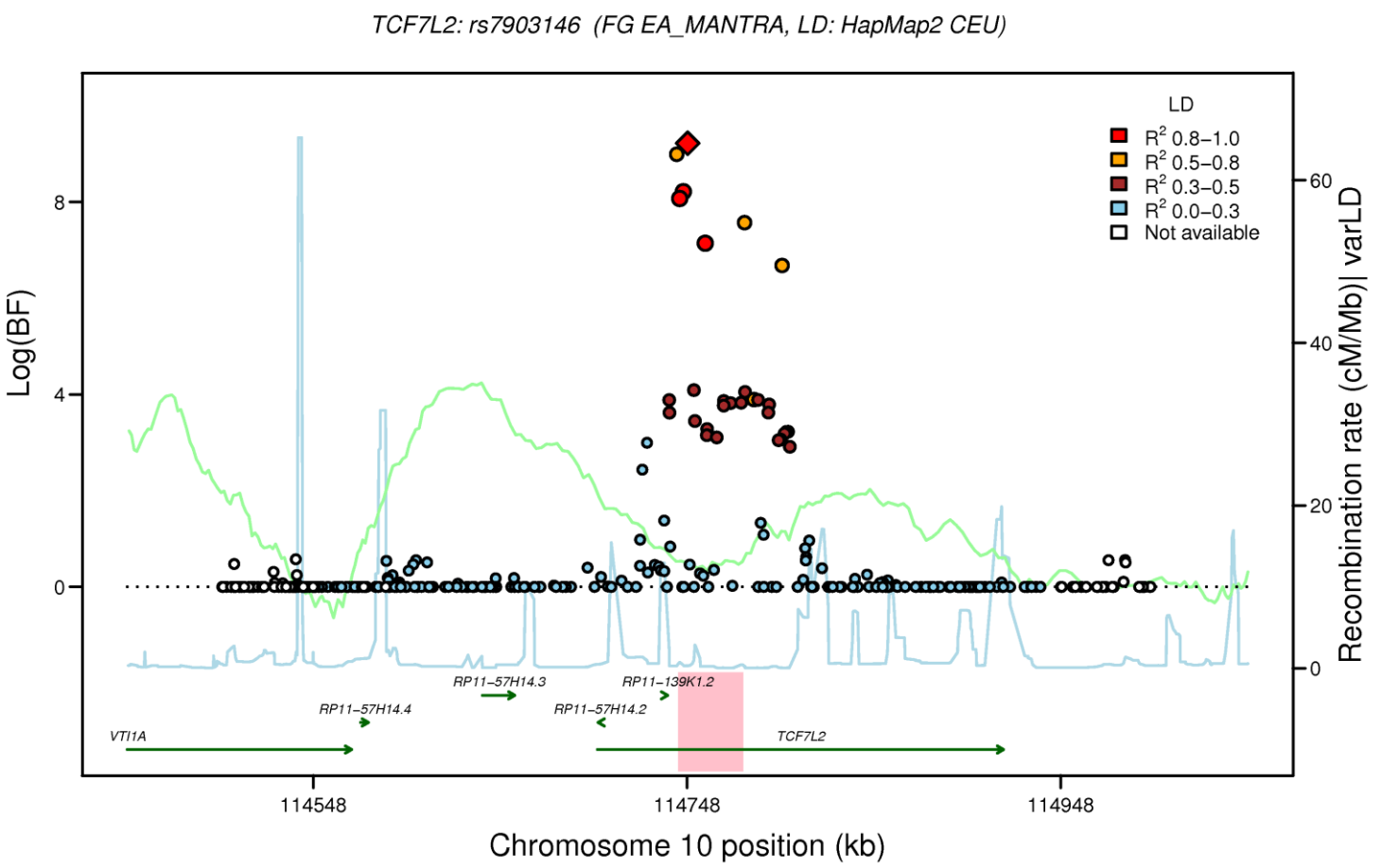
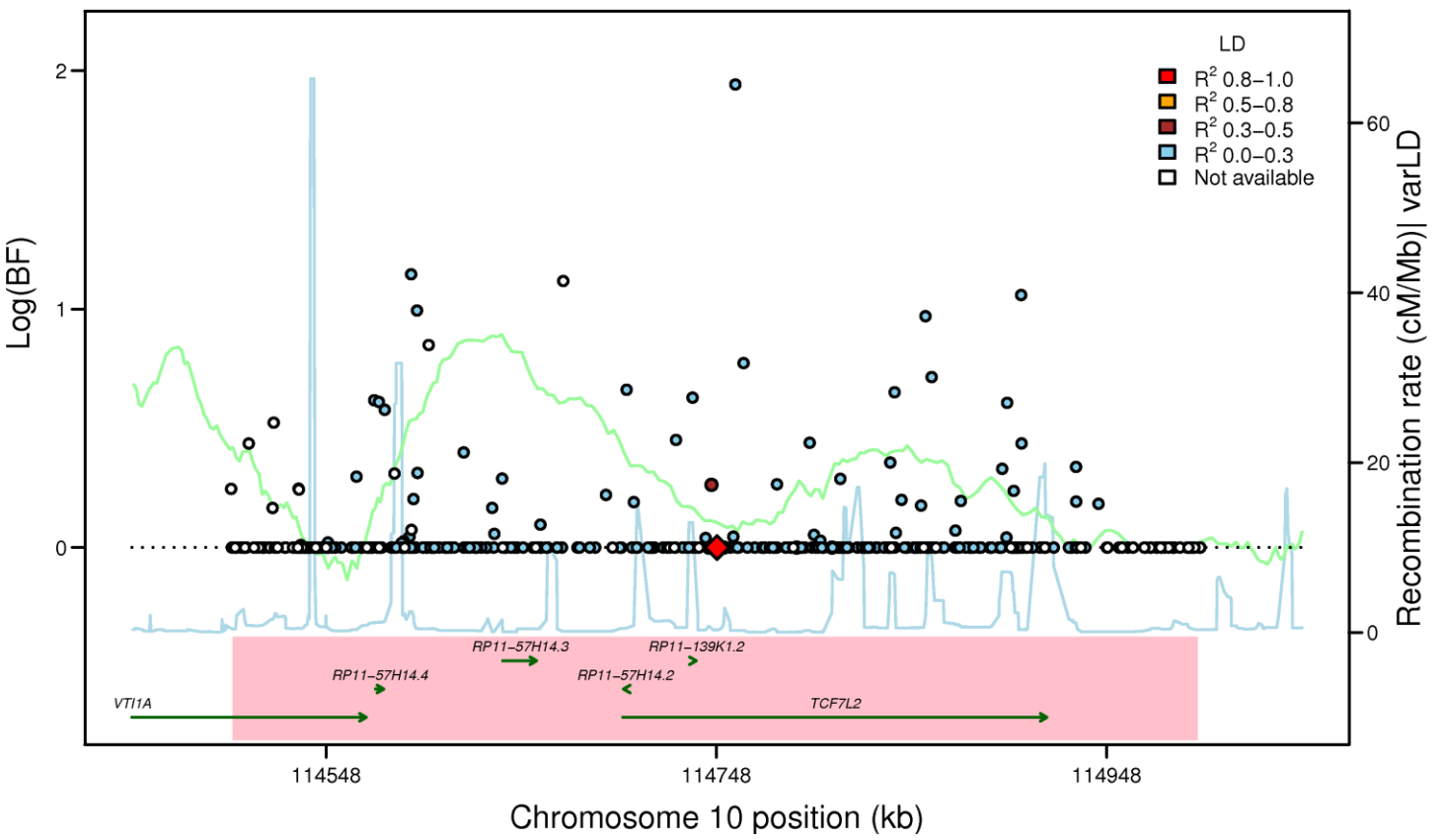

ADRA2A: rs10885122 (FG TE_MANTRA, LD: HapMap2 YRI)
LD
R2 0.8–1.0
R2 0.5–0.8
R2 0.3–0.5
R2 0.0–0.3
Not available
Log(BF)
Recombination rate (cM/Mb)| varLD
ADRA2A
RP11–381K7.1
MIR548E
BTBD7P2
SHOC2
RP11–479A21.1
RP11–309P22.1
112832
113032
113232
Chromosome 10 position (kb)

Figure S3K


TCF7L2: rs7903146 (FG EA_MANTRA, LD: HapMap2 CEU)
LD
R2 0.8–1.0
R2 0.5–0.8
R2 0.3–0.5
R2 0.0–0.3
Not available
Log(BF)
Recombination rate (cM/Mb)| varLD
RP11–57H14.3
RP11–139K1.2
RP11–57H14.4
RP11–57H14.2
VTI1A
TCF7L2
114548
114748
114948
Chromosome 10 position (kb)


TCF7L2: rs7903146 (FG AA_MANTRA, LD: HapMap2 YRI)
LD
R2 0.8–1.0
R2 0.5–0.8
R2 0.3–0.5
R2 0.0–0.3
Not available
Log(BF)
Recombination rate (cM/Mb)| varLD
RP11–57H14.3
RP11–139K1.2
RP11–57H14.4
RP11–57H14.2
VTI1A
TCF7L2
114548
114748
114948
Chromosome 10 position (kb)

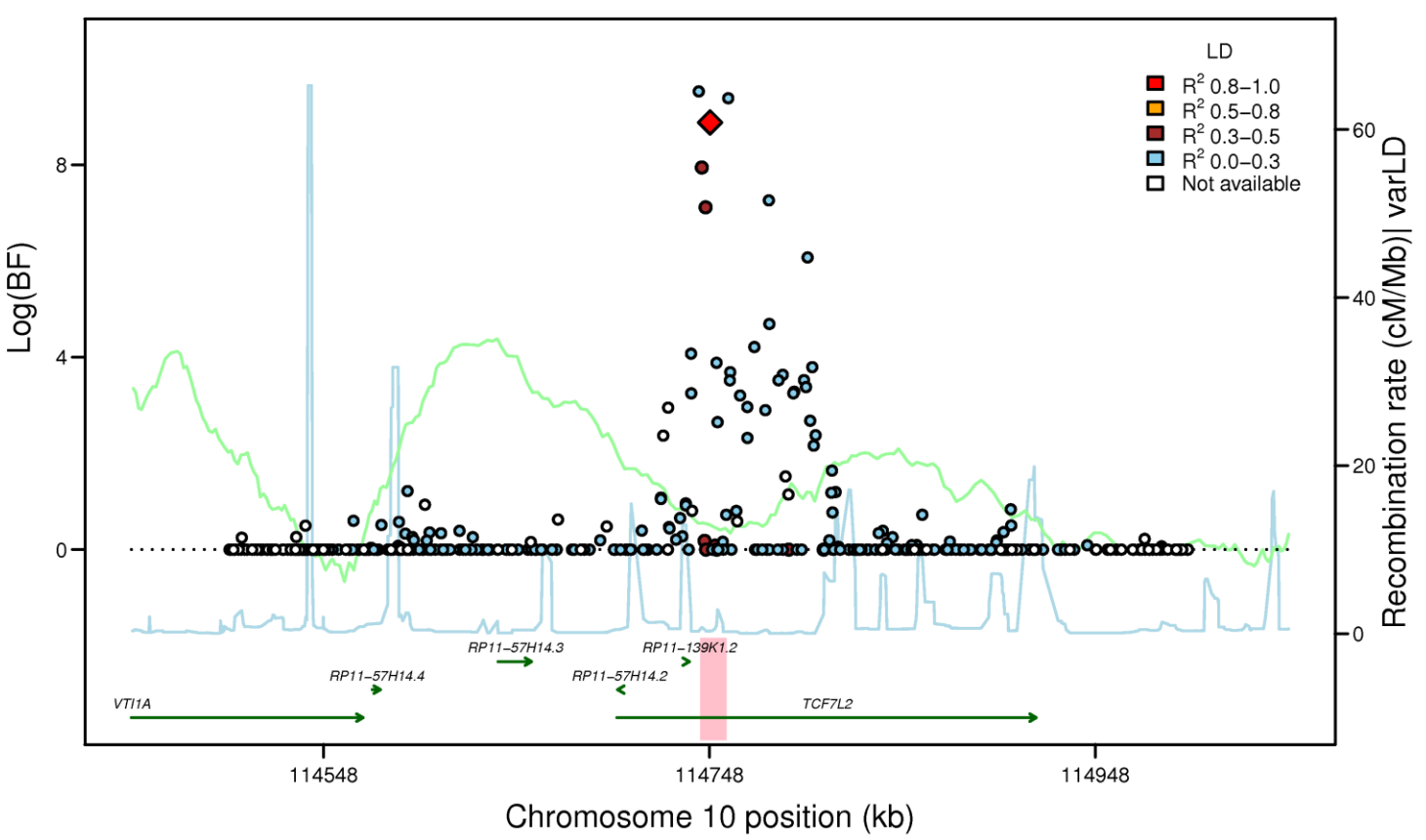
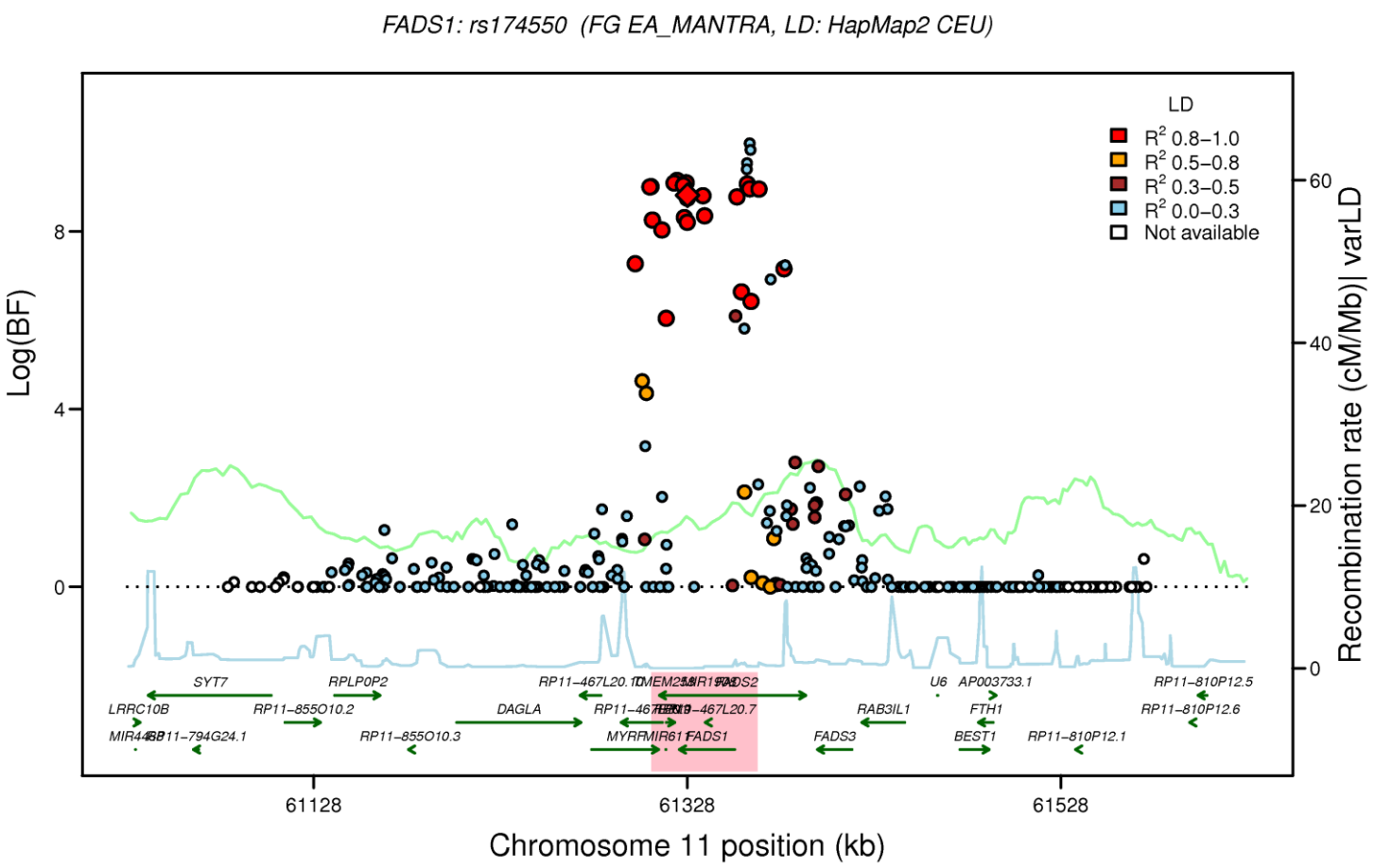
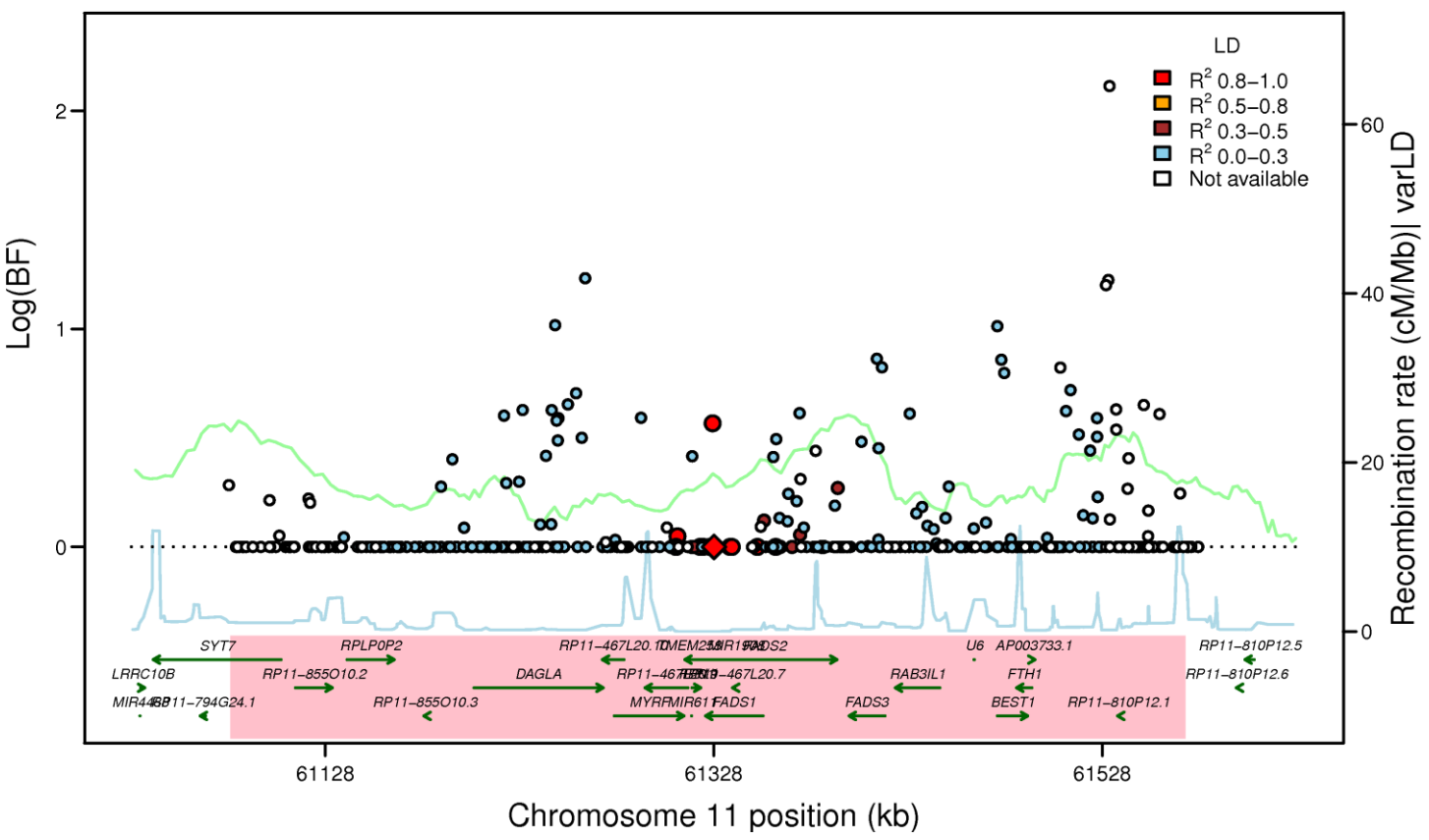

TCF7L2: rs7903146 (FG TE_MANTRA, LD: HapMap2 YRI)
LD
R2 0.8–1.0
R2 0.5–0.8
R2 0.3–0.5
R2 0.0–0.3
Not available
Log(BF)
Recombination rate (cM/Mb)| varLD
RP11–57H14.3
RP11–139K1.2
RP11–57H14.4
RP11–57H14.2
VTI1A
TCF7L2
114548
114748
114948
Chromosome 10 position (kb)

Figure S3L


FADS1: rs174550 (FG EA_MANTRA, LD: HapMap2 CEU)
LD
R2 0.8–1.0
R2 0.5–0.8
R2 0.3–0.5
R2 0.0–0.3
Not available
Log(BF)
Recombination rate (cM/Mb)| varLD
SYT7
RPLP0P2
RP11–467L20.10
DMEM256
MIR1908
FADS2
U6
AP003733.1
RP11–810P12.5
LRRC10B
RP11–855O10.2
DAGLA
RP11–467L20.7
RAB3IL1
FTH1
RP11–810P12.6
MIR4482
RP11–794G24.1
RP11–855O10.3
MYRFL
MIR611
FADS1
FADS3
BEST1
RP11–810P12.1
61128
61328
61528
Chromosome 11 position (kb)


FADS1: rs174550 (FG AA_MANTRA, LD: HapMap2 YRI)
LD
R2 0.8–1.0
R2 0.5–0.8
R2 0.3–0.5
R2 0.0–0.3
Not available
Log(BF)
Recombination rate (cM/Mb)| varLD
SYT7
RPLP0P2
RP11–467L20.10
DMEM256
MIR1908
FADS2
U6
AP003733.1
RP11–810P12.5
LRRC10B
RP11–855O10.2
DAGLA
RP11–467L20.7
RAB3IL1
FTH1
RP11–810P12.6
MIR4482
RP11–794G24.1
RP11–855O10.3
MYRFL
MIR611
FADS1
FADS3
BEST1
RP11–810P12.1
61128
61328
61528
Chromosome 11 position (kb)

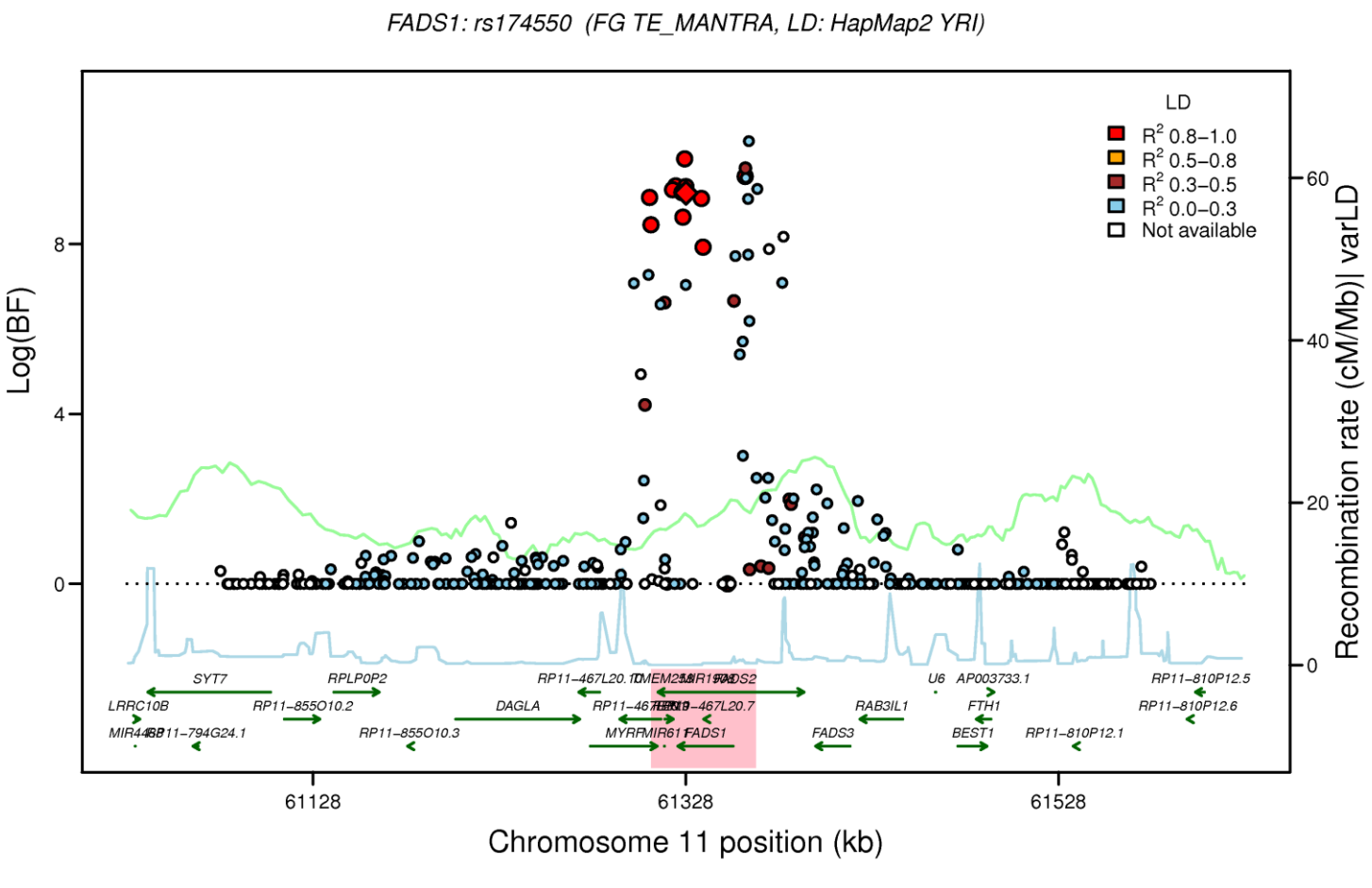
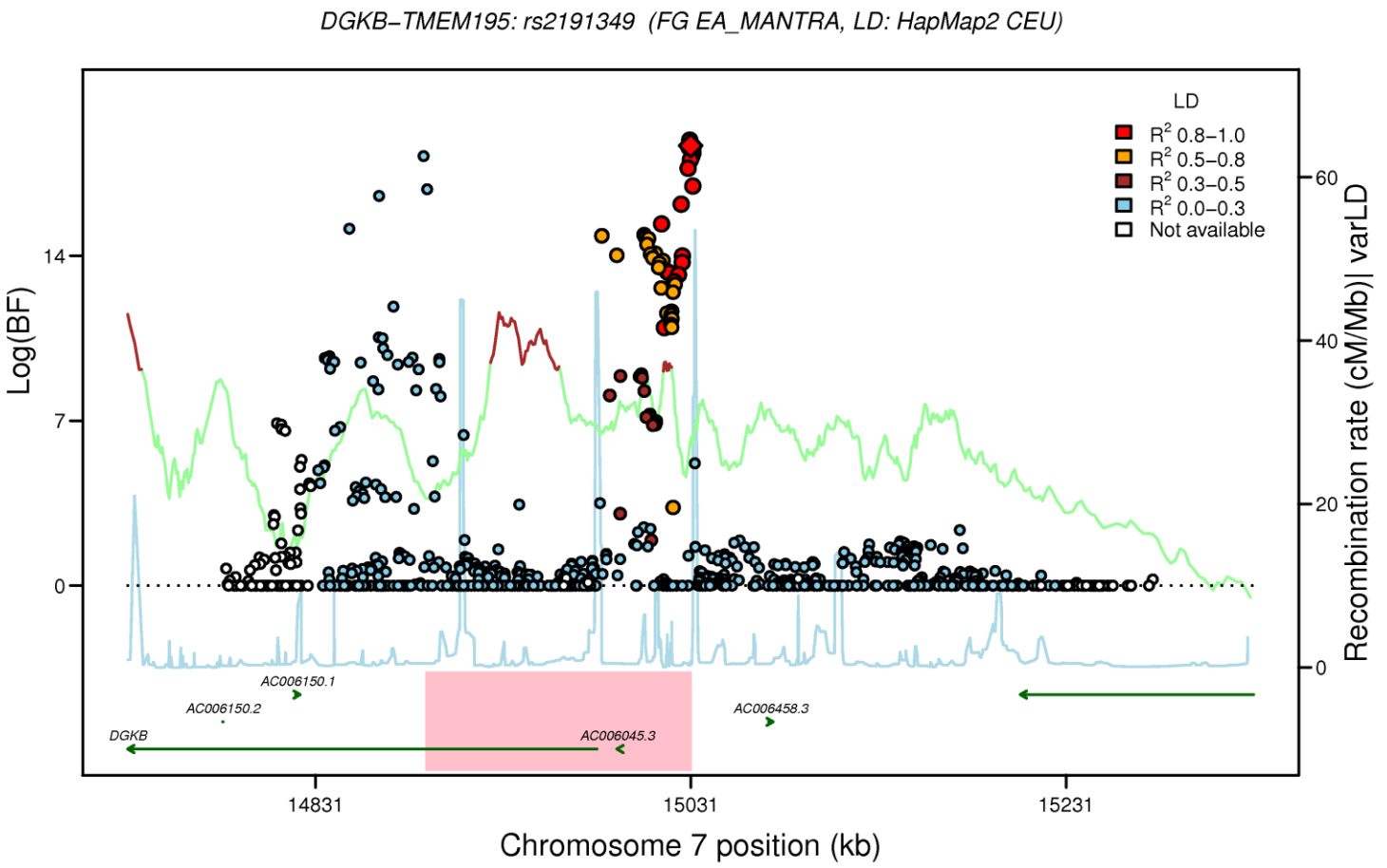
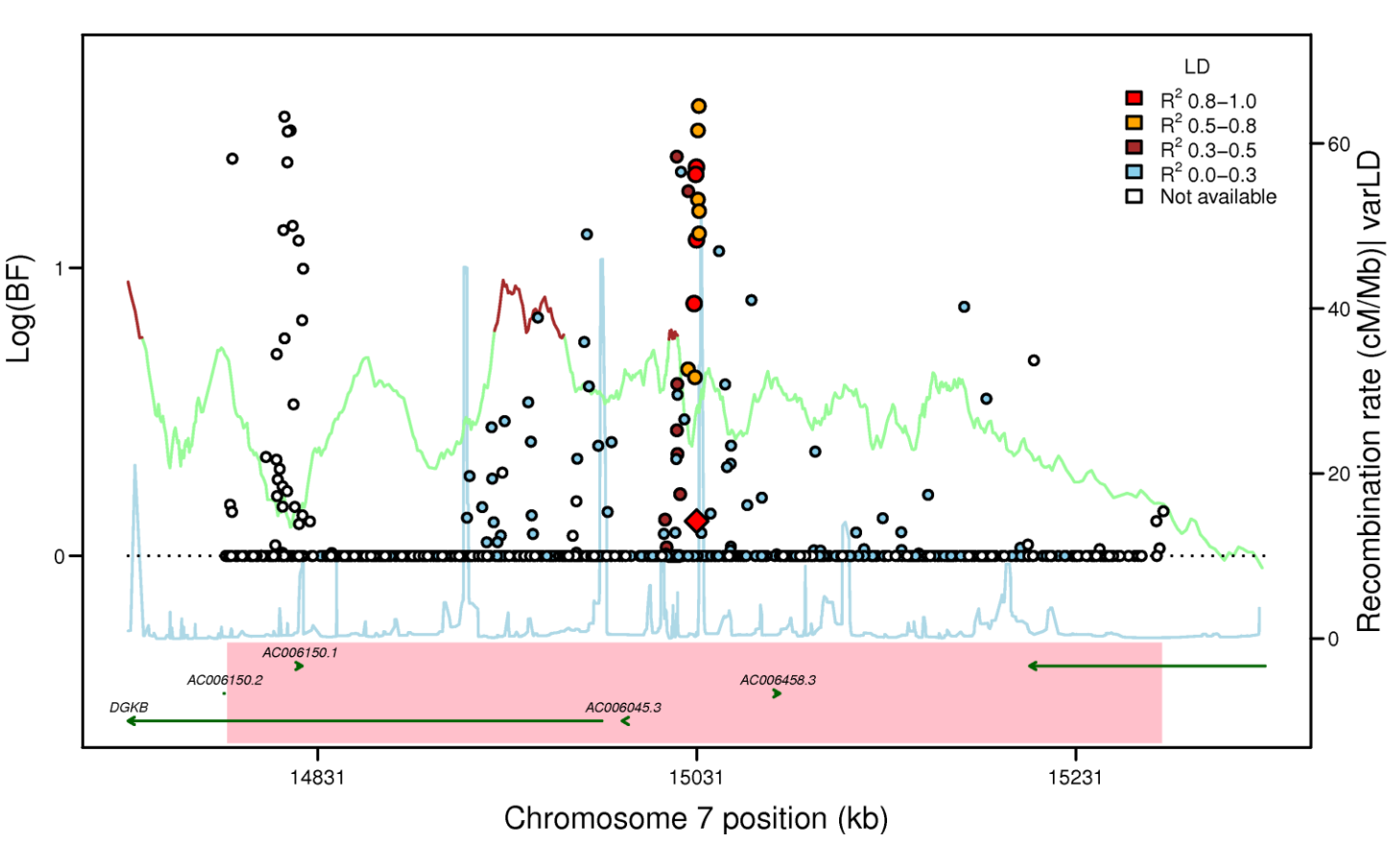

FADS1: rs174550 (FG TE_MANTRA, LD: HapMap2 YRI)
LD
R2 0.8–1.0
R2 0.5–0.8
R2 0.3–0.5
R2 0.0–0.3
Not available
Log(BF)
Recombination rate (cM/Mb)| varLD
SYT7
RPLP0P2
RP11–467L20.10
DMEM256
MIR1908
FADS2
U6
AP003733.1
RP11–810P12.5
LRRC10B
RP11–855O10.2
DAGLA
RP11–467L20.7
RAB3IL1
FTH1
RP11–810P12.6
MIR4482
RP11–794G24.1
RP11–855O10.3
MYRFL
MIR611
FADS1
FADS3
BEST1
RP11–810P12.1
61128
61328
61528
Chromosome 11 position (kb)

Figure S3M


DGKB–TMEM195: rs2191349 (FG EA_MANTRA, LD: HapMap2 CEU)
LD
R2 0.8–1.0
R2 0.5–0.8
R2 0.3–0.5
R2 0.0–0.3
Not available
Log(BF)
14
7
0
Recombination rate (cM/Mb)| varLD
60
40
20
0
AC006150.1
AC006150.2
AC006458.3
DGKB
AC006045.3
14831
15031
15231
Chromosome 7 position (kb)


DGKB–TMEM195: rs2191349 (FG AA_MANTRA, LD: HapMap2 YRI)
LD
R2 0.8–1.0
R2 0.5–0.8
R2 0.3–0.5
R2 0.0–0.3
Not available
Log(BF)
1
0
Recombination rate (cM/Mb)| varLD
60
40
20
0
AC006150.1
AC006150.2
AC006458.3
DGKB
AC006045.3
14831
15031
15231
Chromosome 7 position (kb)

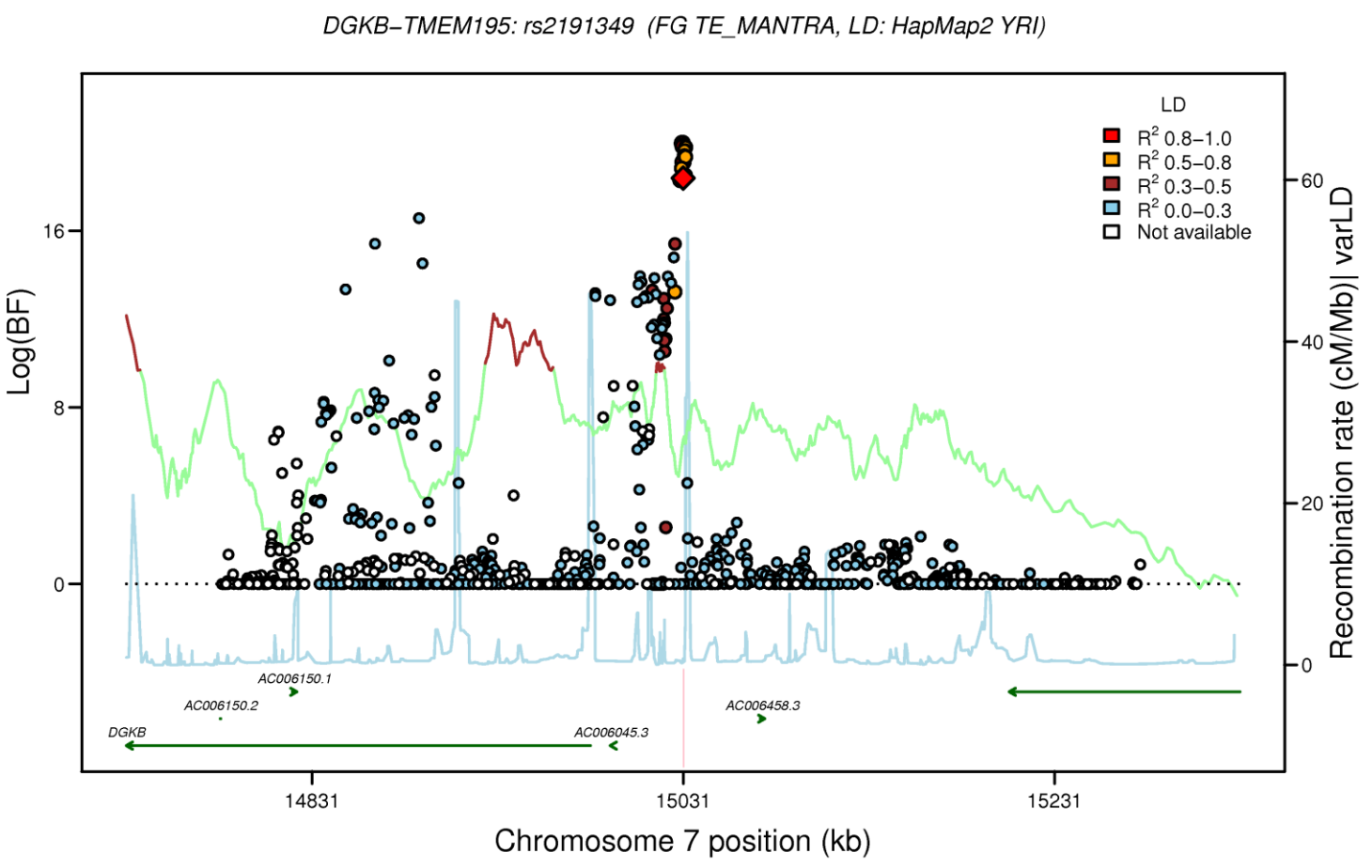
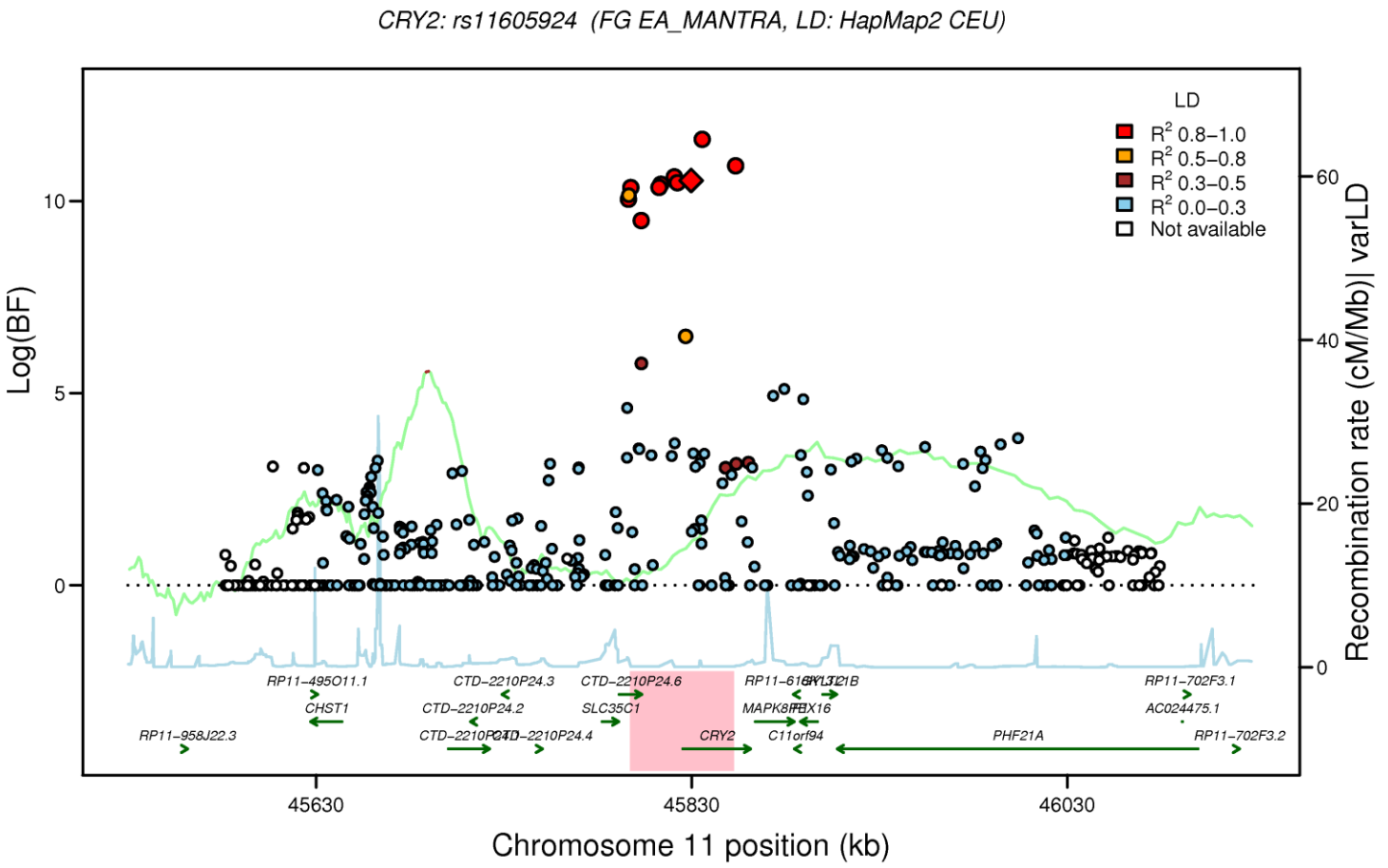
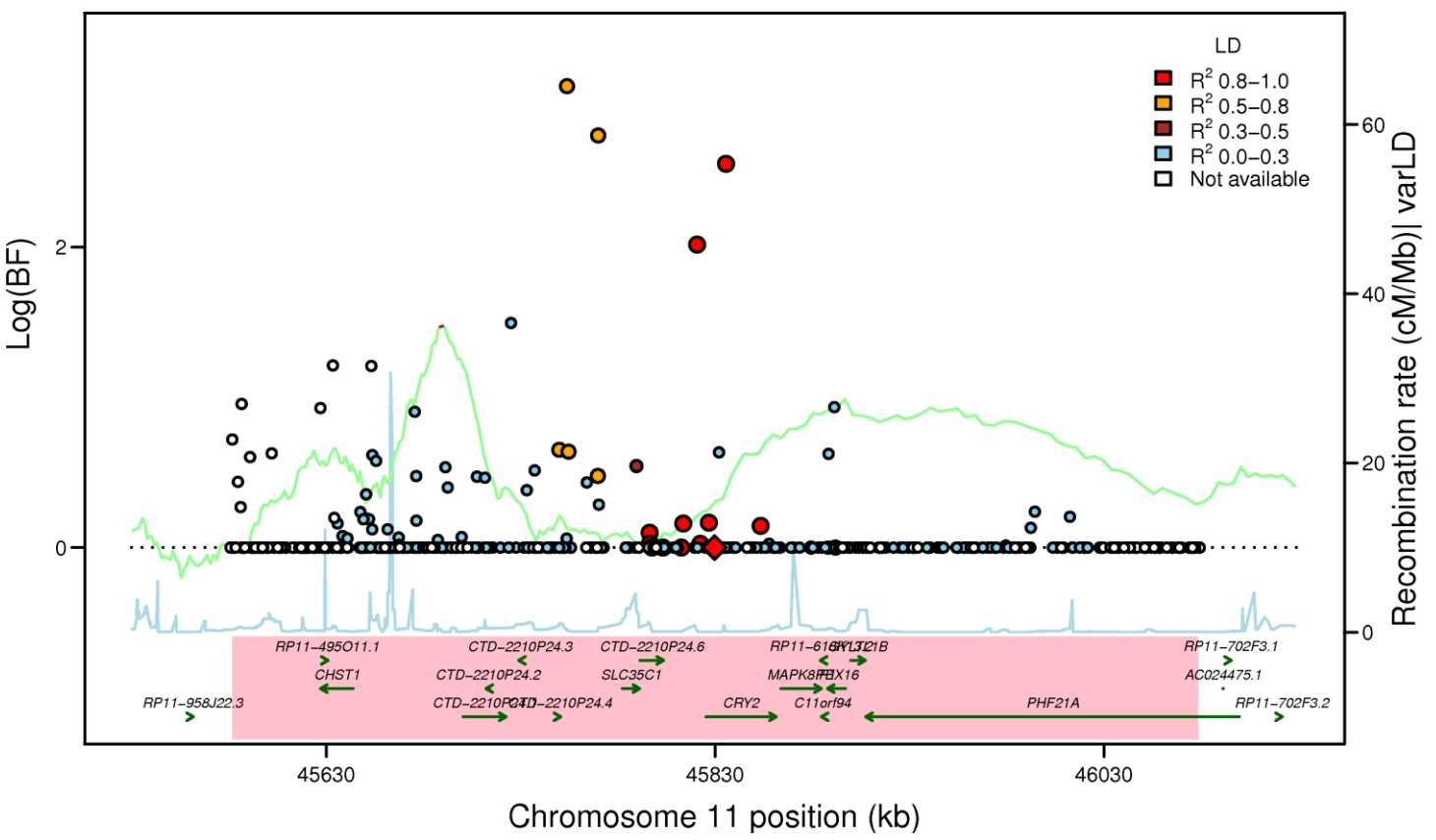

DGKB–TMEM195: rs2191349 (FG TE_MANTRA, LD: HapMap2 YRI)
LD
R2 0.8–1.0
R2 0.5–0.8
R2 0.3–0.5
R2 0.0–0.3
Not available
Log(BF)
16
8
0
Recombination rate (cM/Mb)| varLD
60
40
20
0
AC006150.1
AC006150.2
AC006458.3
DGKB
AC006045.3
14831
15031
15231
Chromosome 7 position (kb)

Figure S3N


CRY2: rs11605924 (FG EA_MANTRA, LD: HapMap2 CEU)
LD
R2 0.8–1.0
R2 0.5–0.8
R2 0.3–0.5
R2 0.0–0.3
Not available
Log(BF)
Recombination rate (cM/Mb)| varLD
RP11–495O11.1
CHST1
RP11–958J22.3
CTD–2210P24.3
CTD–2210P24.2
CTD–2210P24.4
CTD–2210P24.6
SLC35C1
CRY2
C11orf94
PHF21A
RP11–702F3.1
AC024475.1
RP11–702F3.2
45630
45830
46030
Chromosome 11 position (kb)


CRY2: rs11605924 (FG AA_MANTRA, LD: HapMap2 YRI)
LD
R2 0.8–1.0
R2 0.5–0.8
R2 0.3–0.5
R2 0.0–0.3
Not available
Log(BF)
Recombination rate (cM/Mb)| varLD
RP11–495O11.1
CHST1
RP11–958J22.3
CTD–2210P24.3
CTD–2210P24.2
CTD–2210P24.4
CTD–2210P24.6
SLC35C1
CRY2
C11orf94
PHF21A
RP11–702F3.1
AC024475.1
RP11–702F3.2
45630
45830
46030
Chromosome 11 position (kb)

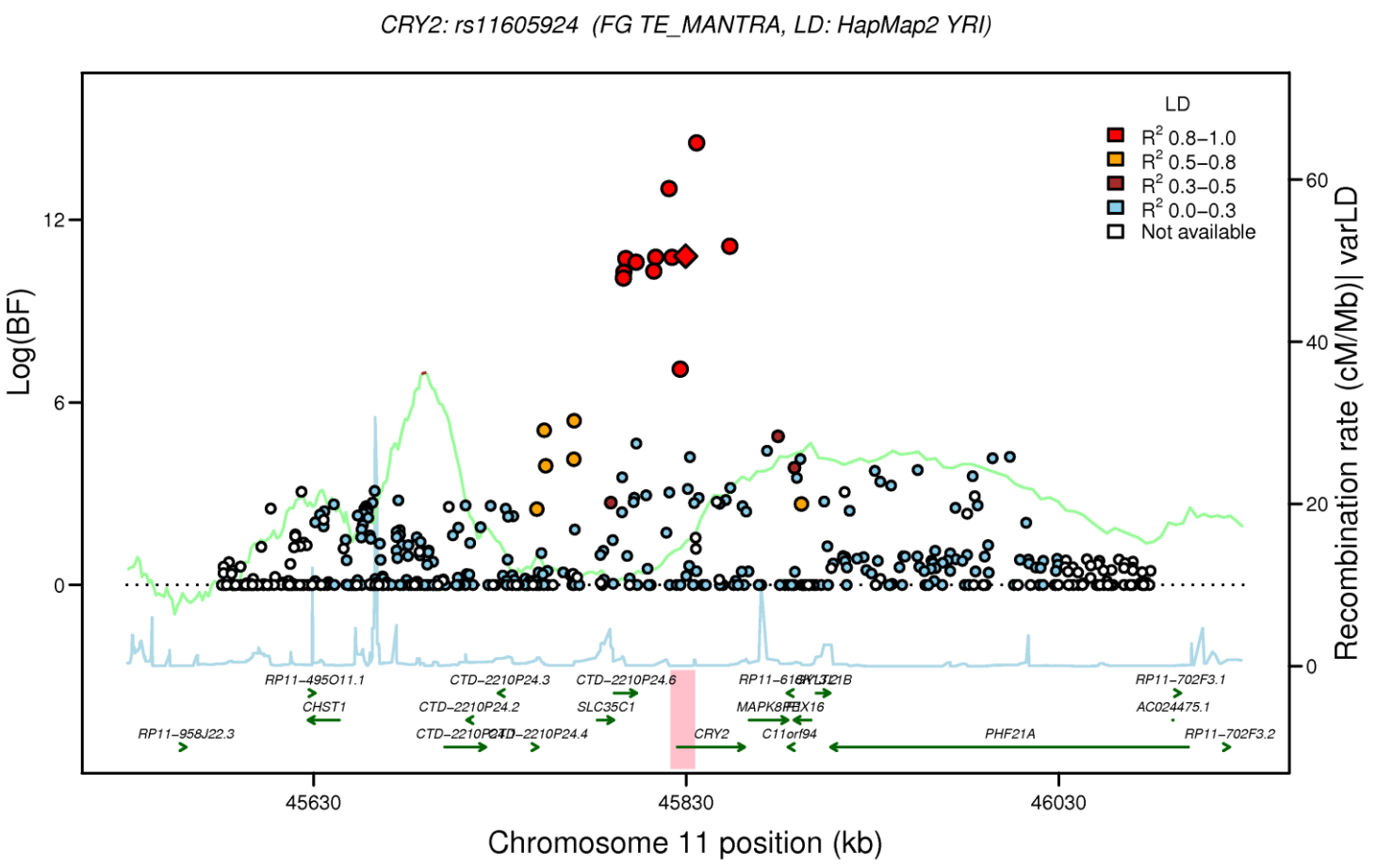
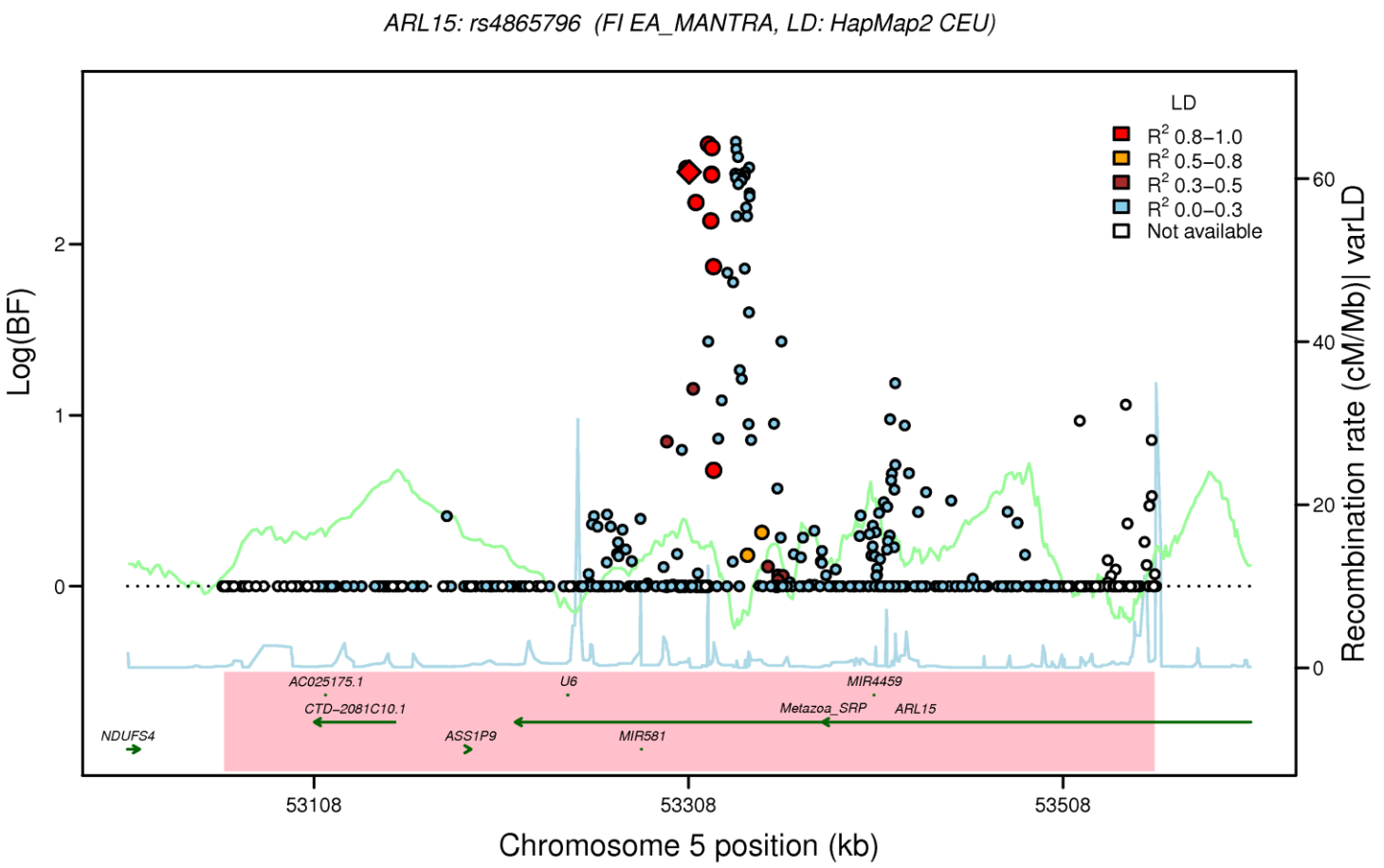
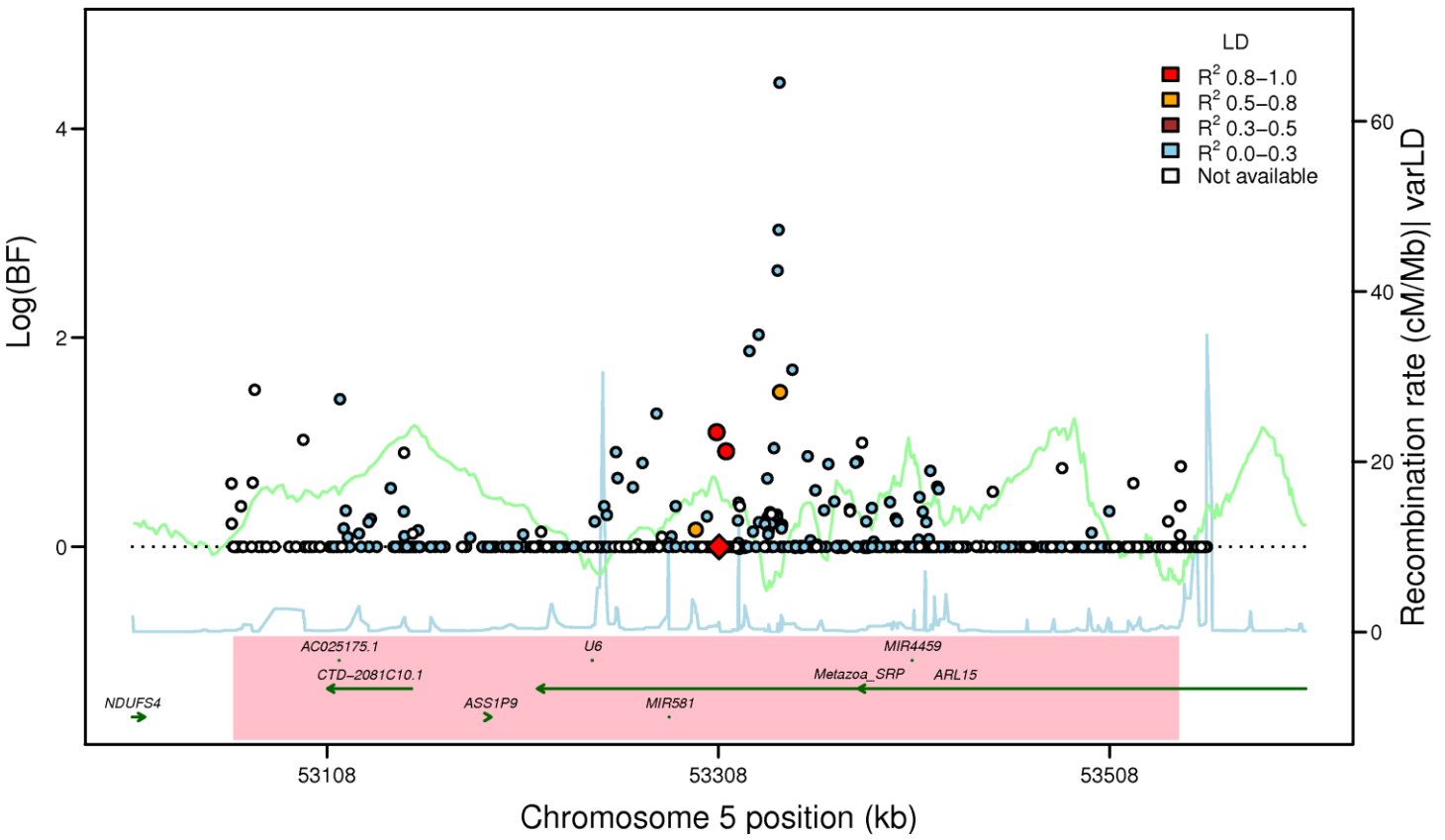

CRY2: rs11605924 (FG TE_MANTRA, LD: HapMap2 YRI)
LD
R2 0.8–1.0
R2 0.5–0.8
R2 0.3–0.5
R2 0.0–0.3
Not available
Log(BF)
Recombination rate (cM/Mb)| varLD
RP11–495O11.1
CHST1
RP11–958J22.3
CTD–2210P24.3
CTD–2210P24.2
CTD–2210P24.4
CTD–2210P24.6
SLC35C1
CRY2
C11orf94
PHF21A
RP11–702F3.1
AC024475.1
RP11–702F3.2
45630
45830
46030
Chromosome 11 position (kb)

Figure S3O


ARL15: rs4865796 (FI EA_MANTRA, LD: HapMap2 CEU)
LD
R2 0.8–1.0
R2 0.5–0.8
R2 0.3–0.5
R2 0.0–0.3
Not available
Log(BF)
Recombination rate (cM/Mb)| varLD
AC025175.1
CTD–2081C10.1
U6
MIR4459
Metazoa_SRP
ARL15
NDUFS4
ASS1P9
MIR581
53108
53308
53508
Chromosome 5 position (kb)


ARL15: rs4865796 (FI AA_MANTRA, LD: HapMap2 YRI)
LD
R2 0.8–1.0
R2 0.5–0.8
R2 0.3–0.5
R2 0.0–0.3
Not available
Log(BF)
Recombination rate
AC025175.1
CTD–2081C10.1
U6
MIR4459
Metazoa_SRP
ARL15
NDUFS4
ASS1P9
MIR581
53108
53308
53508
Chromosome 5 position (kb)

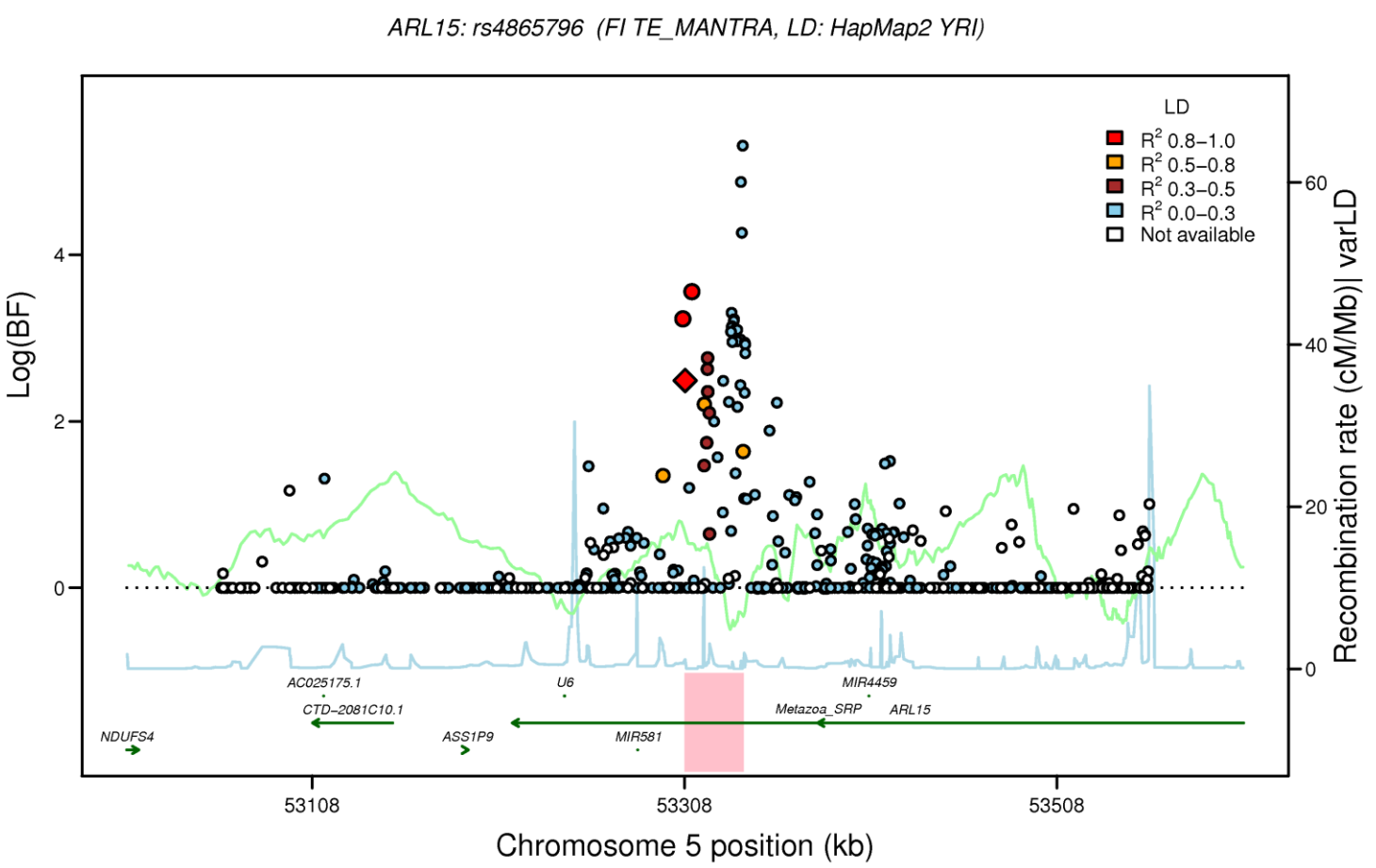
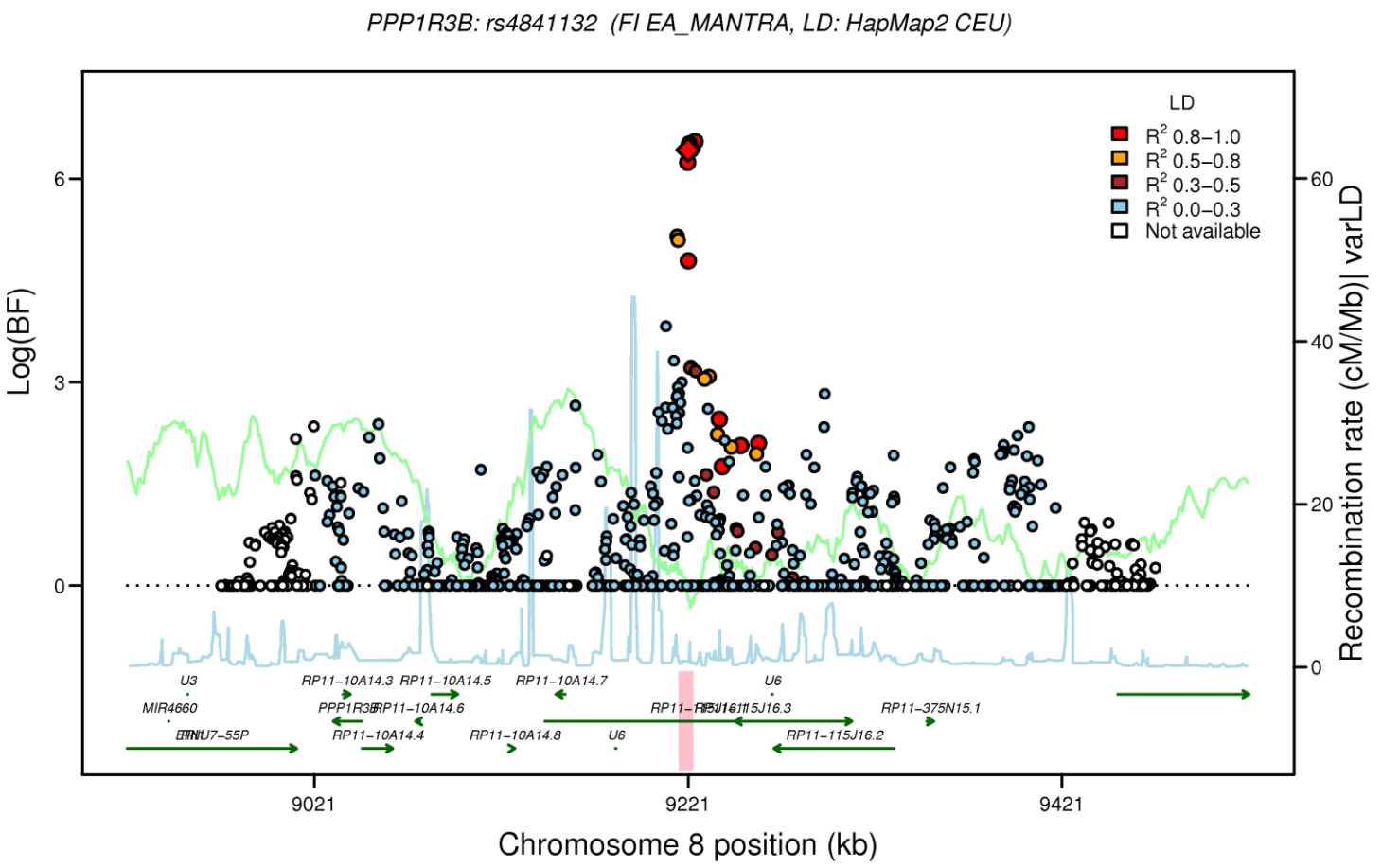
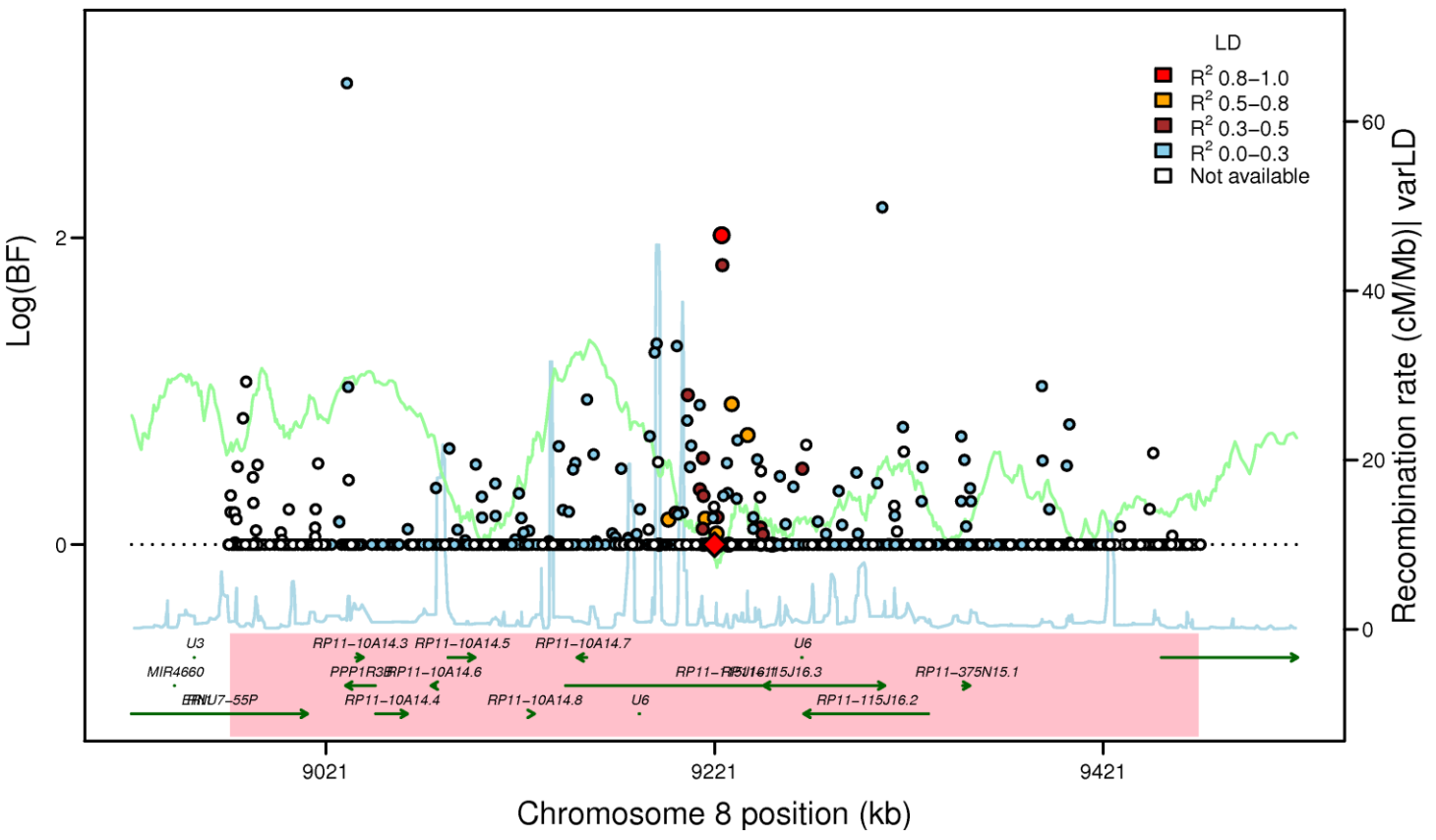

ARL15: rs4865796 (FI TE_MANTRA, LD: HapMap2 YRI)
LD
R2 0.8–1.0
R2 0.5–0.8
R2 0.3–0.5
R2 0.0–0.3
Not available
Log(BF)
Recombination rate (cM/Mb)| varLD
AC025175.1
CTD–2081C10.1
U6
MIR4459
Metazoa_SRP
ARL15
NDUFS4
ASS1P9
MIR581
53108
53308
53508
Chromosome 5 position (kb)

Figure S3P


PPP1R3B: rs4841132 (FI EA_MANTRA, LD: HapMap2 CEU)
LD
R2 0.8–1.0
R2 0.5–0.8
R2 0.3–0.5
R2 0.0–0.3
Not available
Log(BF)
Recombination rate (cM/Mb)| varLD
Chromosome 8 position (kb)
9021
9221
9421
U3
MIR4660
RP11–10A14.3
RP11–10A14.5
RP11–10A14.7
U6
RP11–375N15.1
RP11–10A14.4
RP11–10A14.8
RP11–115J16.2


PPP1R3B: rs4841132 (FI AA_MANTRA, LD: HapMap2 YRI)
LD
R2 0.8–1.0
R2 0.5–0.8
R2 0.3–0.5
R2 0.0–0.3
Not available
Log(BF)
Recombination rate (cM/Mb)| varLD
Chromosome 8 position (kb)
9021
9221
9421
U3
MIR4660
RP11–10A14.3
RP11–10A14.5
RP11–10A14.7
U6
RP11–375N15.1
RP11–10A14.4
RP11–10A14.8
RP11–115J16.2

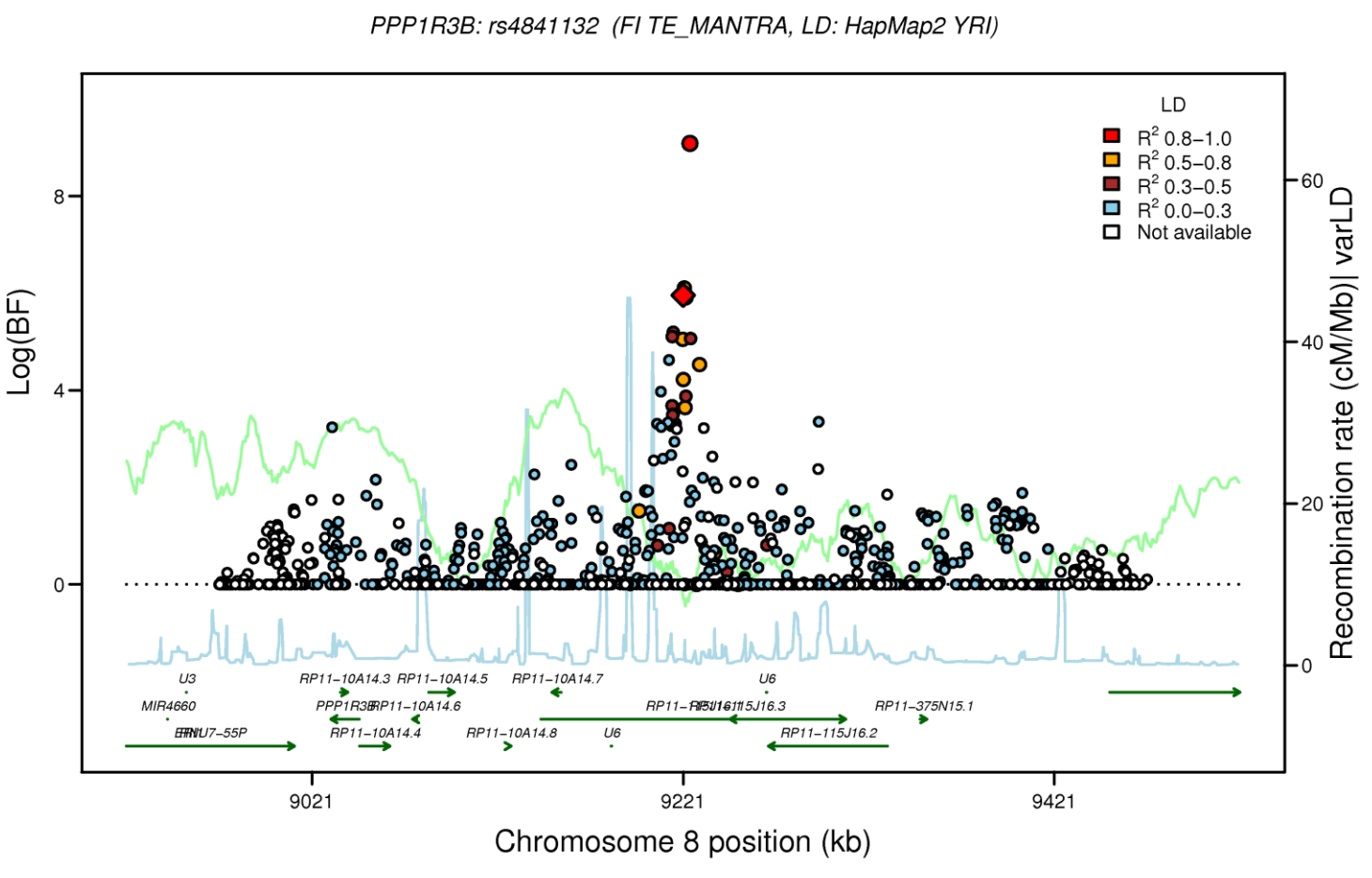
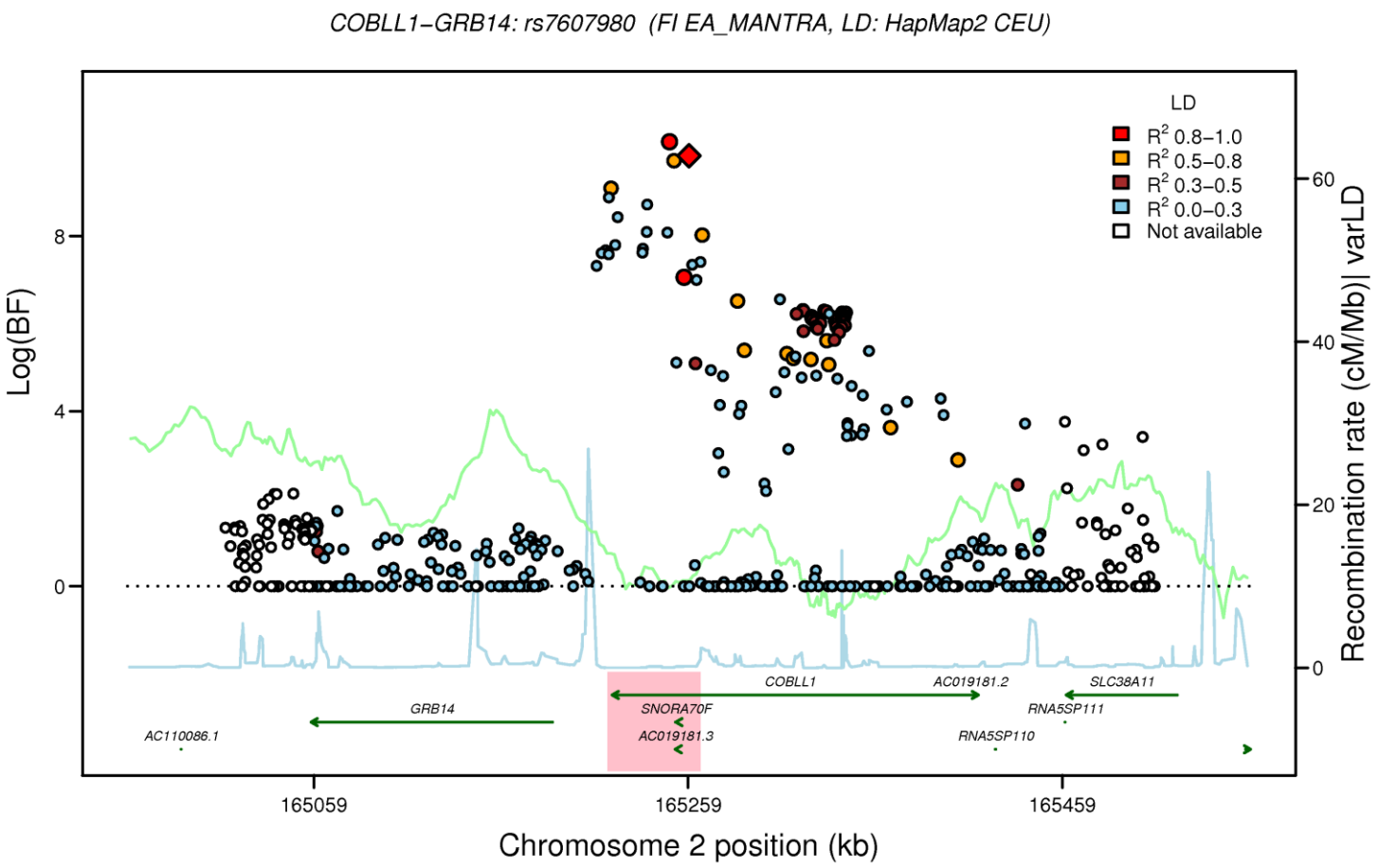
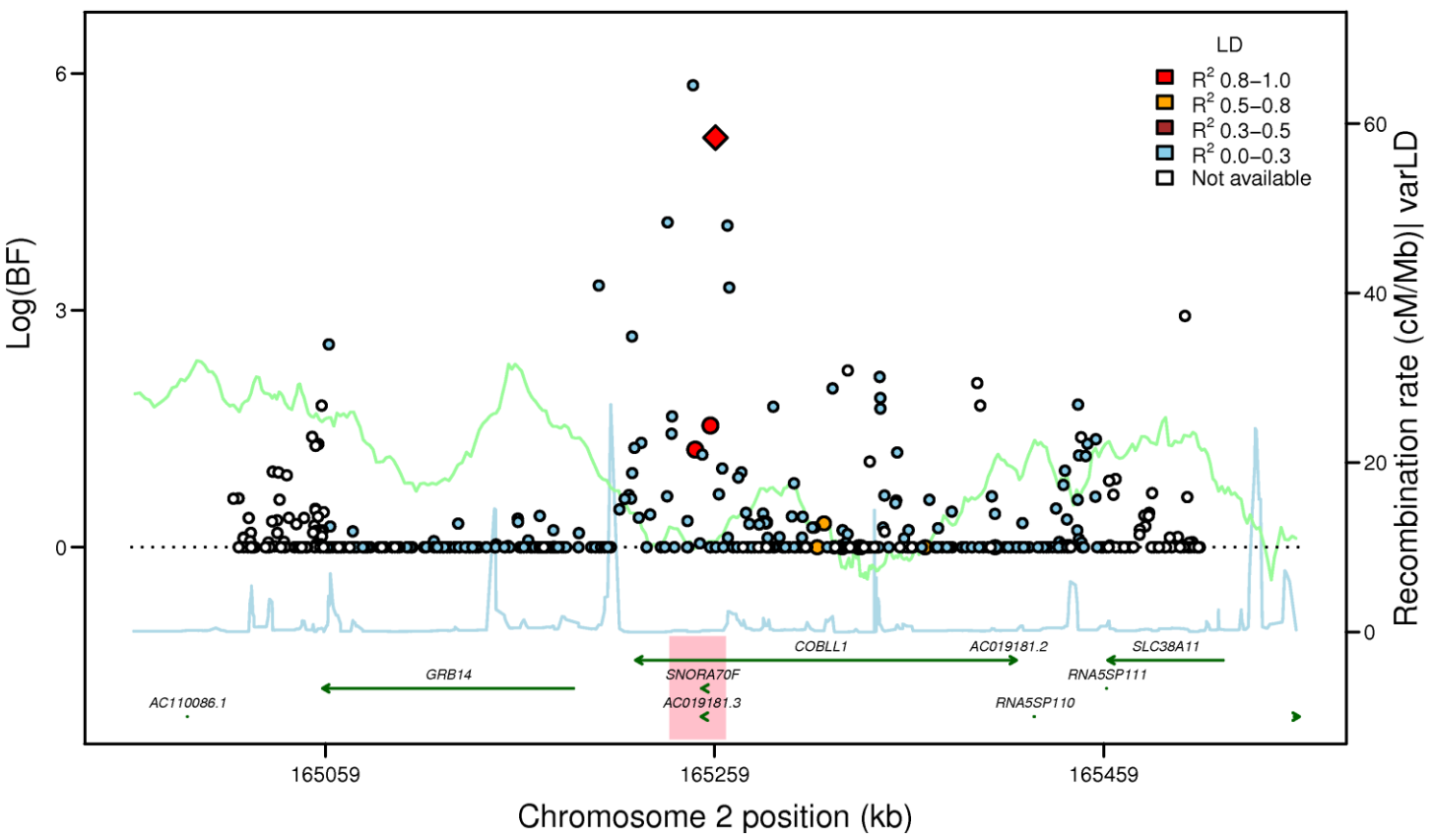

PPP1R3B: rs4841132 (FI TE_MANTRA, LD: HapMap2 YRI)
LD
R2 0.8–1.0
R2 0.5–0.8
R2 0.3–0.5
R2 0.0–0.3
Not available
Log(BF)
Recombination rate (cM/Mb)| varLD
Chromosome 8 position (kb)
9021
9221
9421
U3
MIR4660
RP11–10A14.3
RP11–10A14.5
RP11–10A14.7
U6
RP11–375N15.1
RP11–10A14.4
RP11–10A14.8
RP11–115J16.2

Figure S3Q


COBLL1–GRB14: rs7607980 (FI EA_MANTRA, LD: HapMap2 CEU)
LD
R2 0.8–1.0
R2 0.5–0.8
R2 0.3–0.5
R2 0.0–0.3
Not available
Log(BF)
Recombination rate (cM/Mb)| varLD
COBLL1
AC019181.2
SLC38A11
GRB14
SNORA70F
RNA5SP111
AC110086.1
AC019181.3
RNA5SP110
165059
165259
165459
Chromosome 2 position (kb)


COBLL1–GRB14: rs7607980 (FI AA_MANTRA, LD: HapMap2 YRI)
LD
R2 0.8–1.0
R2 0.5–0.8
R2 0.3–0.5
R2 0.0–0.3
Not available
Log(BF)
Recombination rate (cM/Mb)| varLD
COBLL1
AC019181.2
SLC38A11
GRB14
SNORA70F
RNA5SP111
AC110086.1
AC019181.3
RNA5SP110
165059
165259
165459
Chromosome 2 position (kb)

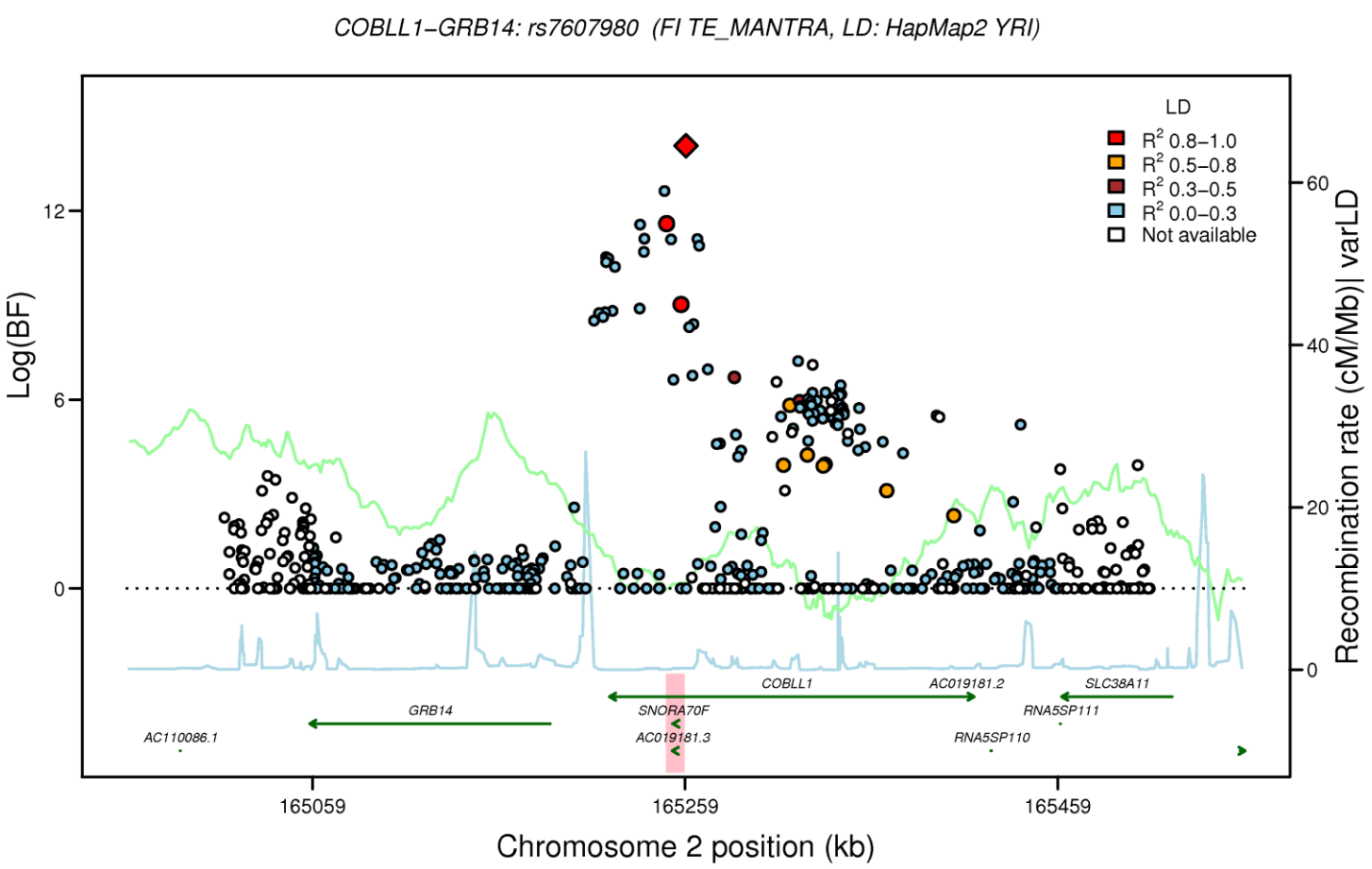
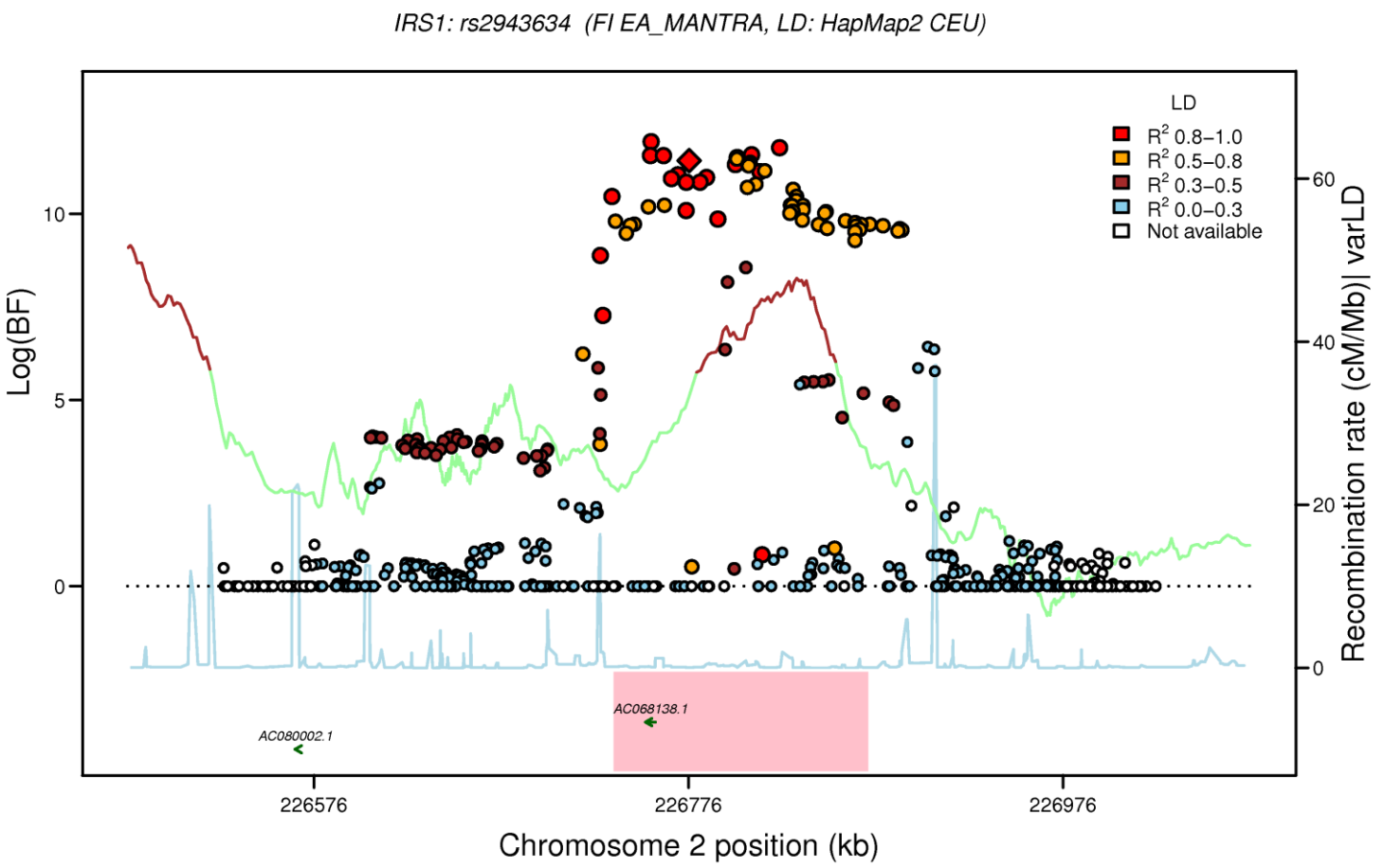
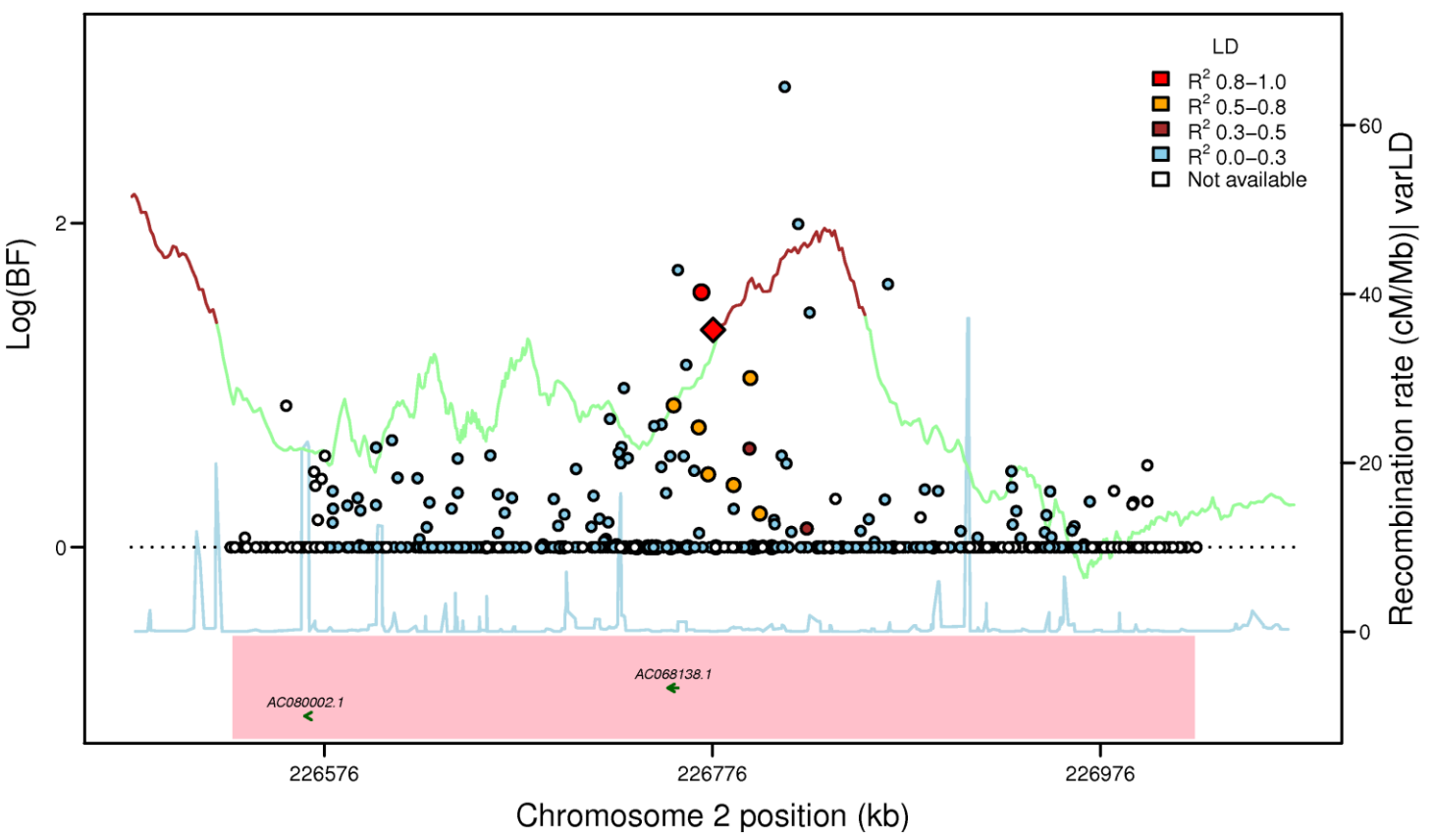

COBLL1–GRB14: rs7607980 (FI TE_MANTRA, LD: HapMap2 YRI)
LD
R2 0.8–1.0
R2 0.5–0.8
R2 0.3–0.5
R2 0.0–0.3
Not available
Log(BF)
Recombination rate (cM/Mb)| varLD
COBLL1
AC019181.2
SLC38A11
GRB14
SNORA70F
RNA5SP111
AC110086.1
AC019181.3
RNA5SP110
165059
165259
165459
Chromosome 2 position (kb)

Figure S3R


IRS1: rs2943634 (FI EA_MANTRA, LD: HapMap2 CEU)
LD
R² 0.8–1.0
R² 0.5–0.8
R² 0.3–0.5
R² 0.0–0.3
Not available
Log(BF)
Recombination rate (cM/Mb)| varLD
AC080002.1
AC068138.1
226576
226776
226976
Chromosome 2 position (kb)


IRS1: rs2943634 (FI AA_MANTRA, LD: HapMap2 YRI)
LD
R² 0.8–1.0
R² 0.5–0.8
R² 0.3–0.5
R² 0.0–0.3
Not available
Log(BF)
Recombination rate (cM/Mb)| varLD
AC080002.1
AC068138.1
226576
226776
226976
Chromosome 2 position (kb)

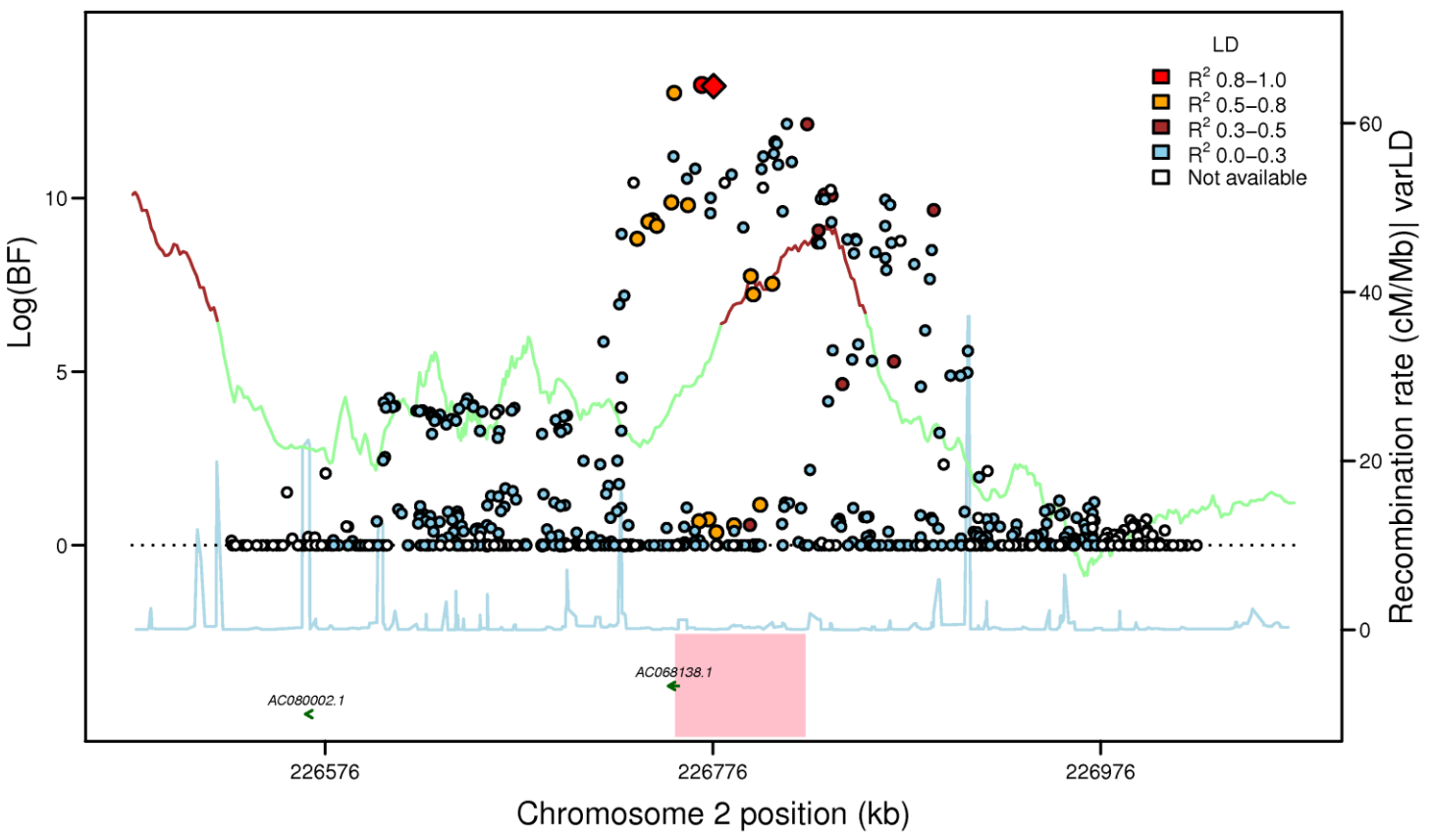
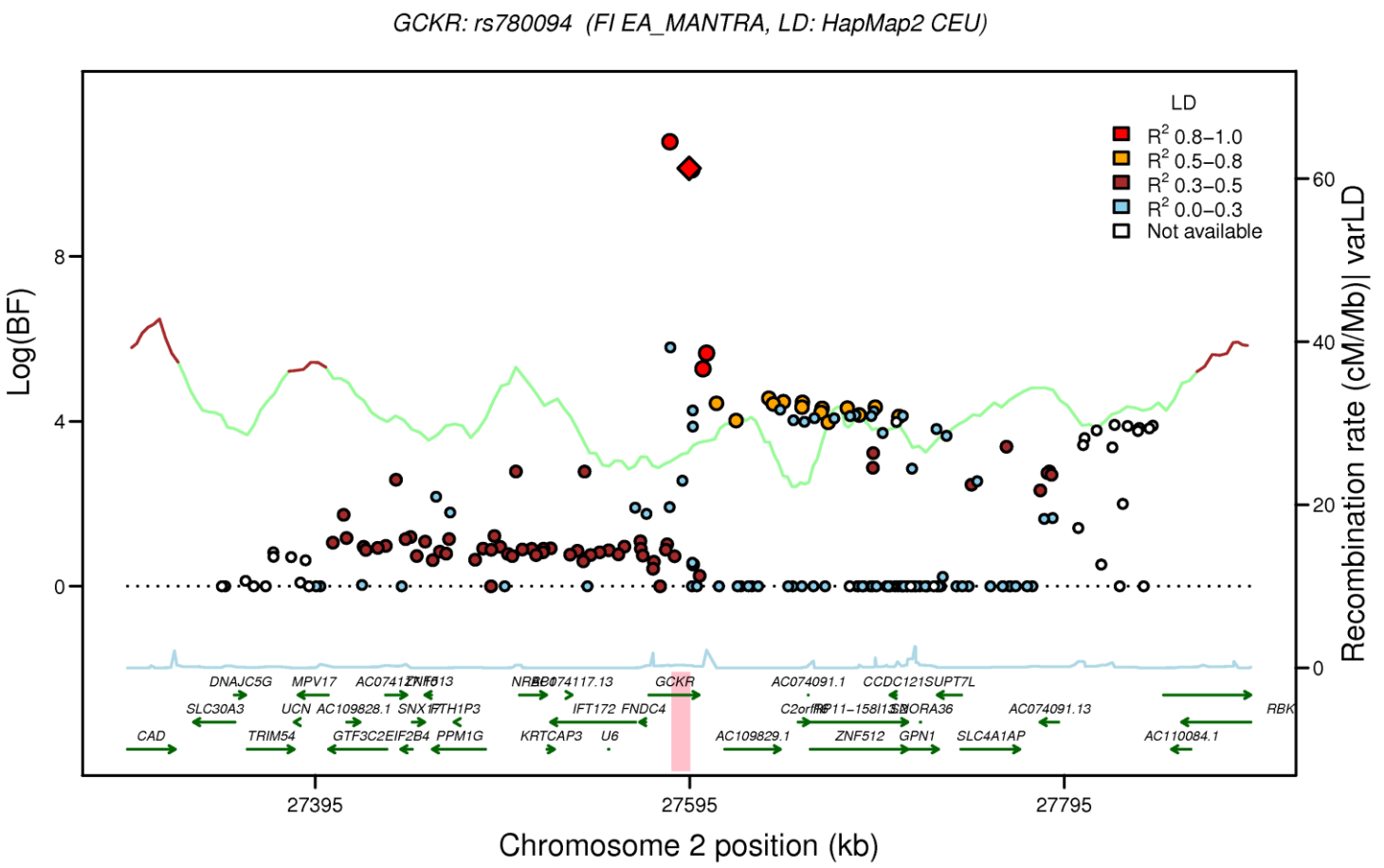
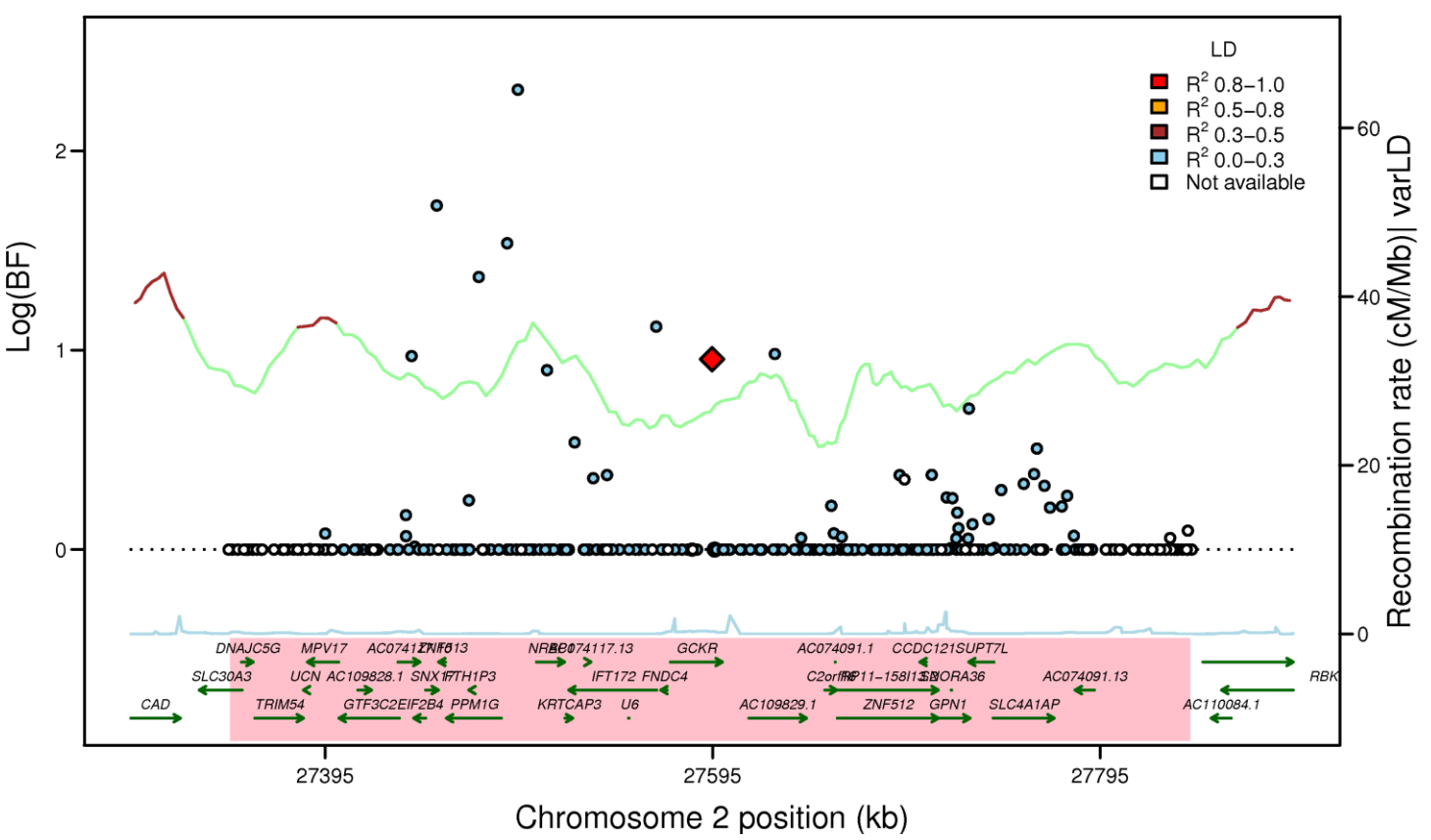

IRS1: rs2943634 (FI TE_MANTRA, LD: HapMap2 YRI)
LD
R² 0.8–1.0
R² 0.5–0.8
R² 0.3–0.5
R² 0.0–0.3
Not available
Log(BF)
Recombination rate (cM/Mb)| varLD
AC080002.1
AC068138.1
226576
226776
226976
Chromosome 2 position (kb)

Figure S3S


GCKR: rs780094 (FI EA_MANTRA, LD: HapMap2 CEU)
LD
R2 0.8–1.0
R2 0.5–0.8
R2 0.3–0.5
R2 0.0–0.3
Not available
Log(BF)
Recombination rate (cM/Mb)| varLD
27395
27595
27795
Chromosome 2 position (kb)
DNAJC5G MPV17 SLC30A3 UCN AC109828.1 SNX17 FTH1P3 IFT172 FNDC4 GCKR AC074091.1 CCDC121SUPT7L SLCO4A1-AS1 SLCORA36 AC074091.13 RBK
CAD TRIM54 GTF3C2 EIF2B4 PPM1G KRTCAP3 U6 AC109829.1 ZNF512 GPN1 SLC4A1AP AC110084.1


GCKR: rs780094 (FI AA_MANTRA, LD: HapMap2 YRI)
LD
R2 0.8–1.0
R2 0.5–0.8
R2 0.3–0.5
R2 0.0–0.3
Not available
Log(BF)
Recombination rate (cM/Mb)| varLD
27395
27595
27795
Chromosome 2 position (kb)
DNAJC5G MPV17 SLC30A3 UCN AC109828.1 SNX17 FTH1P3 IFT172 FNDC4 GCKR AC074091.1 CCDC121SUPT7L SLCO4A1-AS1 SLCORA36 AC074091.13 RBK
CAD TRIM54 GTF3C2 EIF2B4 PPM1G KRTCAP3 U6 AC109829.1 ZNF512 GPN1 SLC4A1AP AC110084.1

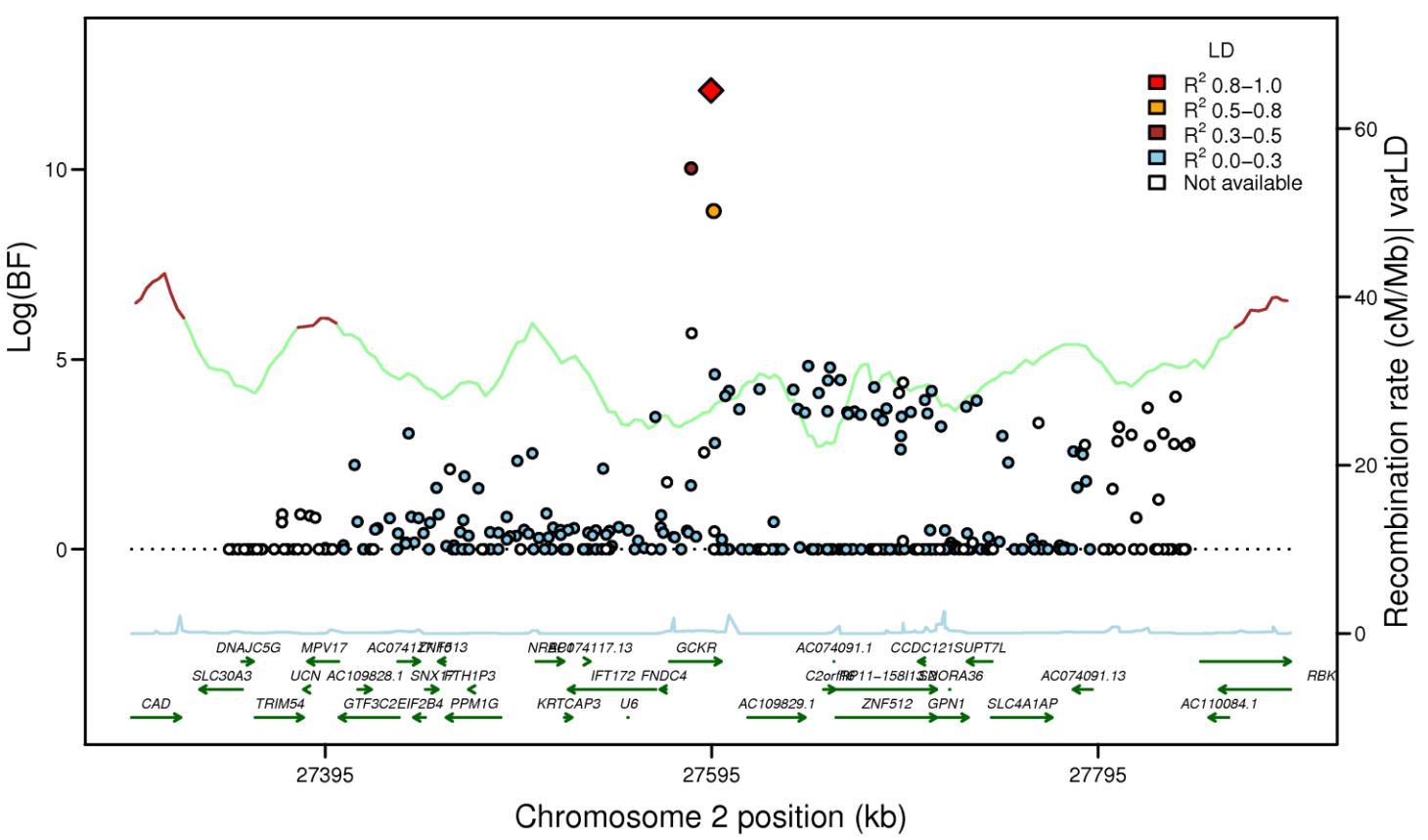
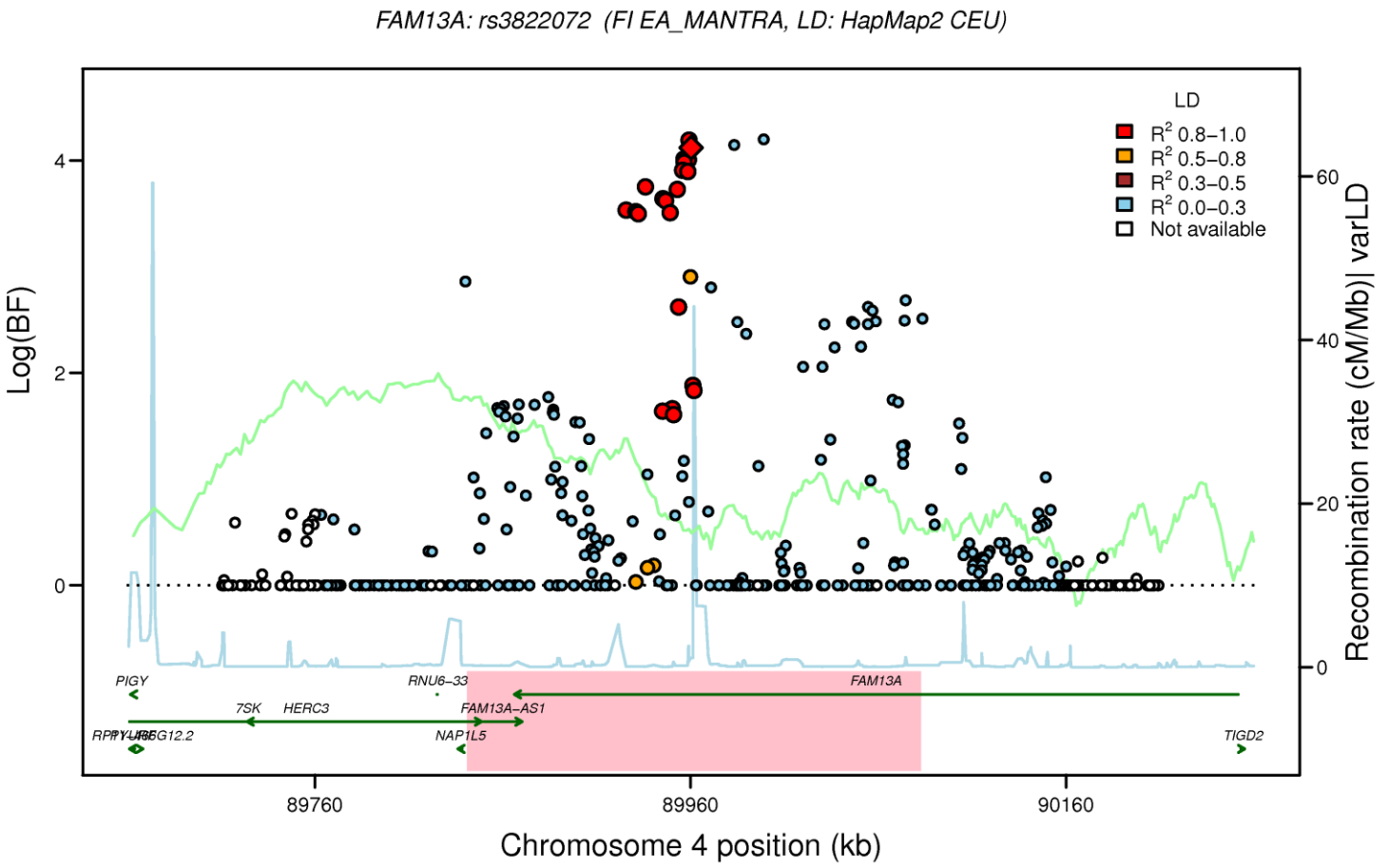
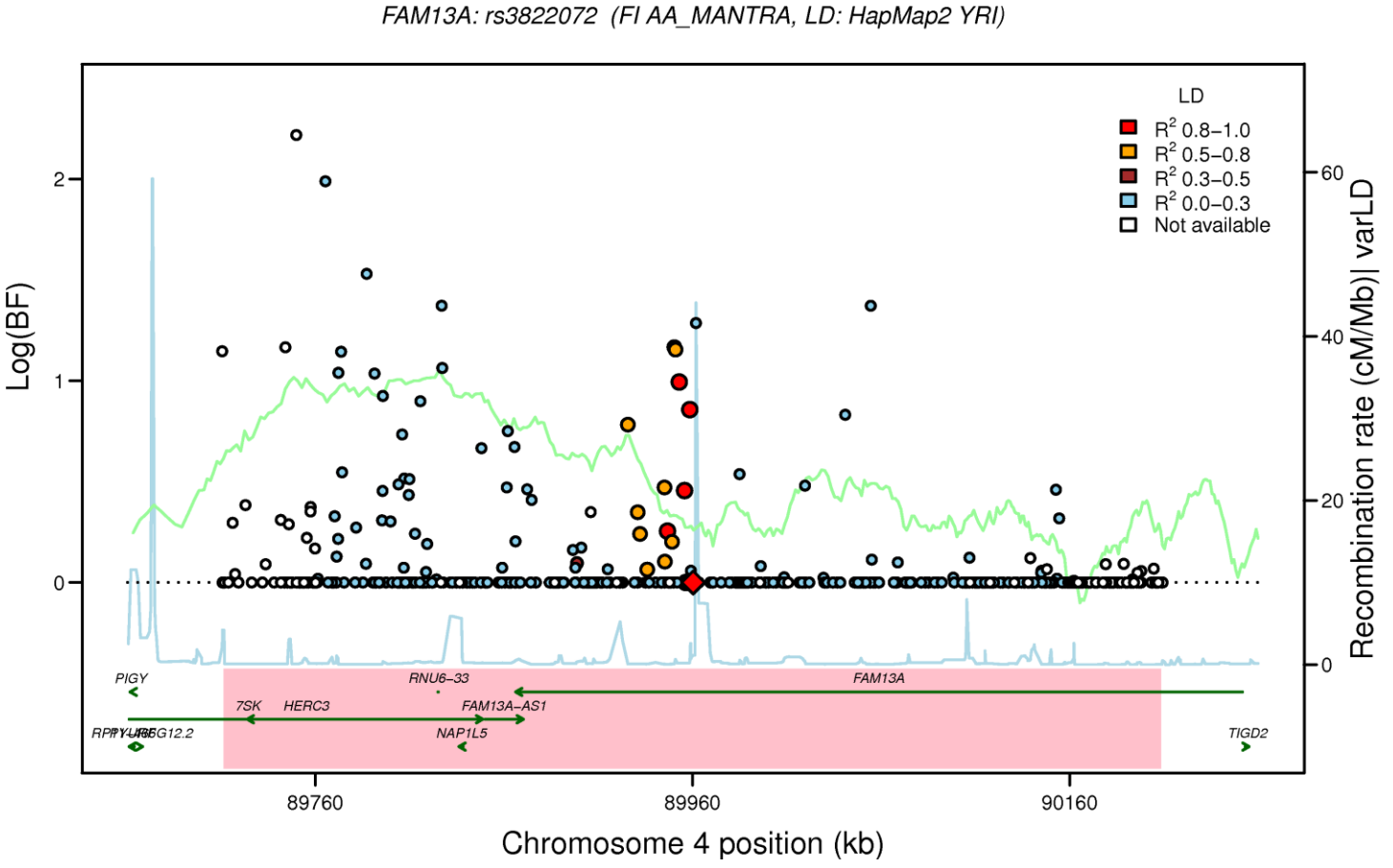

GCKR: rs780094 (FI TE_MANTRA, LD: HapMap2 YRI)
LD
R2 0.8–1.0
R2 0.5–0.8
R2 0.3–0.5
R2 0.0–0.3
Not available
Log(BF)
Recombination rate (cM/Mb)| varLD
27395
27595
27795
Chromosome 2 position (kb)
DNAJC5G MPV17 SLC30A3 UCN AC109828.1 SNX17 FTH1P3 IFT172 FNDC4 GCKR AC074091.1 CCDC121SUPT7L SLCO4A1-AS1 SLCORA36 AC074091.13 RBK
CAD TRIM54 GTF3C2 EIF2B4 PPM1G KRTCAP3 U6 AC109829.1 ZNF512 GPN1 SLC4A1AP AC110084.1

Figure S3T


FAM13A: rs3822072 (FI EA_MANTRA, LD: HapMap2 CEU)
LD
R2 0.8–1.0
R2 0.5–0.8
R2 0.3–0.5
R2 0.0–0.3
Not available
Log(BF)
Recombination rate (cM/Mb)| varLD
PIGY
RNU6–33
FAM13A
7SK
HERC3
FAM13A–AS1
NAP1L5
TIGD2
89760
89960
90160
Chromosome 4 position (kb)


FAM13A: rs3822072 (FI AA_MANTRA, LD: HapMap2 YRI)
LD
R2 0.8–1.0
R2 0.5–0.8
R2 0.3–0.5
R2 0.0–0.3
Not available
Log(BF)
Recombination rate (cM/Mb)| varLD
PIGY
RNU6–33
FAM13A
7SK
HERC3
FAM13A–AS1
NAP1L5
TIGD2
89760
89960
90160
Chromosome 4 position (kb)

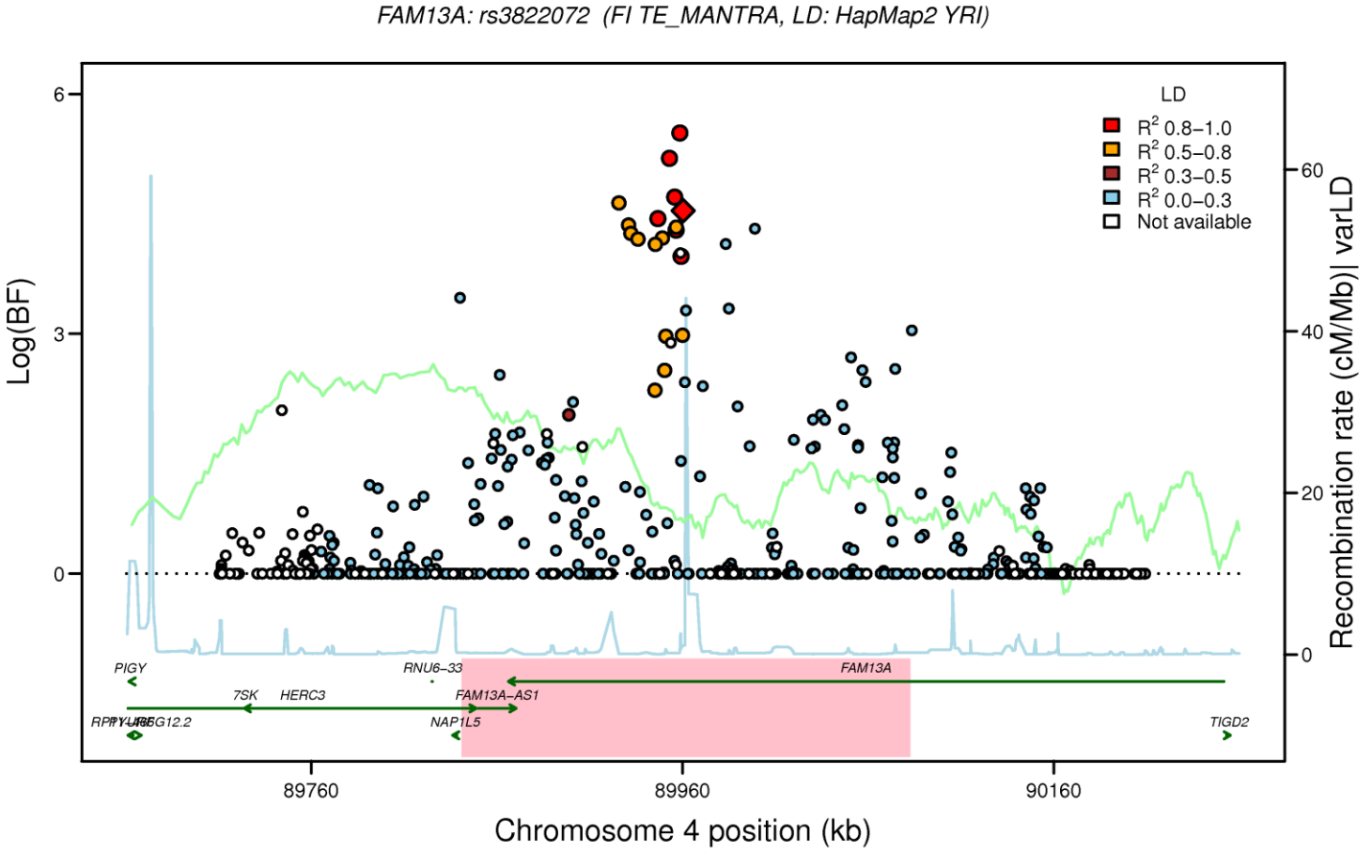
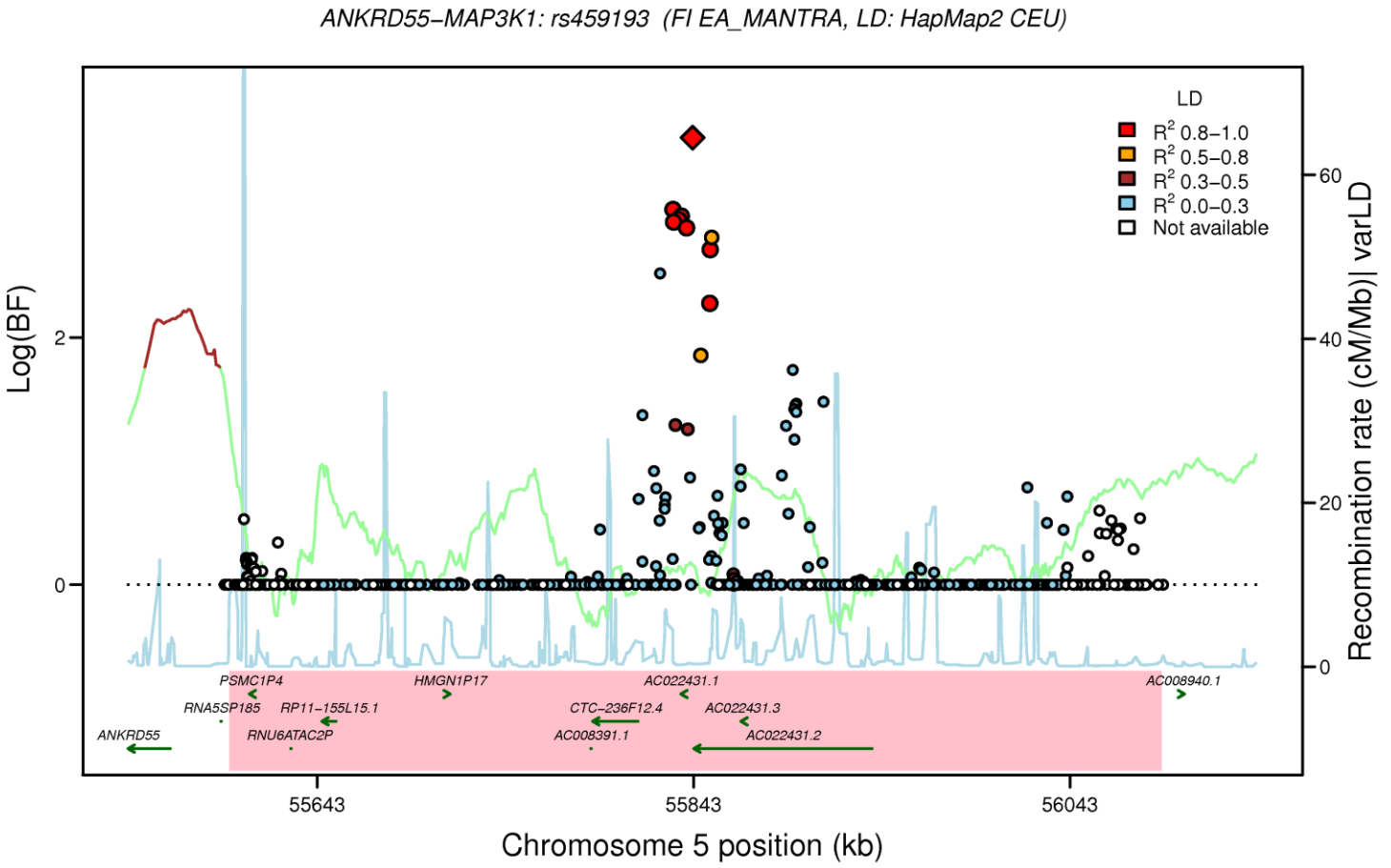
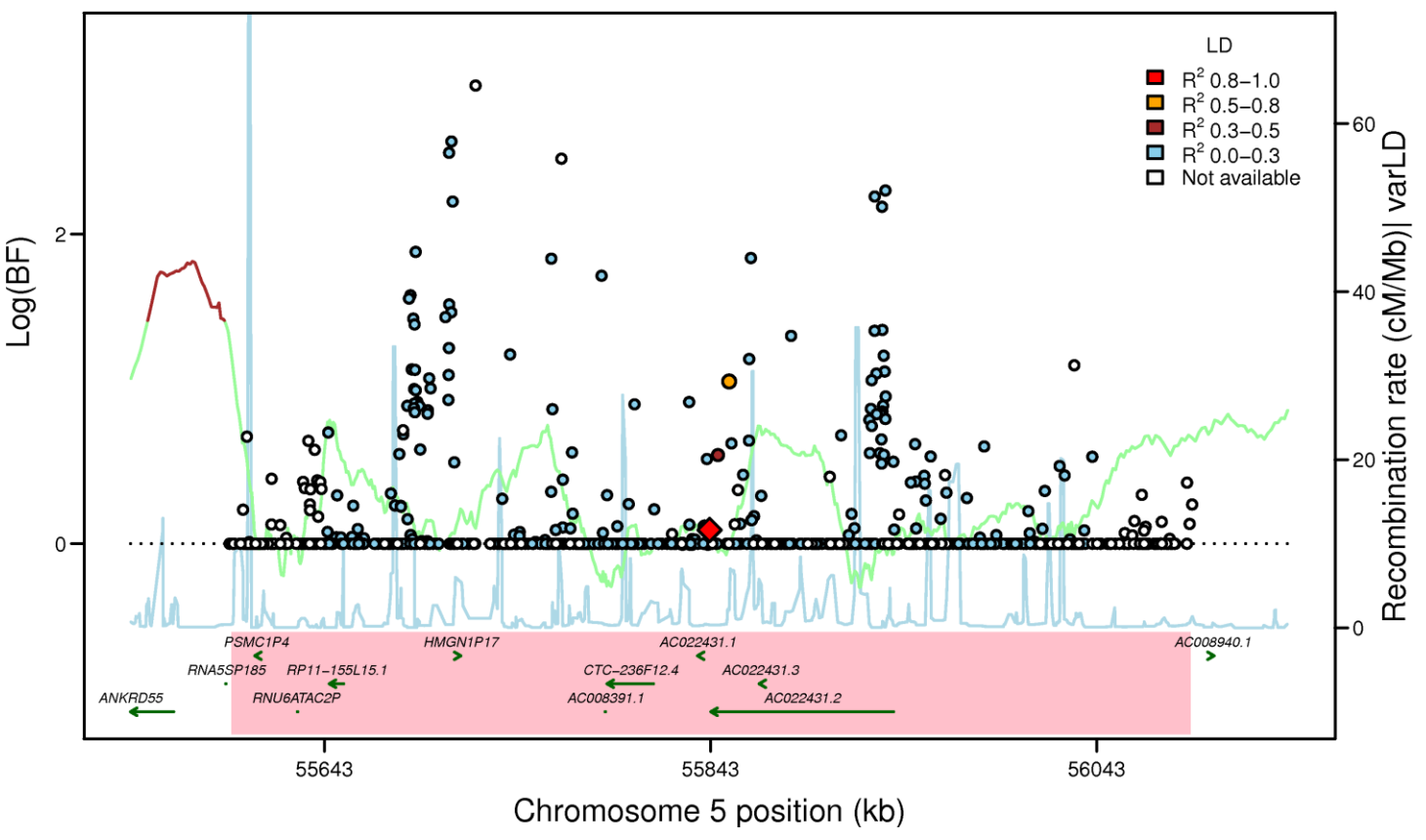

FAM13A: rs3822072 (FI TE_MANTRA, LD: HapMap2 YRI)
LD
R2 0.8–1.0
R2 0.5–0.8
R2 0.3–0.5
R2 0.0–0.3
Not available
Log(BF)
Recombination rate (cM/Mb)| varLD
PIGY
RNU6–33
FAM13A
7SK
HERC3
FAM13A–AS1
NAP1L5
TIGD2
89760
89960
90160
Chromosome 4 position (kb)

Figure S3U


ANKRD55–MAP3K1: rs459193 (FI EA_MANTRA, LD: HapMap2 CEU)
LD
R2 0.8–1.0
R2 0.5–0.8
R2 0.3–0.5
R2 0.0–0.3
Not available
Log(BF)
Recombination rate (cM/Mb)| varLD
PSMC1P4
HMGN1P17
AC022431.1
AC008940.1
RNA5SP185
RP11–155L15.1
CTC–236F12.4
AC022431.3
ANKRD55
RNU6ATAC2P
AC008391.1
AC022431.2
55643
55843
56043
Chromosome 5 position (kb)


ANKRD55–MAP3K1: rs459193 (FI AA_MANTRA, LD: HapMap2 YRI)
LD
R2 0.8–1.0
R2 0.5–0.8
R2 0.3–0.5
R2 0.0–0.3
Not available
Log(BF)
Recombination rate (cM/Mb)| varLD
PSMC1P4
HMGN1P17
AC022431.1
AC008940.1
RNA5SP185
RP11–155L15.1
CTC–236F12.4
AC022431.3
ANKRD55
RNU6ATAC2P
AC008391.1
AC022431.2
55643
55843
56043
Chromosome 5 position (kb)

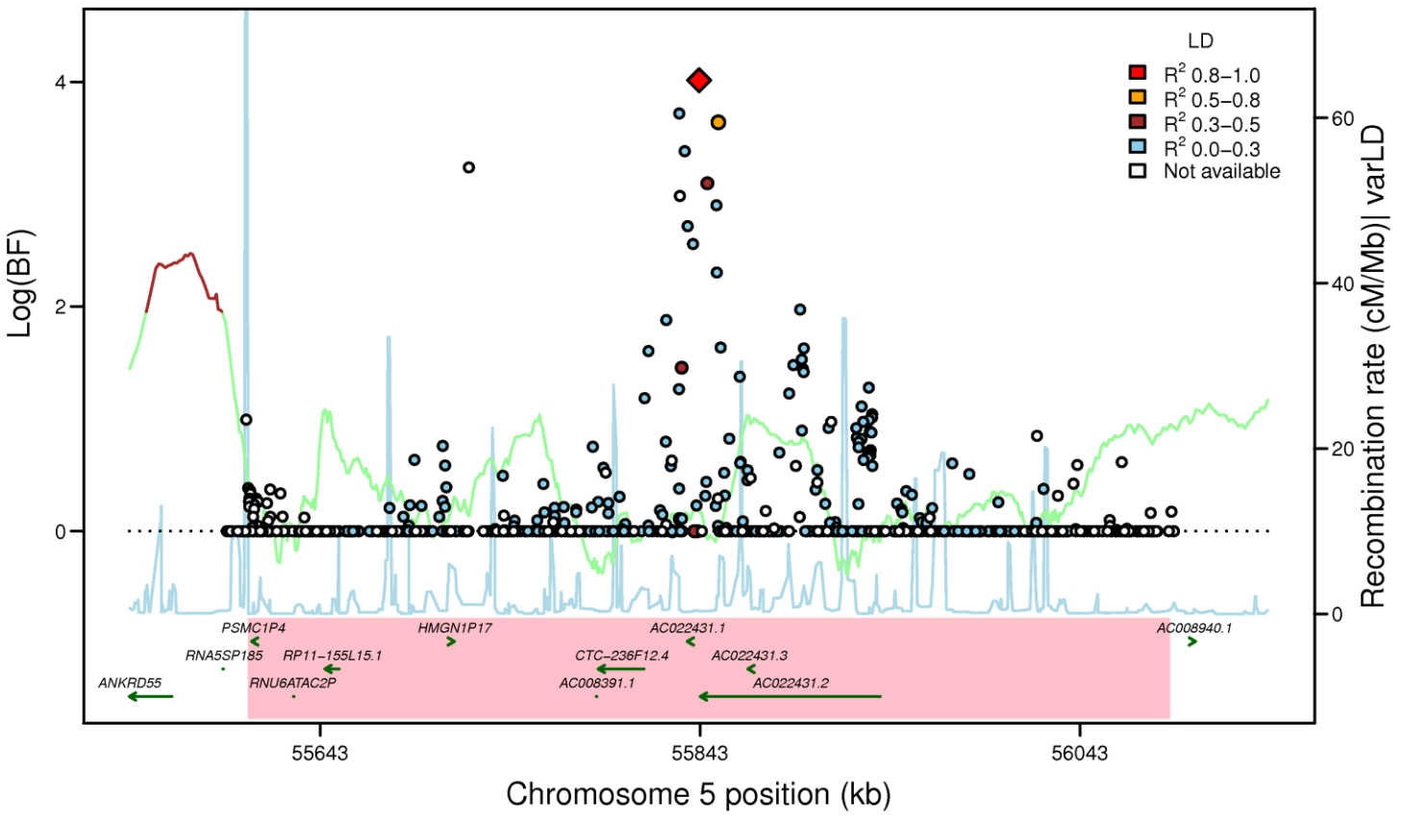
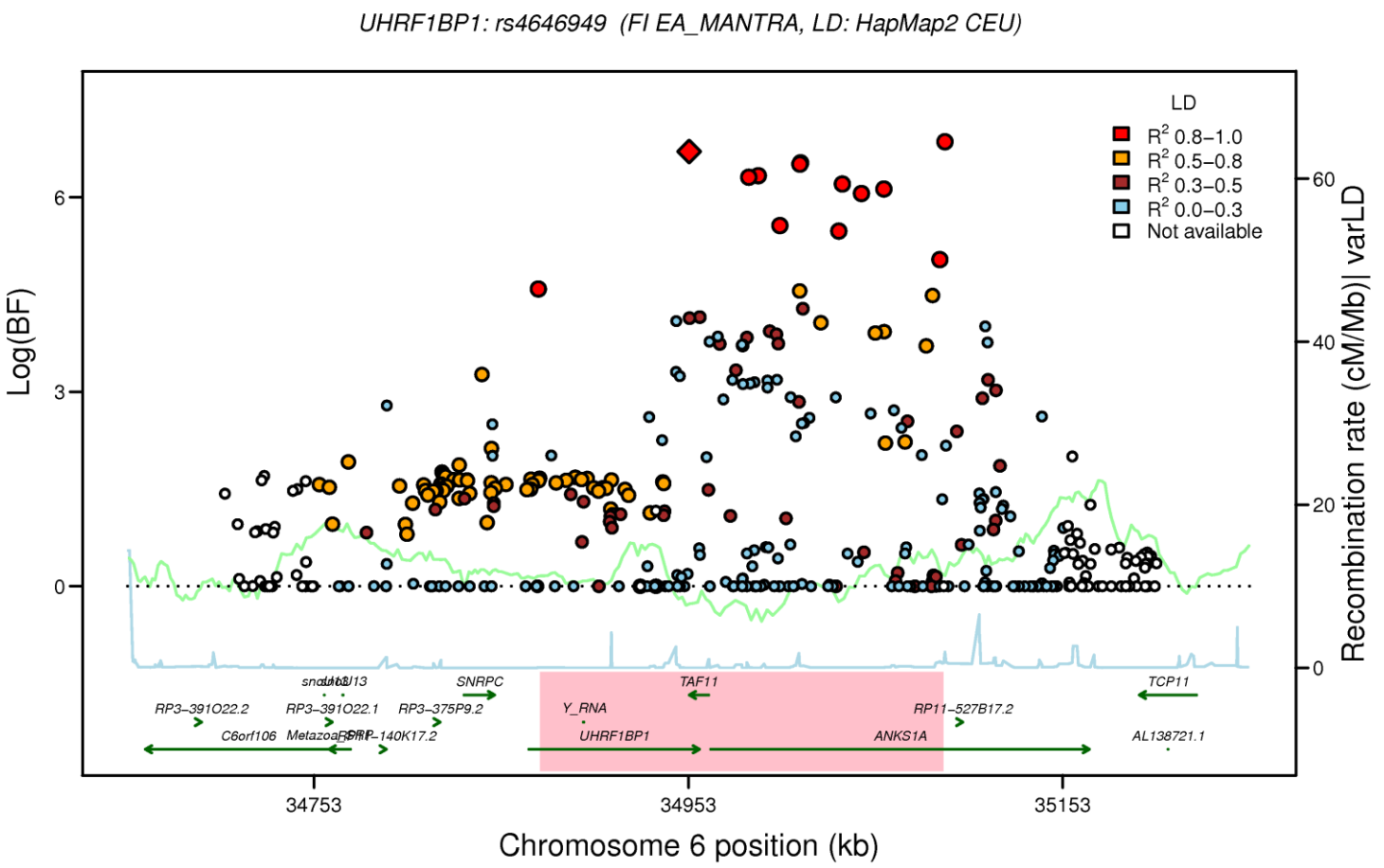
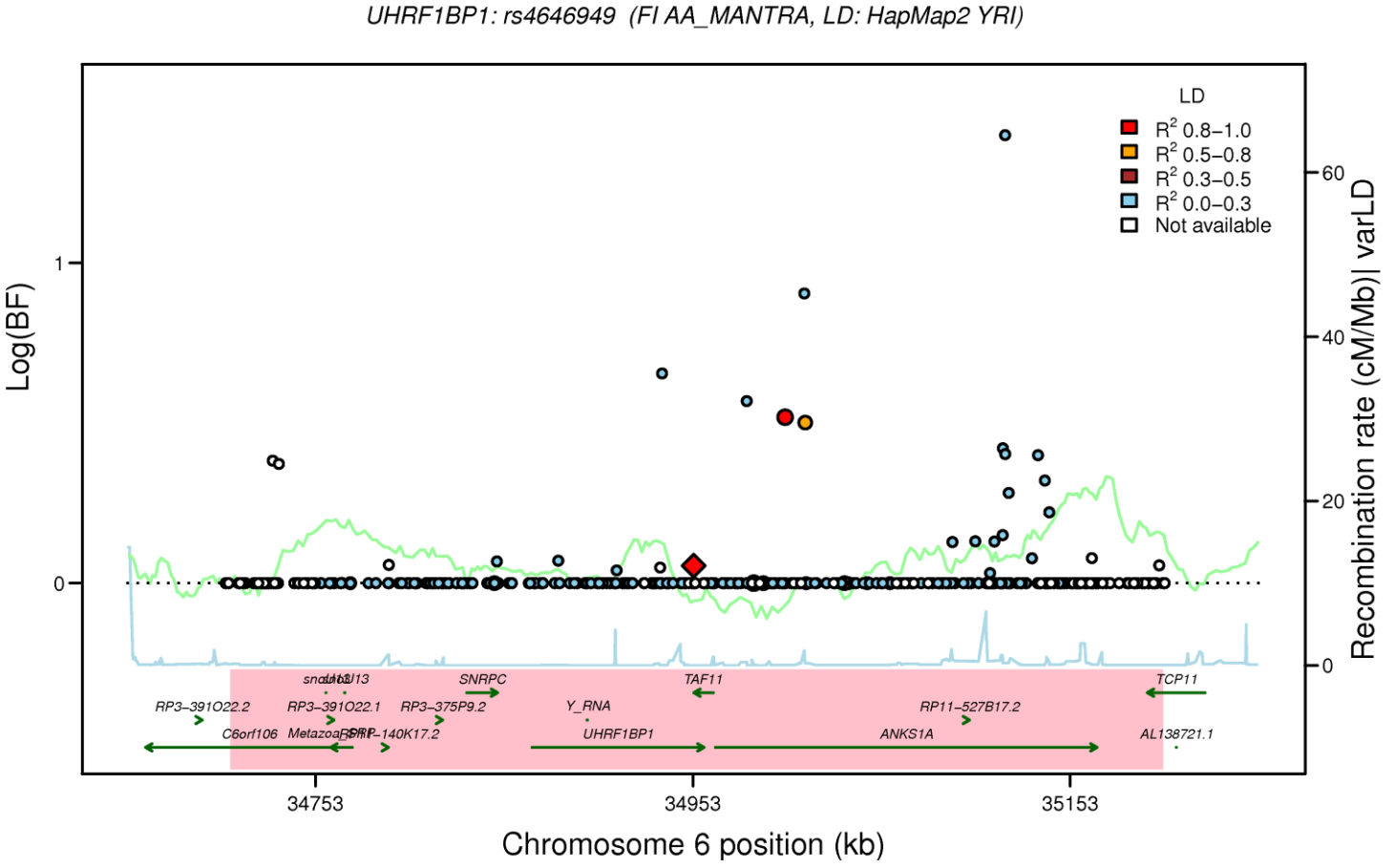

ANKRD55–MAP3K1: rs459193 (FI TE_MANTRA, LD: HapMap2 YRI)
LD
R2 0.8–1.0
R2 0.5–0.8
R2 0.3–0.5
R2 0.0–0.3
Not available
Log(BF)
Recombination rate (cM/Mb)| varLD
PSMC1P4
HMGN1P17
AC022431.1
AC008940.1
RNA5SP185
RP11–155L15.1
CTC–236F12.4
AC022431.3
ANKRD55
RNU6ATAC2P
AC008391.1
AC022431.2
55643
55843
56043
Chromosome 5 position (kb)

Figure S3V


UHRF1BP1: rs4646949 (FI EA_MANTRA, LD: HapMap2 CEU)
LD
R2 0.8–1.0
R2 0.5–0.8
R2 0.3–0.5
R2 0.0–0.3
Not available
Log(BF)
Recombination rate (cM/Mb)| varLD
RP3–391O22.2
RP3–391O22.1
RP3–375P9.2
SNRPC
TAF11
TCP11
Y_RNA
RP11–527B17.2
C6orf106
UHRF1BP1
ANKS1A
AL138721.1
34753
34953
35153
Chromosome 6 position (kb)


UHRF1BP1: rs4646949 (FI AA_MANTRA, LD: HapMap2 YRI)
LD
R2 0.8–1.0
R2 0.5–0.8
R2 0.3–0.5
R2 0.0–0.3
Not available
Log(BF)
Recombination rate (cM/Mb)| varLD
RP3–391O22.2
RP3–391O22.1
RP3–375P9.2
SNRPC
TAF11
TCP11
Y_RNA
RP11–527B17.2
C6orf106
UHRF1BP1
ANKS1A
AL138721.1
34753
34953
35153
Chromosome 6 position (kb)

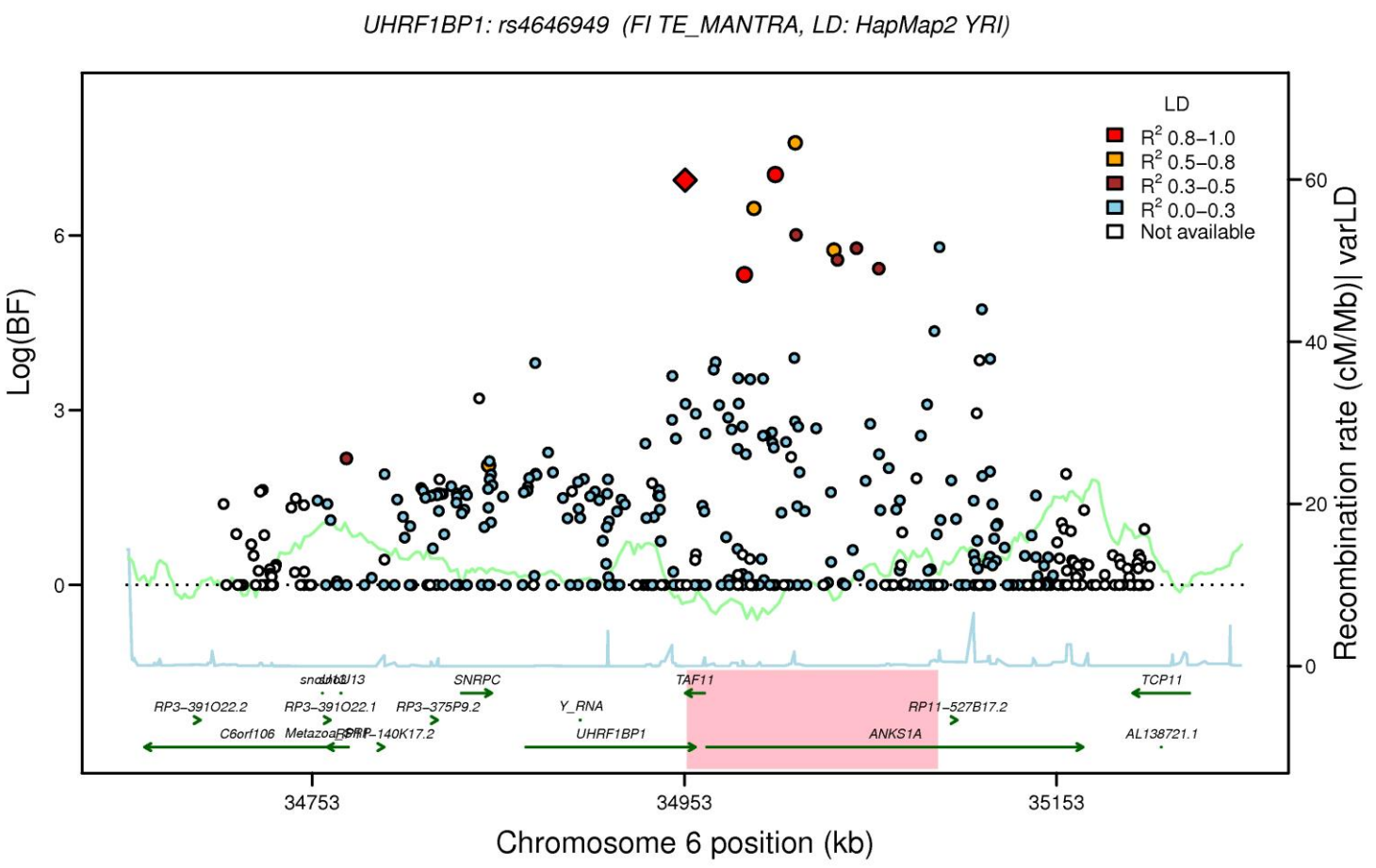
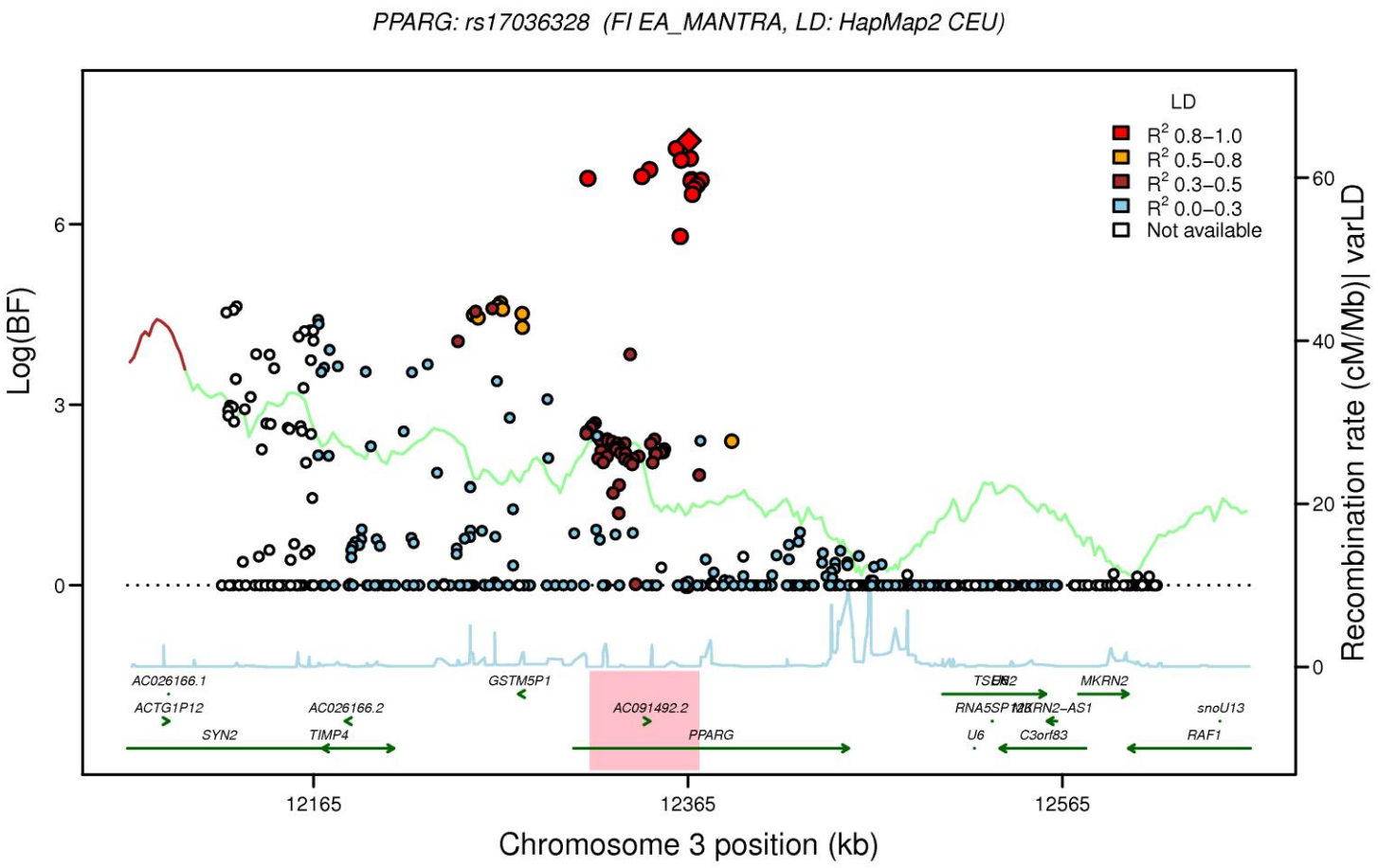
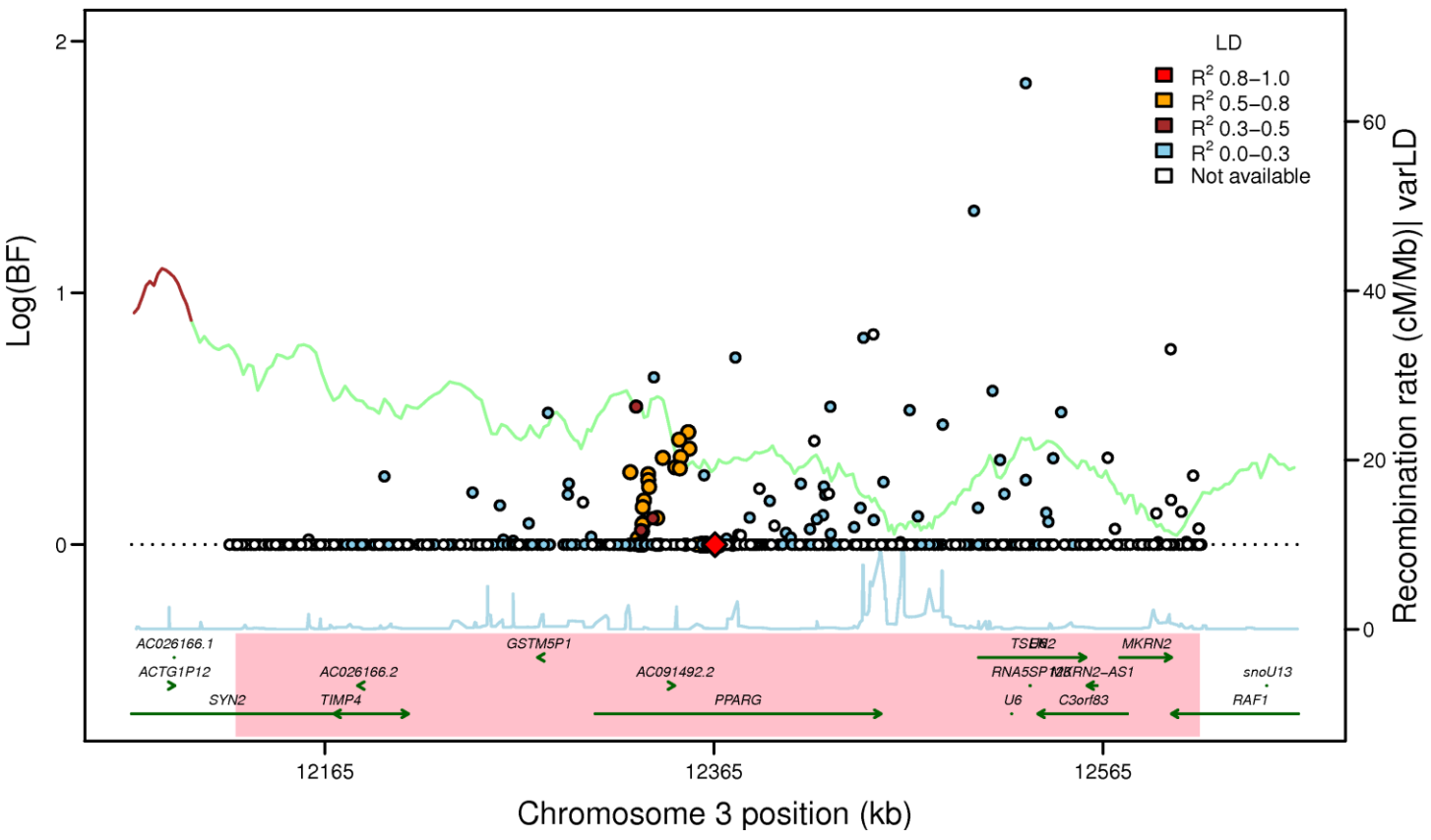

UHRF1BP1: rs4646949 (FI TE_MANTRA, LD: HapMap2 YRI)
LD
R2 0.8–1.0
R2 0.5–0.8
R2 0.3–0.5
R2 0.0–0.3
Not available
Log(BF)
Recombination rate (cM/Mb)| varLD
RP3–391O22.2
RP3–391O22.1
RP3–375P9.2
SNRPC
TAF11
TCP11
Y_RNA
RP11–527B17.2
C6orf106
UHRF1BP1
ANKS1A
AL138721.1
34753
34953
35153
Chromosome 6 position (kb)

Figure S3W


PPARG: rs17036328 (FI EA_MANTRA, LD: HapMap2 CEU)
LD
R² 0.8–1.0
R² 0.5–0.8
R² 0.3–0.5
R² 0.0–0.3
Not available
Log(BF)
Recombination rate (cM/Mb)| varLD
AC026166.1
ACTG1P12
SYN2
AC026166.2
TIMP4
GSTM5P1
AC091492.2
PPARG
MKRN2
snoU13
U6
C3orf83
RAF1
12165
12365
12565
Chromosome 3 position (kb)


PPARG: rs17036328 (FI AA_MANTRA, LD: HapMap2 YRI)
LD
R² 0.8–1.0
R² 0.5–0.8
R² 0.3–0.5
R² 0.0–0.3
Not available
Log(BF)
Recombination rate (cM/Mb)| varLD
AC026166.1
ACTG1P12
SYN2
AC026166.2
TIMP4
GSTM5P1
AC091492.2
PPARG
MKRN2
snoU13
U6
C3orf83
RAF1
12165
12365
12565
Chromosome 3 position (kb)

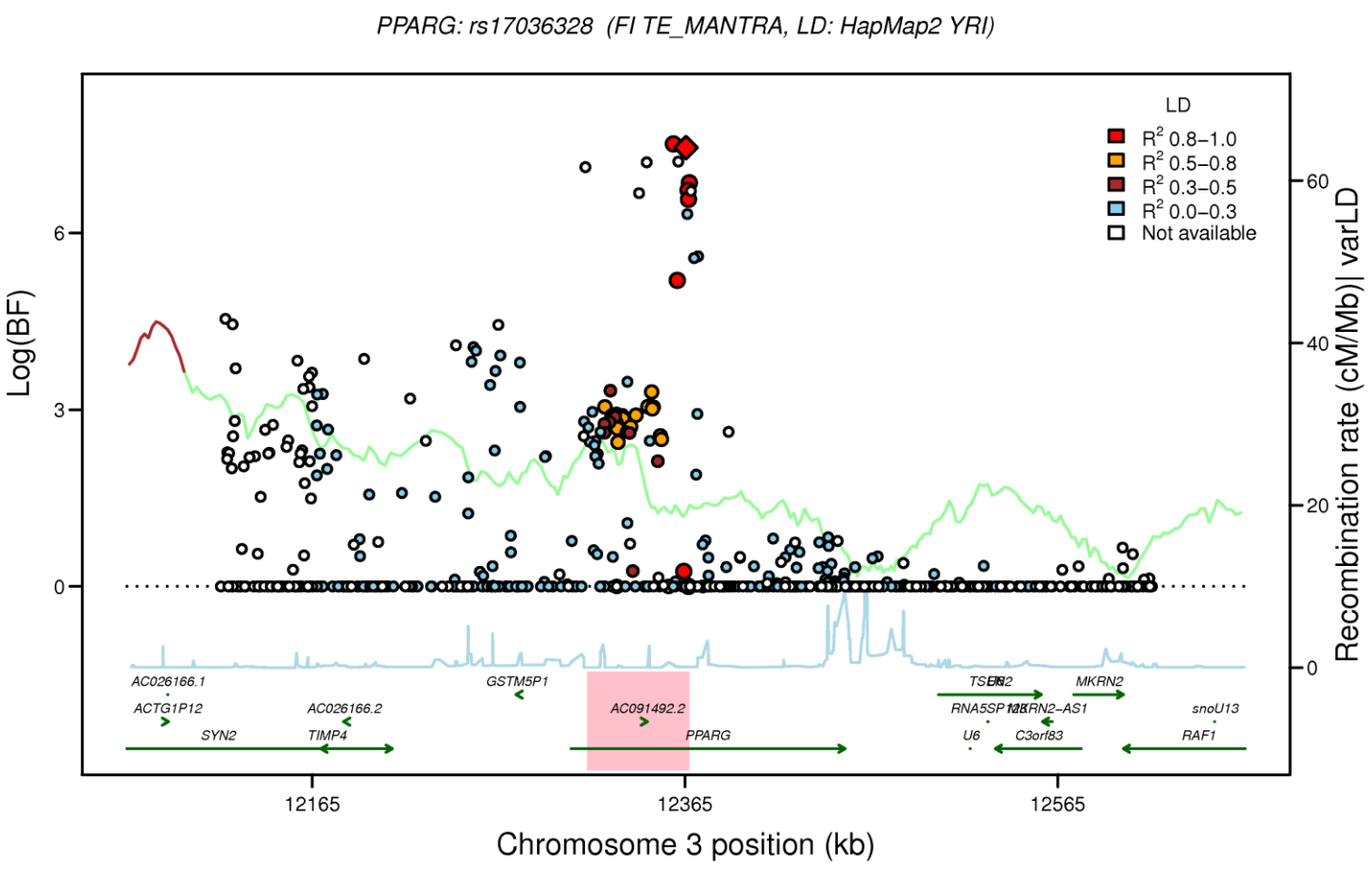
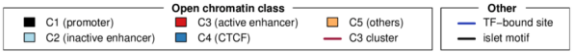
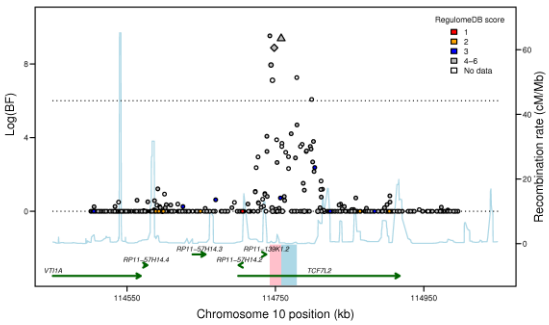

PPARG: rs17036328 (FI TE_MANTRA, LD: HapMap2 YRI)
LD
R² 0.8–1.0
R² 0.5–0.8
R² 0.3–0.5
R² 0.0–0.3
Not available
Log(BF)
Recombination rate (cM/Mb)| varLD
AC026166.1
ACTG1P12
SYN2
AC026166.2
TIMP4
GSTM5P1
AC091492.2
PPARG
MKRN2
snoU13
U6
C3orf83
RAF1
12165
12365
12565
Chromosome 3 position (kb)

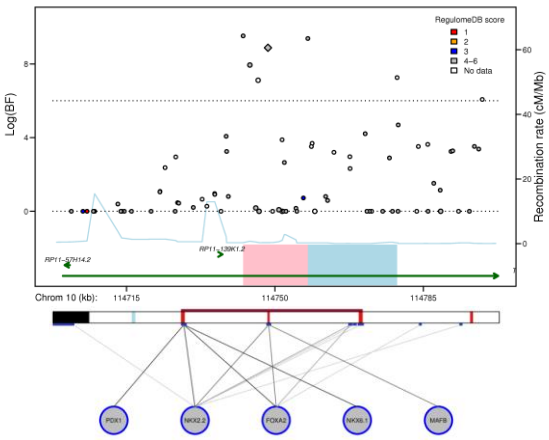
Figure S4. Regional association plots merged with data from RegulomeDB and Islet Regulome Browser at 14 FG and 9 FI loci with substantially narrowed credible sets after trans-ethnic analysis.

FG Loci. Islet Regulome legend:

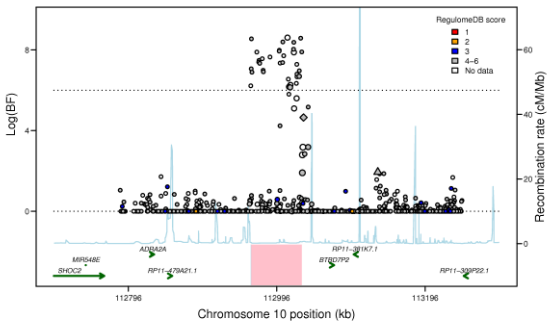


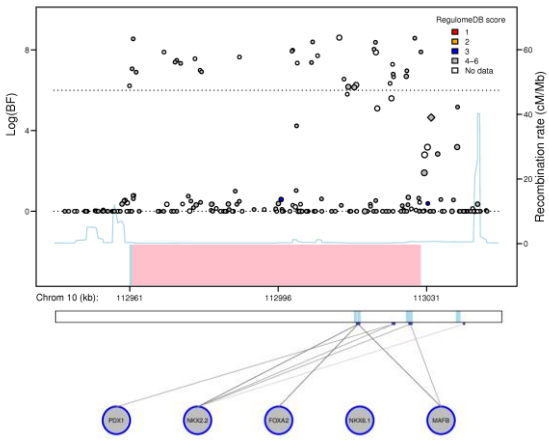
TCF7L2



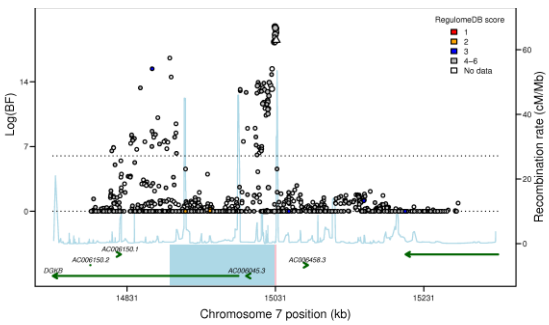


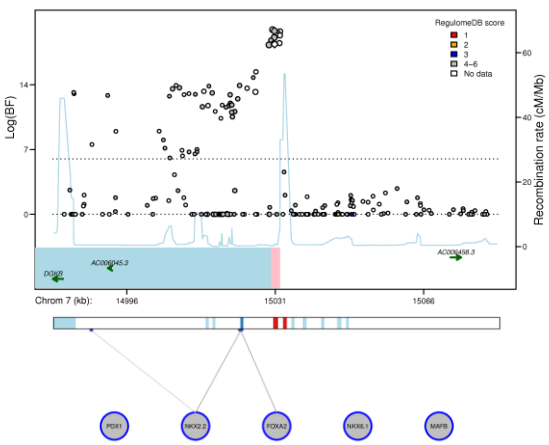
ADRA2



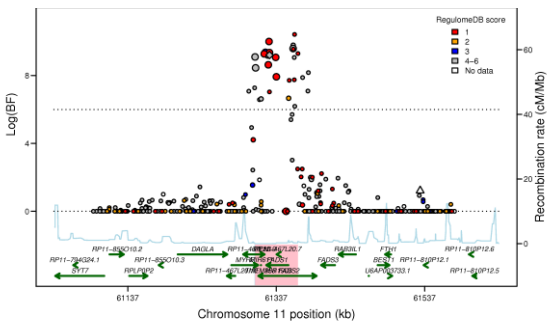


DGKB-TMEM195





FADS1



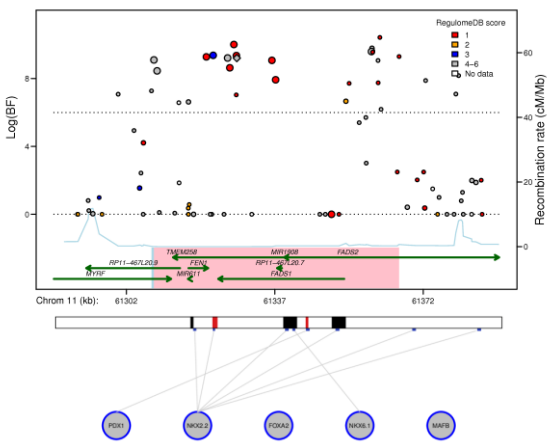
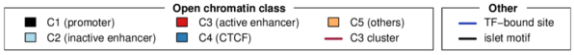
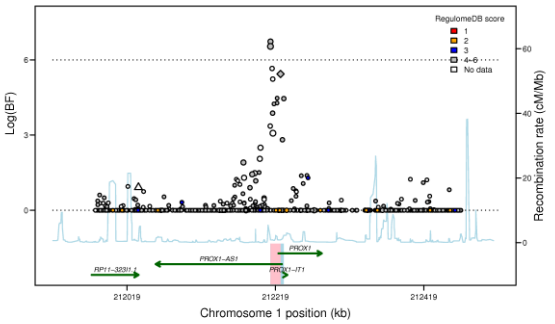


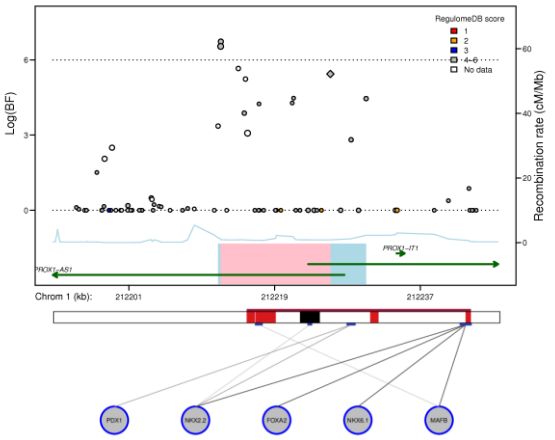
Figure S4, FG Loci continued.

Islet Regulome legend:

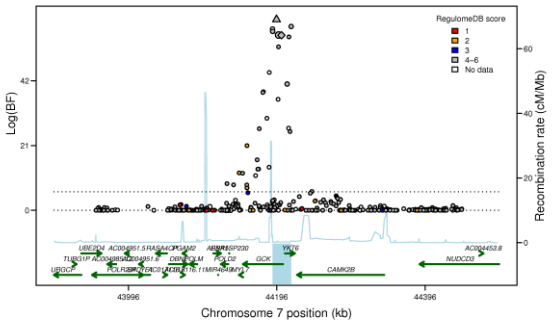


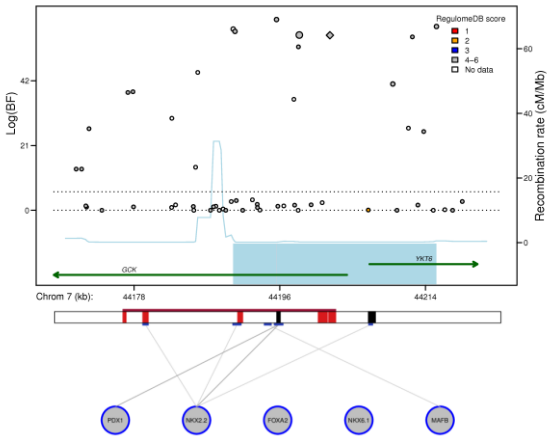
PROX1



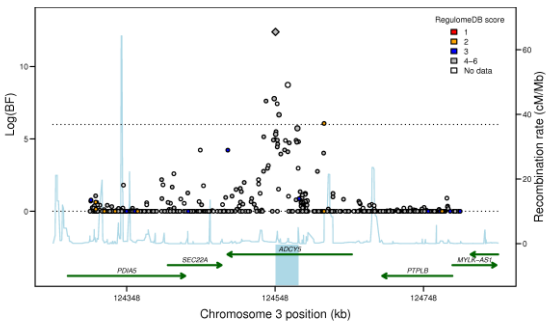


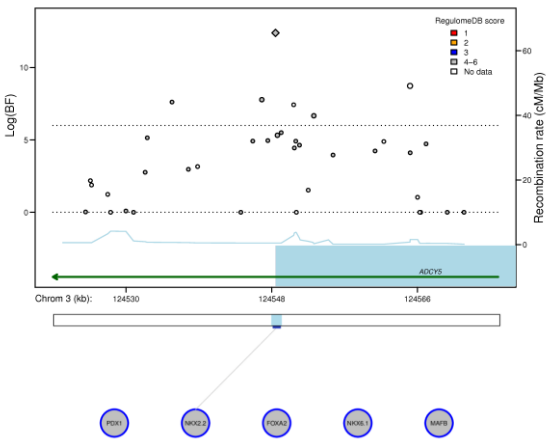
GCK



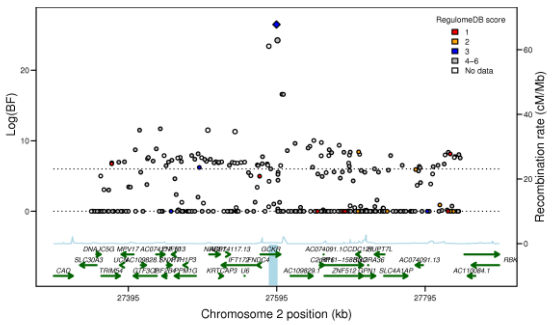


ADCY5





GCKR



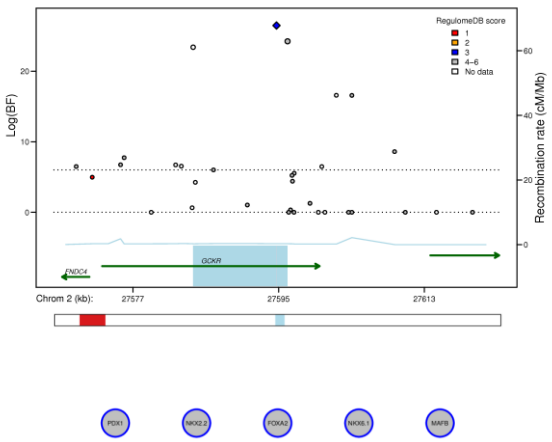
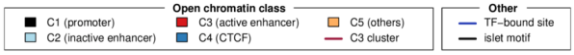
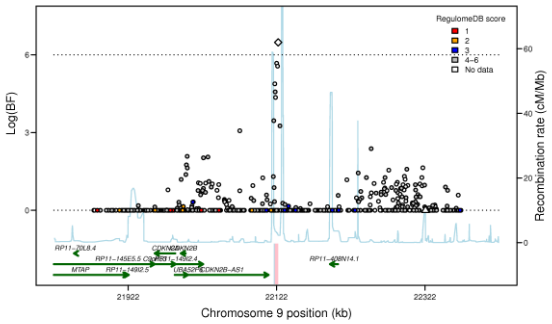


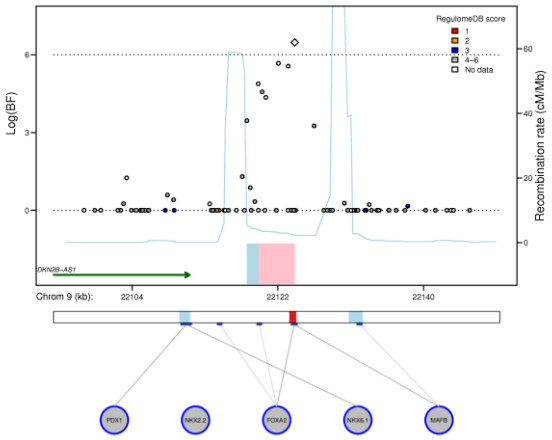
Figure S4, FG Loci continued.

Islet Regulome legend:

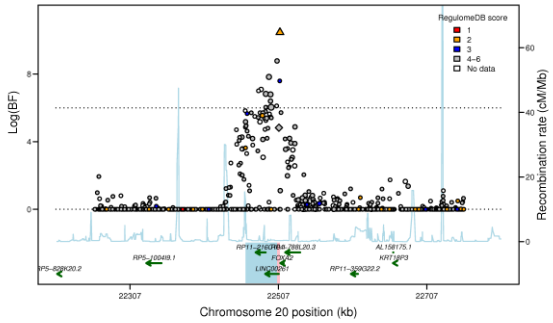


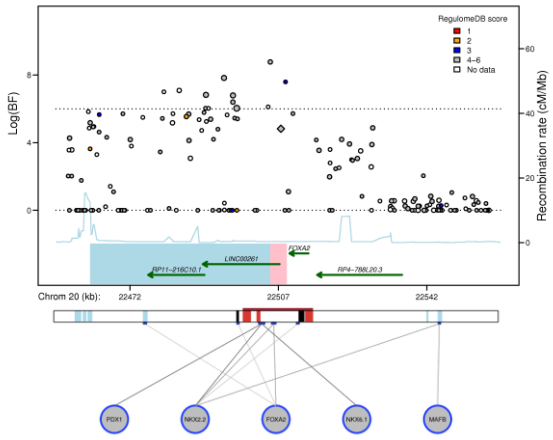
CDKN2B



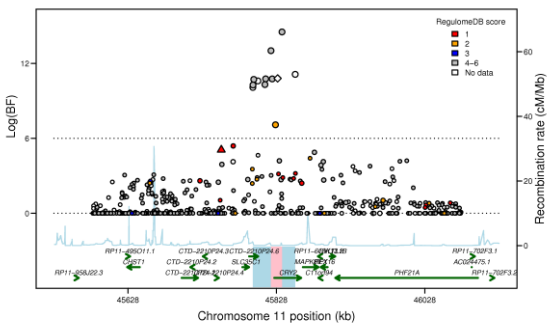


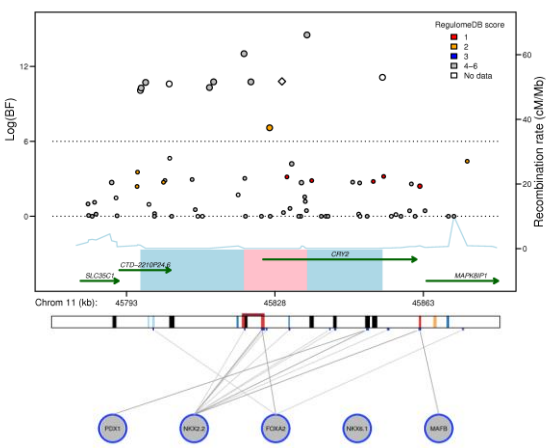
FOXA2





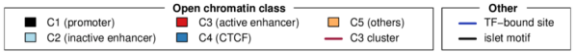
CRY2



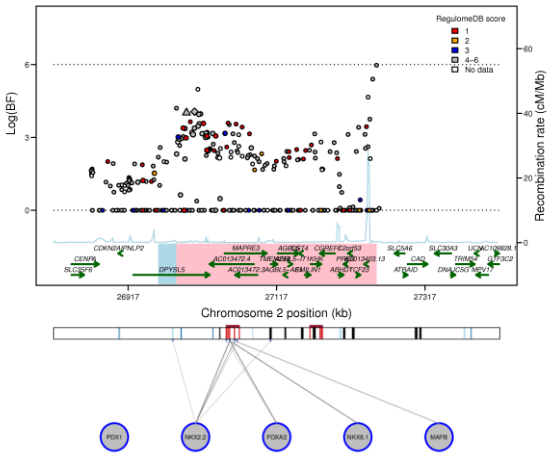


Supplemental Figure 4, FG Loci continued.

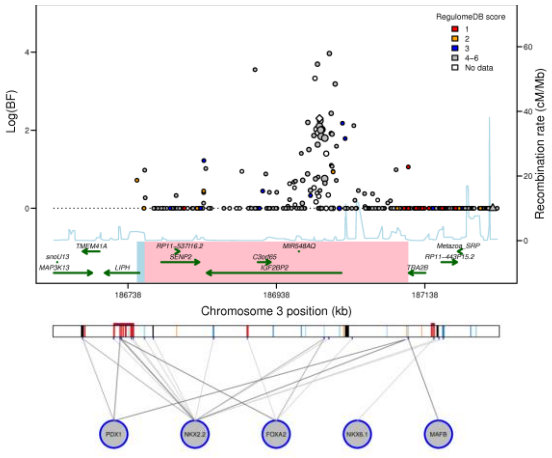
Islet Regulome legend:



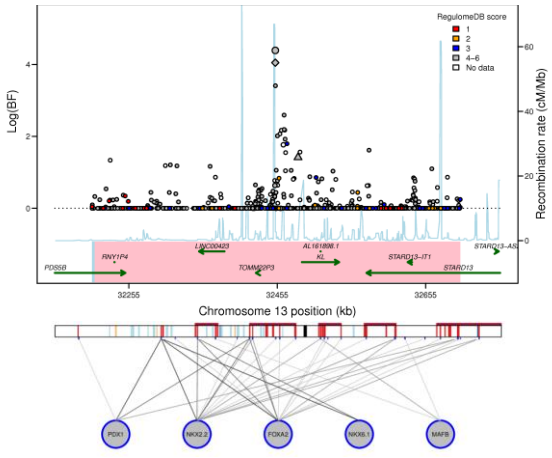
DPYSL5



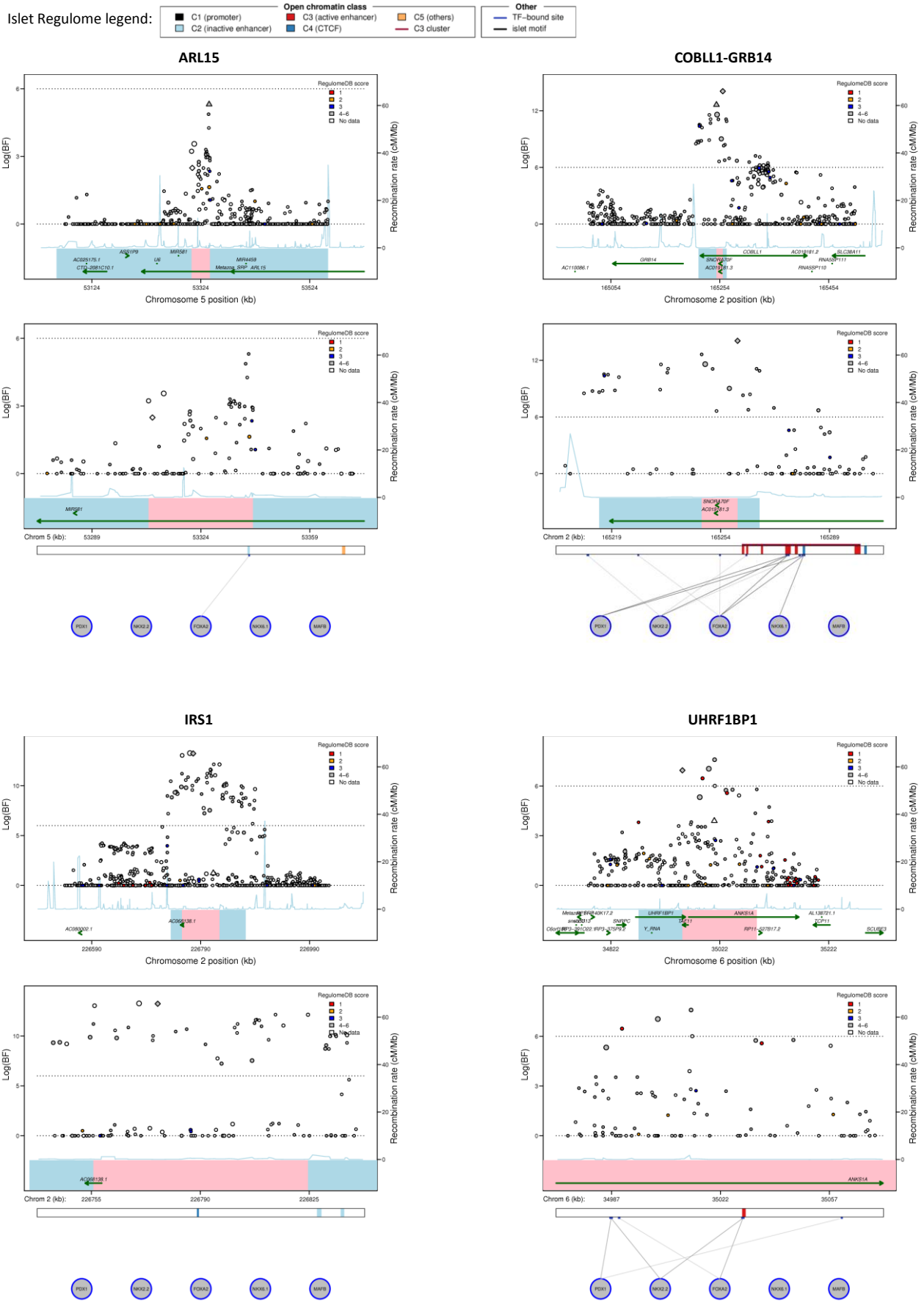
IGF2BP2



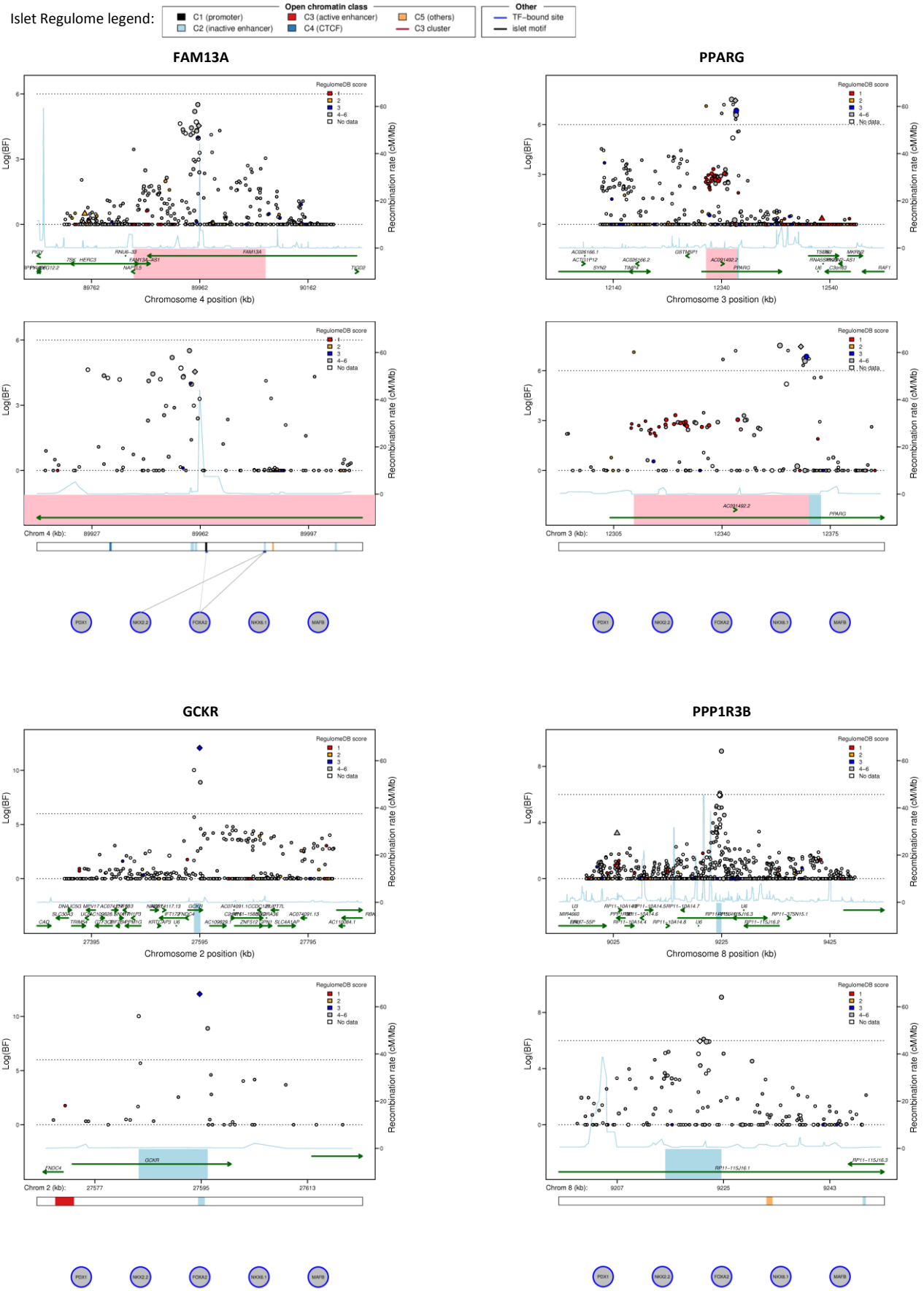
KL



Supplemental Figure 4, FI Loci.



Supplemental Figure 4, FI Loci continued.



Supplemental Figure 4, FI Loci continued.

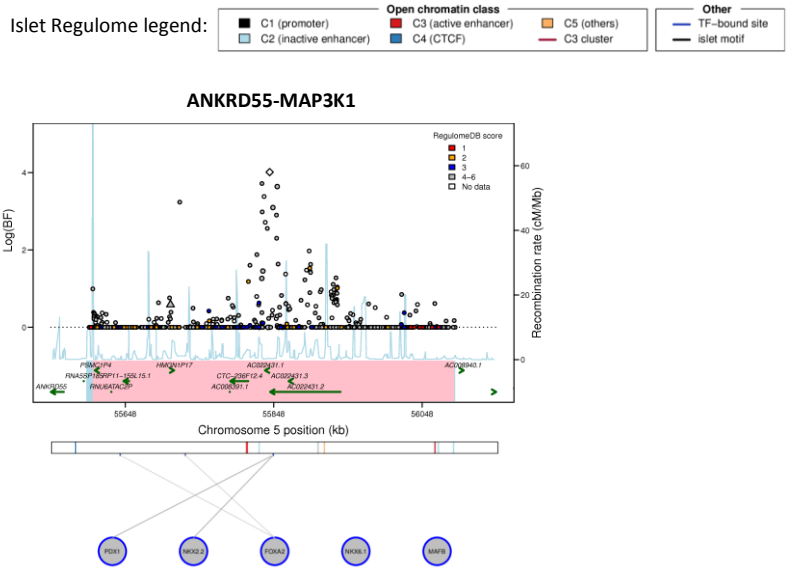
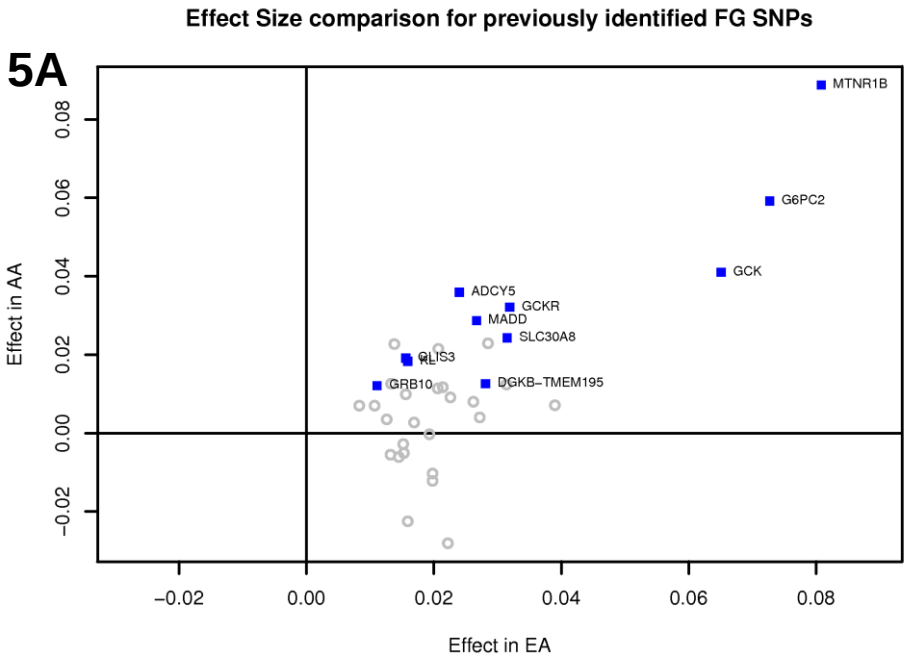
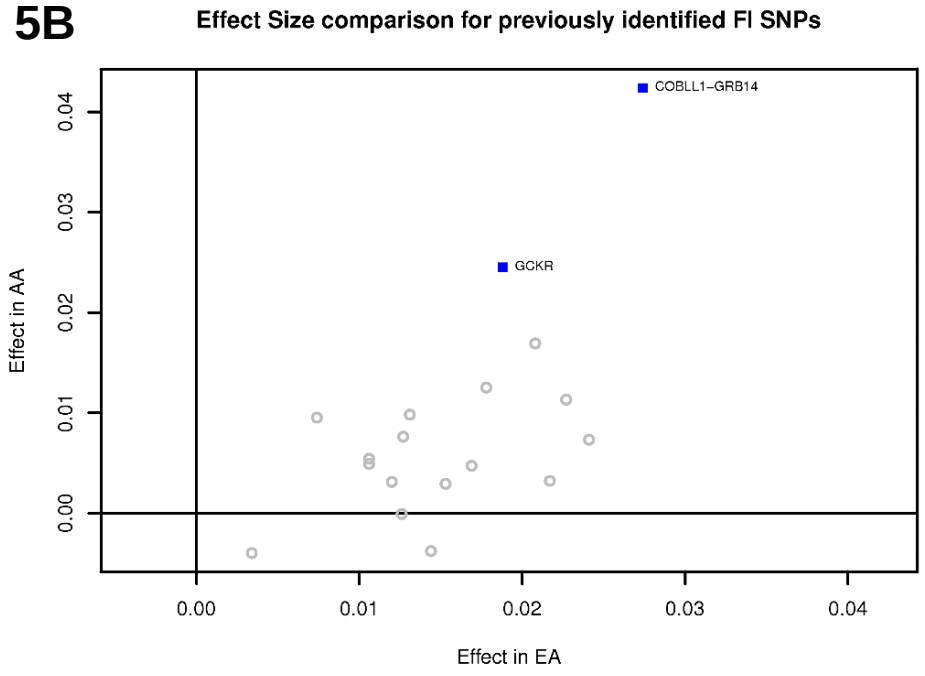
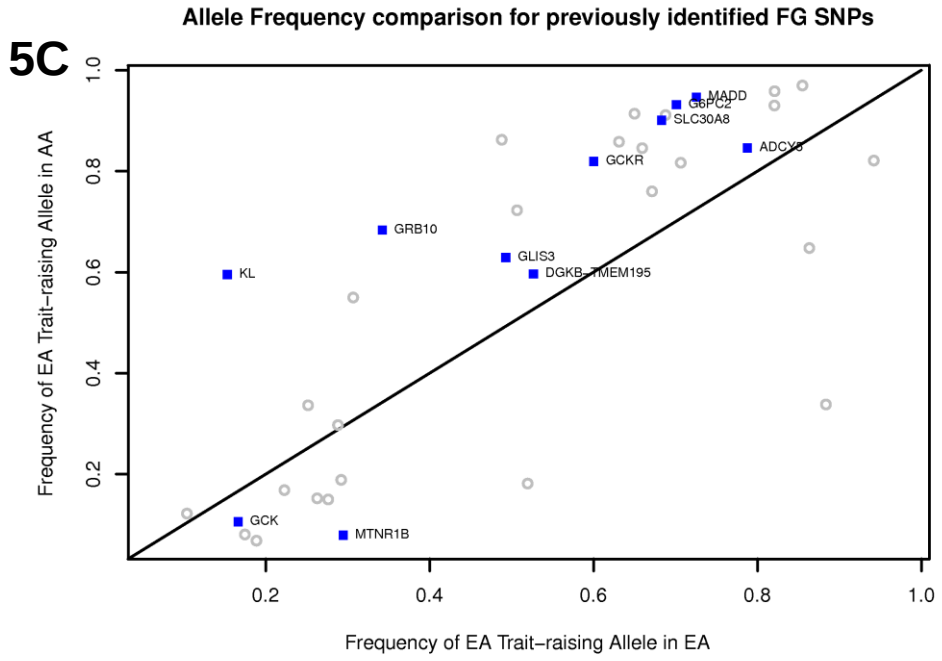
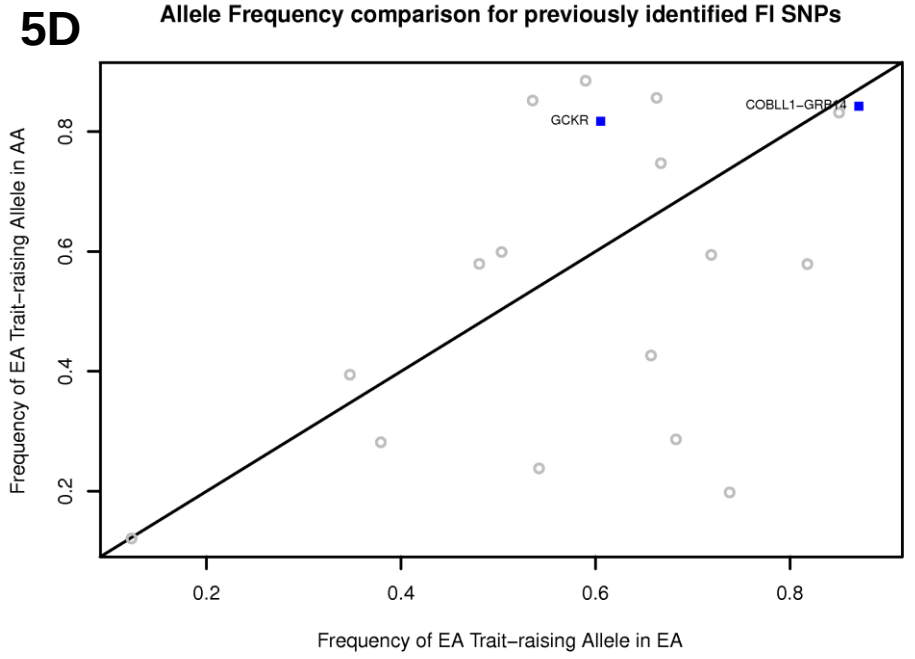


Figure S5









Previously identified EA snp with transferability from EA to AA, i.e P<0.05 in AA and share the same trait-raising allele

Previously identified EA snp without evidence of transferability from EA to AA

Figure S6A

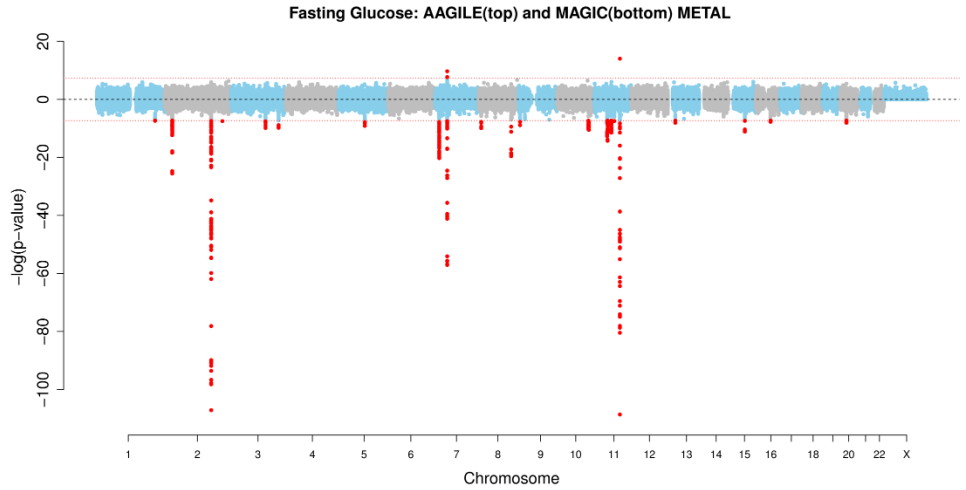


Figure S6B

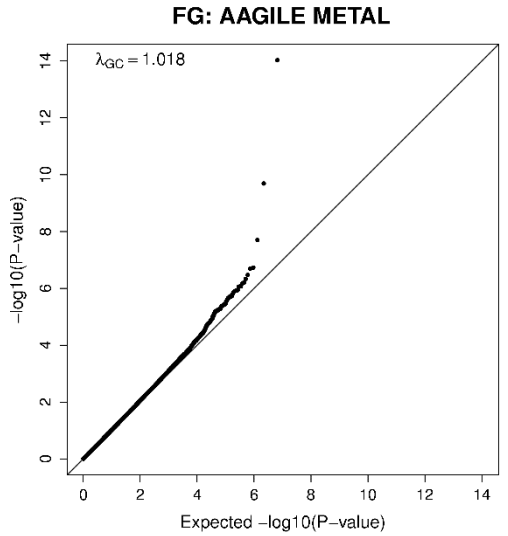


Figure S6C

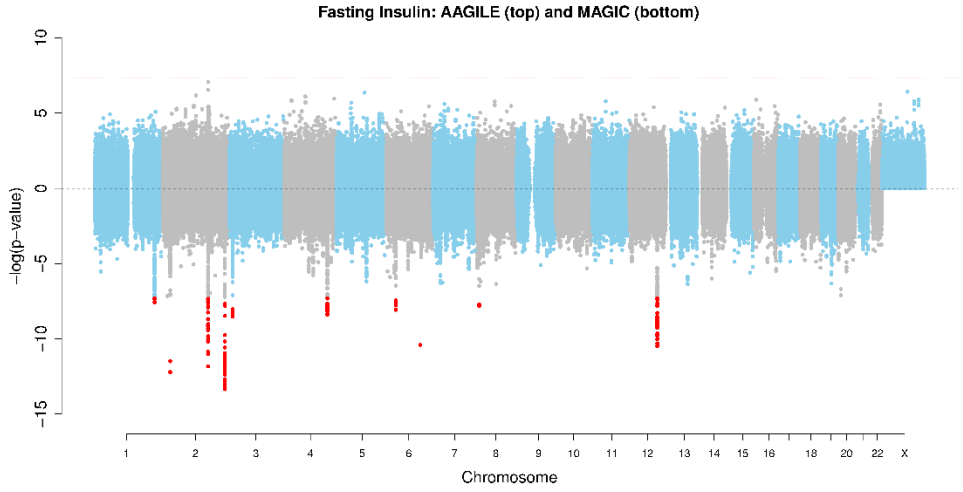


Figure S6D

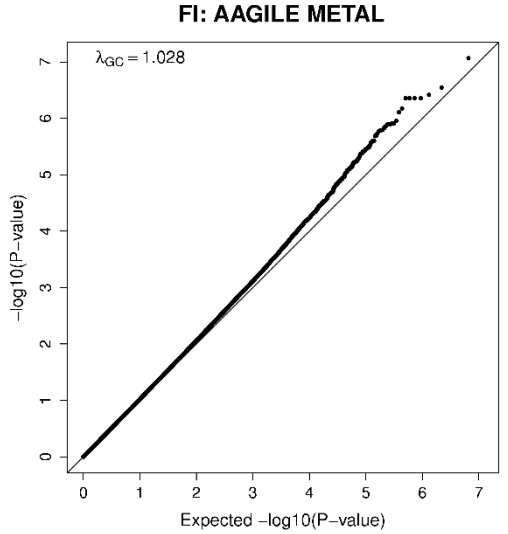


Figure S6E

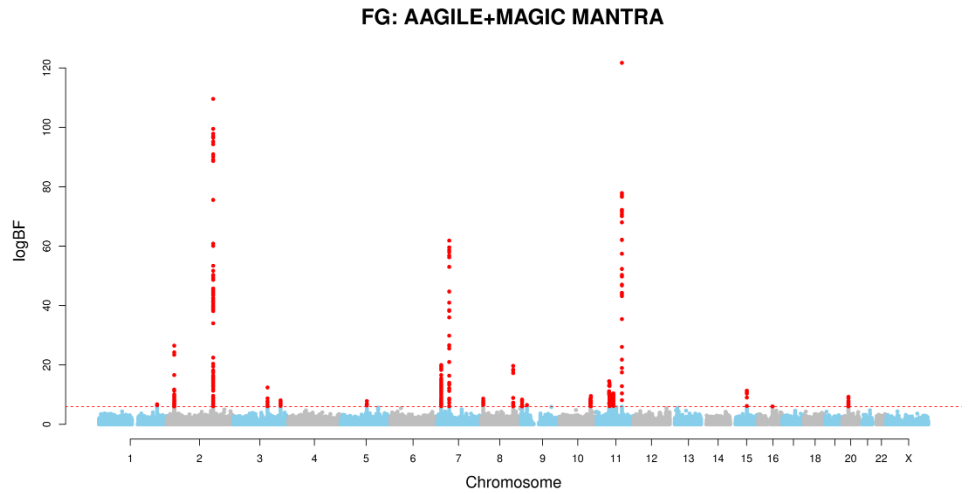


Figure S6F

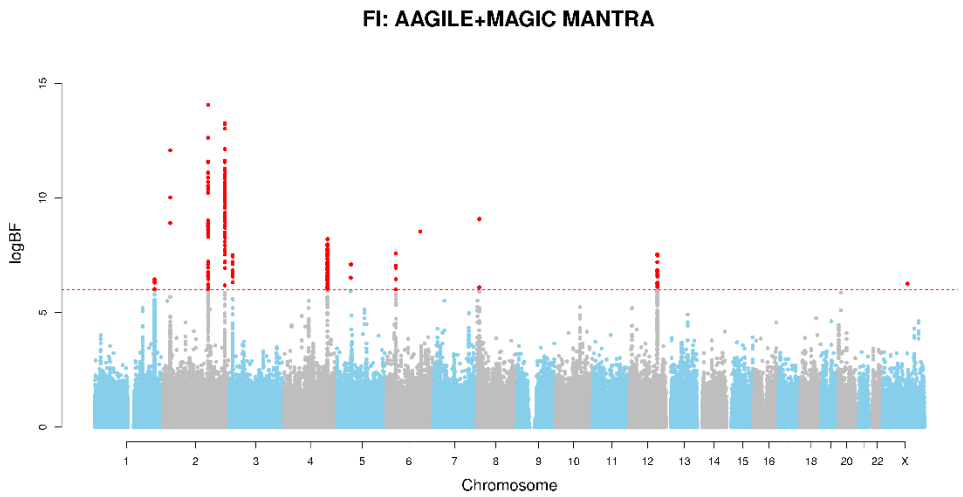


Figure S7

Figure S7A

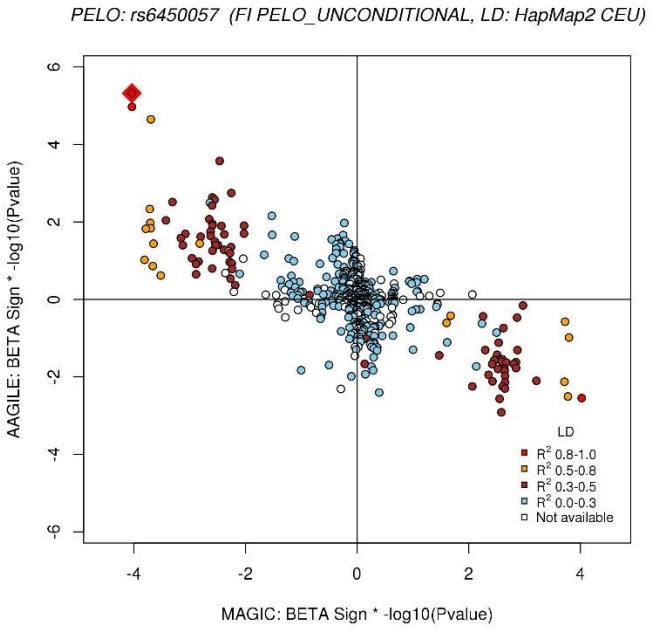


Figure S7B

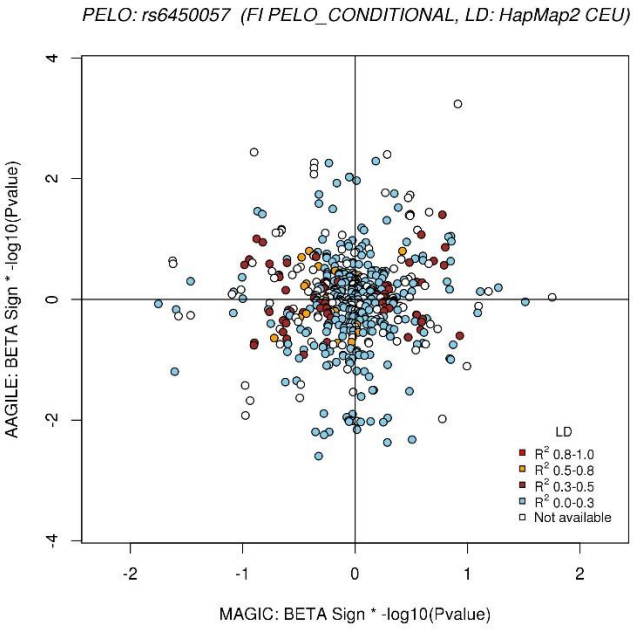


Figure S7C

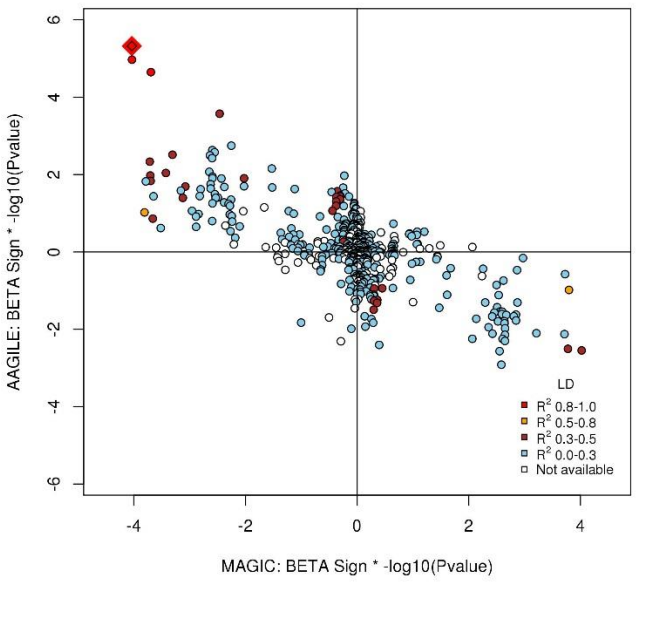
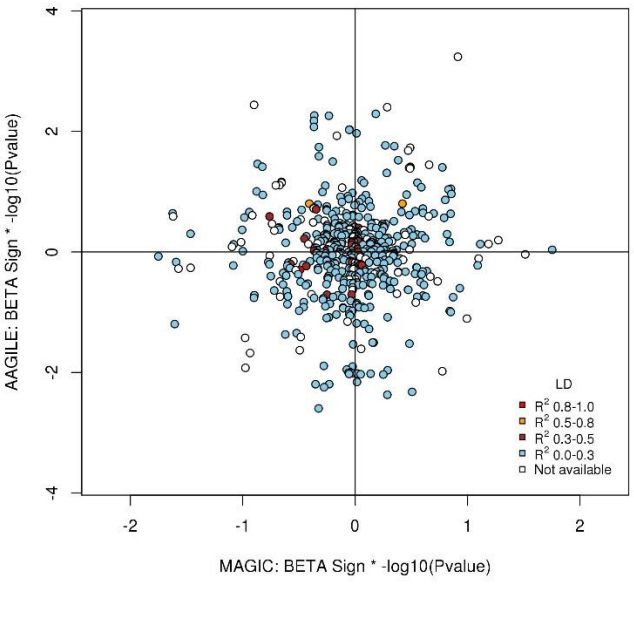


Figure S7D



Legend of Supplemental Figures

Figure S1. **Schematic study diagram.**36 fasting glucose (FG) loci, 16 fasting insulin (FI) loci, and 2 loci associated with both FG and FI previously identified in European ancestry (EA) samples were fine-mapped by combining association statistics from EA and African ancestry (AA) samples using Meta-ANalysis of TRans-ethnic Association studies (MANTRA) software. Substantially reduced trans-ethnic credible sets were further examined for evidence of regulatory annotation or enrichment of regulatory marks. Known loci from EA samples were also analysed for transferability to AA samples, for evidence of independent signals in AA populations, and for evidence of selection. Novel FG- and FI-associated loci were identified in fixed effects GWAS meta-analysis in AA samples alone and in trans-ethnic analyses combining EA and AA samples. Two novel loci associated with FI and 24 loci known from EA samples were examined for association with other cardiometabolic traits.

Figure S2. **Venn diagram of trans-ethnic analysis and transferability results.** Venn diagram showing loci exhibiting transferability between EA and AA in brown, loci at which 99% credible set was reduced by at least 20% shown in blue, and overlapping loci in the central area.

Figure S3. **Trans-ethnic fine-mapping of 22 loci (13 FG, 8 FI, and 1 both FG and FI) with greater than 20% reduction in the 99% credible set**. Trans-ethnic analysis of glycemic quantitative loci provides narrowed intervals spanned by the 99% credible set. 500 kb regional association plots centered at the index SNP identified from EA samples at each locus. The X-axis denotes genomic position and the Y-axis denotes the log (Bayes factor), recombination rate and varLD information. The red diamond data point represents the index SNP within the region previously reported from the EA sample. The color of each data point indicates its LD value (r^2) with the index SNP based on HapMap 2 (YRI for AA results and CEU for EA results): white, r^2 not available*;* blue, r^2=0.0-0.2; brown, r^2=0.2-0.5; orange, r^2=0.5-0.8; red, r^2=0.8-1.0. The blue line represents the recombination rate. The green line shows the varLD score at each SNP and is highlighted with dark brown if the varLD score is > 95th percentile of the genome-wide varLD score, comparing LD information between YRI and CEU HapMap2 samples[52]. The interval spanned by the 99% credible set is highlighted in pink. For each locus, three figures were provided. **Panel A. Association results using EA samples. Panel B. Association results using AA samples. Panel C. Association results using both EA and AA samples.**

Figure S4. **Overlay of regional association plots with regulatory annotation at 22 loci (13 FG, 8 FI, and 1 both FG and FI) with greater than 20% reduction in the 99% credible set**. **Top panel:** 500 Kb genomic span showing the top SNP in EA (MAGIC) with a diamond, the top SNP in AA (AAGILE) with a triangle, the EA-only 99% credible set bounded by the blue and pink boxes, and the narrowed trans-ethnic 99% credible set indicated by the pink boxes. SNPs are colored according to score assigned in RegulomeDB with lower score corresponding to stronger level of evidence supporting regulatory function. **Lower panel (where shown):** A zoomed in region of the locus, showing either a 100 Kb or a 50 Kb genomic span. Again, the top SNP in EA (MAGIC) is represented by a diamond, SNPs are colored according to score assigned in RegulomeDB with lower score corresponding to stronger level of evidence supporting regulatory function, and the blue box indicates the span of the EA-only 99% credible set while the pink box indicates the narrowed trans-ethnic 99% credible set. Data from the Islet Regulome Browser for the genomic interval are shown below the regional association plots.

Figure S5. Concordance of effect size and Comparison of EA trait-raising allele Frequency in EA and AA. We show the concordance of effect (**Figures S5A and S5B**) and comparison of

frequency (**Figures S5C and S5D**) for each EA trait-raising allele between EA and AA samples. The blue rectangles represent the SNPs meeting SNP transferability criteria from EA to AA, i.e. association P < 0.05 in AA and sharing the trait-raising allele in EA and AA. Grey circles represent SNPs without evidence of SNP transferability. In **Figures S5A** and S**5B**, X-axis is the effect size in EA and Y-axis is the effect size in AA for the EA trait-raising allele. There is evidence of excess concordance of effect between EA and AA. Specifically, of 36 EA FG index SNPs, 28 SNPs share the same direction in AA (binomial test P of 5.96×10^{-4}), (**Figure S5A**); of 18 EA FI index SNPs, 14 SNPs share the same direction in AA (binomial test P of 1.544E-2), (**Figure S5B**). Also, for both traits, SNPs that meet the transferability criteria tended to have larger effect size of similar magnitudes in both the EA and AA samples than those not meeting criteria. In **Figures S5C** and **S5D,** X-axis is the frequency in EA and Y-axis is the frequency in AA for the EA trait-raising allele. There is wide variation in the frequency of the EA trait-raising allele between EA and AA; the majority of SNPs with locus transferability from EA to AA exhibit higher frequency of the EA trait-raising allele in AA than in EA.

Figure S6. Genome-wide association plots and quantile-quantile (QQ) plots for FG and FI. **Figures S6A** and S**6C** display the Miami plots of association, which mirror the results of AA (on the top) and EA (on the bottom). The X-axis is the chromosome and position and the Y-axis is the –log-scale of association *P*. **Figures S6B** and S**6D** are the QQ plots for FG and FI, respectively. Both associations of FG and FI are minimally inflated with lambda of 1.018 and 1.028, respectively. **Figures S6E** and S**6F** are the genome-wide association plots of trans-ethnic meta-analysis results for FG and FI, respectively. The X-axis is the chromosome and position and the Y-axis is the –log-scale of association *P*.

Figure S7. Conditional analysis at *PELO*/rs6450057. The product of the sign of the beta-coefficient for FI level and –log(*P*-value) for each SNP residing ±250kb of the top SNP, rs6450057, at the locus in EA samples (MAGIC) and in AA samples (AAGILE) plotted on the X- and Y-axis, respectively. **Figures S7A** and **7C** show the comparison for unconditional association results with HapMap 2 CEU and YRI LD information, respectively. **Figures S7B** and **S7D** show the comparison for conditional association results with HapMap 2 CEU and YRI LD information, respectively. There is a clear pattern in the discordant directions of effect between EA and AA. However, after conditional analysis on the top SNP, this discordant pattern disappears, implying that the top SNP rs6450057 drives the original association signal.

Supplemental Tables

Table of Contents (All Tables are available in a separate file.)

References

1. Manning, A.K. *et al.* A genome-wide approach accounting for body mass index identifies genetic variants influencing fasting glycemic traits and insulin resistance. *Nature genetics* **44**, 659-69 (2012).
2. Wang, Y.J. *et al.* The association of the vanin-1 N131S variant with blood pressure is mediated by endoplasmic reticulum-associated degradation and loss of function. *Plos Genetics* **10**, e1004641 (2014).
3. McCarty, C.A. *et al.* The eMERGE Network: a consortium of biorepositories linked to electronic medical records data for conducting genomic studies. *BMC medical genomics* **4**, 13 (2011).
4. Lettre, G. *et al.* Genome-wide association study of coronary heart disease and its risk factors in 8,090 African Americans: the NHLBI CARe Project. *Plos Genetics* **7**, e1001300 (2011).
5. Ng, M.C. *et al.* Meta-analysis of genome-wide association studies in African Americans provides insights into the genetic architecture of type 2 diabetes. *Plos Genetics* **10**, e1004517 (2014).
6. Monda, K.L. *et al.* A meta-analysis identifies new loci associated with body mass index in individuals of African ancestry. *Nature genetics* **45**, 690-6 (2013).
7. Liu, C.T. *et al.* Genome-wide association of body fat distribution in African ancestry populations suggests new loci. *Plos Genetics* **9**, e1003681 (2013).
8. Ward, L.D. & Kellis, M. HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. *Nucleic acids research* **40**, D930-4 (2012).