University of Southern Denmark

# Deciphering colorectal cancer genetics through multi-omic analysis of 100,204 cases and 154,587 controls of European and east Asian ancestries

Fernandez-Rozadilla, Ceres; Timofeeva, Maria; Chen, Zhishan; Law, Philip; Thomas, Minta; Schmit, Stephanie; Díez-Obrero, Virginia; Hsu, Li; Fernandez-Tajes, Juan; Palles, Claire; Sherwood, Kitty; Briggs, Sarah; Svinti, Victoria; Donnelly, Kevin; Farrington, Susan; Blackmur, James; Vaughan-Shaw, Peter; Shu, Xiao ou; Long, Jirong; Cai, Qiuyin; Guo, Xingyi; Lu, Yingchang; Broderick, Peter; Studd, James; Huyghe, Jeroen; Harrison, Tabitha; Conti, David; Dampier, Christopher; Devall, Mathew; Schumacher, Fredrick; Melas, Marilena; Rennert, Gad; Obón-Santacana, Mireia; Martín-Sánchez, Vicente; Moratalla-Navarro, Ferran; Oh, Jae Hwan; Kim, Jeongseon; Jee, Sun Ha; Jung, Keum Ji; Kweon, Sun Seog; Shin, Min Ho; Shin, Aesun; Ahn, Yoon Ok; Kim, Dong Hyun; Oze, Isao; Wen, Wanqing; Matsuo, Keitaro; Matsuda, Koichi; Tanikawa, Chizu; Ren, Zefang; Gao, Yu Tang; Jia, Wei Hua; Hopper, John; Jenkins, Mark; Win, Aung Ko; Pai, Rish; Figueiredo, Jane; Haile, Robert; Gallinger, Steven; Woods, Michael; Newcomb, Polly; Duggan, David; Cheadle, Jeremy; Kaplan, Richard; Maughan, Timothy; Kerr, Rachel; Kerr, David; Kirac, Iva; Böhm, Jan; Mecklin, Lukka Pekka; Jousilahti, Pekka; Knekt, Paul; Aaltonen, Lauri; Rissanen, Harri; Pukkala, Eero; Eriksson, Johan; Cajuso, Tatiana; Hänninen, Ulrika; Kondelin, Johanna; Palin, Kimmo; Tanskanen, Tomas; Renkonen-Sinisalo, Laura; Zanke, Brent; Männistö, Satu; Albanes, Demetrius; Weinstein, Stephanie; Ruiz-Narvaez, Edward; Palmer, Julie; Buchanan, Daniel; Platz, Elizabeth; Visvanathan, Kala; Ulrich, Cornelia; Siegel, Erin; Brezina, Stefanie; Gsur, Andrea; Campbell, Peter; Chang-Claude, Jenny; Hoffmeister, Michael; Brenner, Hermann; Slattery, Martha; Potter, John; Tsilidis, Konstantinos; Schulze, Matthias; Gunter, Marc; Murphy, Neil; Castells, Antoni; Castellví-Bel, Sergi; Moreira, Leticia; Arndt, Volker; Shcherbina, Anna; Stern, Mariana; Pardamean, Bens; Bishop, Timothy; Giles, Graham; Southey, Melissa; Idos, Gregory; McDonnell, Kevin; Abu-Ful, Zomoroda; Greenson, Joel; Shulman, Katerina; Lejbkowicz, Flavio; Offit, Kenneth; Su, Yu Ru; Steinfelder, Robert; Keku, Temitope; van Guelpen, Bethany; Hudson, Thomas; Hampel, Heather; Pearlman, Rachel; Berndt, Sonja; Hayes, Richard; Martinez, Marie Elena; Thomas, Sushma; Corley, Douglas; Pharoah, Paul; Larsson, Susanna; Yen, Yun; Lenz, Heinz Josef; White, Emily; Li, Li; Doheny, Kimberly; Pugh, Elizabeth; Shelford, Tameka; Chan, Andrew; Cruz-Correa, Marcia; Lindblom, Annika; Hunter, David; Joshi, Amit; Schafmayer, Clemens; Scacheri, Peter; Kundaje, Anshul; Nickerson, Deborah; Schoen, Robert; Hampe, Jochen; Stadler, Zsofia; Vodicka, Pavel; Vodickova, Ludmila; Vymetalkova, Veronika; Papadopoulos, Nickolas; Edlund, Chistopher; Gauderman, William; Thomas, Duncan; Shibata, David; Toland, Amanda; Markowitz, Sanford; Kim, Andre; Chanock, Stephen; van Duijnhoven, Franzel; Feskens, Edith; Sakoda, Lori; Gago-Dominguez, Manuela; Wolk, Alicja; Naccarati, Alessio; Pardini, Barbara; FitzGerald, Liesel; Lee, Soo Chin; Ogino, Shuji; Bien, Stephanie; Kooperberg, Charles; Li, Christopher; Lin, Yi; Prentice, Ross; Qu, Conghui; Bézieau, Stéphane; Tangen, Catherine; Mardis, Elaine; Yamaji, Taiki; Sawada, Norie; Iwasaki, Motoki; Haiman, Christopher; Le Marchand, Loic; Wu, Anna; Qu, Chenxu; McNeil, Caroline; Coetzee, Gerhard; Hayward, Caroline; Deary, Ian; Harris, Sarah; Theodoratou, Evropi; Reid, Stuart; Walker, Marion; Ooi, Li Yin; Moreno, Victor; Casey, Graham; Gruber, Stephen; Tomlinson, Ian; Zheng, Wei; Dunlop, Malcolm; Houlston, Richard; Peters, Ulrike

*Published in:*
Nature Genetics

Supplementary Note

**URLs**

BarcUVa-seq - https://barcuvaseq.org/

ENCODE ATAC-seq pipeline - https://github.com/ENCODE-DCC/atac-seq-pipeline

ENCODE histone ChIP-seq pipeline - https://github.com/ENCODE-DCC/chip-seq-pipeline2

Enrichr - https://maayanlab.cloud/Enrichr/

GCTA-COJO- https://cnsgenomics.com/https://cnsgenomics.com/content/softwarecontent/software

GENCODE 24 - https://www.gencodegenes.org/human/release_24.html

GENESIS software - https://github.com/yandorazhang/GENESIS

GTEx - https://www.gtexportal.org/home/

GTEx Pipeline - https://github.com/broadinstitute/gtex-pipeline

LD Hub - http://ldsc.broadinstitute.org/ldhub/

META - https://mathgen.stats.ox.ac.uk/genetics_software/meta/meta.html

Michigan Imputation Server - https://imputationserver.sph.umich.edu/index.html

Open Targets - https://platform.opentargets.org/

PredictDB, HakyImLab Team 2020 - http://predictdb.org/

pyGenomeTracks Python library - https://pygenometracks.readthedocs.io

The NHGRI-EBI Catalog of human genome-wide association studies - https://www.ebi.ac.uk/gwas/

The NIH Roadmap Epigenomics Mapping Consortium - http://www.roadmapepigenomics.org/

***GWAS patient and data sets***

We analyzed GWAS data from previously published GWAS (**Supplementary Table 1, Supplementary Table 3**), grouping participants into analytical units by study or genotyping platform[1-4]. Studies that contributed to more than one prior GWAS were analyzed only once in the current analysis. In total, there were 31 analytical units (17 from European descent populations and 14 from Asian descent populations), totaling 100,204 CRC cases and 154,587 controls. Comprehensive details on the subjects, genotyping and standard quality control (QC) procedures have been previously reported and are summarized in **Supplementary Table 1**. As reference for LD estimation, we made use of genotyping data from 6,684 unrelated East Asian samples genotyped with MEGA array (interindividual genetic relationships < 0.025, 453 from Aichi1, 162 from HCES1, 1,764 from HCES2, 832 from Korea_NCC, 312 from Korea_NCC2, 405 from Korea_seoul, 1,833 from Shanghai4, 70 from SBCS1 (PMID: 32139696), 426 from SBCS2 (PMID: 32139696), 427 from the lung cancer Asian study cohort). To evaluate the polygenic risk score (PRS) based on all significant

GWAS loci we used independent studies not included in the discovery: the Genetic Epidemiology Research on Adult Health and Aging (GERA) cohort[5,6] for PRS evaluation in the European population (1,401 cases and 75,611 cohort participants) and several Asian studies (2,324 cases and 2,331 controls) as described before for PRS evaluation in the Asian population[7]. All study protocols were approved by the relevant Institutional Review Boards, and informed consent was obtained from all study participants in accordance with the Helsinki accord.

***Transcriptome and methylome samples and data***

INTERMPHEN study: 109 individuals of self-reported European ancestry undergoing colonoscopy[8]. Of these, 58% were males and average age was 58 years. Colorectal biopsies were obtained from the caecum, sigmoid colon, and rectum in each individual, together with an EDTA-venous blood sample. DNA and RNA were extracted and quantified using standard methodologies. 75bp paired-end RNA sequencing of samples was performed on an Illumina HiSeq4000 platform (Illumina Inc. San Diego, USA) to achieve a median of 50M reads per sample. Genotyping of DNA was performed using Illumina Infinium Human Core Exome arrays. Methylation profiling of caecal, sigmoid, and rectal samples from 89 of the 109 individuals was performed using Illumina EPIC methylation arrays on bisulfite modified DNA samples from histologically normal colorectal mucosa (n=267). Ethical approval for the study was obtained from the Oxfordshire Research Ethics Committee A and all participants provided informed written consent.

The Study of Colorectal Cancer in Scotland (SOCCS): SOCCS is a prospective study to identify genetic and environmental factors influencing CRC risk and survival outcome[9]. Histologically normal colorectal mucosa was sampled from a single site from freshly resected surgical specimens or through rectal biopsy from cancer, non-cancer patients, and healthy participants within SOCCS (n=221). Of these, 56% were males and average age was 62 years. Most of the individuals were of European ancestry (98%), with five participants showing evidence of admixture based on PCA plots. RNA was extracted and purified from histologically normal mucosa using A Ribopure kits (Applied Biosystems, Foster City, USA) according to the manufacturer's protocol. Paired-end RNA sequencing was performed in two batches on an Illumina HiSeq 4000 platform on 150bp and 75bp paired-end reads to achieve a median of 100M reads per sample. Epigenome-wide methylation analysis was performed for a subset of SOCCS samples (n=93, 55% males, mean age of 67). DNA was extracted from 93 histologically normal colonic mucosa samples using Qiagen (Qiagen, Hilden, Germany) DNeasy Blood and Tissue Kits according to the manufacturer's protocol. Genomic DNA was bisulfite modified using the Zymo EZ DNA Methylation Gold DNA methylation kit (ZymoResearch, Irvine, USA). Methylation analysis was performed using Illumina Infinium HumanMethylation450 (450K) BeadChip array (Illumina Inc). Genotyping of matched peripheral blood

DNA was conducted using Illumina OmniExpressExome8 or HumanOmni5 arrays. All participants provided informed written consent and the study was approved by the local (13/SS/0248) research ethics committees and National Health Service management (2014/0058).

BarcUVa-Seq: The BarcUVa-Seq (University of Barcelona and University of Virginia RNA sequencing project) data is based on 191 individuals (70 male, mean age = 60, and 93% of European ancestry) from the Barcelona province of Spain who had no personal history of CRC and underwent negative screening colonoscopies[10]. Mucosal biopsies from the ascending (n=68), transverse (n=47) and descending (n=76) colon were obtained together with peripheral blood samples. Total RNA was extracted using mirVana kits (Thermo Fisher Scientific) without miRNA enrichment. Libraries were prepared using Illumina TruSeq Stranded Total RNA Library Prep Gold kits, which include Ribo-Zero Plus rRNA Depletion kits for depletion of ribosomal RNA. Paired-end, 101bp or 51bp (depending on batch), RNA sequencing was performed using an Illumina HiSeq 2500 sequencer in High Output mode. Genotyping of lymphocytes was performed using the Illumina OncoArray 500K beadchip. The study protocol was approved by the Bellvitge University Hospital Ethics Committee (PR073/11 and PR286/15).

COLONOMICS: Methylation profiles of histologically normal colon tissue samples from the Colonomics cohort, which comprises both colon cancer patients and healthy individuals of European ancestry[11]. It included 128 samples from 128 individuals (85 males, mean age of 69), 92 adjacent to tumors from CRC patients and 36 from healthy individuals. DNA was extracted with phenol-chloroform and quantified by Nanodrop (Thermo Scientific, Wilmington, DE). Bisulfite conversion of DNA (200-500 ng) was performed using Illumina Infinium Assays (EZ DNA methylation kit. Zymo Research. Cat. No. D5004), and samples were profiled using Illumina Infinium HumanMethylation450 BeadChips. DNA genotyping was performed using Affymetrix Genome-Wide Human SNP 6.0 arrays. The study protocol was approved by the Bellvitge University Hospital Ethics Committee (PR074/11).

GTEx v8: This included data from 368 samples derived from colonic mucosa and underlying muscularis propria (*"Colon Transverse"*), as well as an additional 14,833 samples from 48 other tissues. eQTL data and gene models for S-PrediXcan and S-MultiXcan analyses were retrieved from the PredictDB resources repository (http://predictdb.org/)[12,13]. DGN: Whole blood data from 922 samples from the Depression Genes and Networks (DGN) cohort[14] was obtained from the PredictDB website.

No participants were compensated to participate in the presented studies.

***Regional GWAS association plots***

To examine overlap between the reported CRC association signals and CRC-relevant regulatory genomic annotations, we used the pyGenomeTracks Python library[15] to plot genomic data tracks displaying

epigenomic annotations for histologically normal colonic crypt epithelium and colonic mucosa tissue, and diverse CRC cell lines or CRC tissue[1]. Specifically, we examined overlap between the lead variant(s) or variants in linkage disequilibrium ($r^2 \geq 0.8$) with the lead variant and active enhancer regions identified by histone marks H3K27ac and H3K4me1 along with active regulatory regions identified by accessible chromatin identified through DNase I hypersensitive sites (DHSs) and ATAC-seq. ChIP-seq data were processed with the ENCODE histone ChIP-seq pipeline, and DHS and ATAC-seq data were processed with the ENCODE ATAC-seq pipeline. Peak calls visualized in pyGenomeTracks are optimal overlap peak calls.

***Pathway analysis***

We used Data-driven Expression-Prioritized Integration for Complex Traits (DEPICT)[16] to predict gene targets based on gene functions that are shared across genome-wide significant risk loci, as well as those associated at $P < 10^{-5}$ as advocated to mitigate against type II error.

***Drugability analysis***

Drug target information were obtained through the Open Targets platform (21.04 version). The drug tractability data is based on the Open Targets Tractability pipeline (version 2)[17]. Briefly, the pipeline assigns genes to tractability buckets based on evidence from diverse sources, including the Druggable Genome, ChEMBL, PDB, DrugEBIlity, UniProt, PROTAC, and Gene Ontology. Depending on the quality of the data, the gene is annotated with a tractability score, from strongest to weakest: "Clinical_Precedence", "Discovery_Precedence", "Predicted_Tractable". Data were accessed via OncoEnrichR[18].

**References**

1. Huyghe JR, Bien SA, Harrison TA, et al. Discovery of common and rare genetic risk variants for colorectal cancer. Nat Genet. 2019;51(1):76-87.

2. Law PJ, Timofeeva M, Fernandez-Rozadilla C, et al. Association analyses identify 31 new risk loci for colorectal cancer susceptibility. Nat Commun. 2019;10(1):2154.

3. Lu Y, Kweon SS, Cai Q, et al. Identification of Novel Loci and New Risk Variant in Known Loci for Colorectal Cancer Risk in East Asians. Cancer Epidemiol Biomarkers Prev. 2020;29(2):477-486.

4. Schmit SL, Edlund CK, Schumacher FR, et al. Novel Common Genetic Susceptibility Loci for Colorectal Cancer. J Natl Cancer Inst. 2019;111(2):146-157.

5. Thomas M, Sakoda LC, Hoffmeister M, et al. Genome-wide Modeling of Polygenic Risk Score in Colorectal Cancer Risk. Am J Hum Genet. 2020;107(3):432-444.

6. Kvale MN, Hesselson S, Hoffmann TJ, et al. Genotyping Informatics and Quality Control for 100,000 Subjects in the Genetic Epidemiology Research on Adult Health and Aging (GERA) Cohort. Genetics. 2015;200(4):1051-1060.

7. Wang H, Burnett T, Kono S, et al. Trans-ethnic genome-wide association study of colorectal cancer identifies a new susceptibility locus in VTI1A. Nat Commun. 2014;5:4613.

8. Fernandez-Rozadilla C, Kartsonaki C, Woolley C, et al. Telomere length and genetics are independent colorectal tumour risk factors in an evaluation of biomarkers in normal bowel. Br J Cancer. 2018;118(5):727-732.

9. Theodoratou E, McNeill G, Cetnarskyj R, et al. Dietary fatty acids and colorectal cancer: a case-control study. Am J Epidemiol. 2007;166(2):181-195.

10. Dampier CH, Devall M, Jennelle LT, et al. Oncogenic Features in Histologically Normal Mucosa: Novel Insights Into Field Effect From a Mega-Analysis of Colorectal Transcriptomes. Clin Transl Gastroenterol. 2020;11(7):e00210.

11. Sanz-Pamplona R, Berenguer A, Cordero D, et al. Aberrant gene expression in mucosa adjacent to tumor reveals a molecular crosstalk in colon cancer. Mol Cancer. 2014;13:46.

12. Barbeira AN, Melia OJ, Liang Y, et al. Fine-mapping and QTL tissue-sharing information improves the reliability of causal gene identification. Genet Epidemiol. 2020.

13. Barbeira AN, Bonazzola R, Gamazon ER, et al. Exploiting the GTEx resources to decipher the mechanisms at GWAS loci. Genome Biol. 2021;22(1):49.

14. Battle A, Mostafavi S, Zhu X, et al. Characterizing the genetic basis of transcriptome diversity through RNA-sequencing of 922 individuals. Genome Res. 2014;24(1):14-24.

15. Lopez-Delisle L, Rabbani L, Wolff J, et al. pyGenomeTracks: reproducible plots for multivariate genomic datasets. Bioinformatics. 2021;37(3):422-423.

16. Pers TH, Karjalainen JM, Chan Y, et al. Biological interpretation of genome-wide association studies using predicted gene functions. Nat Commun. 2015;6:5890.

17. Brown KK, Hann MM, Lakdawala AS, Santos R, Thomas PJ, Todd K. Approaches to target tractability assessment - a practical perspective. Medchemcomm. 2018;9(4):606-613.

18. Nakken S, Gundersen S, Bernal FLM, Hovig E, Wesche J. OncoEnrichR: cancer-dedicated gene set interpretation. 2021.

**Funding and acknowledgements**

UK: SEB is supported by an MRC Clinical Research Training Fellowship (MR/P001106/1). ET was supported by Cancer Research UK CDF (C31250/A22804). PV-S was supported by a MRC Clinical Research Training Fellowship (MR/M004007/1). CH was supported by an MRC Human Genetics Unit programme grant 'Quantitative traits in health and disease' (U. MC_UU_00007/10). The views expressed are not necessarily those of the NHS, the NIHR or the Department of Health. COIN translational studies were supported by Tenovus Cancer Care and Cancer Research Wales. The COIN and COIN-B trials were funded by Cancer Research UK, the Medical Research Council and an unrestricted educational grant from Merck-Serono and were conducted with the support of the National Institute of Health Research Cancer Research Network. We thank the Lothian Birth Cohorts' members, investigators, research associates, and other team members. We thank the Edinburgh Clinical Research Facility (ECRF) Genetics Core, Western General Hospital, Edinburgh, for genotyping. Lothian Birth Cohorts' data collection is supported by the Disconnected Mind project (funded by Age UK), and the Biotechnology and Biological Sciences Research Council (BBSRC, for genotyping; BB/F019394/1) and undertaken within the University of Edinburgh Centre for Cognitive Ageing and Cognitive Epidemiology (funded by the BBSRC and Medical Research Council RC as part of the LLHW [MR/K026992/1]). Generation Scotland received core support from the Chief Scientist Office of the Scottish Government Health Directorates [CZD/16/6] and the Scottish Funding Council [HR03006] and is currently supported by the Wellcome Trust [216767/Z/19/Z]. Genotyping of the GS:SFHS samples was carried out by the Genetics Core Laboratory at the Clinical Research Facility, University of Edinburgh and was funded by the Medical Research Council UK and the Wellcome Trust (Wellcome Trust Strategic Award "STratifying Resilience and Depression Longitudinally" (STRADL) [104036/Z/14/Z]). We acknowledge use of genotype data from the British 1958 Birth Cohort DNA collection, which was funded by the Medical Research Council Grant G0000934 and the Wellcome Trust Grant 068545/Z/02. A full list of the investigators who contributed to the generation of the data is available from http://www.wtccc.org.uk.

The Asia Colorectal Cancer Consortium is primarily funded by R01 CA188214 from US National Institutes of Health. Studies (listed with grant support) participating in the Asia Colorectal Cancer Consortium include the Shanghai Women's Health Study (US NIH, R37CA070867, UM1CA182910), the Shanghai Men's Health Study (US NIH, R01CA082729, UM1CA173640), the Shanghai Breast and Endometrial Cancer Studies (US NIH, R01CA064277 and R01CA092585; contributing only controls), the Shanghai Colorectal Cancer Study 3 (US NIH, R37CA070867,R01CA188214 and Anne Potter Wilson Chair funds), the Guangzhou Colorectal Cancer Study (National Key Scientific and Technological Project, 2011ZX09307-001-04; the National Basic Research Program, 2011CB504303, contributing only controls, the Natural Science Foundation of China, 81072383, contributing only controls), the Hwasun Cancer Epidemiology Study–Colon and Rectum Cancer (HCES-CRC; grants from Chonnam National University Hwasun Hospital

Biomedical Research Institute,HCRI18007), the Japan BioBank Colorectal Cancer Study (grant from the Ministry of Education, Culture, Sports, Science and Technology of the Japanese government), the Aichi Colorectal Cancer Study (Grant-in-Aid for Cancer Research, grant for the Third Term Comprehensive Control Research for Cancer and Grants-in-Aid for Scientific Research from the Japanese Ministry of Education, Culture, Sports, Science and Technology, 17015018 and 221S0001), the Korea-NCC (National Cancer Center) Colorectal Cancer Study (Basic Science Research Program through the National Research Foundation of Korea, 2010-0010276 and 2013R1A1A2A10008260; National Cancer Center Korea, 0910220 and 1710882), the KCPS-II Colorectal Cancer Study (National R&D Program for Cancer Control, 1631020; Seoul R&D Program, 10526).. Studies (listed with grant support) participating in the Asia Colorectal Cancer Consortium include the Shanghai Women's Health Study (US NIH, R37CA070867, UM1CA182910), the Shanghai Men's Health Study (US NIH, R01CA082729, UM1CA173640), the Shanghai Breast and Endometrial Cancer Studies (US NIH, R01CA064277 and R01CA092585; contributing only controls), the Shanghai Colorectal Cancer Study 3 (US NIH, R37CA070867, R01CA188214 and Anne Potter Wilson Chair funds), the Guangzhou Colorectal Cancer Study (National Key Scientific and Technological Project, 2011ZX09307-001-04; the National Basic Research Program, 2011CB504303, contributing only controls, the Natural Science Foundation of China, 81072383, contributing only controls), the Hwasun Cancer Epidemiology Study–Colon and Rectum Cancer (HCES-CRC; grants from Chonnam National University Hwasun Hospital Biomedical Research Institute, HCRI18007), the Japan BioBank Colorectal Cancer Study (grant from the Ministry of Education, Culture, Sports, Science and Technology of the Japanese government), the Aichi Colorectal Cancer Study (Grant-in-Aid for Cancer Research, grant for the Third Term Comprehensive Control Research for Cancer and Grants-in-Aid for Scientific Research from the Japanese Ministry of Education, Culture, Sports, Science and Technology, 17015018 and 221S0001), the Korea-NCC (National Cancer Center) Colorectal Cancer Study (Basic Science Research Program through the National Research Foundation of Korea, 2010-0010276 and 2013R1A1A2A10008260; National Cancer Center Korea, 0910220), and the KCPS-II Colorectal Cancer Study (National R&D Program for Cancer Control, 1631020; Seoul R&D Program, 10526).

Finland: In Finland, this work was supported by grants from the Academy of Finland [Finnish Center of Excellence Program 2012–2017, 250345 and 2018–2025, 312041], the Jane and Aatos Erkko Foundation, the Finnish Cancer Society [also a personal grant to K.P.], the Sigrid Juselius Foundation, SYSCOL, the Nordic Information for Action eScience Center (NIASC), the Nordic Center of Excellence financed by NordForsk [project 62721, personal grant to K.P.] and State Research Funding of Kuopio University Hospital [B1401]. We acknowledge the computational resources provided by the ELIXIR node, hosted at the CSC–IT Center for Science, Finland, and funded by the Academy of Finland [grants 271642 and

263164], the Ministry of Education and Culture, Finland. V.S. was supported by the Finnish Academy [grant number 139635] and the Finnish Foundation for Cardiovascular Research. J.-P.M. was funded by The Finnish Cancer Foundation and The Jane and Aatos Erkko Foundation. Sample collection and genotyping in the Finnish Twin Cohort has been supported by the Wellcome Trust Sanger Institute, ENGAGE—European Network for Genetic and Genomic Epidemiology, FP7-HEALTH-F4–2007; [grant agreement number 201413], the National Institute of Alcohol Abuse and Alcoholism [grants AA-12502 and AA-00145; to R.J.R. and K02AA018755 to D.M.D.] and the Academy of Finland [grants 100499, 205585, 265240 and 263278 to J.K.].

Croatia: The Croatian study was supported through the 10,001 Dalmatians Project, and institutional support of University Hospital for Tumours, Sestre Milosrdnice University Hospital Center.

BCAC: The BCAC study would not have been possible without the contributions of the following: Manjeet K. Bolla, Qin Wang, Kyriaki Michailidou and Joe Dennis. BCAC is funded by Cancer Research UK (C1287/A10118, C1287/A16563). For the BBCS study, we thank Eileen Williams, Elaine Ryder-Mills, Kara Sargus. The BBCS is funded by Cancer Research UK and Breast Cancer Now and acknowledges NHS funding to the National Institute of Health Research (NIHR) Biomedical Research Centre (BRC) and the National Cancer Research Network (NCRN). We thank the participants and the investigators of EPIC (European Prospective Investigation into Cancer and Nutrition). The coordination of EPIC is financially supported by the European Commission (DG-SANCO) and the International Agency for Research on Cancer. The national cohorts are supported by: Ligue Contre le Cancer, Institut Gustave Roussy, Mutuelle Générale de l'Education Nationale, Institut National de la Santé et de la Recherche Médicale (INSERM) (France); German Cancer Aid, German Cancer Research Center (DKFZ), Federal Ministry of Education and Research (BMBF) (Germany); the Hellenic Health Foundation, the Stavros Niarchos Foundation (Greece); Associazione Italiana per la Ricerca sul Cancro-AIRC-Italy and National Research Council (Italy); Dutch Ministry of Public Health, Welfare and Sports (VWS), Netherlands Cancer Registry (NKR), LK Research Funds, Dutch Prevention Funds, Dutch ZON (Zorg Onderzoek Nederland), World Cancer Research Fund (WCRF), Statistics Netherlands (The Netherlands); Health Research Fund (FIS), PI13/00061 to Granada, PI13/01162 to EPIC-Murcia, Regional Governments of Andalucía, Asturias, Basque Country, Murcia and Navarra, ISCIII RETIC (RD06/0020) (Spain); Cancer Research UK (14136 to EPIC-Norfolk; C570/A16491 and C8221/A19170 to EPIC-Oxford), Medical Research Council (1000143 to EPIC-Norfolk, MR/M012190/1 to EPIC-Oxford) (United Kingdom). We thank the SEARCH and EPIC teams, which were funded by a programme grant from Cancer Research UK (C490/A10124) and supported by the UK NIHR BRC at the University of Cambridge. We thank Breast Cancer Now and the Institute of Cancer Research (ICR) for support and funding of the UKBGS, and the study participants, study staff, and the doctors, nurses and other health-care providers and health information sources who have contributed to the study.

PRACTICAL: Genotyping of the PRACTICAL consortium OncoArray was funded by the US National Institutes of Health (NIH) [U19 CA 148537 for ELucidating associations Involved in Prostate cancer SuscEptibility (ELLIPSE) project and X01HG007492 to the Center for Inherited Disease Research (CIDR) under contract number HHSN268201200008I]. Additional analytic support was provided by NIH NCI U01 CA188392 (PI: Schumacher). The PRACTICAL consortium was supported by Cancer Research UK Grants C5047/A7357, C1287/A10118, C1287/A16563, C5047/A3354, C5047/A10692, C16913/A6135, European Commission's Seventh Framework Programme grant agreement n° 223175 (HEALTH-F2–2009–223175), and The National Institute of Health (NIH) Cancer Post-Cancer GWAS initiative grant: No. 1 U19 CA 148537–01 (the GAME-ON initiative). We would also like to thank the following for funding support: The Institute of Cancer Research and The Everyman Campaign, The Prostate Cancer Research Foundation, Prostate Research Campaign UK (now Prostate Action), The Orchid Cancer Appeal, The National Cancer Research Network UK, The National Cancer Research Institute (NCRI) UK. We are grateful for support of NIHR funding to the NIHR Biomedical Research Centre at The Institute of Cancer Research and The Royal Marsden NHS Foundation Trust, the Spanish Instituto de Salud Carlos III (ISCIII) an initiative of the Spanish Ministry of Economy and Innovation (Spain), and the Xunta de Galicia (Spain). The APBC BioResource, which forms part of the PRACTICAL consortium, consists of the following members: Wayne Tilley, Gail Risbridger, Renea Taylor, Lisa Horvath, Vanessa Hayes, Lisa Butler, Trina Yeadon, Allison Eckert, Anne-Maree Haynes, Melissa Papargiris.

LD hub data: We gratefully acknowledge all the studies and databases that made GWAS summary data available through LD Hub database: ADIPOGen (Adiponectin genetics consortium), C4D (Coronary Artery Disease Genetics Consortium), CARDIoGRAM (Coronary ARtery DIsease Genome wide Replication and Meta-analysis), CKDGen (Chronic Kidney Disease Genetics consortium), dbGAP (database of Genotypes and Phenotypes), DIAGRAM (DIAbetes Genetics Replication And Meta-analysis), ENIGMA (Enhancing Neuro Imaging Genetics through Meta Analysis), EAGLE (EArly Genetics & Lifecourse Epidemiology Eczema Consortium, excluding 23andMe), EGG (Early Growth Genetics Consortium), GABRIEL (A Multidisciplinary Study to Identify the Genetic and Environmental Causes of Asthma in the European Community), GCAN (Genetic Consortium for Anorexia Nervosa), GEFOS (GEnetic Factors for OSteoporosis Consortium), GIANT (Genetic Investigation of ANthropometric Traits), GIS (Genetics of Iron Status consortium), GLGC (Global Lipids Genetics Consortium), GPC (Genetics of Personality Consortium), GUGC (Global Urate and Gout consortium), HaemGen (haematological and platelet traits genetics consortium), HRgene (Heart Rate consortium), IIBDGC (International Inflammatory Bowel Disease Genetics Consortium), ILCCO (International Lung Cancer Consortium), IMSGC (International Multiple Sclerosis Genetic Consortium), MAGIC (Meta-Analyses of Glucose and Insulin-related traits Consortium), MESA (Multi-Ethnic Study of Atherosclerosis), PGC (Psychiatric Genomics Consortium), Project MinE consortium,

ReproGen (Reproductive Genetics Consortium), SSGAC (Social Science Genetics Association Consortium) and TAG (Tobacco and Genetics Consortium), TRICL (Transdisciplinary Research in Cancer of the Lung consortium), UK Biobank. We gratefully acknowledge the contributions of Alkes Price (the systemic lupus erythematosus GWAS and primary biliary cirrhosis GWAS) and Johannes Kettunen (lipids metabolites GWAS).

GTEx: The Genotype-Tissue Expression (GTEx) Project was supported by the Common Fund of the Office of the Director of the National Institutes of Health, and by NCI, NHGRI, NHLBI, NIDA, NIMH, and NINDS. The data used for the analyses described in this manuscript were obtained from PredictDB Data Repository. Finally, the authors gratefully acknowledge the participation of all individuals in the relevant studies.

ASTERISK: A Hospital Clinical Research Program (PHRC-BRD09/C) from the University Hospital Center of Nantes (CHU de Nantes) and supported by the Regional Council of Pays de la Loire, the Groupement des Entreprises Françaises dans la Lutte contre le Cancer (GEFLUC), the Association Anne de Bretagne Génétique and the Ligue Régionale Contre le Cancer (LRCC). We are very grateful to Dr. Bruno Buecher without whom this project would not have existed. We also thank all those who agreed to participate in this study, including the patients and the healthy control persons, as well as all the physicians, technicians and students.

ATBC: The Alpha-Tocopherol, Beta-Carotene Cancer Prevention Study is supported by the Intramural Research Program of the U.S. National Cancer Institute, National Institutes of Health, Department of Health and Human Services.

Black Women's Health Study (BWHS); funding from the National Cancer Institute (R01CA058420, R01CA098663, and U01CA164974).

CLUE II funding was from the National Cancer Institute (U01 CA86308, Early Detection Research Network; P30 CA006973), National Institute on Aging (U01 AG18033), and the American Institute for Cancer Research. The content of this publication does not necessarily reflect the views or policies of the Department of Health and Human Services, nor does mention of trade names, commercial products, or organizations imply endorsement by the US government. Maryland Cancer Registry (MCR) Cancer data was provided by the Maryland Cancer Registry, Center for Cancer Prevention and Control, Maryland Department of Health, with funding from the State of Maryland and the Maryland Cigarette Restitution Fund. The collection and availability of cancer registry data is also supported by the Cooperative Agreement NU58DP006333, funded by the Centers for Disease Control and Prevention. Its contents are solely the responsibility of the authors and do not necessarily represent the official views of the Centers for Disease Control and Prevention or the Department of Health and Human Services.

ColoCare: This work was supported by the National Institutes of Health (grant numbers R01 CA189184 (Li/Ulrich), U01 CA206110 (Ulrich/Li/Siegel/Figueireido/Colditz), 2P30CA015704- 40 (Gilliland), R01 CA207371 (Ulrich/Li), P30-CA076292 (Cleveland)), the Matthias Lackas-Foundation, the German Consortium for Translational Cancer Research, and the EU TRANSCAN initiative.

The Colon Cancer Family Registry (CCFR, www.coloncfr.org) is supported in part by funding from the National Cancer Institute (NCI), National Institutes of Health (NIH) (award U01 CA167551). Support for case ascertainment was provided in part from the Surveillance, Epidemiology, and End Results (SEER) Program and the following U.S. state cancer registries: AZ, CO, MN, NC, NH; and by the Victoria Cancer Registry (Australia) and Ontario Cancer Registry (Canada). The CCFR Set-1 (Illumina 1M/1M-Duo) and Set-2 (Illumina Omni1-Quad) scans were supported by NIH awards U01 CA122839 and R01 CA143237 (to GC). The CCFR Set-3 (Affymetrix Axiom CORECT Set array) was supported by NIH award U19 CA148107 and R01 CA81488 (to SBG). The CCFR Set-4 (Illumina OncoArray 600K SNP array) was supported by NIH award U19 CA148107 (to SBG) and by the Center for Inherited Disease Research (CIDR), which is funded by the NIH to the Johns Hopkins University, contract number HHSN268201200008I. Additional funding for the OFCCR/ARCTIC was through award GL201-043 from the Ontario Research Fund (to BWZ), award 112746 from the Canadian Institutes of Health Research (to TJH), through a Cancer Risk Evaluation (CaRE) Program grant from the Canadian Cancer Society (to SG), and through generous support from the Ontario Ministry of Research and Innovation. The SFCCR Illumina HumanCytoSNP array was supported in part through NCI/NIH awards U01 CA074794 (to JDP) and /U24 CA074794 and R01 CA076366 (to PAN). The content of this manuscript does not necessarily reflect the views or policies of the NCI, NIH or any of the collaborating centers in the Colon Cancer Family Registry (CCFR), nor does mention of trade names, commercial products, or organizations imply endorsement by the US Government, any cancer registry, or the CCFR.

COLON: The COLON study is sponsored by Wereld Kanker Onderzoek Fonds, including funds from grant 2014/1179 as part of the World Cancer Research Fund International Regular Grant Programme, by Alpe d'Huzes and the Dutch Cancer Society (UM 2012–5653, UW 2013-5927, UW2015-7946), and by TRANSCAN (JTC2012-MetaboCCC, JTC2013-FOCUS). The Nqplus study is sponsored by a ZonMW investment grant (98-10030); by PREVIEW, the project PREVention of diabetes through lifestyle intervention and population studies in Europe and around the World (PREVIEW) project which received funding from the European Union Seventh Framework Programme (FP7/2007–2013) under grant no. 312057; by funds from TI Food and Nutrition (cardiovascular health theme), a public–private partnership on pre-competitive research in food and nutrition; and by FOODBALL, the Food Biomarker Alliance, a project from JPI Healthy Diet for a Healthy Life.

CORSA: The CORSA study was funded by Austrian Research Funding Agency (FFG) BRIDGE (grant 829675, to A. G.), the “Herzfeldersche Familienstiftung” (grant to A.G.) and was supported by COST Action BM1206.

CPS-II: The American Cancer Society funds the creation, maintenance, and updating of the Cancer Prevention Study-II (CPS-II) cohort. This study was conducted with Institutional Review Board approval.

CRCGEN: Colorectal Cancer Genetics & Genomics, Spanish study was supported by Instituto de Salud Carlos III, co-funded by FEDER funds –a way to build Europe– (grants PI14-613 and PI09-1286), Agency for Management of University and Research Grants (AGAUR) of the Catalan Government (grant 2017SGR723), Junta de Castilla y León (grant LE22A10-2), Spanish Association Against Cancer (AECC) Scientific Foundation grant GCTRA18022MORE and POSTD037OBON, and the Consortium for Biomedical Research in Epidemiology and Public Health (CIBERESP), action Genrisk. Sample collection of this work was supported by the Xarxa de Bancs de Tumors de Catalunya sponsored by Pla Director d’Oncología de Catalunya (XBTC), Plataforma Biobancos PT13/0010/0013 and ICOBIOBANC, sponsored by the Catalan Institute of Oncology.

Czech Republic CCS: This work was supported by the Czech Science Foundation (20-03997S) and by the Grant Agency of the Ministry of Health of the Czech Republic (grants NV18/03/00199 and NU21-07-00247).

DACHS: This work was supported by the German Research Council (BR 1704/6-1, BR 1704/6-3, BR 1704/6-4, CH 117/1-1, HO 5117/2-1, HE 5998/2-1, KL 2354/3-1, RO 2270/8-1 and BR 1704/17-1), the Interdisciplinary Research Program of the National Center for Tumor Diseases (NCT), Germany, and the German Federal Ministry of Education and Research (01KH0404, 01ER0814, 01ER0815, 01ER1505A and 01ER1505B).

DALS: National Institutes of Health (R01 CA48998 to M. L. Slattery).

EDRN: This work is funded and supported by the NCI, EDRN Grant (U01 CA 84968-06).

EPIC: The coordination of EPIC is financially supported by the European Commission (DGSANCO) and the International Agency for Research on Cancer. The national cohorts are supported by Danish Cancer Society (Denmark); Ligue Contre le Cancer, Institut Gustave Roussy, Mutuelle Générale de l’Education Nationale, Institut National de la Santé et de la Recherche Médicale (INSERM) (France); German Cancer Aid, German Cancer Research Center (DKFZ), Federal Ministry of Education and Research (BMBF), Deutsche Krebshilfe, Deutsches Krebsforschungszentrum and Federal Ministry of Education and Research (Germany); the Hellenic Health Foundation (Greece); Associazione Italiana per la Ricerca sul Cancro-AIRCItaly and National Research Council (Italy); Dutch Ministry of Public Health, Welfare and Sports (VWS), Netherlands Cancer Registry (NKR), LK Research Funds, Dutch Prevention Funds, Dutch ZON (Zorg Onderzoek Nederland), World Cancer Research Fund (WCRF), Statistics Netherlands (The Netherlands);

ERC-2009-AdG 232997 and Nordforsk, Nordic Centre of Excellence programme on Food, Nutrition and Health (Norway); Health Research Fund (FIS), PI13/00061 to Granada, PI13/01162 to EPIC-Murcia, Regional Governments of Andalucía, Asturias, Basque Country, Murcia and Navarra, ISCIII RETIC (RD06/0020) (Spain); Swedish Cancer Society, Swedish Research Council, Region Skåne and Region Västerbotten (Sweden); Cancer Research UK (14136 to EPIC-Norfolk; C570/A16491 and C8221/A19170 to EPIC-Oxford), Medical Research Council (1000143 to EPIC-Norfolk, MR/M012190/1 to EPICOxford) (United Kingdom).

The EPIC-Norfolk study (https://doi.org/10.22025/2019.10.105.00004) has received funding from the Medical Research Council (MR/N003284/1 and MC-UU_12015/1) and Cancer Research UK (C864/A14136). The genetics work in the EPIC-Norfolk study was funded by the Medical Research Council (MC_PC_13048). Metabolite measurements in the EPIC-Norfolk study were supported by the MRC Cambridge Initiative in Metabolic Science (MR/L00002/1) and the Innovative Medicines Initiative Joint Undertaking under EMIF grant agreement no. 115372.

EPICOLON: This work was supported by grants from Fondo de Investigación Sanitaria/FEDER (PI08/0024, PI08/1276, PS09/02368, PI11/00219, PI11/00681, PI14/00173, PI14/00230, PI17/00509, 17/00878, PI20/00113, PI20/00226, Acción Transversal de Cáncer), Xunta de Galicia (PGIDIT07PXIB9101209PR), Ministerio de Economia y Competitividad (SAF07-64873, SAF 2010-19273, SAF2014-54453R), Fundación Científica de la Asociación Española contra el Cáncer (GCB13131592CAST), Beca Grupo de Trabajo "Oncología" AEG (Asociación Española de Gastroenterología), Fundación Privada Olga Torres, FP7 CHIBCHA Consortium, Agència de Gestió d'Ajuts Universitaris i de Recerca (AGAUR, Generalitat de Catalunya, 2014SGR135, 2014SGR255, 2017SGR21, 2017SGR653), Catalan Tumour Bank Network (Pla Director d'Oncologia, Generalitat de Catalunya), PERIS (SLT002/16/00398, Generalitat de Catalunya), Marató TV3 (202008-10), CERCA Programme (Generalitat de Catalunya) and COST Actions BM1206 and CA17118. CIBERehd is funded by the Instituto de Salud Carlos III.

ESTHER/VERDI. This work was supported by grants from the Baden-Württemberg Ministry of Science, Research and Arts and the German Cancer Aid.

Harvard cohorts (HPFS, NHS, PHS): HPFS is supported by the National Institutes of Health (P01 CA055075, UM1 CA167552, U01 CA167552, R01 CA137178, R01 CA151993, and R35 CA197735), NHS by the National Institutes of Health (R01 CA137178, P01 CA087969, UM1 CA186107, R01 CA151993, and R35 CA197735) and PHS by the National Institutes of Health (R01 CA042182).

Hawaii Adenoma Study: NCI grants R01 CA72520.

HCES-CRC: the Hwasun Cancer Epidemiology Study–Colon and Rectum Cancer (HCES-CRC; grants from Chonnam National University Hwasun Hospital, HCRI15011-1).

IARC: disclaimer: where authors are identified as personnel of the INternational Agency for Research on Cancer/World Health Organization, the authors alone are responsible for the views expressed in this article and they do not necessarily represent the decisions, policy or views of the International Agency for Research on Cancer/World Health Organization.

Kentucky: This work was supported by the following grant support: Clinical Investigator Award from Damon Runyon Cancer Research Foundation (CI-8); NCI R01CA136726.

LCCS: The Leeds Colorectal Cancer Study was funded by the Food Standards Agency and Cancer Research UK Programme Award (C588/A19167).

Melbourne Collaborative Cohort Study (MCCS) cohort recruitment was funded by VicHealth and Cancer Council Victoria. The MCCS was further augmented by Australian National Health and Medical Research Council grants 209057, 396414 and 1074383 and by infrastructure provided by Cancer Council Victoria. Cases and their vital status were ascertained through the Victorian Cancer Registry and the Australian Institute of Health and Welfare, including the National Death Index and the Australian Cancer Database.

Multiethnic Cohort (MEC) Study: National Institutes of Health (R37 CA54281, P01 CA033619, R01 CA063464 and U01 CA164973).

MECC: This work was supported by the National Institutes of Health, U.S. Department of Health and Human Services (R01 CA81488 to SBG and GR).

MSKCC: The work at Sloan Kettering in New York was supported by the Robert and Kate Niehaus Center for Inherited Cancer Genomics and the Romeo Milio Foundation. Moffitt: This work was supported by funding from the National Institutes of Health (grant numbers R01 CA189184, P30 CA076292), Florida Department of Health Bankhead-Coley Grant 09BN-13, and the University of South Florida Oehler Foundation. Moffitt contributions were supported in part by the Total Cancer Care Initiative, Collaborative Data Services Core, and Tissue Core at the H. Lee Moffitt Cancer Center & Research Institute, a National Cancer Institute-designated Comprehensive Cancer Center (grant number P30 CA076292).

NCCCS I & II: We acknowledge funding support for this project from the National Institutes of Health, R01 CA66635 and P30 DK034987.

NFCCR: This work was supported by an Interdisciplinary Health Research Team award from the Canadian Institutes of Health Research (CRT 43821); the National Institutes of Health, U.S. Department of Health and Human Services (U01 CA74783); and National Cancer Institute of Canada grants (18223 and 18226). The authors wish to acknowledge the contribution of Alexandre Belisle and the genotyping team of the McGill University and Génome Québec Innovation Centre, Montréal, Canada, for genotyping the Sequenom panel in the NFCCR samples. Funding was provided to Michael O. Woods by the Canadian Cancer Society Research Institute.

NSHDS: Swedish Research Council; Swedish Cancer Society; Cutting-Edge Research Grant and other grants from Region Västerbotten; Knut and Alice Wallenberg Foundation; Lion's Cancer Research Foundation at Umeå University; the Cancer Research Foundation in Northern Sweden; and the Faculty of Medicine, Umeå University, Umeå, Sweden.

OSUMC: OCCPI funding was provided by Pelotonia and HNPCC funding was provided by the NCI (CA16058 and CA67941).

PLCO: Intramural Research Program of the Division of Cancer Epidemiology and Genetics and supported by contracts from the Division of Cancer Prevention, National Cancer Institute, NIH, DHHS. Funding was provided by National Institutes of Health (NIH), Genes, Environment and Health Initiative (GEI) Z01 CP 010200, NIH U01 HG004446, and NIH GEI U01 HG 004438.

SEARCH: The University of Cambridge has received salary support in respect of PDPP from the NHS in the East of England through the Clinical Academic Reserve. Cancer Research UK (C490/A16561); the UK National Institute for Health Research Biomedical Research Centres at the University of Cambridge.

SELECT: Research reported in this publication was supported in part by the National Cancer Institute of the National Institutes of Health under Award Numbers U10 CA37429 (CD Blanke), and UM1 CA182883 (CM Tangen/IM Thompson). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

SMS and REACH: This work was supported by the National Cancer Institute (grant P01 CA074184 to J.D.P. and P.A.N., grants R01 CA097325, R03 CA153323, and K05 CA152715 to P.A.N., and the National Center for Advancing Translational Sciences at the National Institutes of Health (grant KL2 TR000421 to A.N.B.-H.)

The Swedish Low-risk Colorectal Cancer Study: The study was supported by grants from the Swedish research council; K2015-55X-22674-01-4, K2008-55X-20157-03-3, K2006-72X-20157-01-2 and the Stockholm County Council (ALF project).

Swedish Mammography Cohort and Cohort of Swedish Men: This work is supported by the Swedish Research Council /Infrastructure grant, the Swedish Cancer Foundation, and the Karolinska Institute´s Distinguished Professor Award to Alicja Wolk.

UK Biobank: This research has been conducted using the UK Biobank Resource under Application Number 8614

VITAL: National Institutes of Health (K05 CA154337).

WHI: The WHI program is funded by the National Heart, Lung, and Blood Institute, National Institutes of Health, U.S. Department of Health and Human Services through contracts HHSN268201100046C, HHSN268201100001C, HHSN268201100002C, HHSN268201100003C, HHSN268201100004C, and HHSN271201100004C.

CCFR: The Colon CFR graciously thanks the generous contributions of their study participants, dedication of study staff, and the financial support from the U.S. National Cancer Institute, without which this important registry would not exist. The authors would like to thank the study participants and staff of the Seattle Colon Cancer Family Registry and the Hormones and Colon Cancer study (CORE Studies).

CLUE II: We thank the participants of Clue II and appreciate the continued efforts of the staff at the Johns Hopkins George W. Comstock Center for Public Health Research and Prevention in the conduct of the Clue II Cohort Study.

COLON and NQplus: the authors would like to thank the COLON and NQplus investigators at Wageningen University & Research and the involved clinicians in the participating hospitals.

CORSA: We kindly thank all those who contributed to the screening project Burgenland against CRC. Furthermore, we are grateful to Doris Mejri and Monika Hunjadi for laboratory assistance.

CPS-II: The authors thank the CPS-II participants and Study Management Group for their invaluable contributions to this research. The authors would also like to acknowledge the contribution to this study from central cancer registries supported through the Centers for Disease Control and Prevention National Program of Cancer Registries, and cancer registries supported by the National Cancer Institute Surveillance Epidemiology and End Results program.

Czech Republic CCS: We are thankful to all clinicians in major hospitals in the Czech Republic, without whom the study would not be practicable. We are also sincerely grateful to all patients participating in this study.

DACHS: We thank all participants and cooperating clinicians, and Ute Handte-Daub, Utz Benscheid, Muhabbet Celik and Ursula Eilber for excellent technical assistance.

EDRN: We acknowledge all the following contributors to the development of the resource: University of Pittsburgh School of Medicine, Department of Gastroenterology, Hepatology and Nutrition: Lynda Dzubinski; University of Pittsburgh School of Medicine, Department of Pathology: Michelle Bisceglia; and University of Pittsburgh School of Medicine, Department of Biomedical Informatics.

EPIC: Where authors are identified as personnel of the International Agency for Research on Cancer/World Health Organization, the authors alone are responsible for the views expressed in this article and they do not necessarily represent the decisions, policy or views of the International Agency for Research on Cancer/World Health Organization.

The EPIC-Norfolk study: we are grateful to all the participants who have been part of the project and to the many members of the study teams at the University of Cambridge who have enabled this research.

EPICOLON: We are sincerely grateful to all patients participating in this study who were recruited as part of the EPICOLON project. We acknowledge the Spanish National DNA Bank, Biobank of Hospital Clínic–

IDIBAPS and Biobanco Vasco for the availability of the samples. The work was carried out (in part) at the Esther Koplowitz Centre, Barcelona.

Harvard cohorts (HPFS, NHS, PHS): The study protocol was approved by the institutional review boards of the Brigham and Women's Hospital and Harvard T.H. Chan School of Public Health, and those of participating registries as required. We would like to thank the participants and staff of the HPFS, NHS and PHS for their valuable contributions as well as the following state cancer registries for their help: AL, AZ, AR, CA, CO, CT, DE, FL, GA, ID, IL, IN, IA, KY, LA, ME, MD, MA, MI, NE, NH, NJ, NY, NC, ND, OH, OK, OR, PA, RI, SC, TN, TX, VA, WA, WY. The authors assume full responsibility for analyses and interpretation of these data.

Interval: A complete list of the investigators and contributors to the INTERVAL trial is provided in reference (32). The academic coordinating centre would like to thank blood donor centre staff and blood donors for participating in the INTERVAL trial.

Kentucky: We would like to acknowledge the staff at the Kentucky Cancer Registry.

LCCS: We acknowledge the contributions of co-investigators: David Forman, Roland Wolf, Jennifer Barrett, Robin Waxman, Gillian Smith and Emma Northwood in conducting this study as well as all the interviewers.

NCCCS I & II: We would like to thank the study participants, and the NC Colorectal Cancer Study staff.

NSHDS investigators thank the Biobank Research Unit at Umeå University, the Västerbotten Intervention Programme, the Northern Sweden MONICA study and Region Västerbotten for providing data and samples and acknowledge the contribution from Biobank Sweden, supported by the Swedish Research Council (VR 2017-00650).

PLCO: The authors thank the PLCO Cancer Screening Trial screening center investigators and the staff from Information Management Services Inc and Westat Inc. Most importantly, we thank the study participants for their contributions that made this study possible.

SEARCH: We thank the SEARCH team.

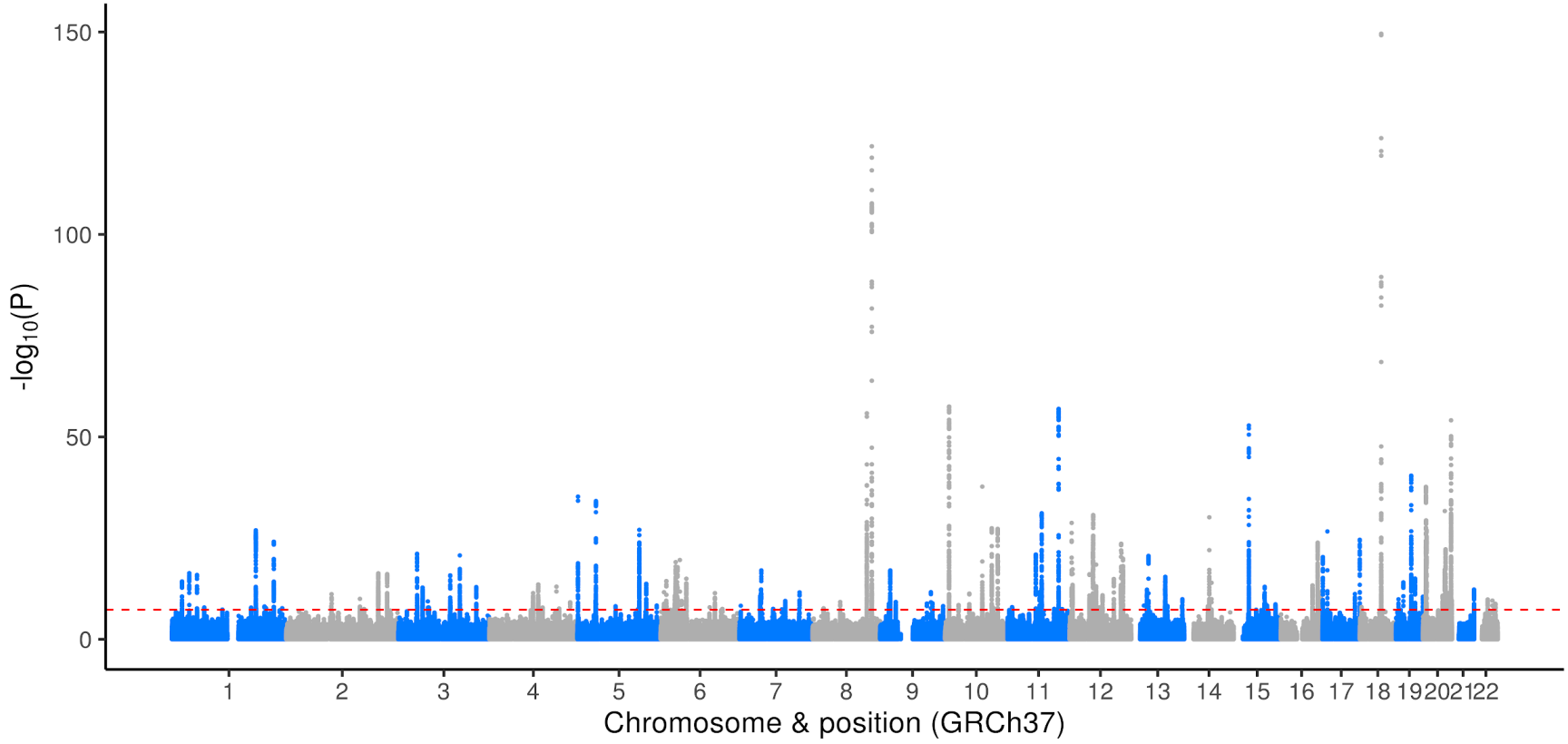
SELECT: We thank the research and clinical staff at the sites that participated on SELECT study, without whom the trial would not have been successful. We are also grateful to the 35,533 dedicated men who participated in SELECT.

UK Biobank: This research has been conducted using the UK Biobank Resource under Application Number 8614 and 7441. We would like to thank the participants and researchers UK Biobank for their participation and acquisition of data.

WHI: The authors thank the WHI investigators and staff for their dedication, and the study participants for making the program possible. A full listing of WHI investigators can be found at:

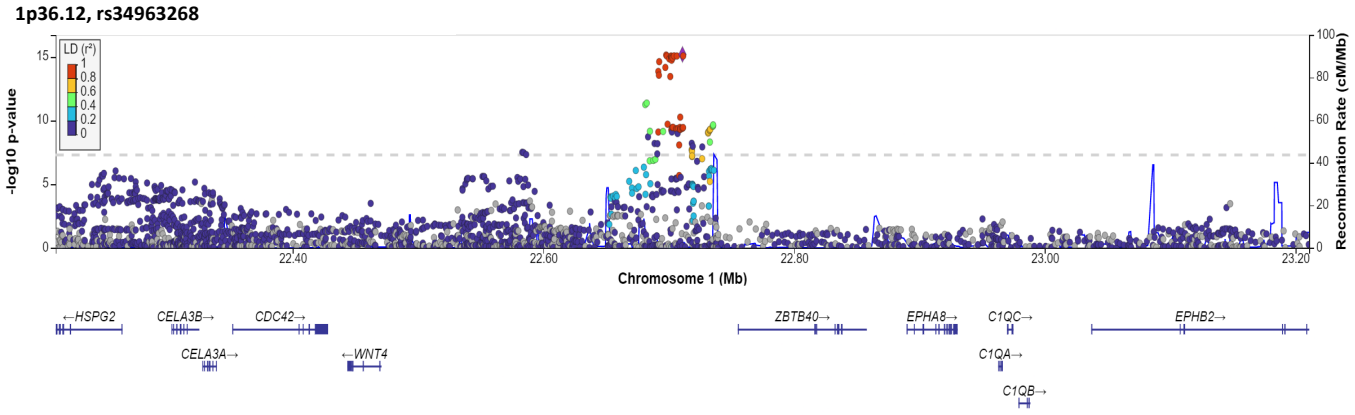
http://www.whi.org/researchers/Documents%20%20Write%20a%20Paper/WHI%20Investigator%20Short%20List.pdf

**Supplementary Figure 1: Manhattan plot showing all lead SNPs independently associated with colorectal cancer risk at $P < 5 \times 10^{-8}$.** The red line indicates the genome-wide significance threshold. The x-axis represents the chromosomal positions and the y-axis represents the $-\log_{10}$ *P*-values of the SNPs. P-values obtained from the meta-analysis.

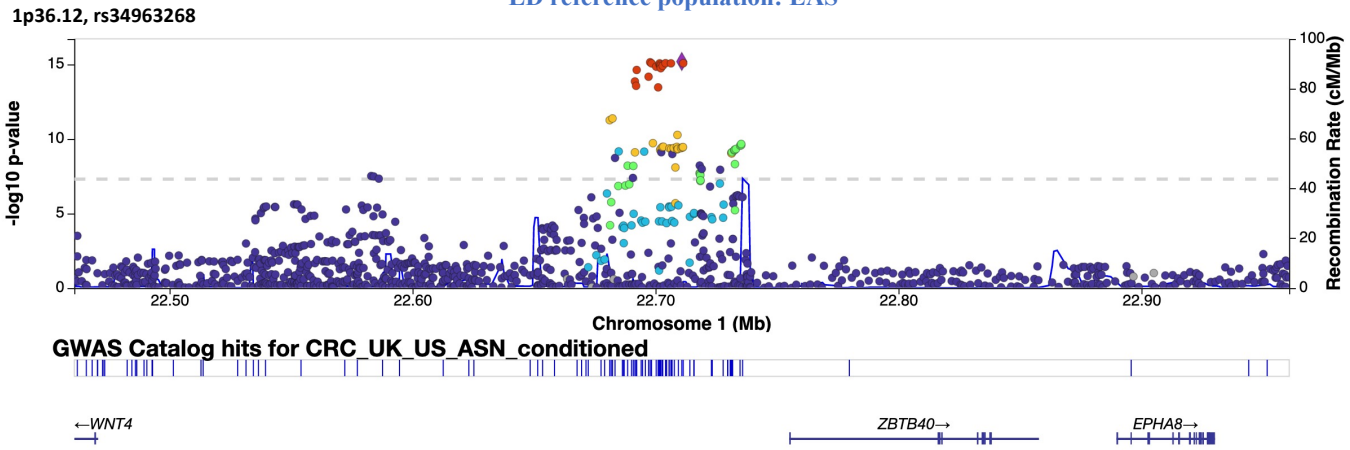


**Supplementary Figure 2: Regional association plots for the new colorectal cancer risk loci reaching genome-wide significance ($P < 5 \times 10^{-8}$).**
P-values obtained from the meta-analysis.

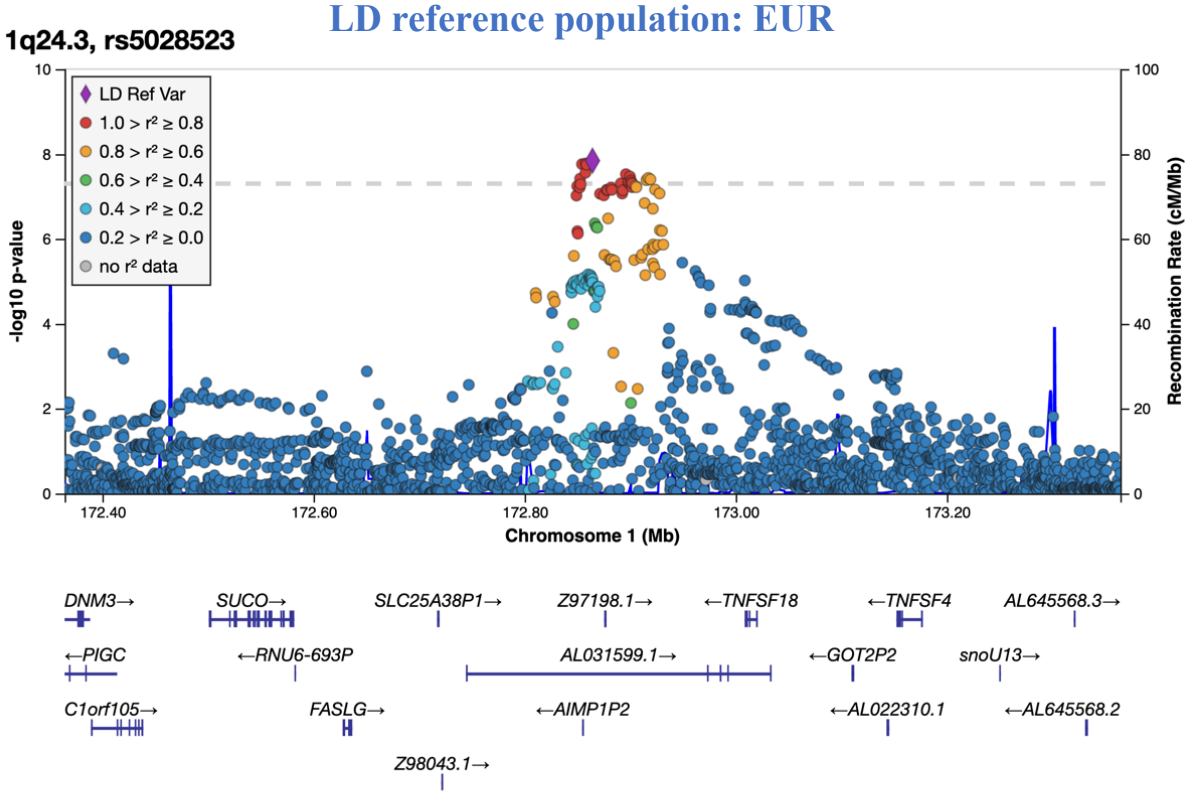
## LD reference population: EUR



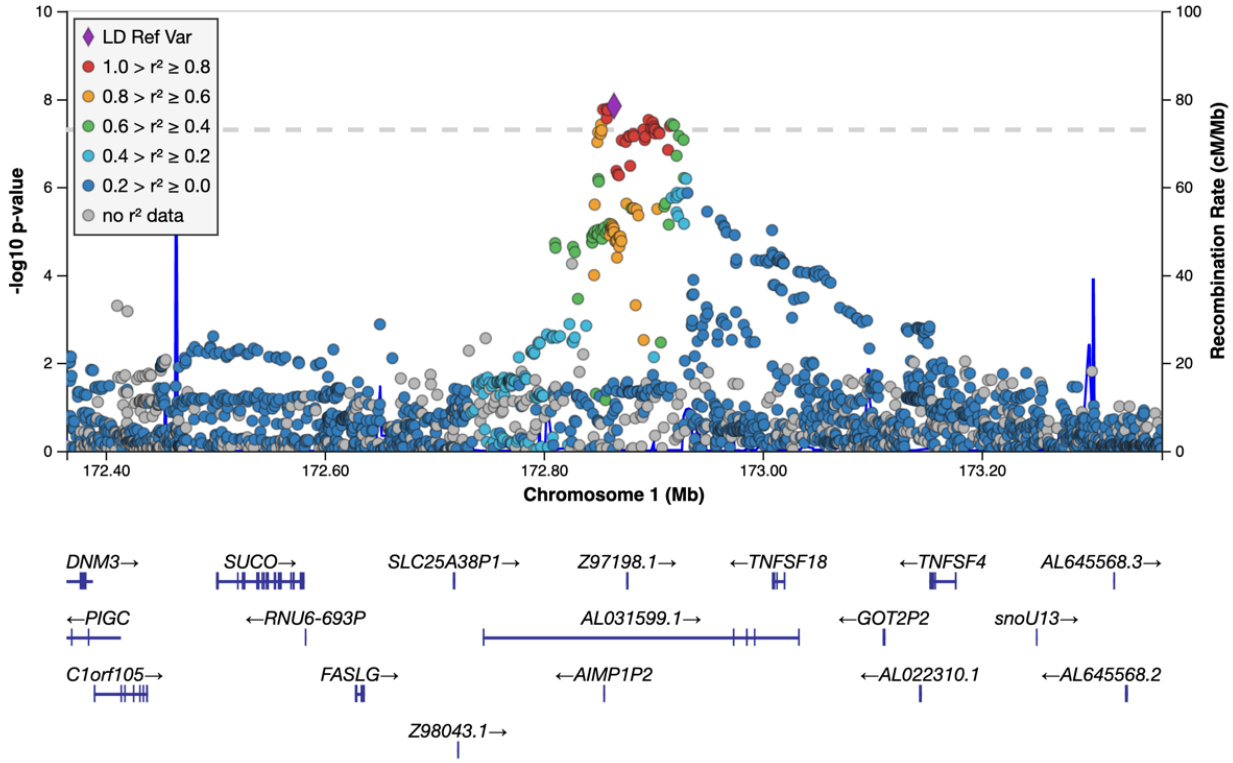

## LD reference population: EAS



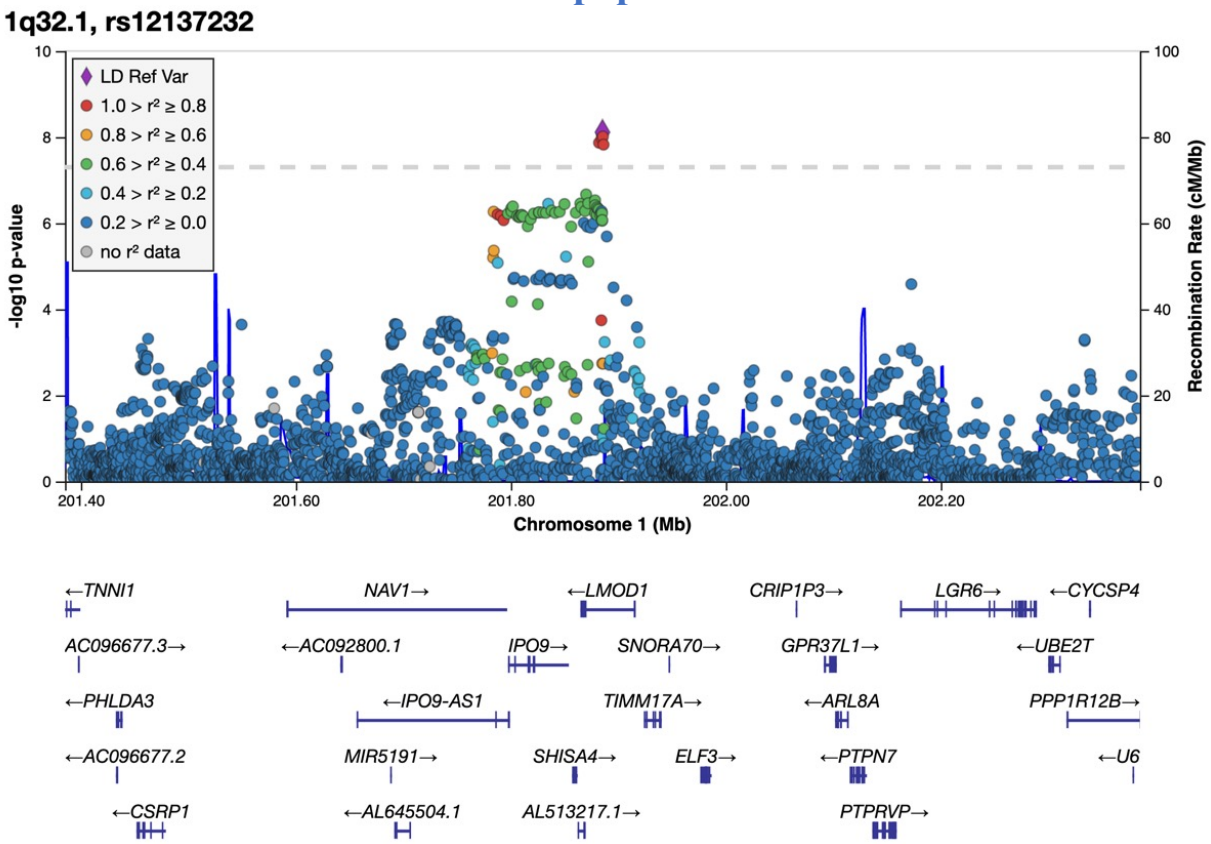
## LD reference population: EUR



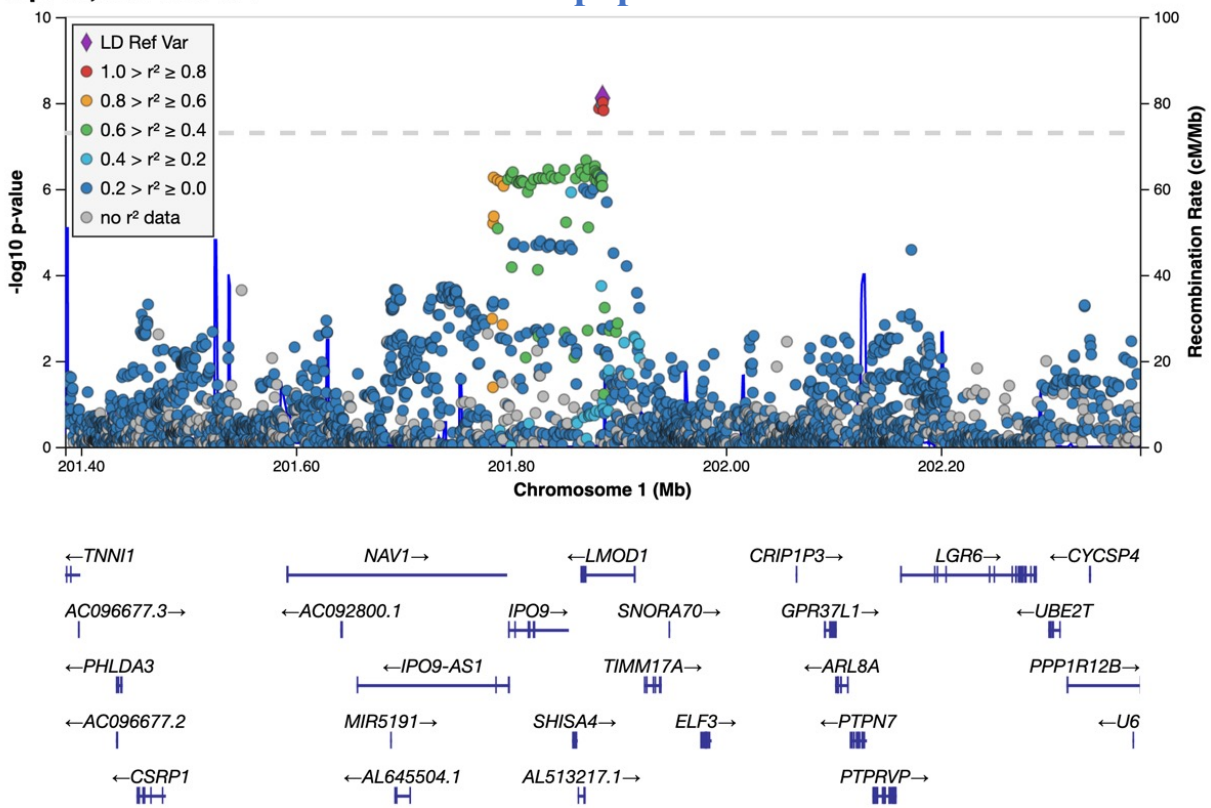

## LD reference population: EAS



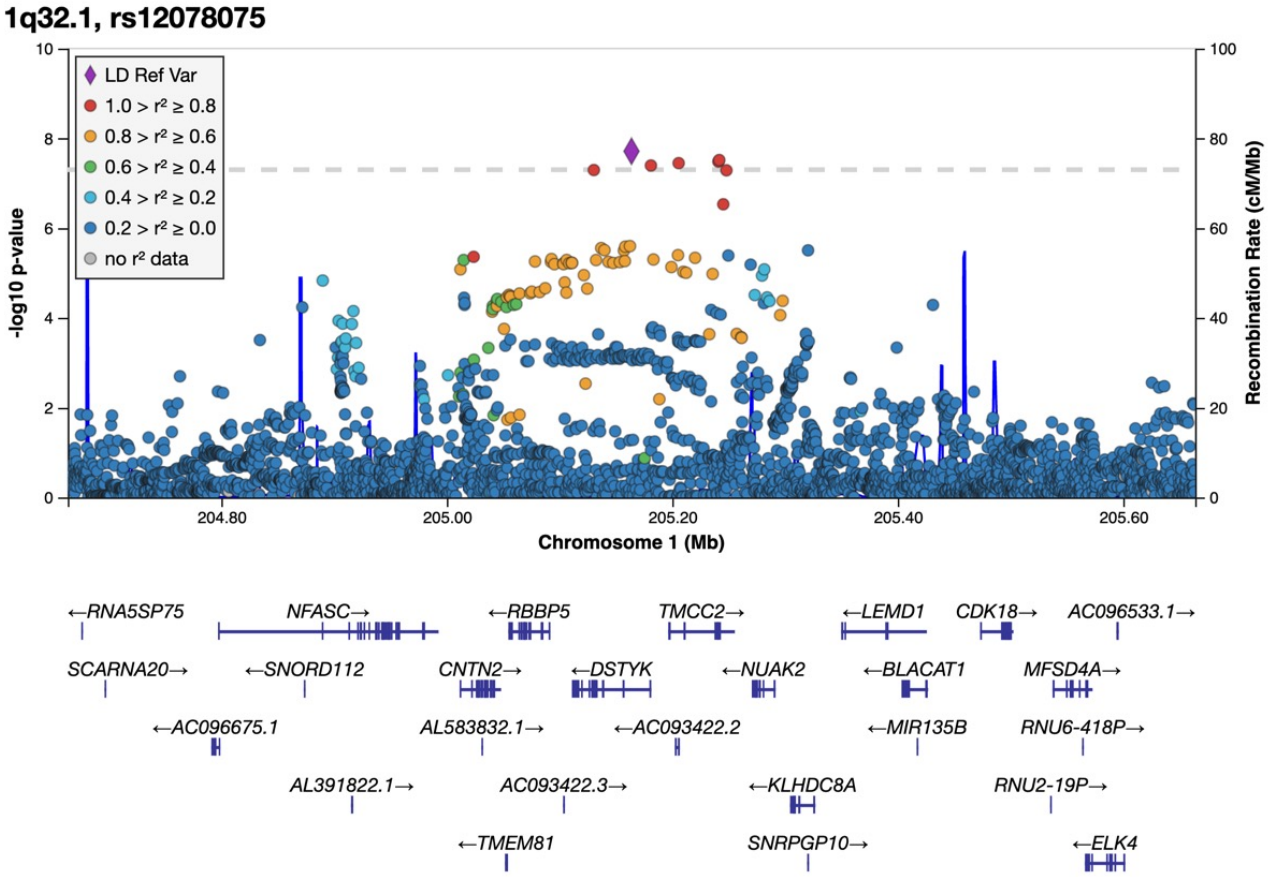
## LD reference population: EUR




## LD reference population: EAS



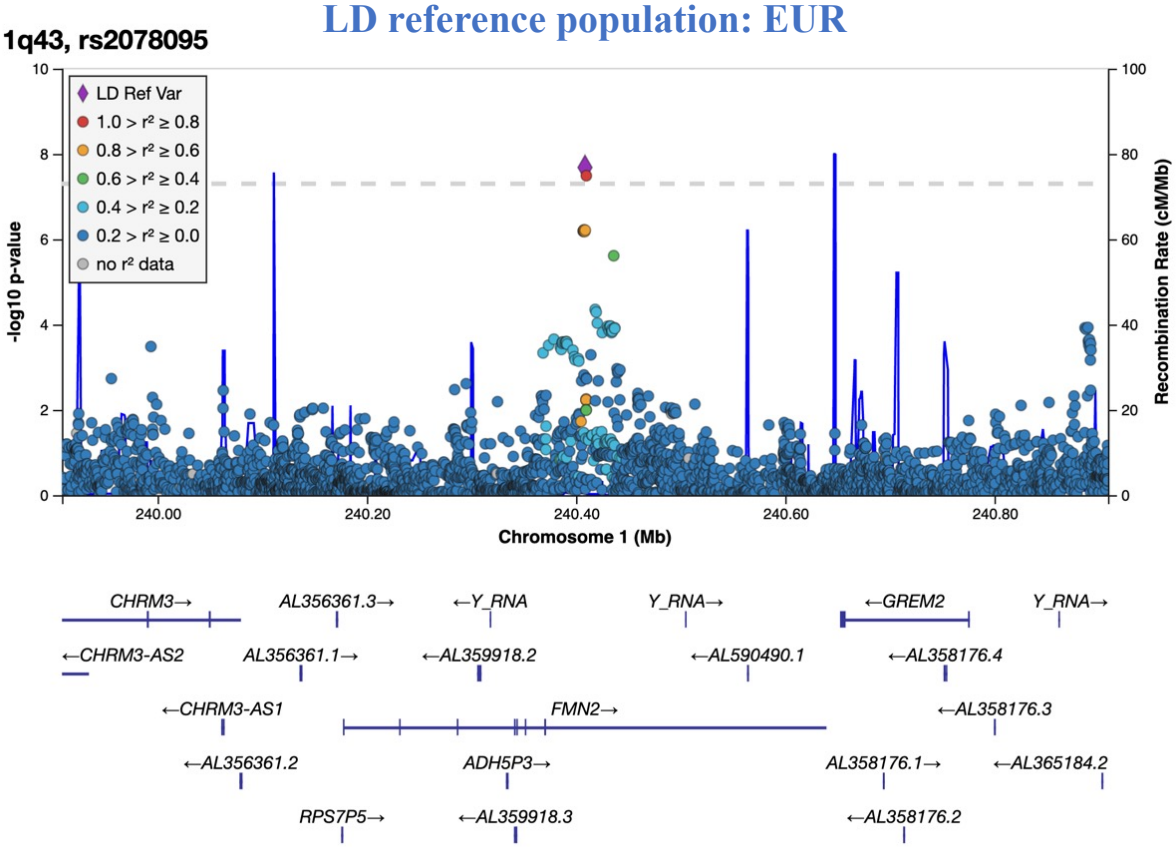
## LD reference population: EUR



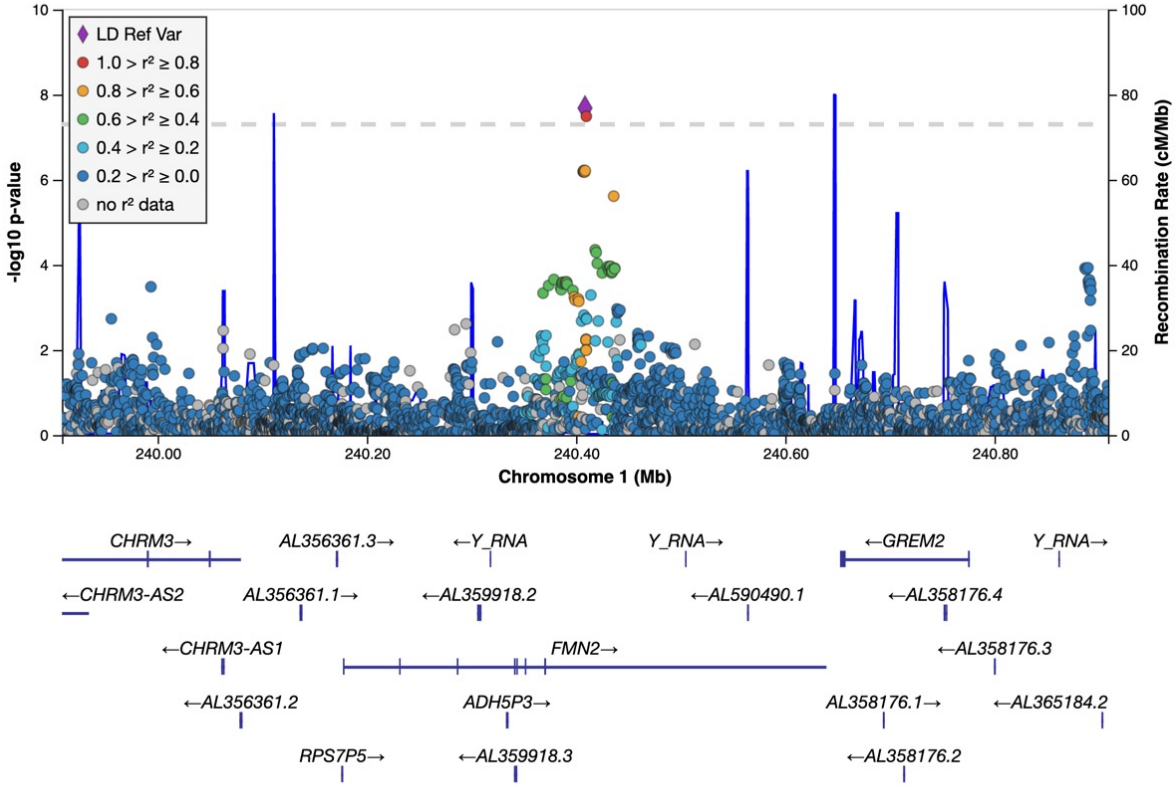

## LD reference population: EAS

rs12078075 is monoallelic in the EAS population

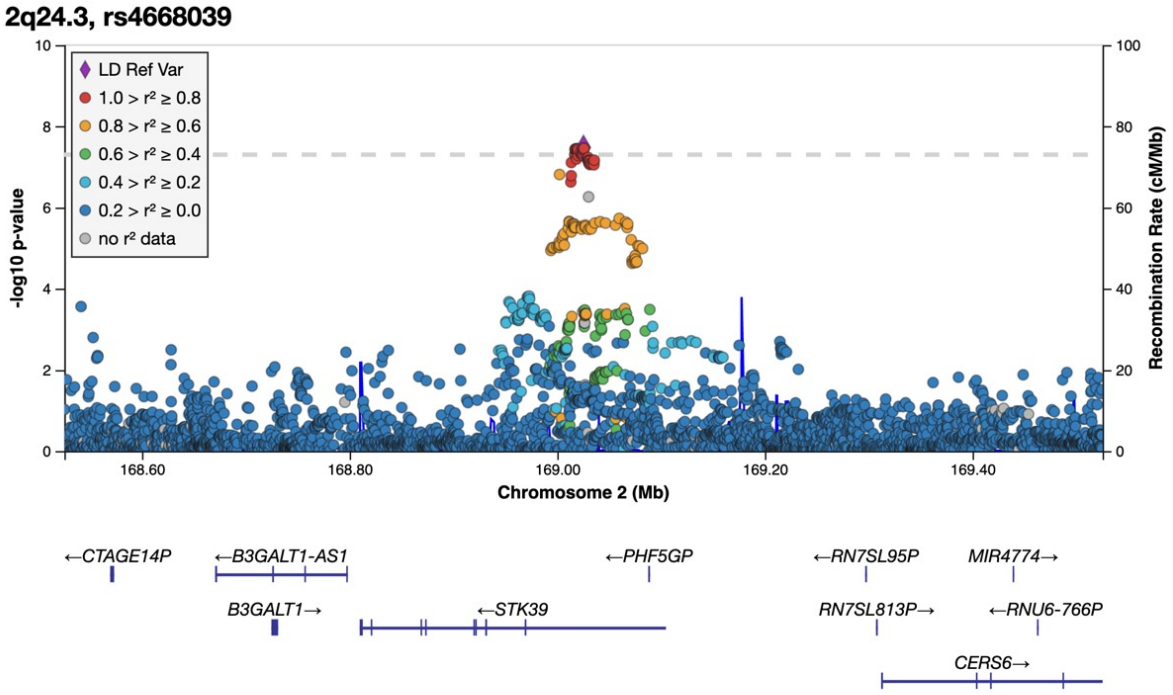
## LD reference population: EUR



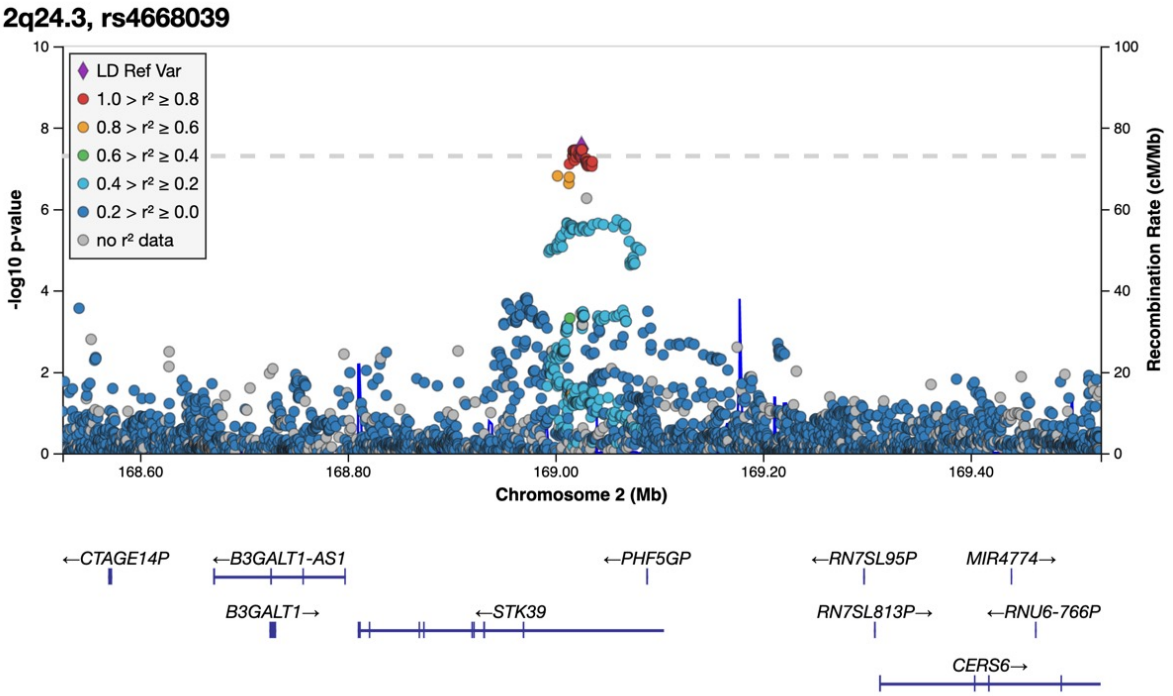

## LD reference population: EAS



## LD reference population: EUR

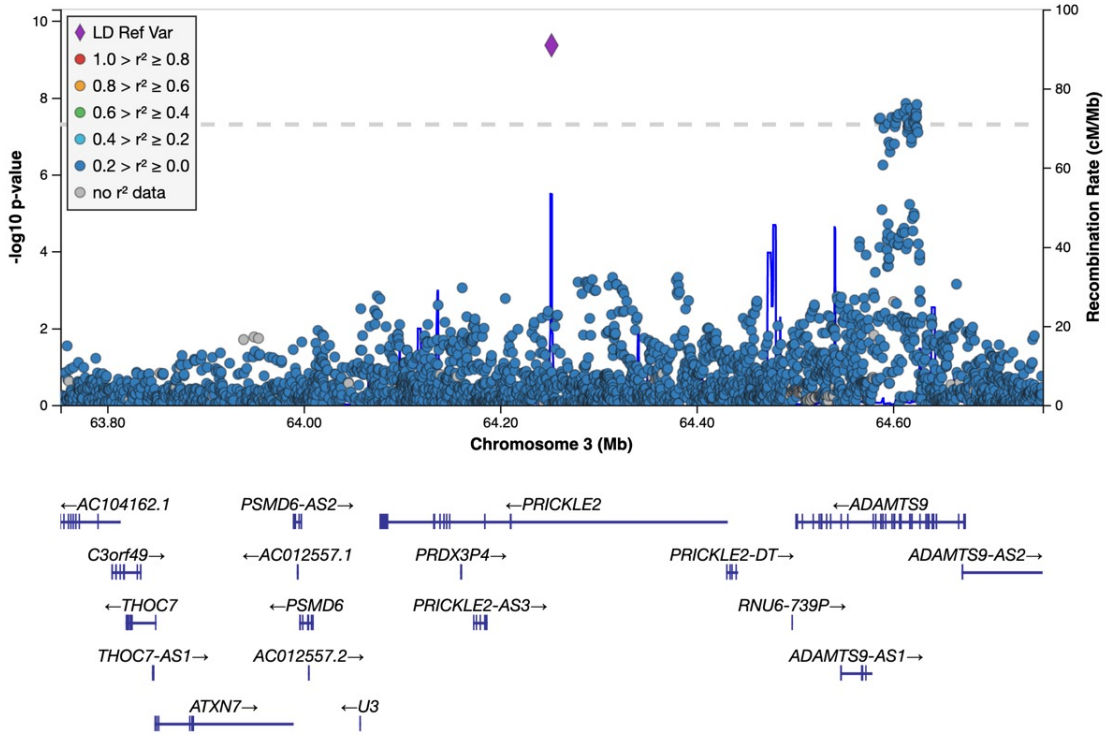



## LD reference population: EAS



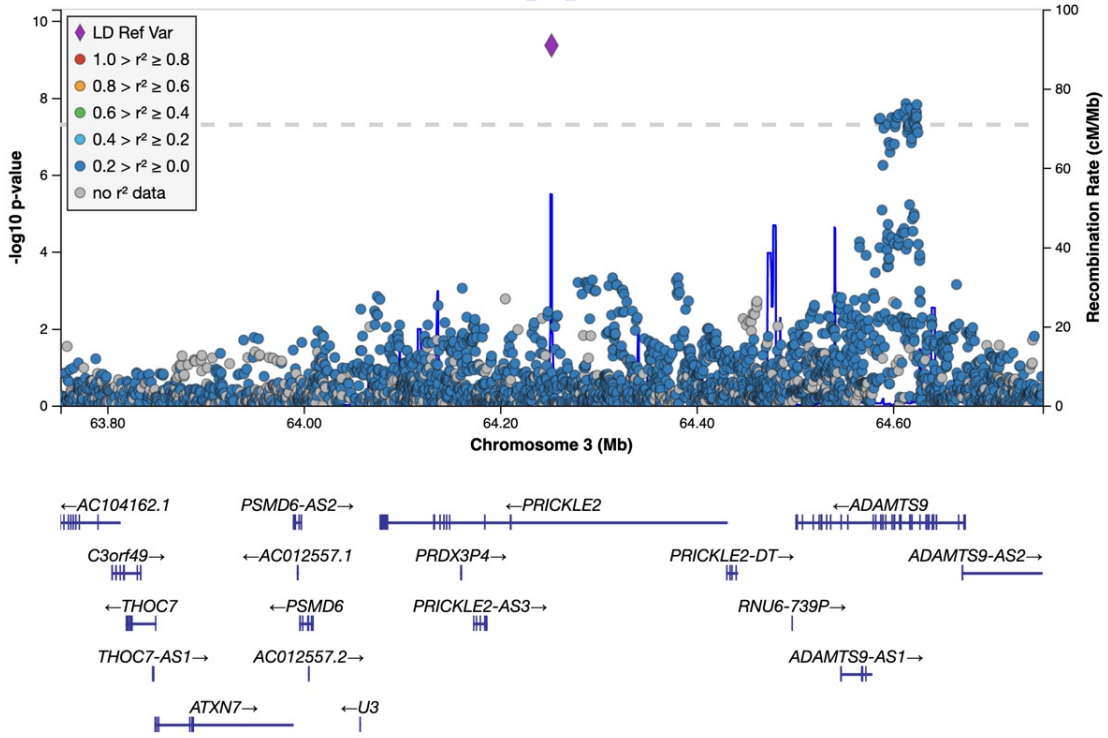
3p14.1, rs704417

## LD reference population: EUR

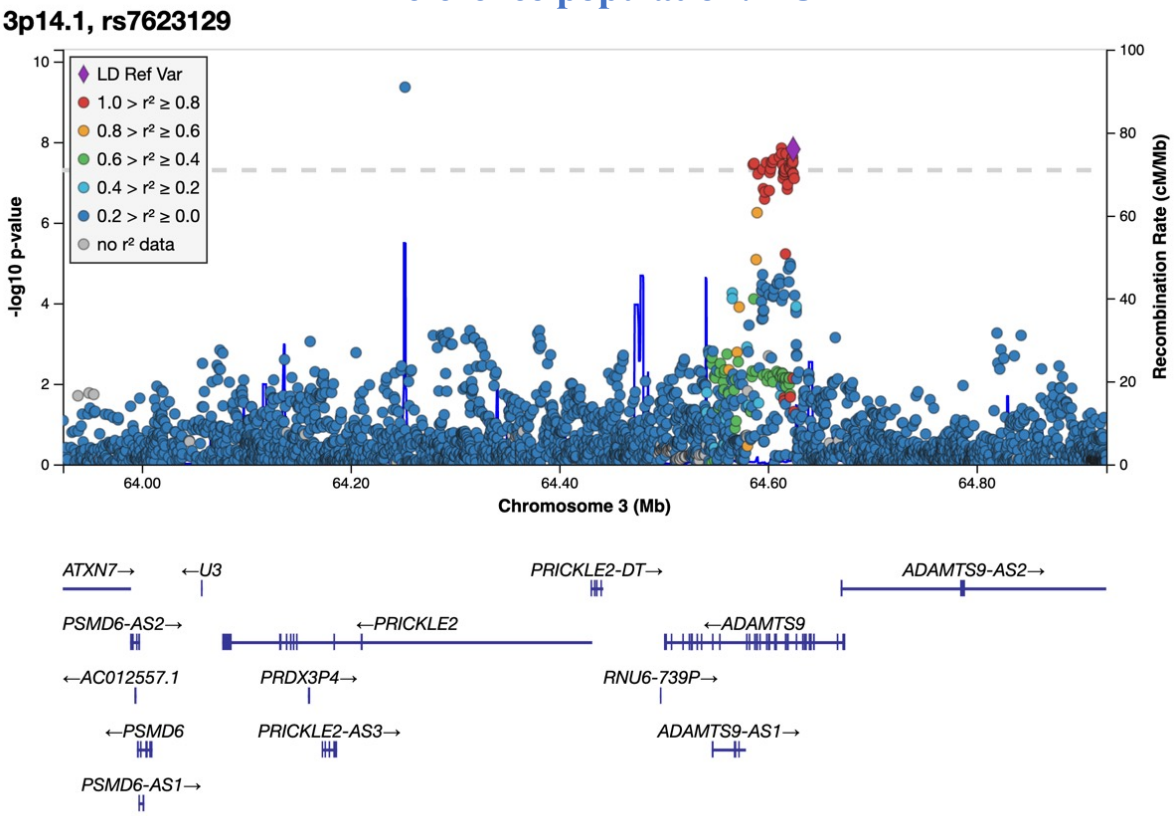



3p14.1, rs704417

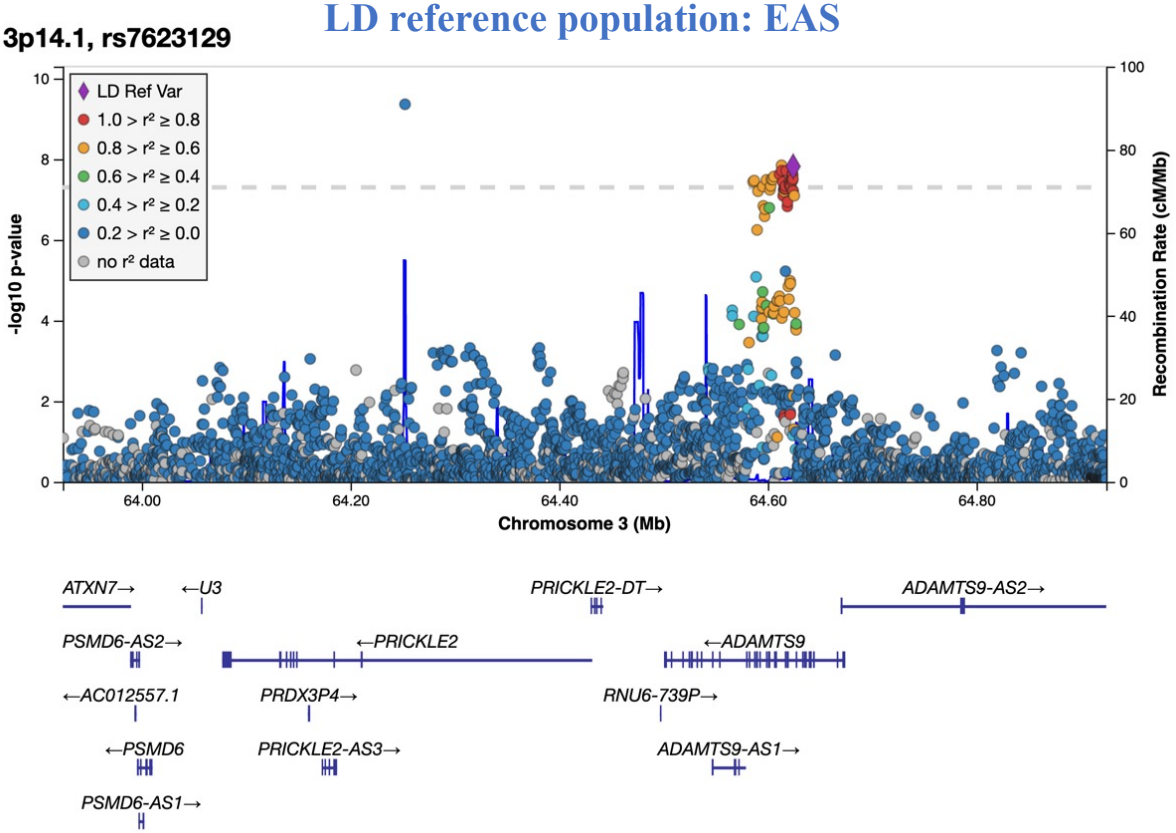
## LD reference population: EAS



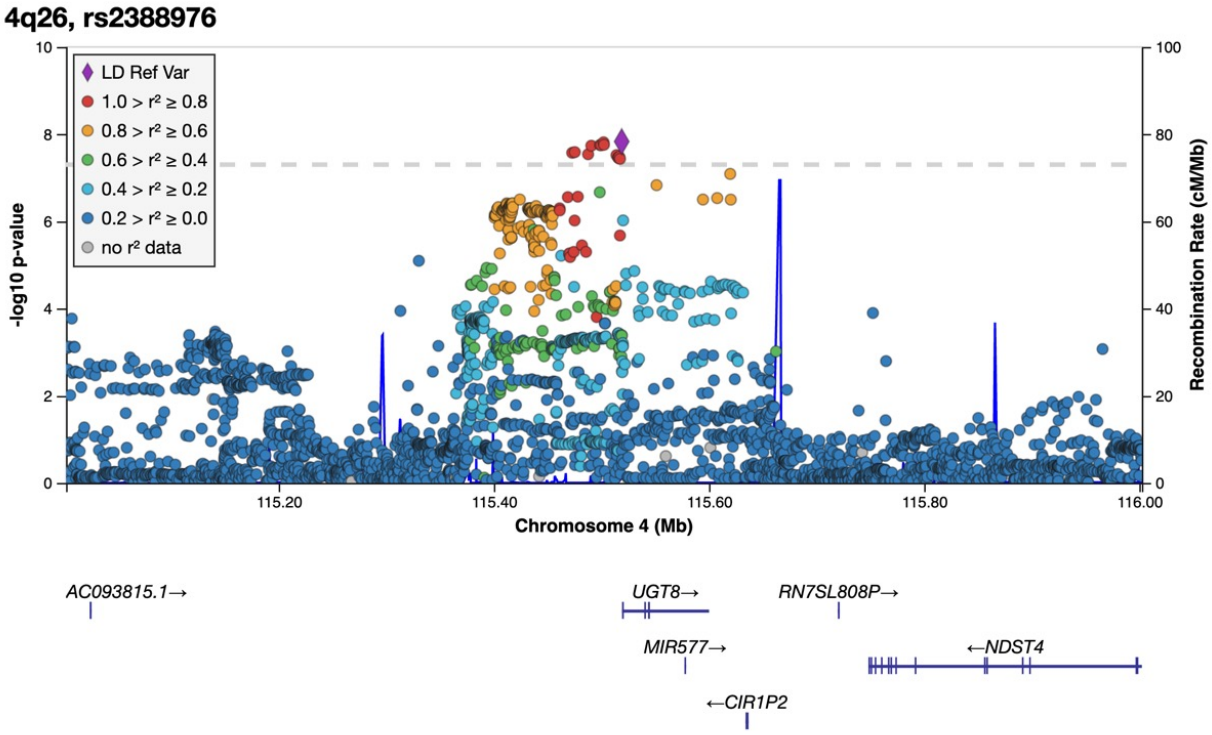
## LD reference population: EUR



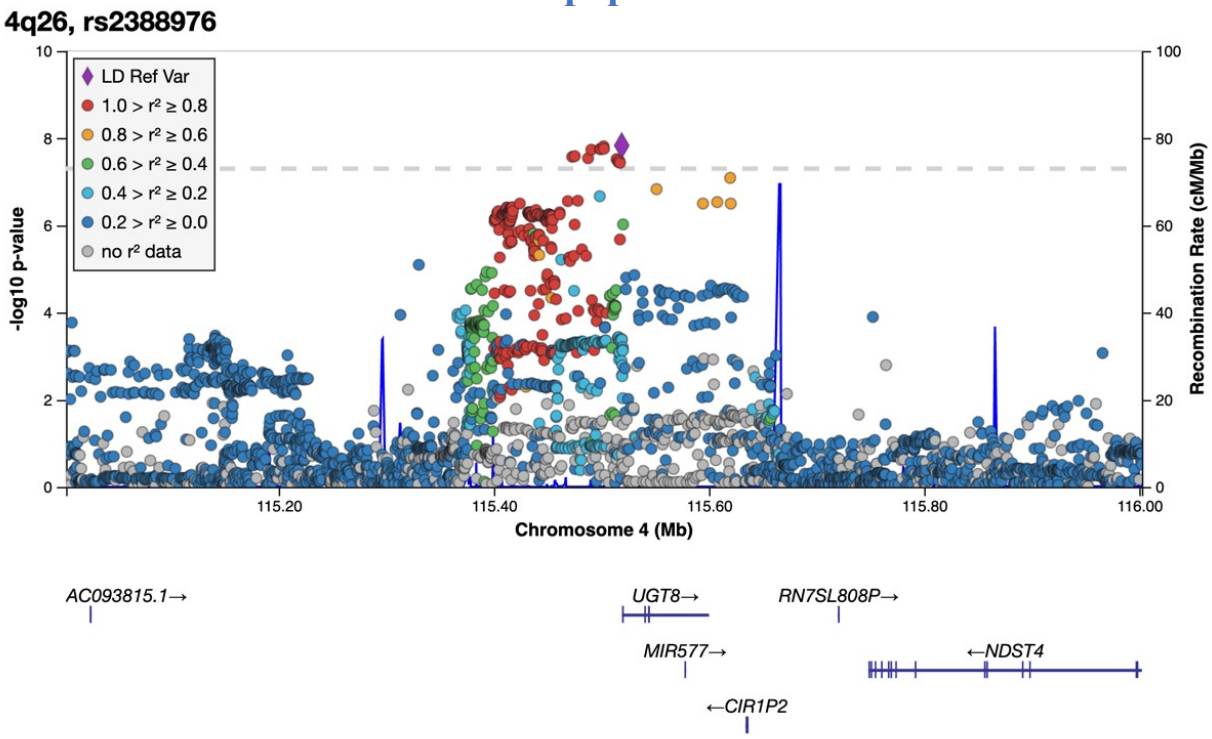

## LD reference population: EAS



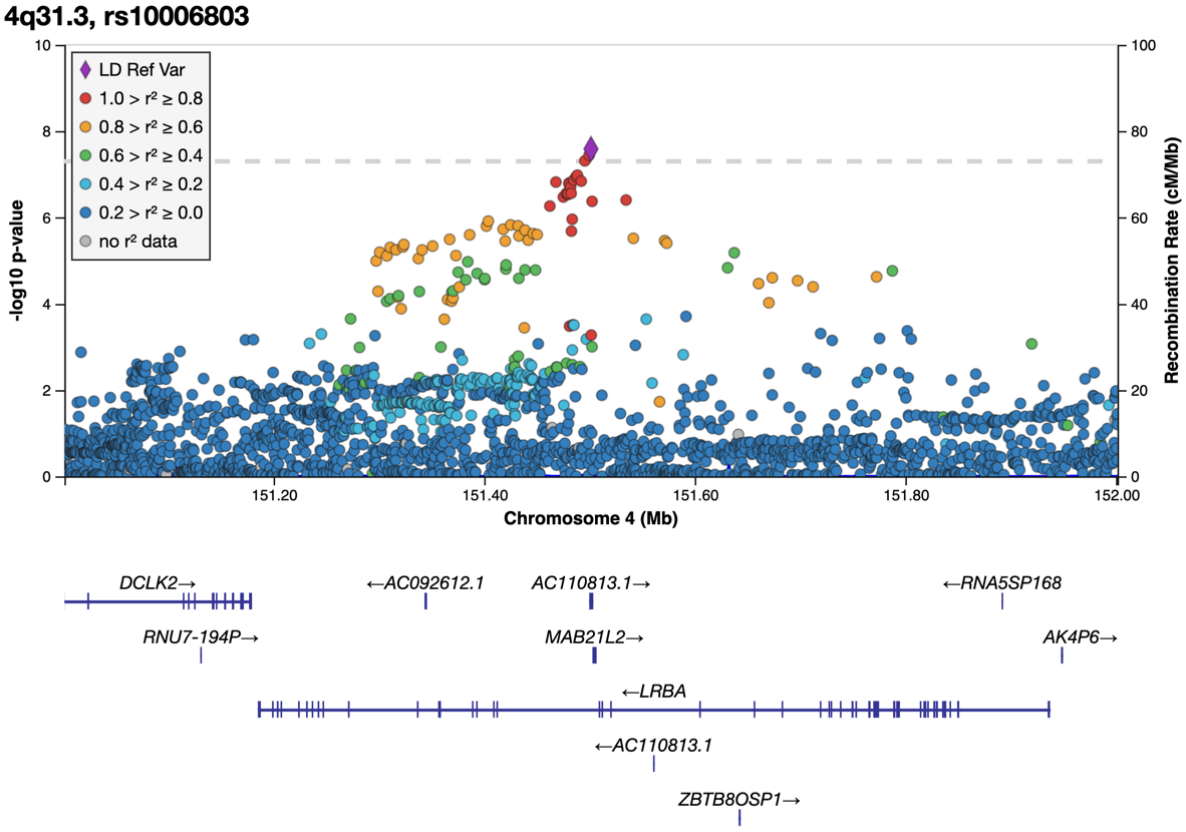
## LD reference population: EUR



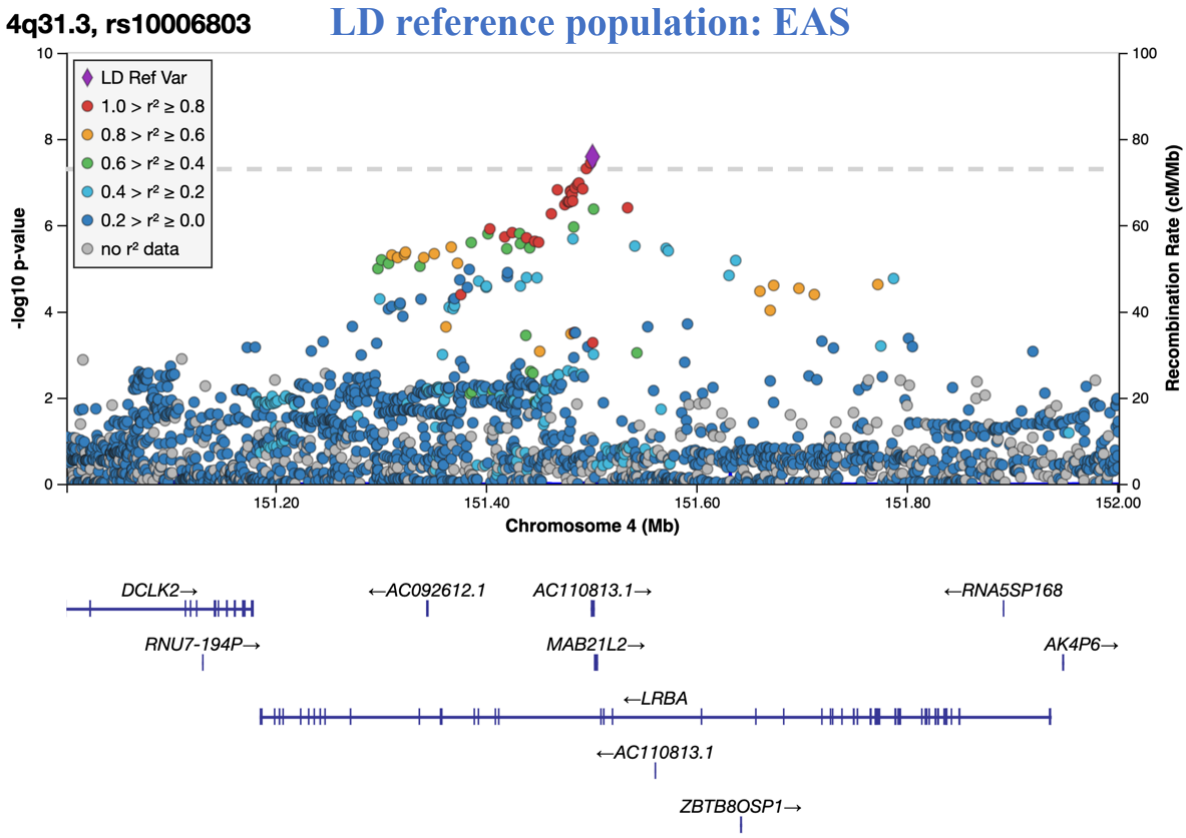

## LD reference population: EAS



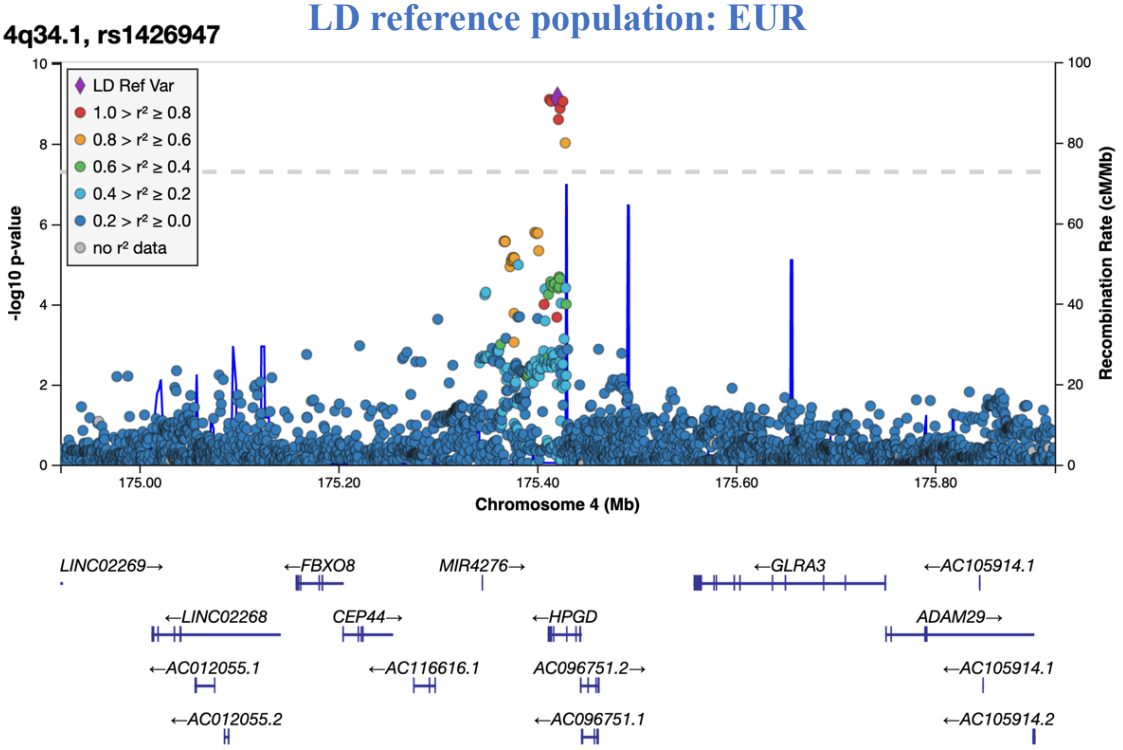
## LD reference population: EUR



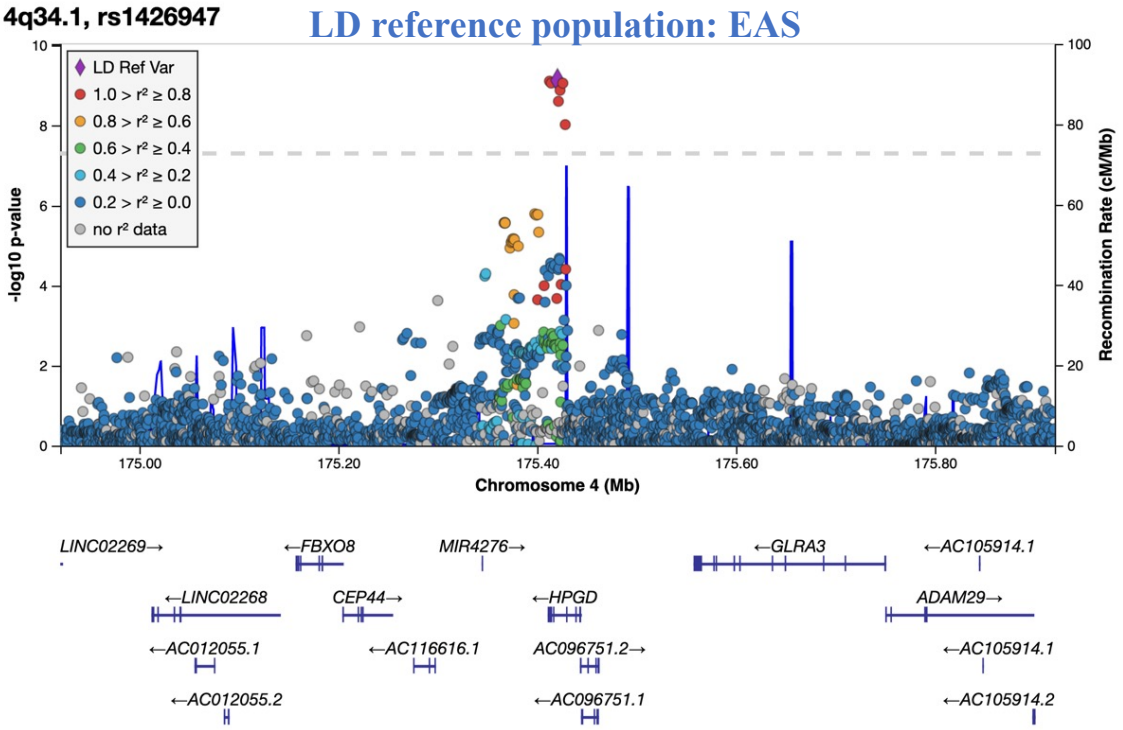

## LD reference population: EAS

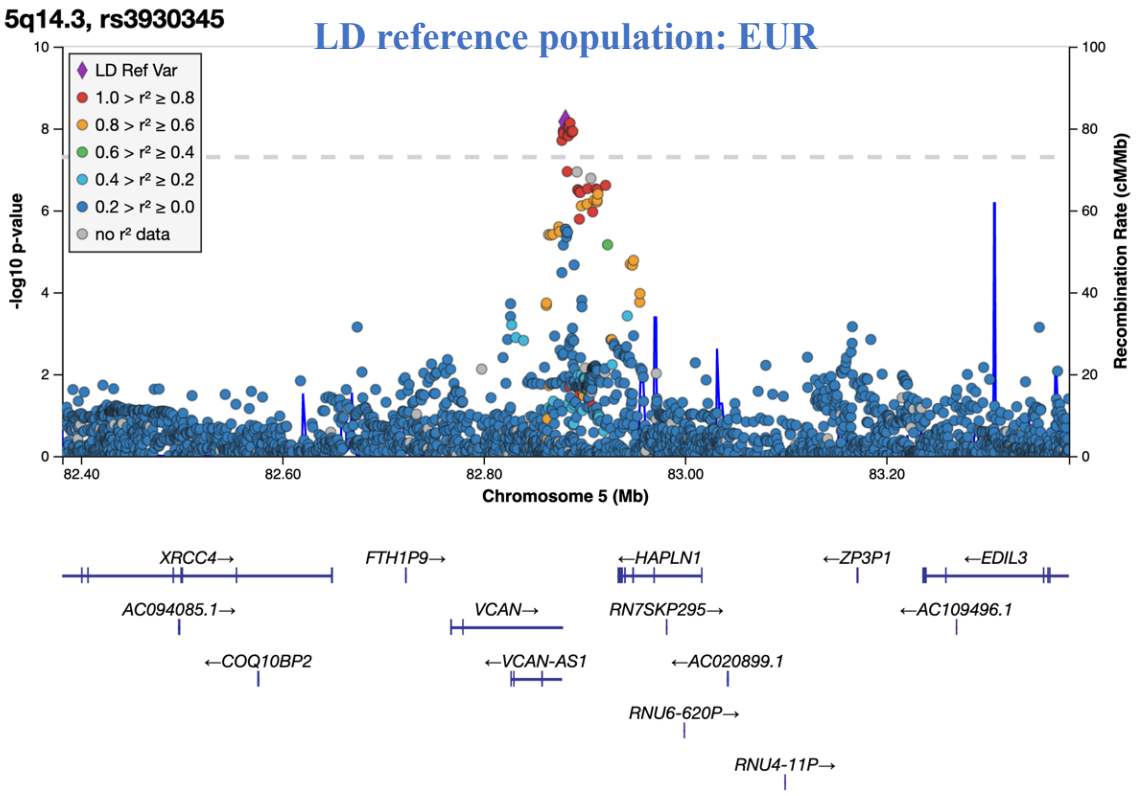


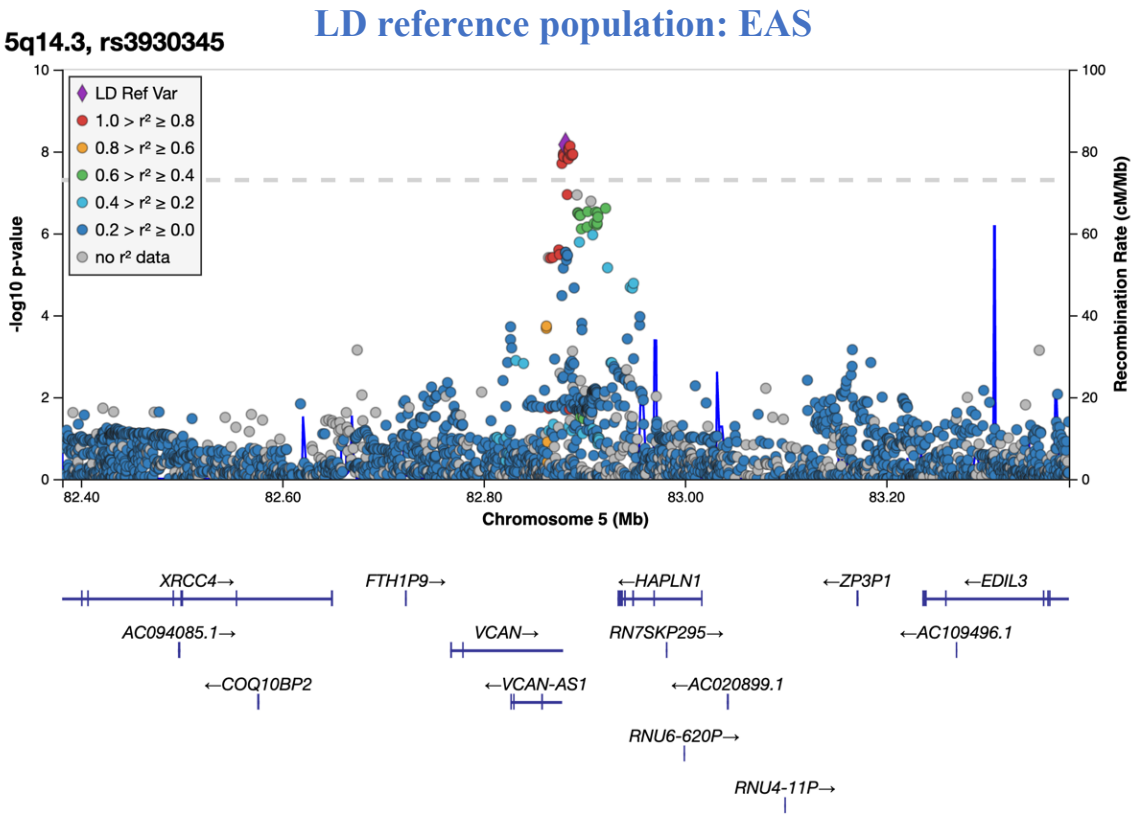
## LD reference population: EUR



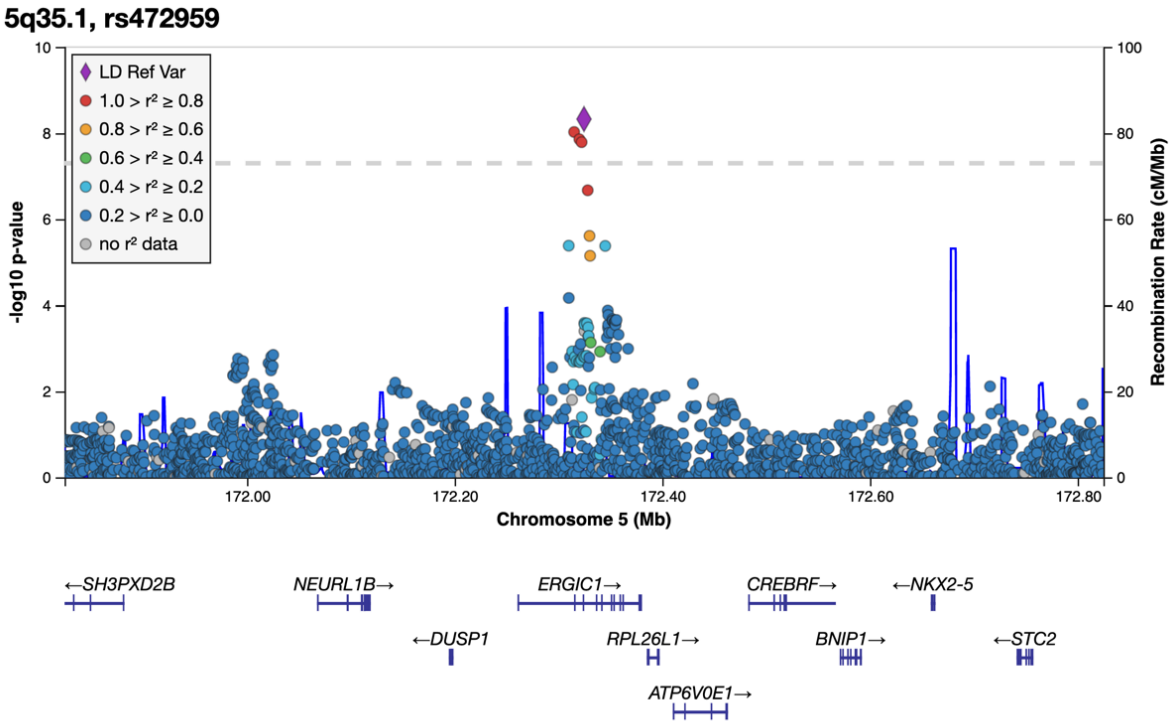

## LD reference population: EAS



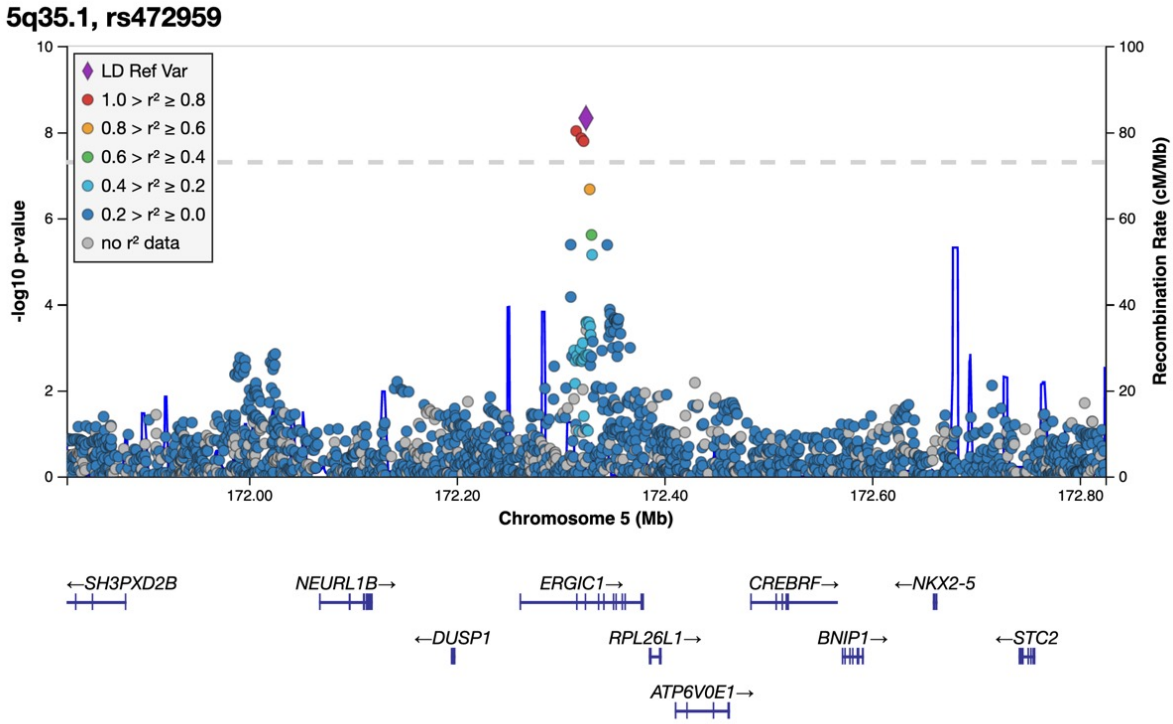

5q14.3, rs3930345
LD reference population: EUR
LD Ref Var
1.0 > r² ≥ 0.8
0.8 > r² ≥ 0.6
0.6 > r² ≥ 0.4
0.4 > r² ≥ 0.2
0.2 > r² ≥ 0.0
no r² data
-log10 p-value
Recombination Rate (cM/Mb)
Chromosome 5 (Mb)
82.40
82.60
82.80
83.00
83.20
XRCC4→
FTH1P9→
←HAPLN1
←ZP3P1
←EDIL3
AC094085.1→
VCAN→
RN7SKP295→
←AC109496.1
←COQ10BP2
←VCAN-AS1
←AC020899.1
RNU6-620P→
RNU4-11P→


5q14.3, rs3930345
LD reference population: EAS
LD Ref Var
1.0 > r² ≥ 0.8
0.8 > r² ≥ 0.6
0.6 > r² ≥ 0.4
0.4 > r² ≥ 0.2
0.2 > r² ≥ 0.0
no r² data
-log10 p-value
Recombination Rate (cM/Mb)
Chromosome 5 (Mb)
82.40
82.60
82.80
83.00
83.20
XRCC4→
FTH1P9→
←HAPLN1
←ZP3P1
←EDIL3
AC094085.1→
VCAN→
RN7SKP295→
←AC109496.1
←COQ10BP2
←VCAN-AS1
←AC020899.1
RNU6-620P→
RNU4-11P→

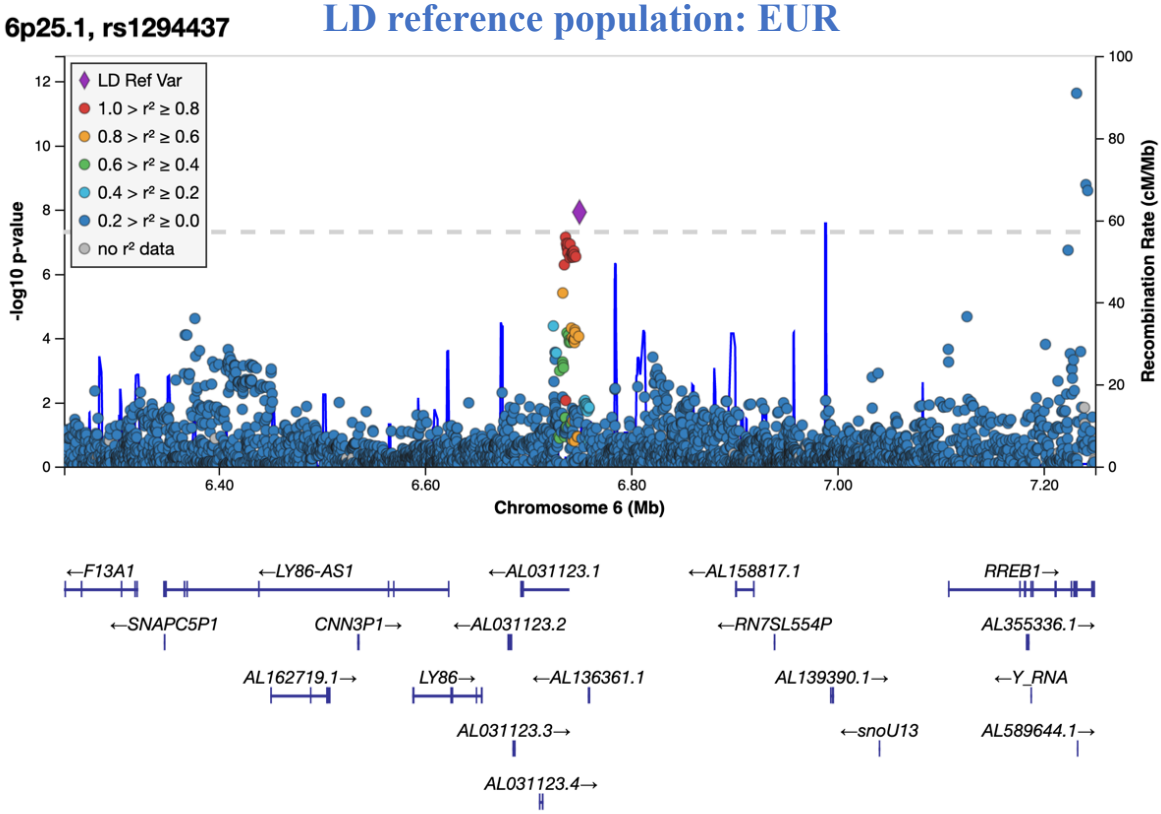
## LD reference population: EUR



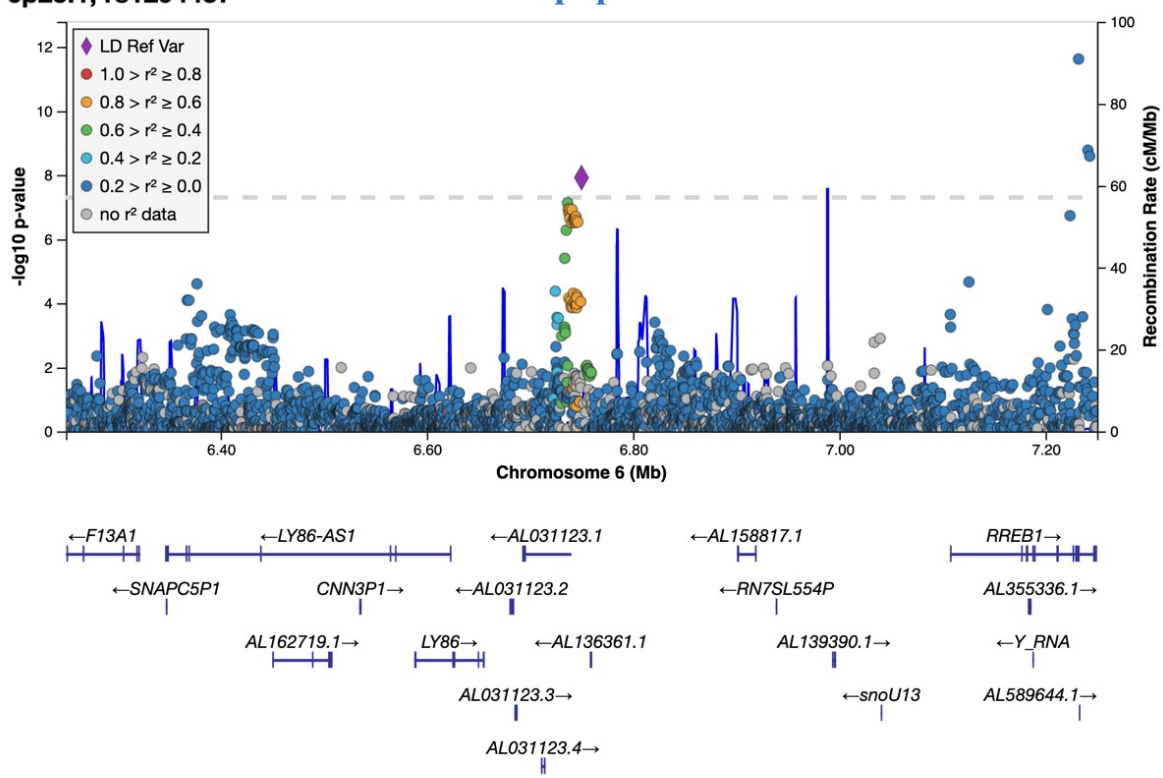

## LD reference population: EAS



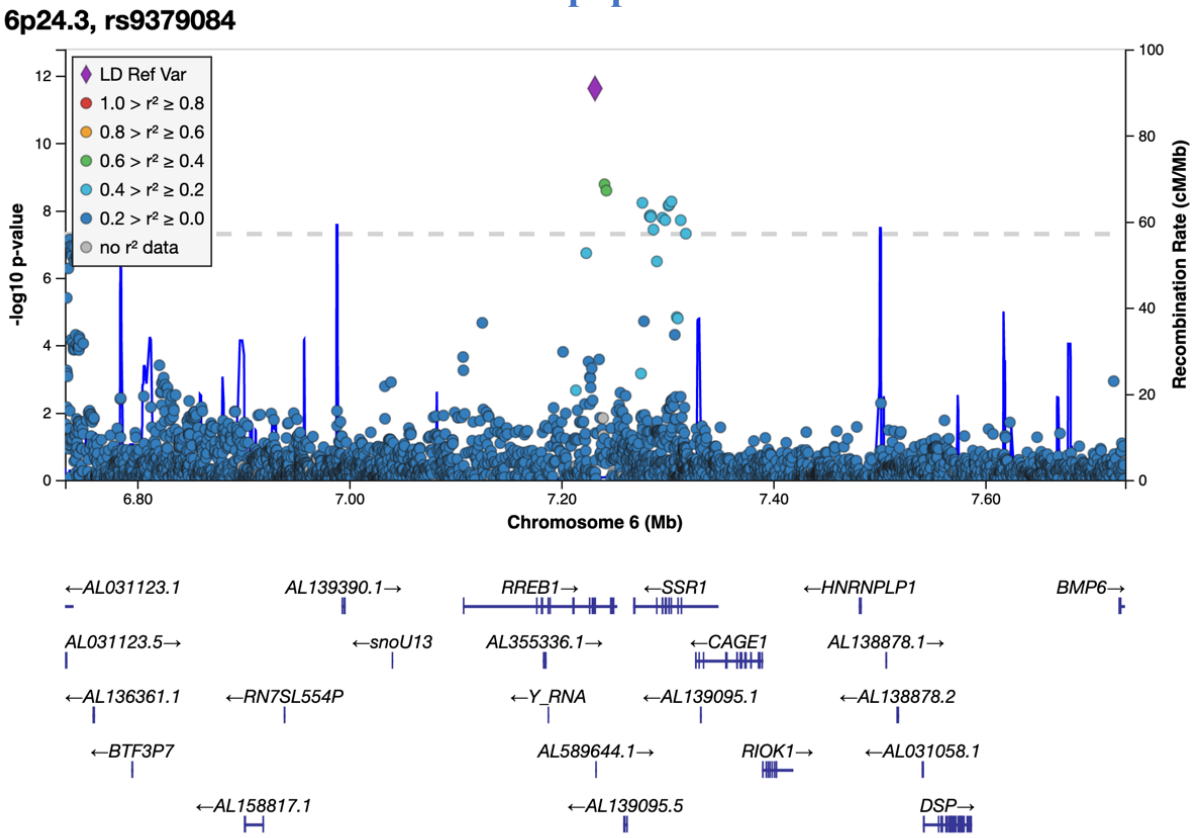
## LD reference population: EUR



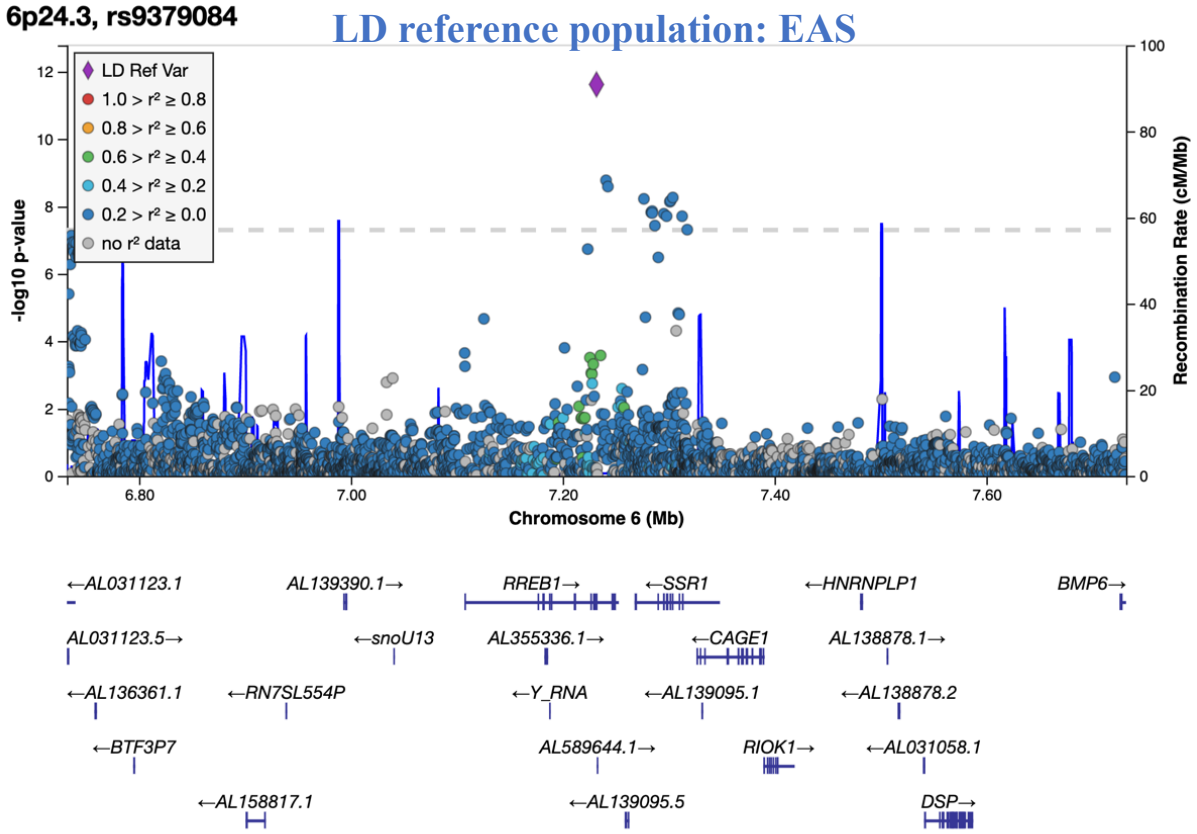

## LD reference population: EAS

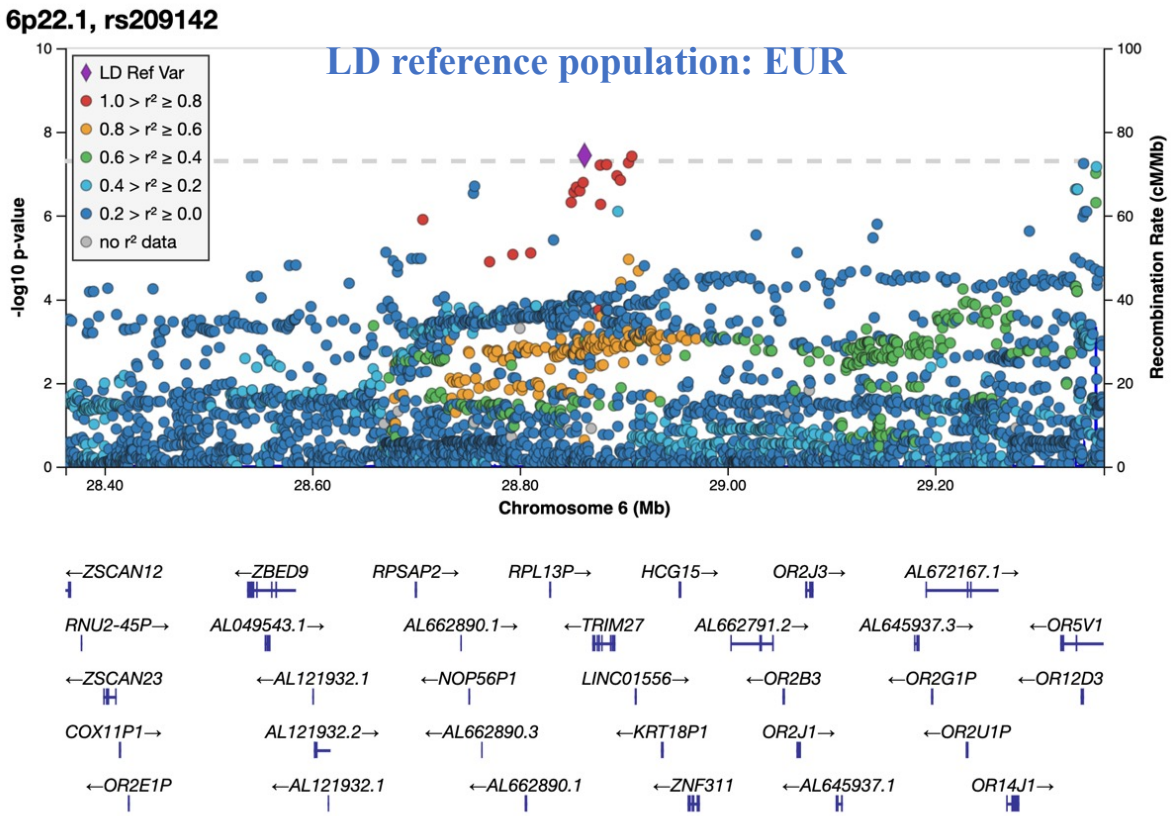


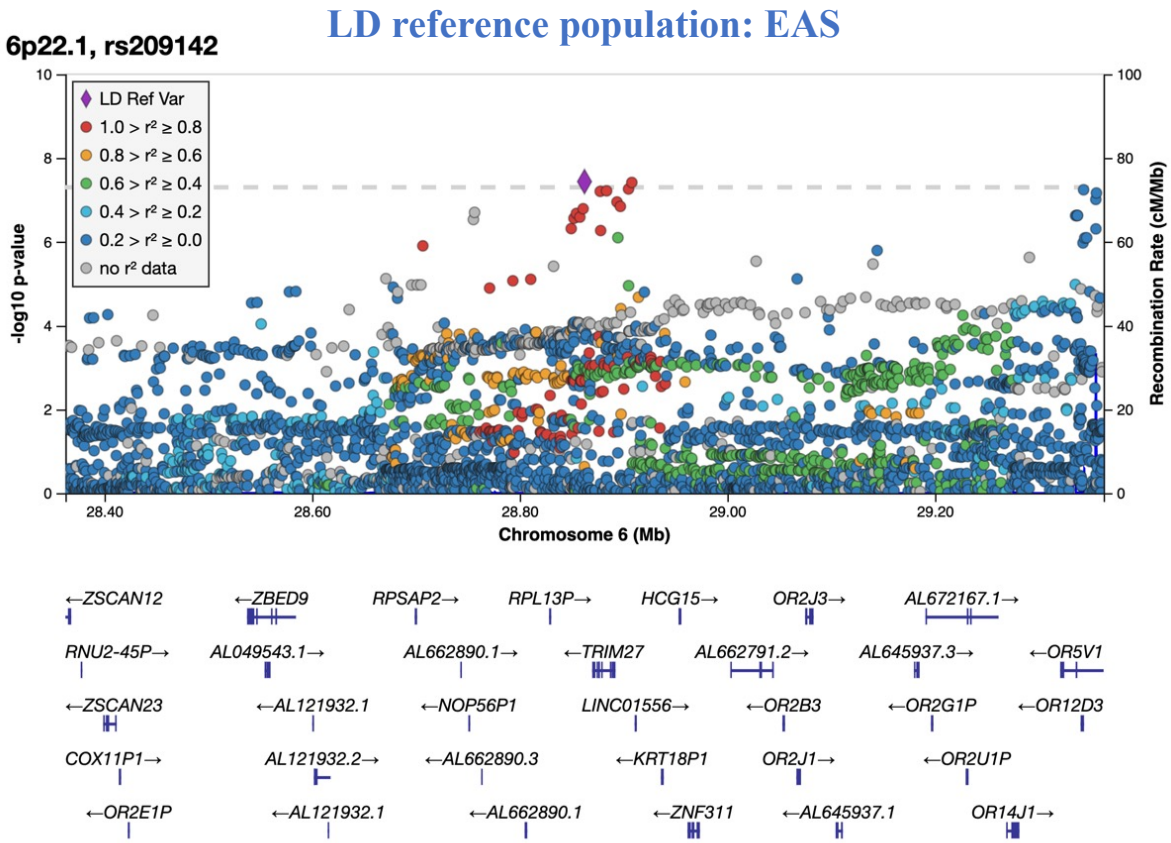
## LD reference population: EUR



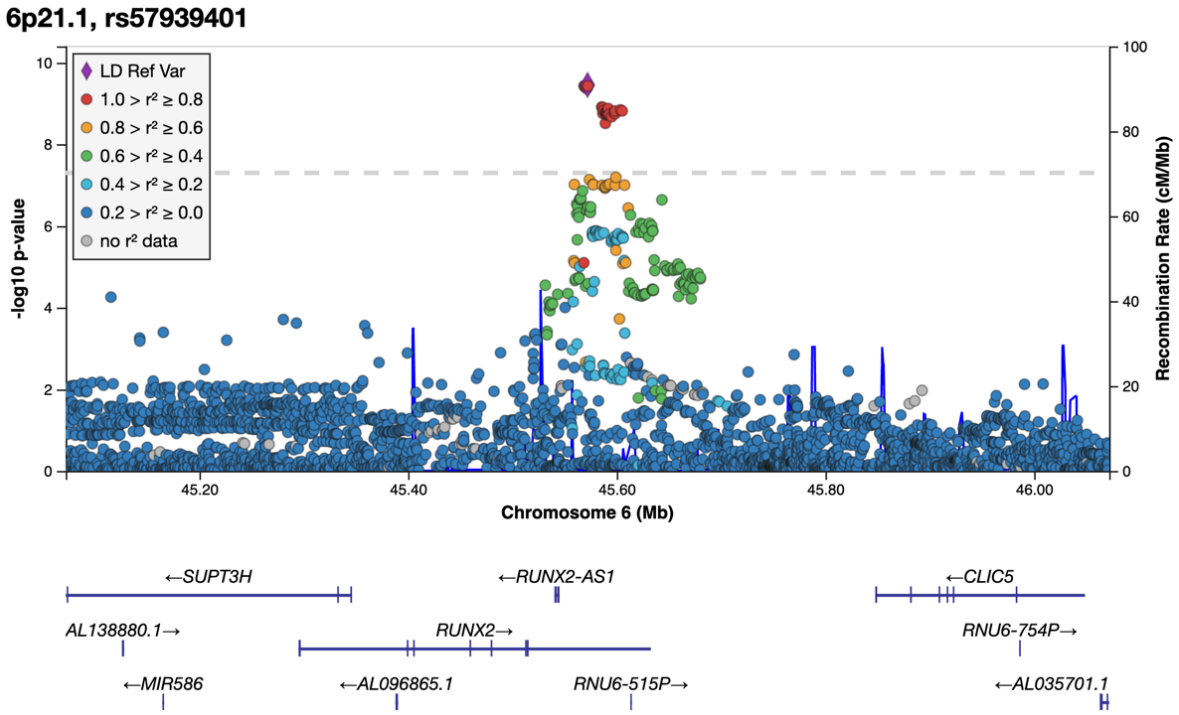

## LD reference population: EAS



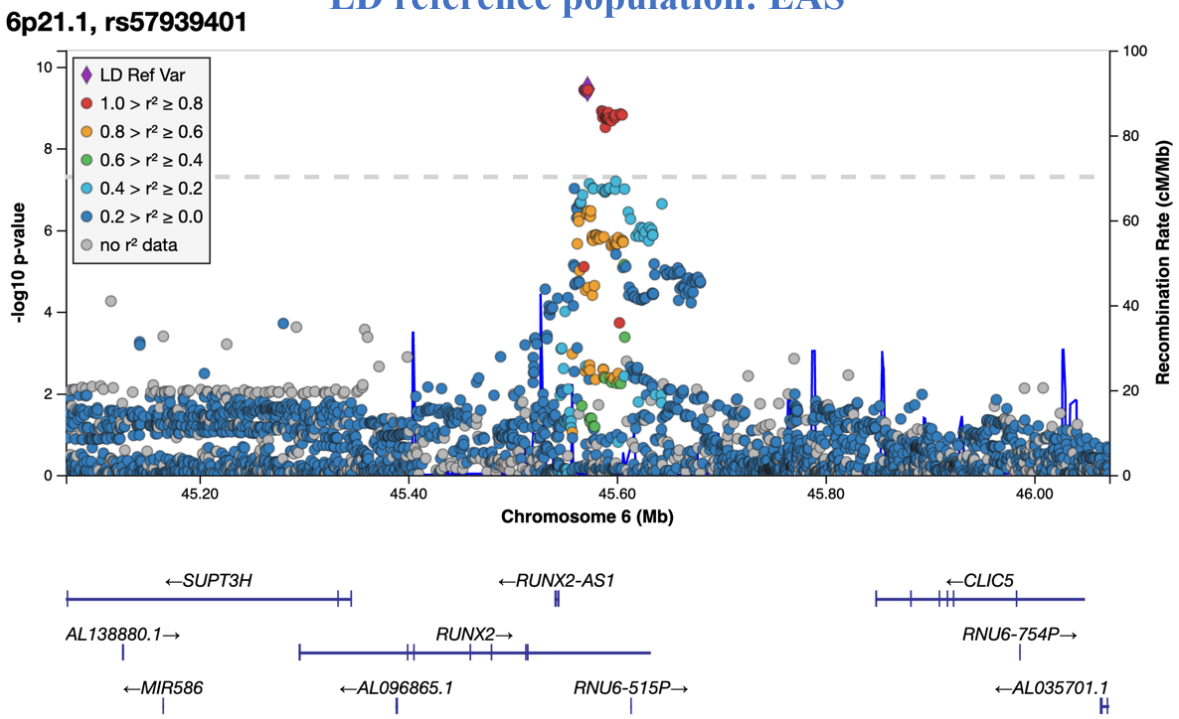

6p22.1, rs209142
LD reference population: EUR
LD Ref Var
1.0 > r² ≥ 0.8
0.8 > r² ≥ 0.6
0.6 > r² ≥ 0.4
0.4 > r² ≥ 0.2
0.2 > r² ≥ 0.0
no r² data
-log10 p-value
Recombination Rate (cM/Mb)
Chromosome 6 (Mb)
28.40
28.60
28.80
29.00
29.20
←ZSCAN12
←ZBED9
RPSAP2→
RPL13P→
HCG15→
OR2J3→
AL672167.1→
RNU2-45P→
AL049543.1→
AL662890.1→
←TRIM27
AL662791.2→
AL645937.3→
←OR5V1
←ZSCAN23
←AL121932.1
←NOP56P1
LINC01556→
←OR2B3
←OR2G1P
←OR12D3
COX11P1→
AL121932.2→
←AL662890.3
←KRT18P1
OR2J1→
←OR2U1P
←OR2E1P
←AL121932.1
←AL662890.1
←ZNF311
←AL645937.1
OR14J1→


LD reference population: EAS
6p22.1, rs209142
LD Ref Var
1.0 > r² ≥ 0.8
0.8 > r² ≥ 0.6
0.6 > r² ≥ 0.4
0.4 > r² ≥ 0.2
0.2 > r² ≥ 0.0
no r² data
-log10 p-value
Recombination Rate (cM/Mb)
Chromosome 6 (Mb)
28.40
28.60
28.80
29.00
29.20
←ZSCAN12
←ZBED9
RPSAP2→
RPL13P→
HCG15→
OR2J3→
AL672167.1→
RNU2-45P→
AL049543.1→
AL662890.1→
←TRIM27
AL662791.2→
AL645937.3→
←OR5V1
←ZSCAN23
←AL121932.1
←NOP56P1
LINC01556→
←OR2B3
←OR2G1P
←OR12D3
COX11P1→
AL121932.2→
←AL662890.3
←KRT18P1
OR2J1→
←OR2U1P
←OR2E1P
←AL121932.1
←AL662890.1
←ZNF311
←AL645937.1
OR14J1→

## LD reference population: EUR

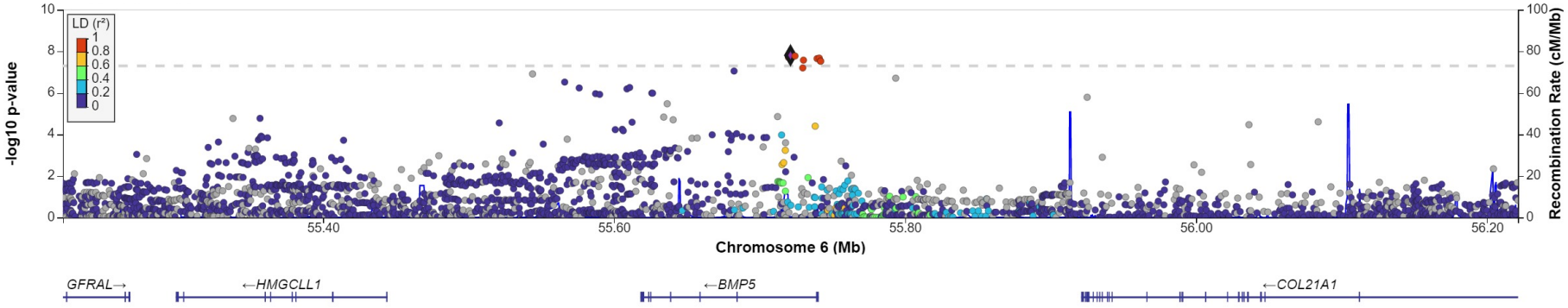



## LD reference population: EAS



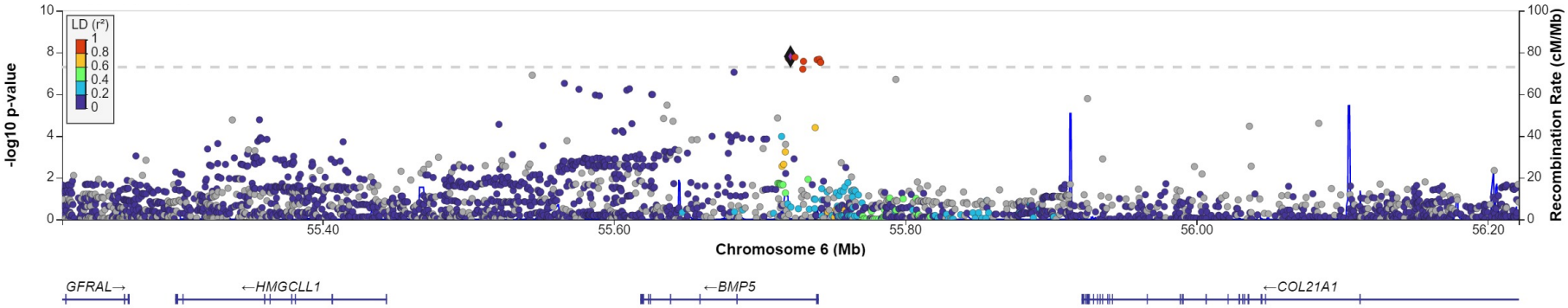
## LD reference population: EUR

**6p12.1, rs6912214**

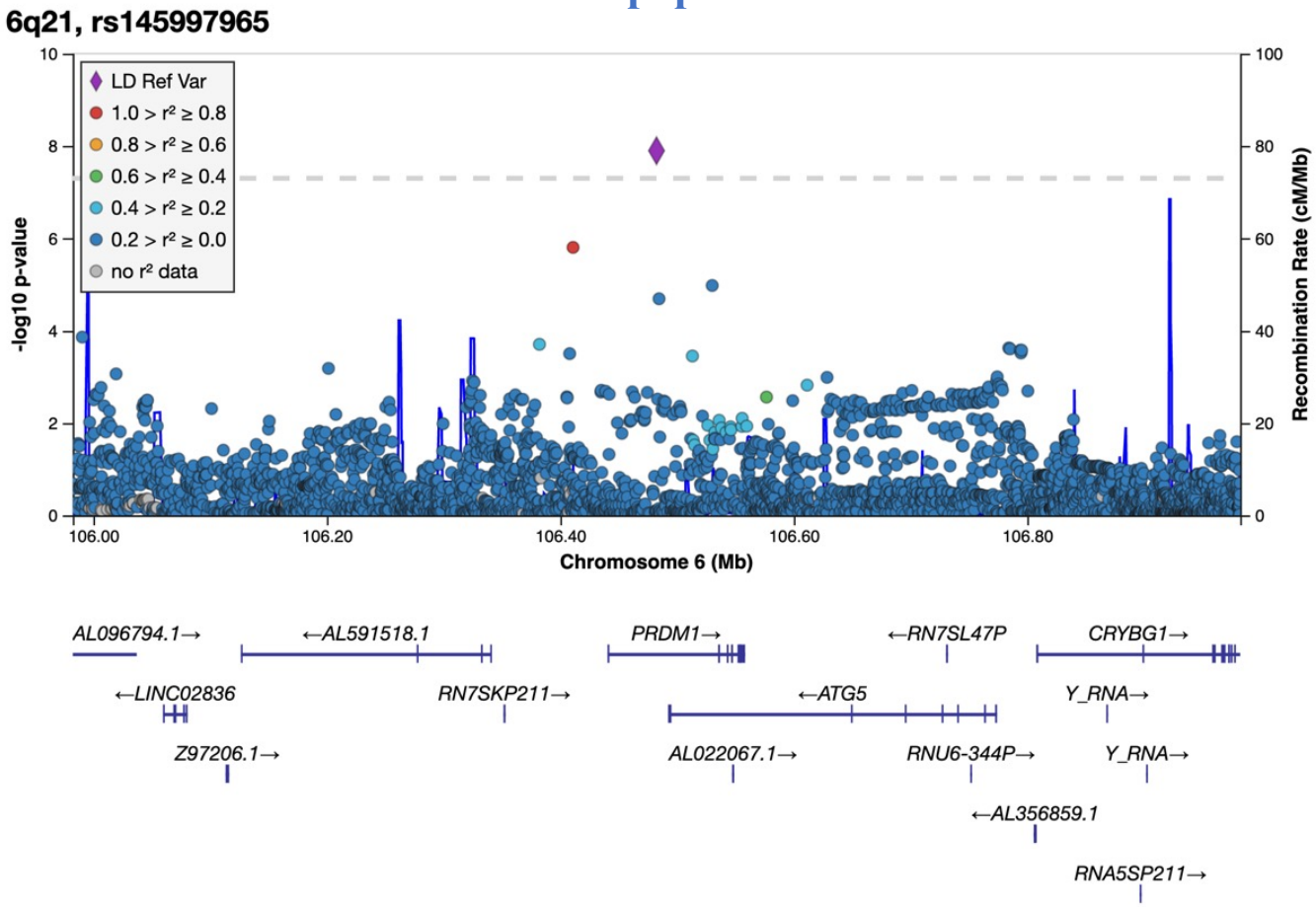



## LD reference population: EAS

**6p12.1, rs6912214**



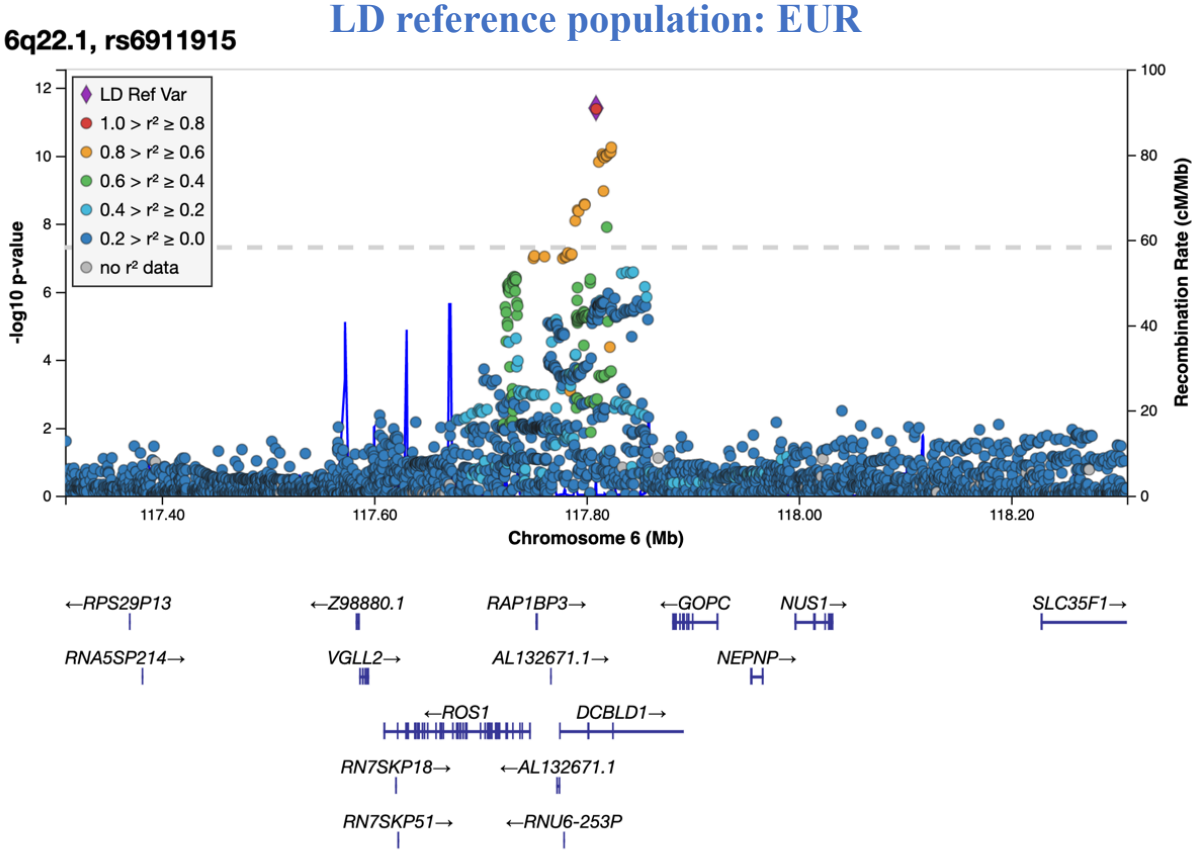
## LD reference population: EUR



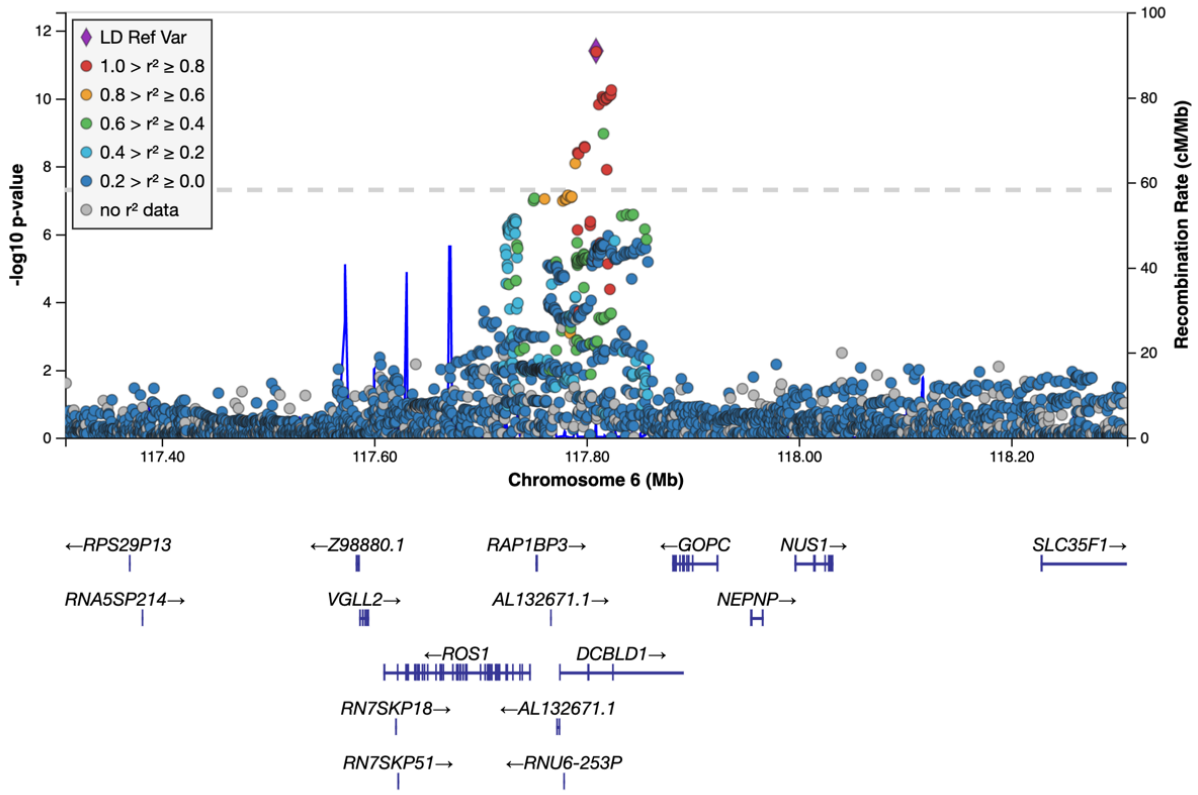

## LD reference population: EAS

rs145997965 is monoallelic in the EAS population

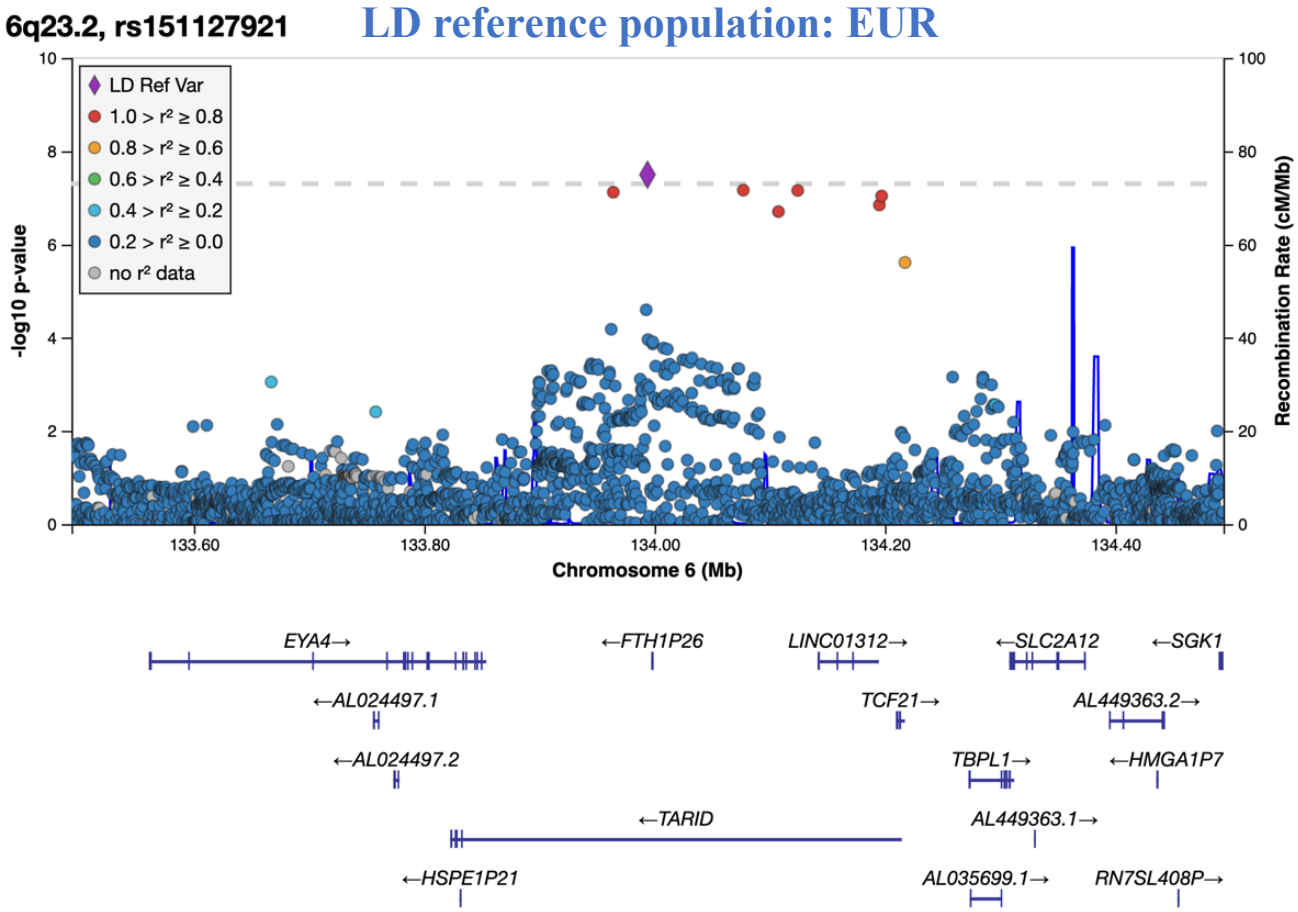
## LD reference population: EUR




## LD reference population: EAS



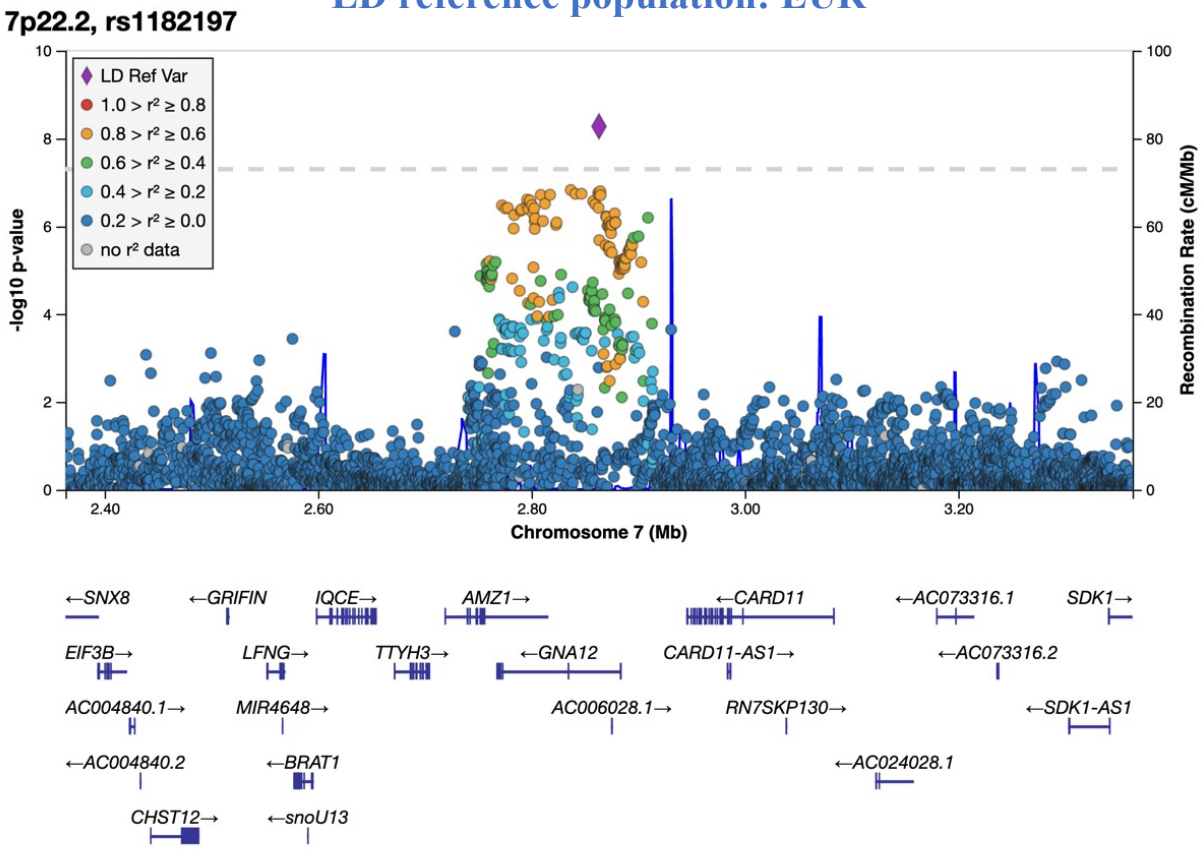
## LD reference population: EUR



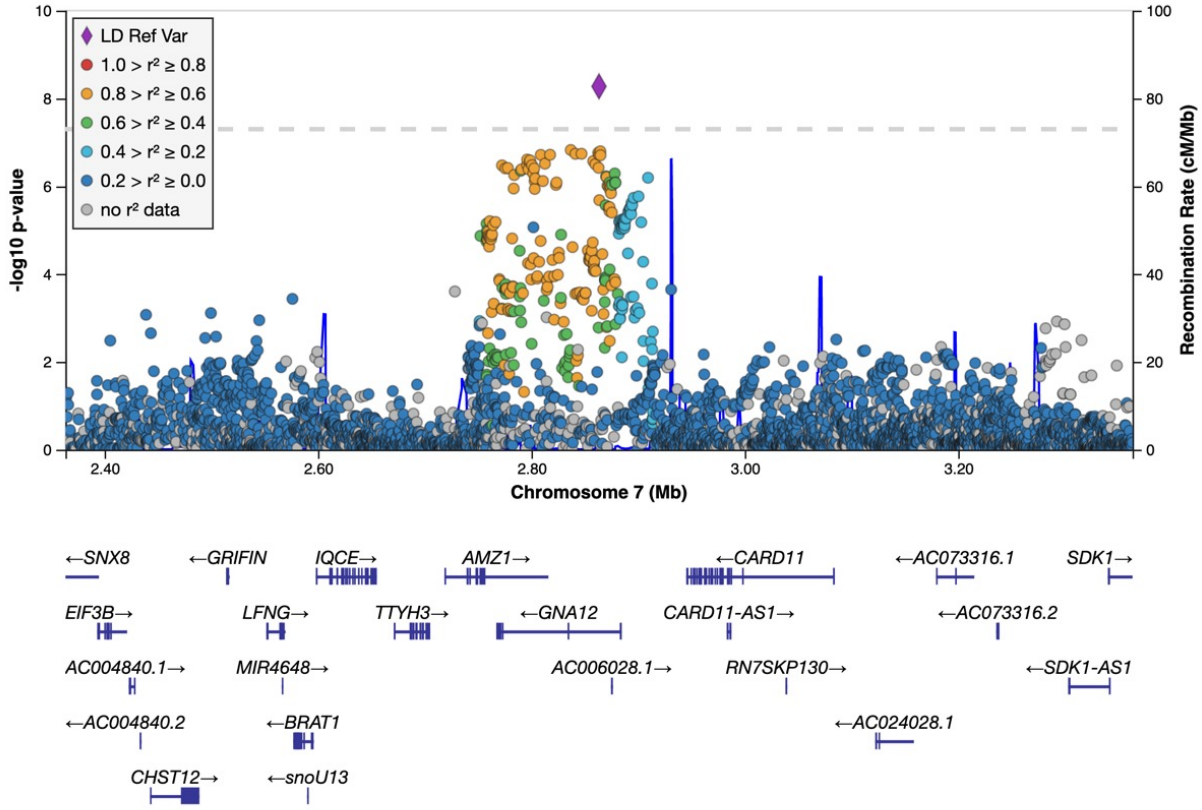

## LD reference population: EAS

rs151127921 is monoallelic in the EAS population

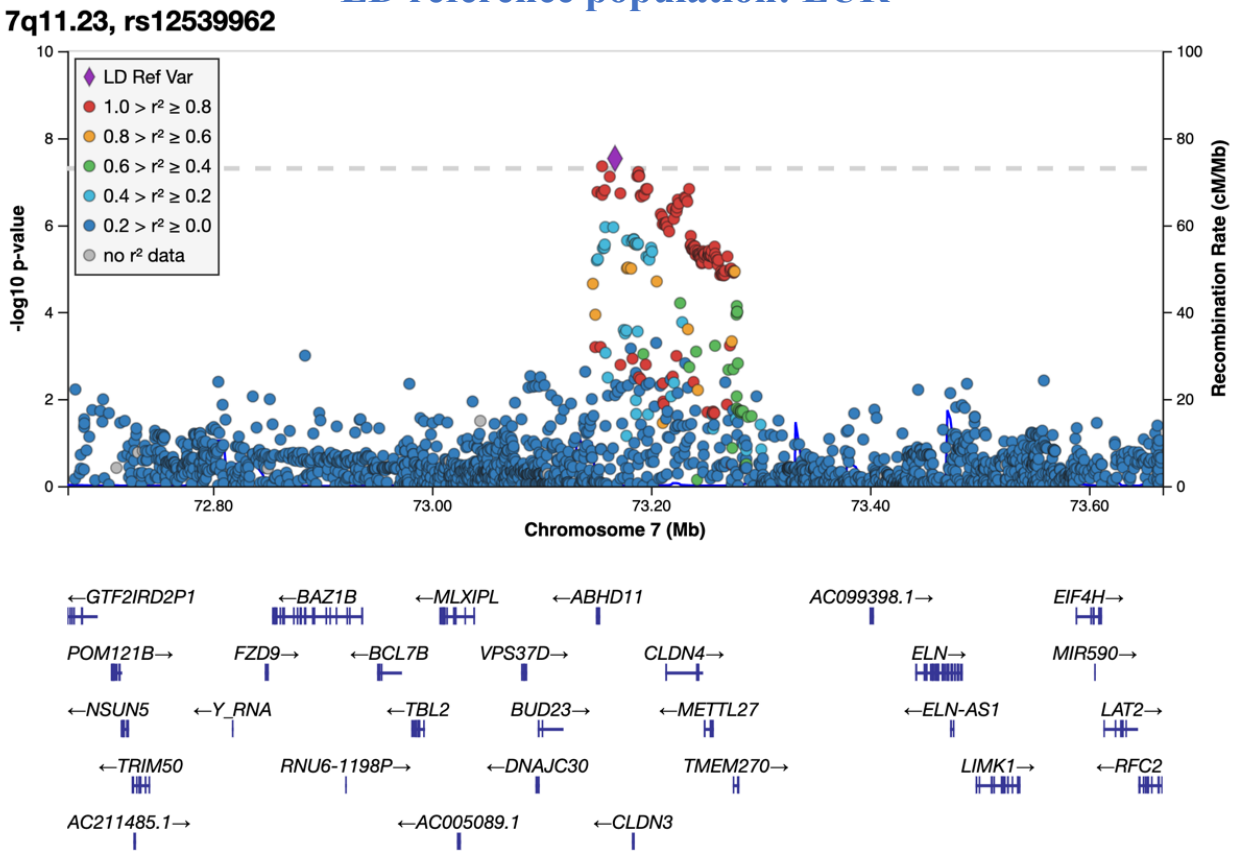
## LD reference population: EUR



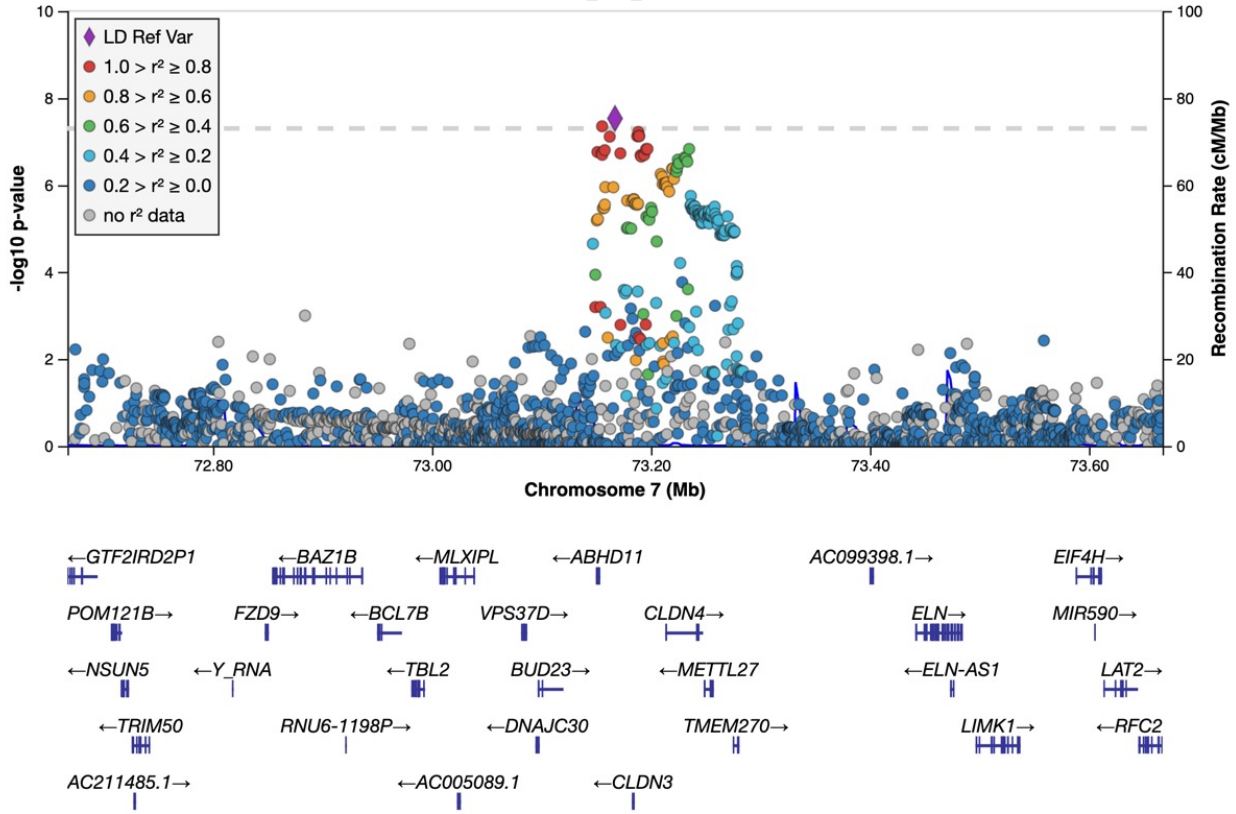

## LD reference population: EAS



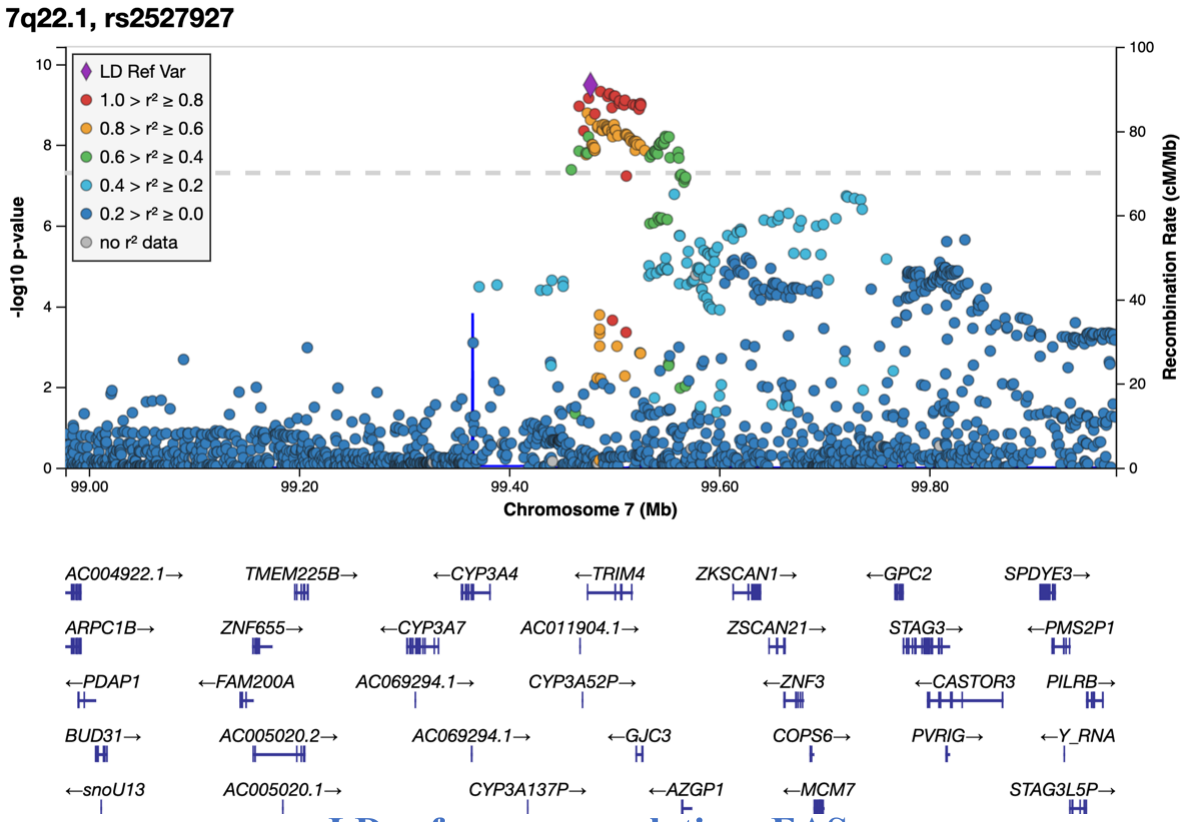
## LD reference population: EUR



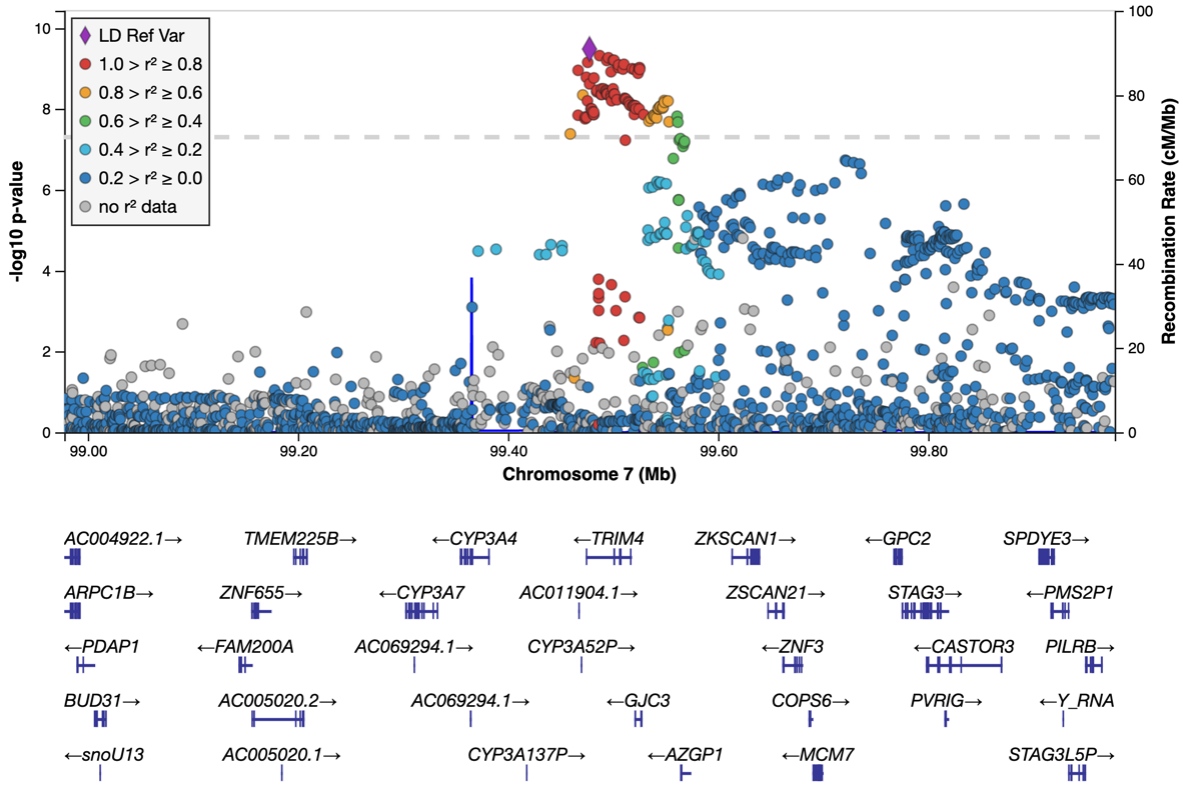

## LD reference population: EAS



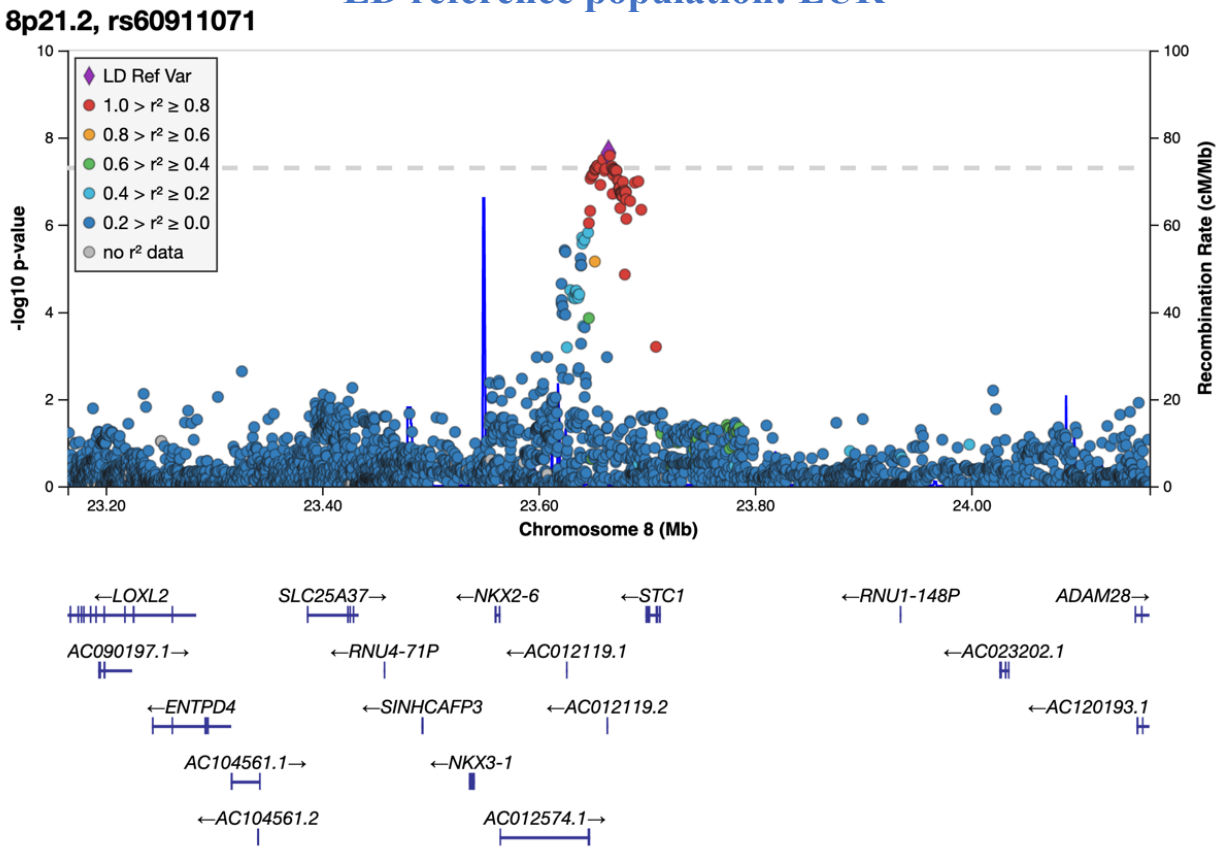
## LD reference population: EUR



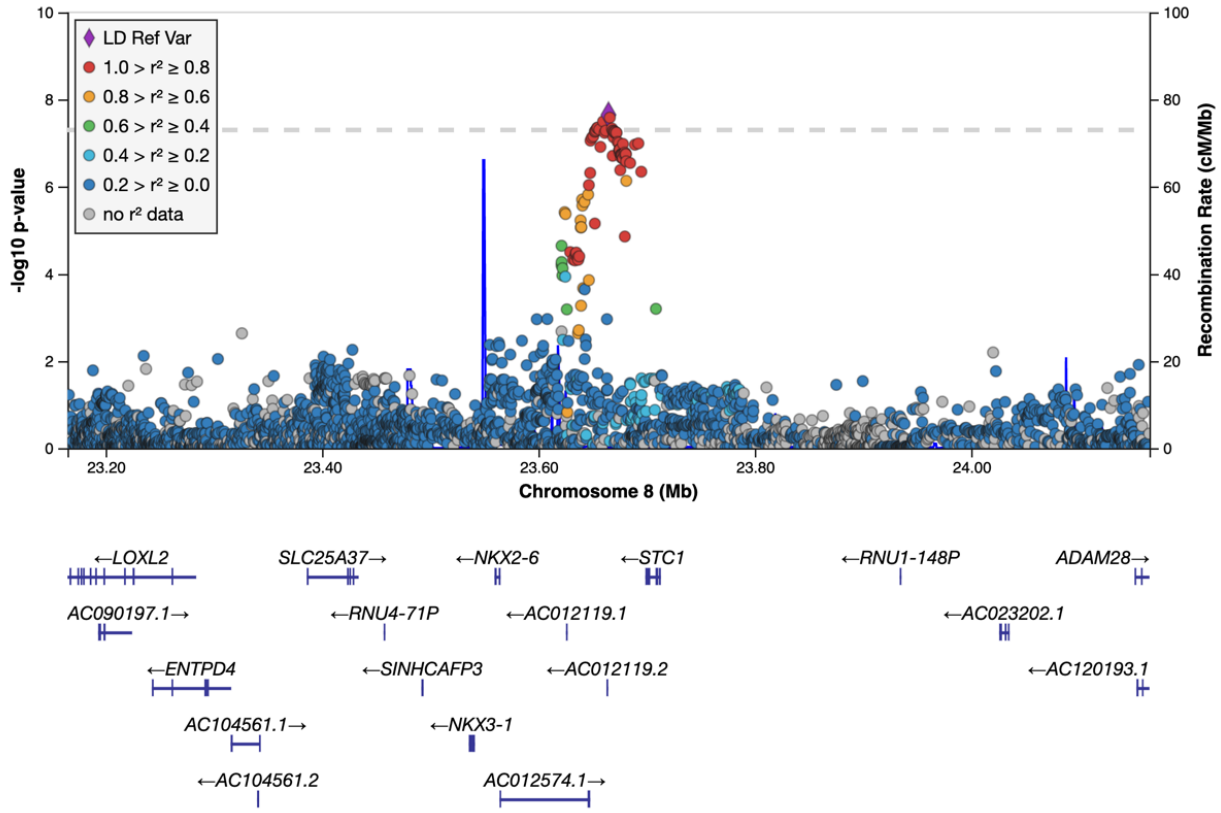

## LD reference population: EAS



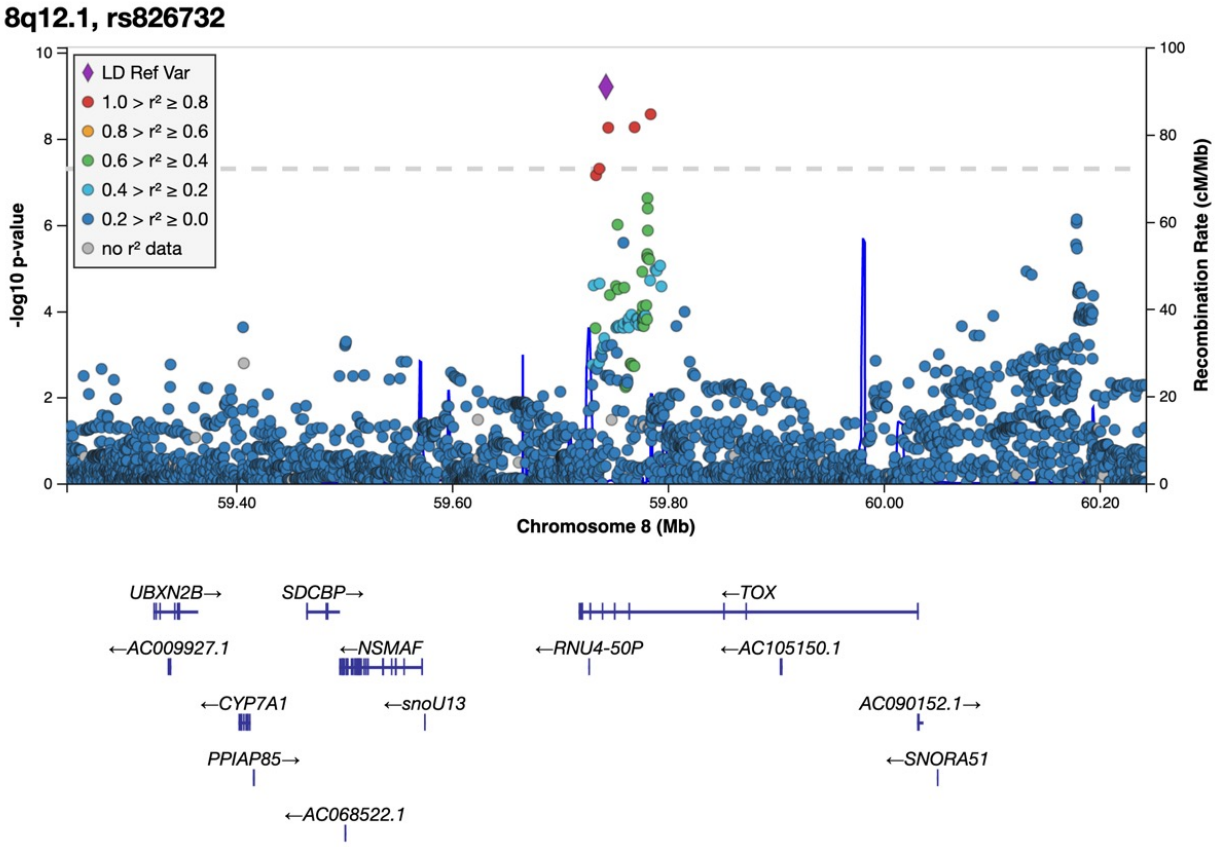
## LD reference population: EUR



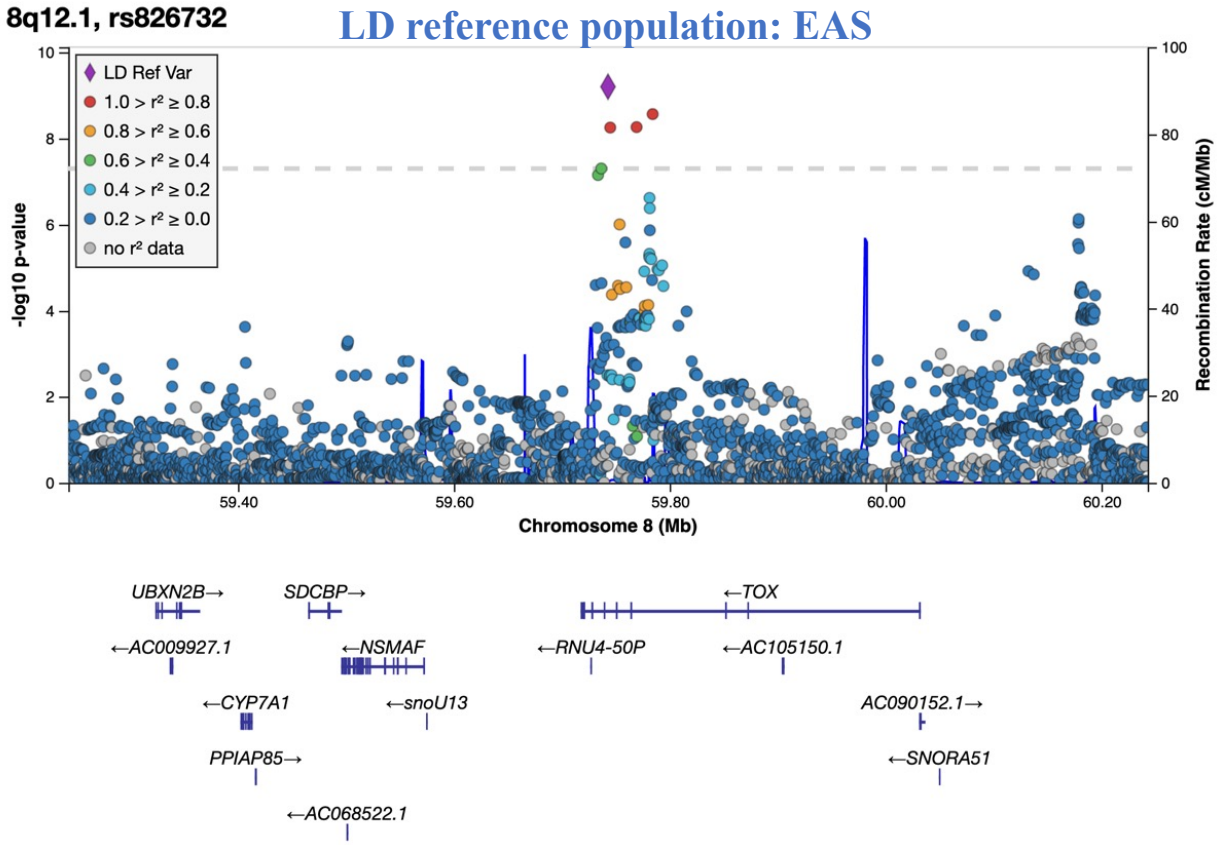

## LD reference population: EAS



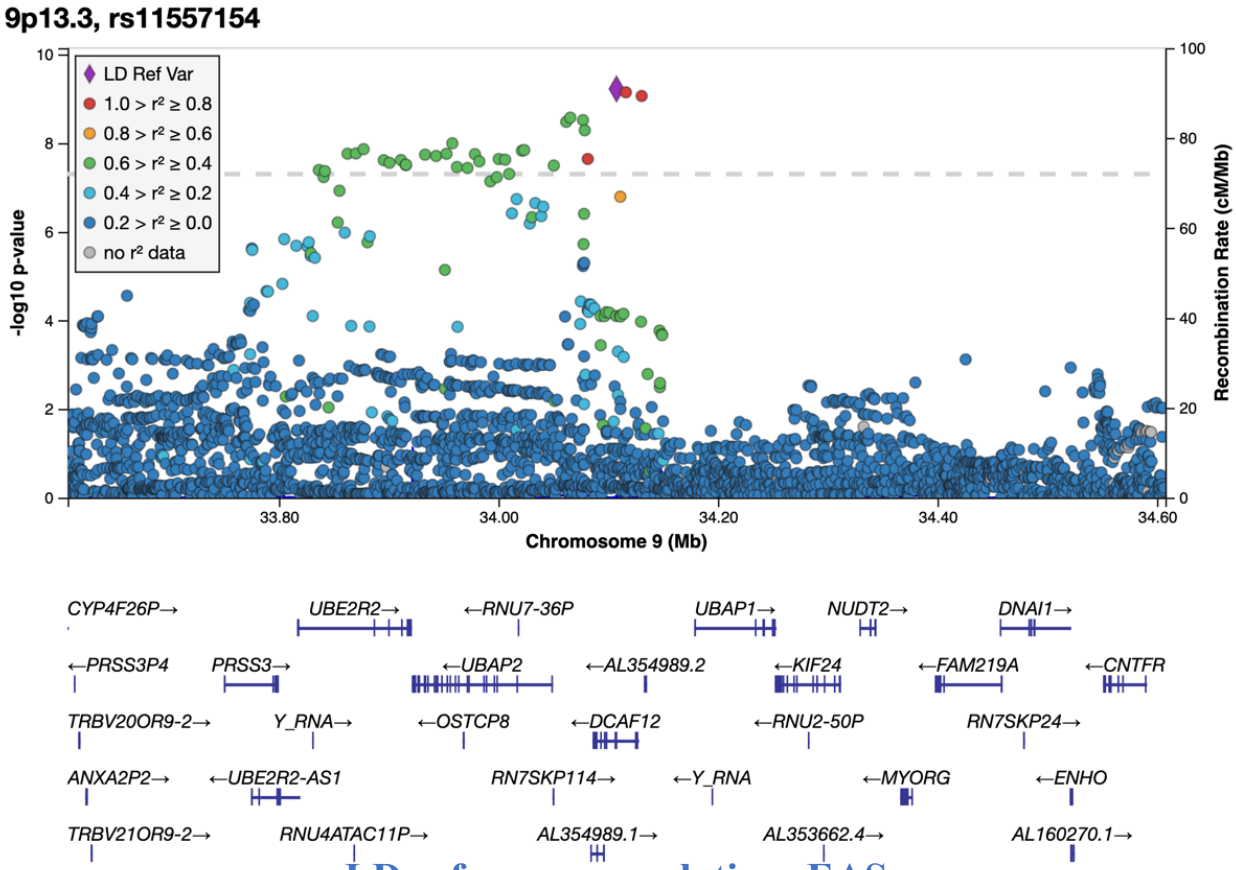
## LD reference population: EUR



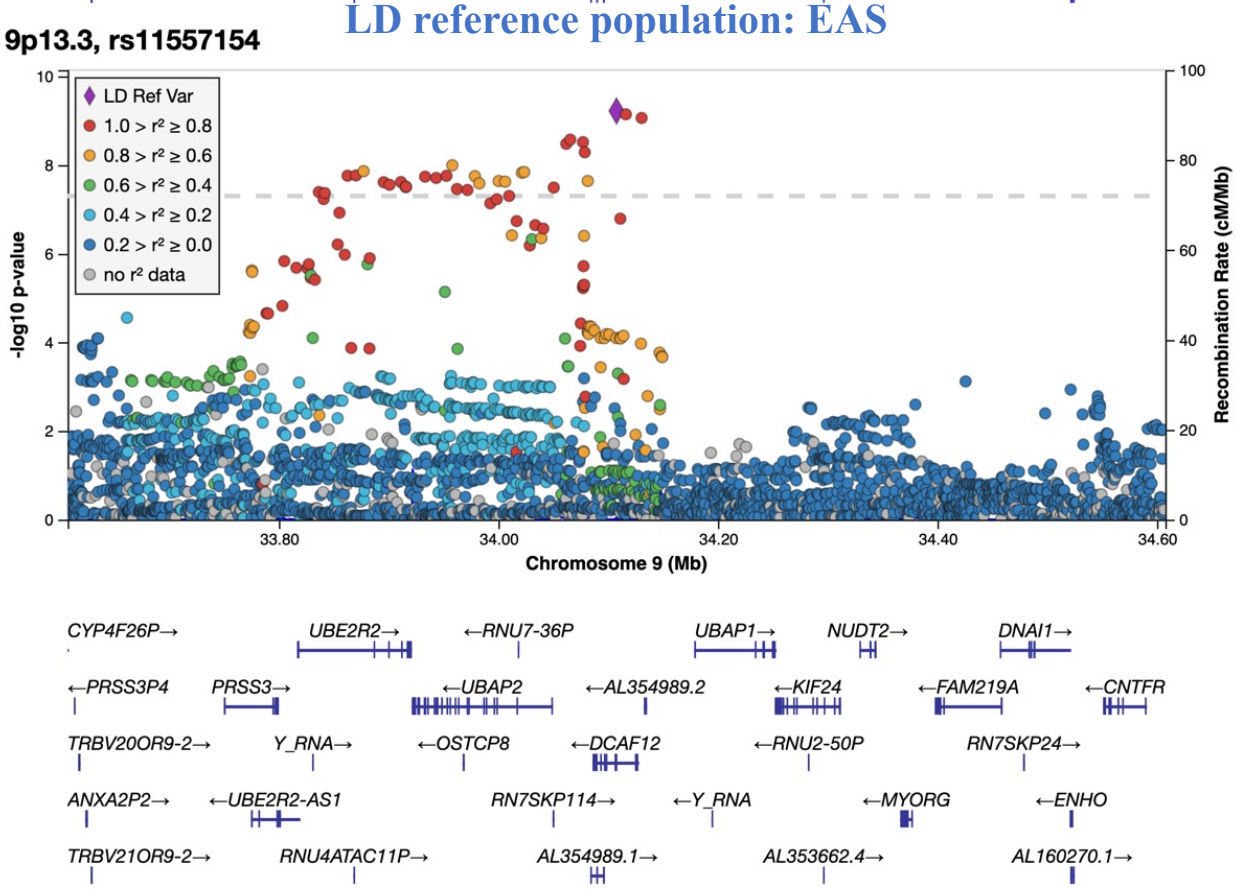

## LD reference population: EAS



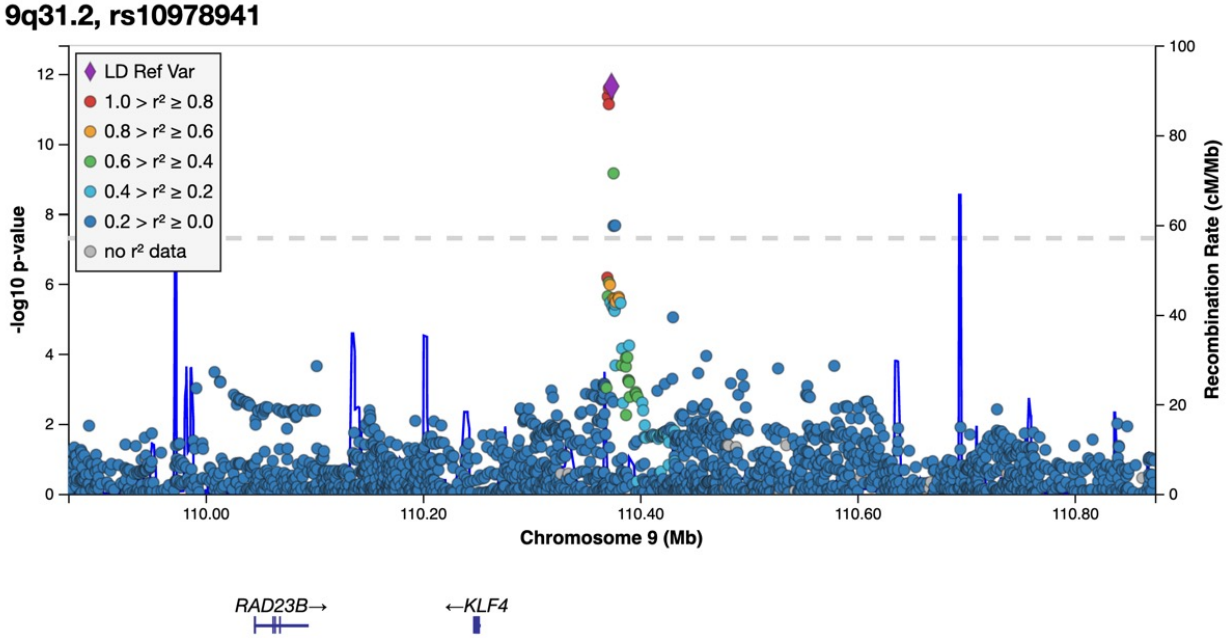
## LD reference population: EUR



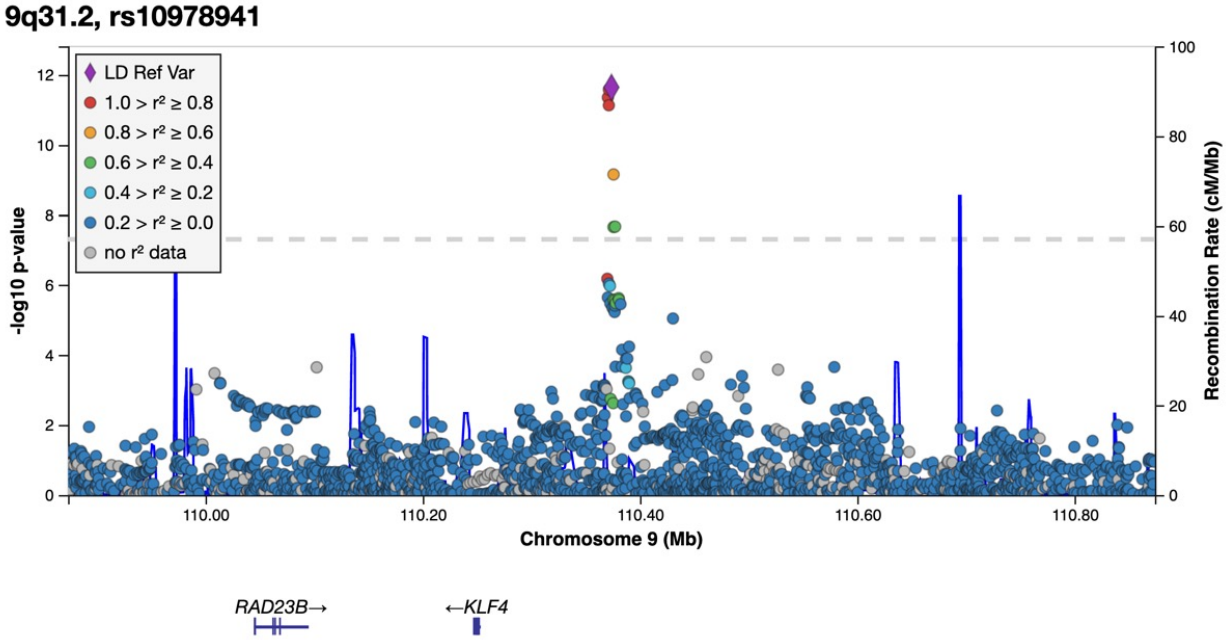

## LD reference population: EAS



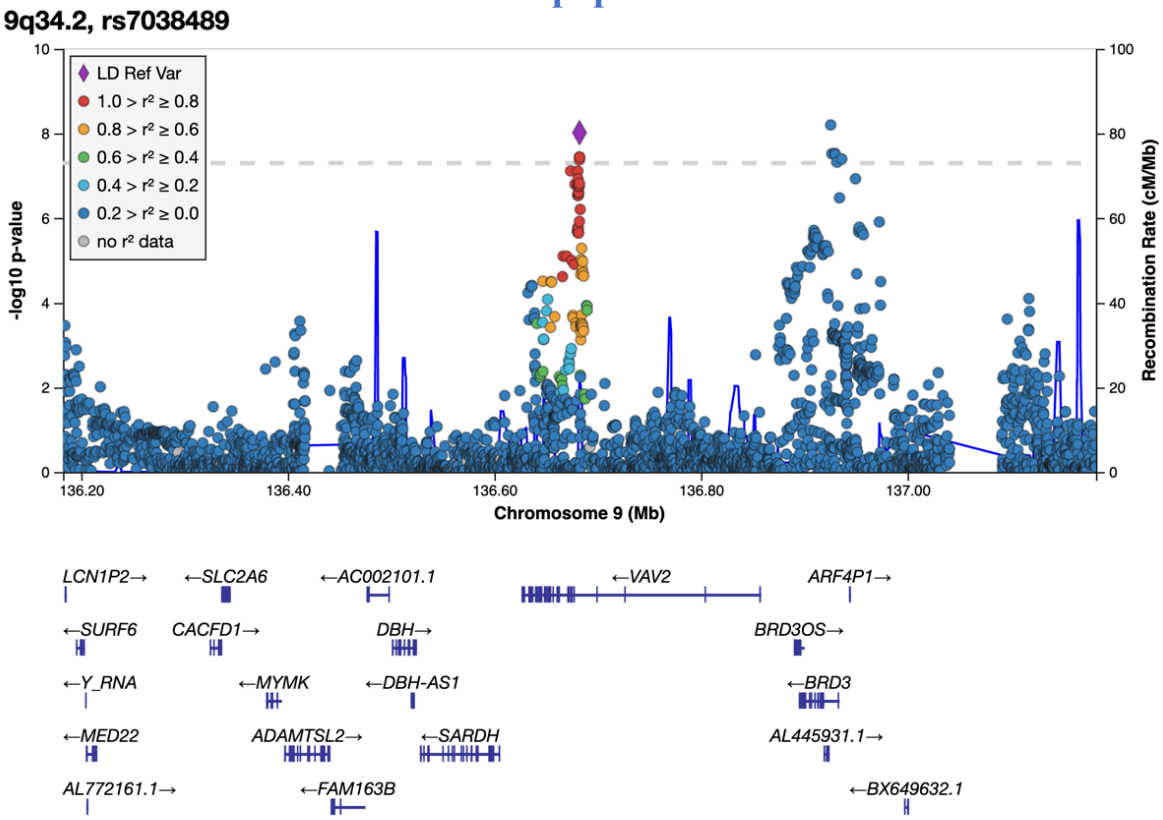
## LD reference population: EUR



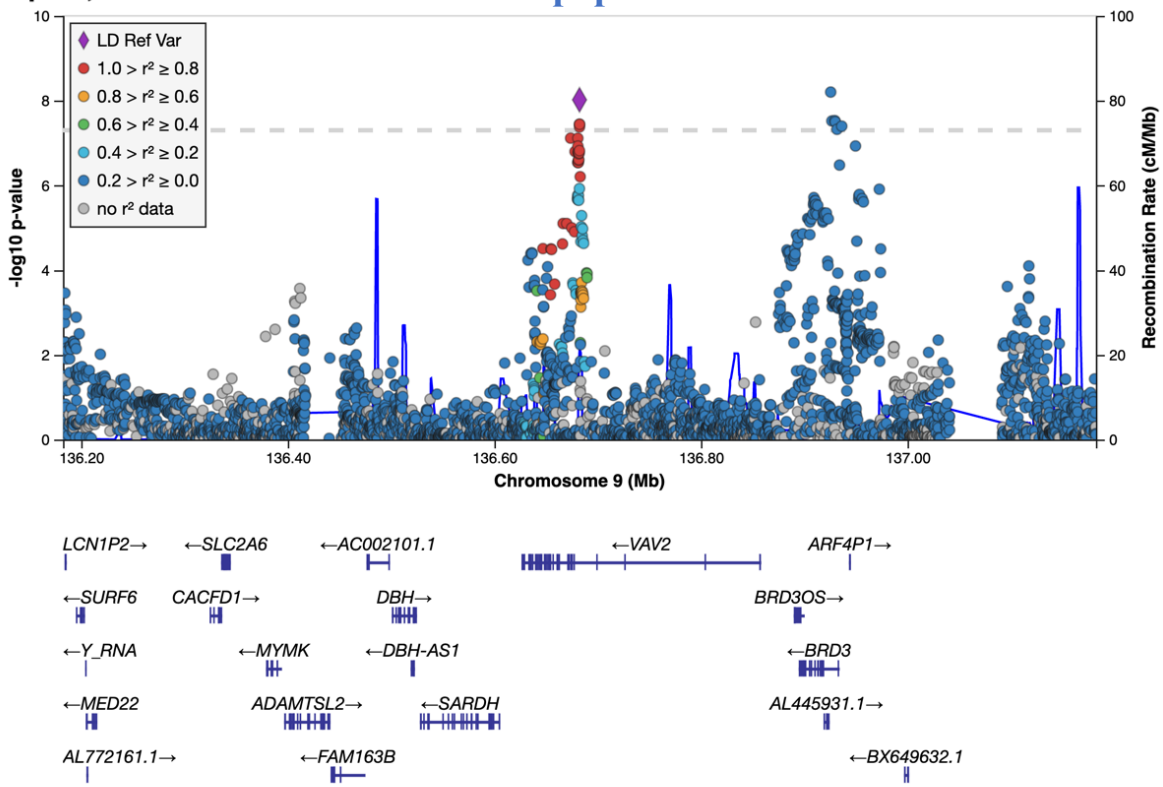

## LD reference population: EAS

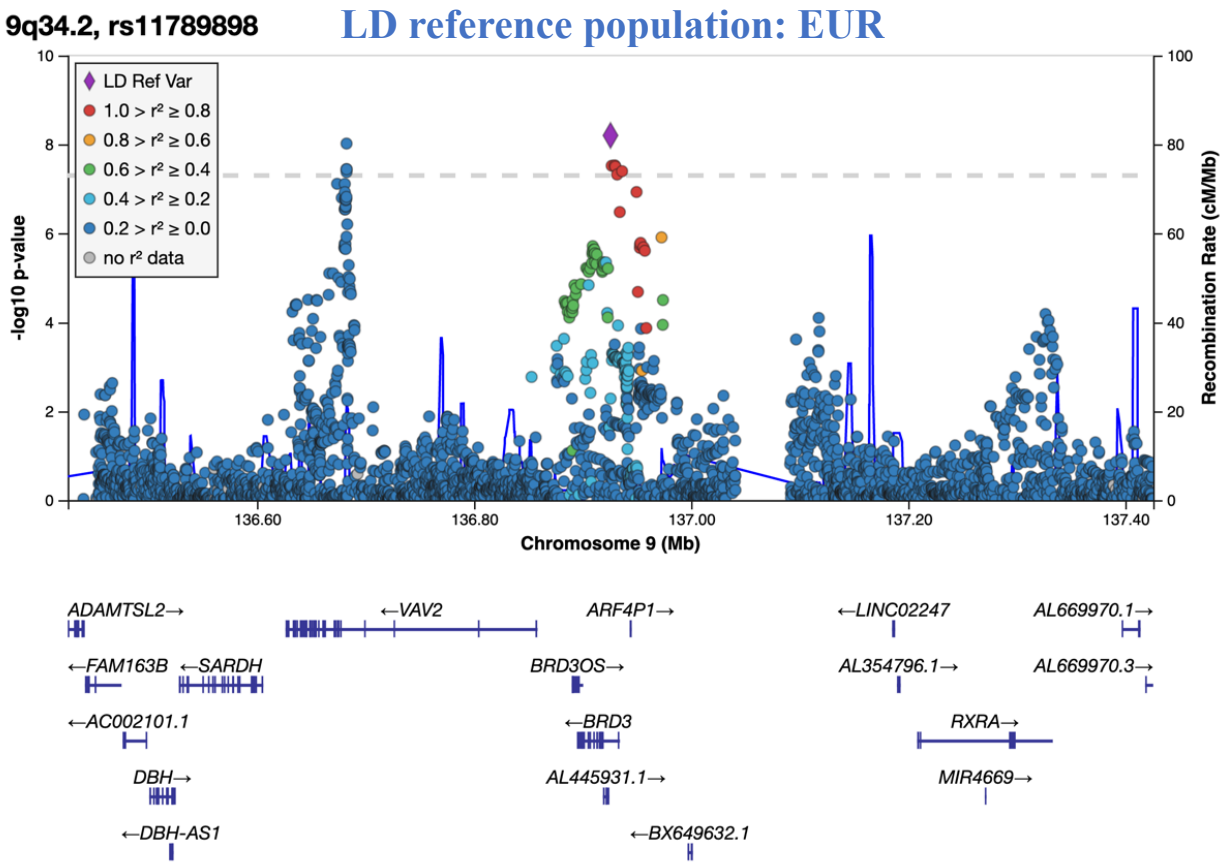


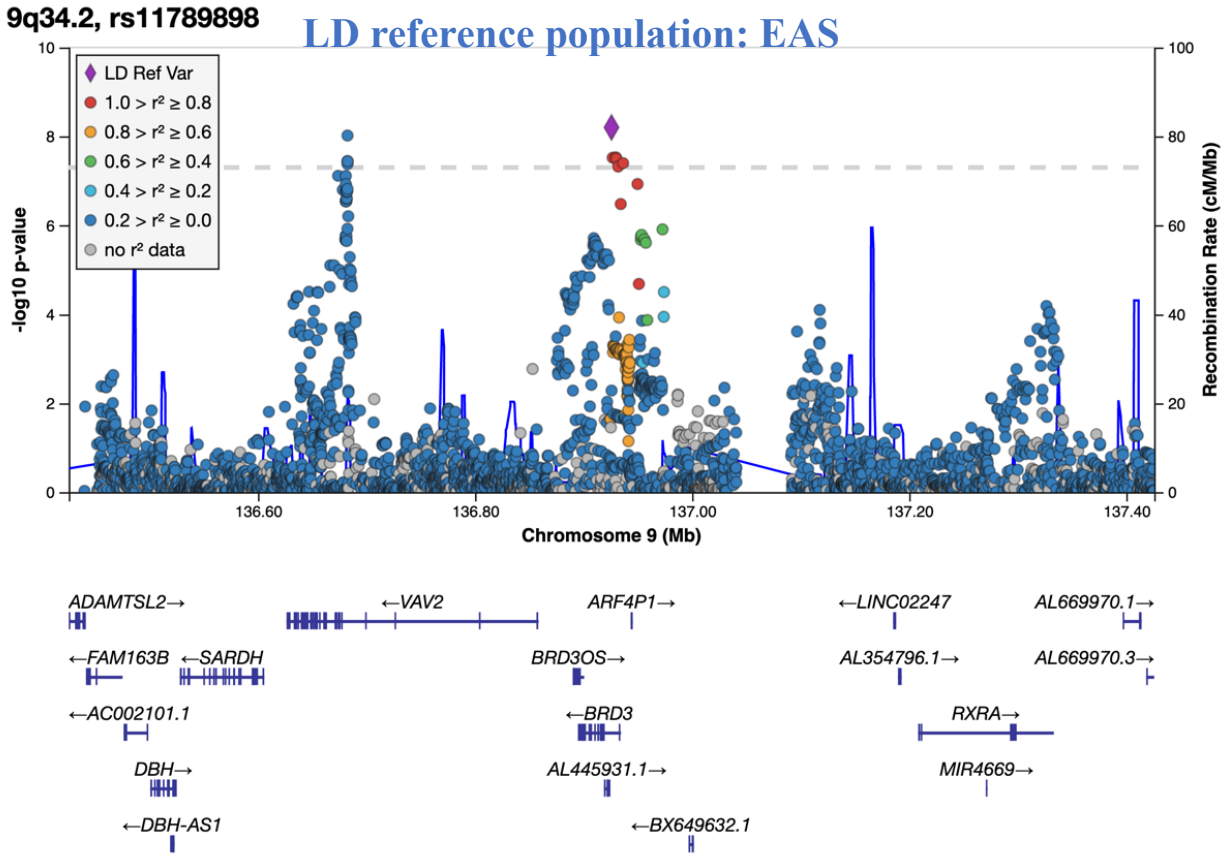
## LD reference population: EUR




## LD reference population: EAS

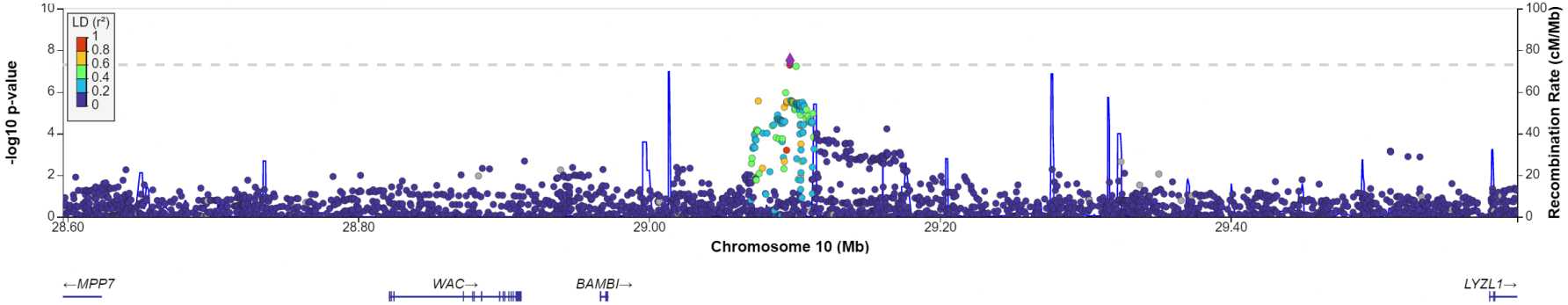



9q34.2, rs11789898
LD reference population: EUR
LD Ref Var
1.0 > r² ≥ 0.8
0.8 > r² ≥ 0.6
0.6 > r² ≥ 0.4
0.4 > r² ≥ 0.2
0.2 > r² ≥ 0.0
no r² data
-log10 p-value
Recombination Rate (cM/Mb)
Chromosome 9 (Mb)
136.60
136.80
137.00
137.20
137.40
ADAMTSL2→
←VAV2
ARF4P1→
←LINC02247
AL669970.1→
←FAM163B
←SARDH
BRD3OS→
AL354796.1→
AL669970.3→
←AC002101.1
←BRD3
RXRA→
DBH→
AL445931.1→
MIR4669→
←DBH-AS1
←BX649632.1


9q34.2, rs11789898
LD reference population: EAS
LD Ref Var
1.0 > r² ≥ 0.8
0.8 > r² ≥ 0.6
0.6 > r² ≥ 0.4
0.4 > r² ≥ 0.2
0.2 > r² ≥ 0.0
no r² data
-log10 p-value
Recombination Rate (cM/Mb)
Chromosome 9 (Mb)
136.60
136.80
137.00
137.20
137.40
ADAMTSL2→
←VAV2
ARF4P1→
←LINC02247
AL669970.1→
←FAM163B
←SARDH
BRD3OS→
AL354796.1→
AL669970.3→
←AC002101.1
←BRD3
RXRA→
DBH→
AL445931.1→
MIR4669→
←DBH-AS1
←BX649632.1

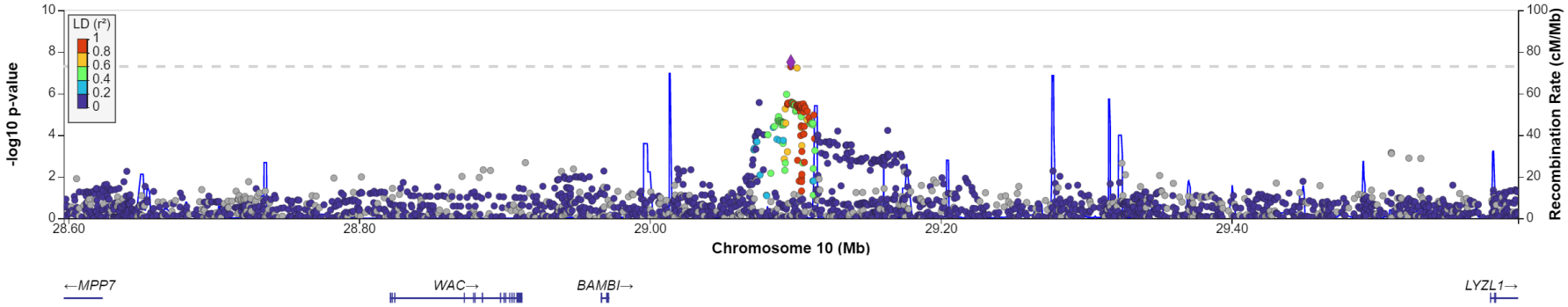
## LD reference population: EUR

**10p12.1, rs1775910**

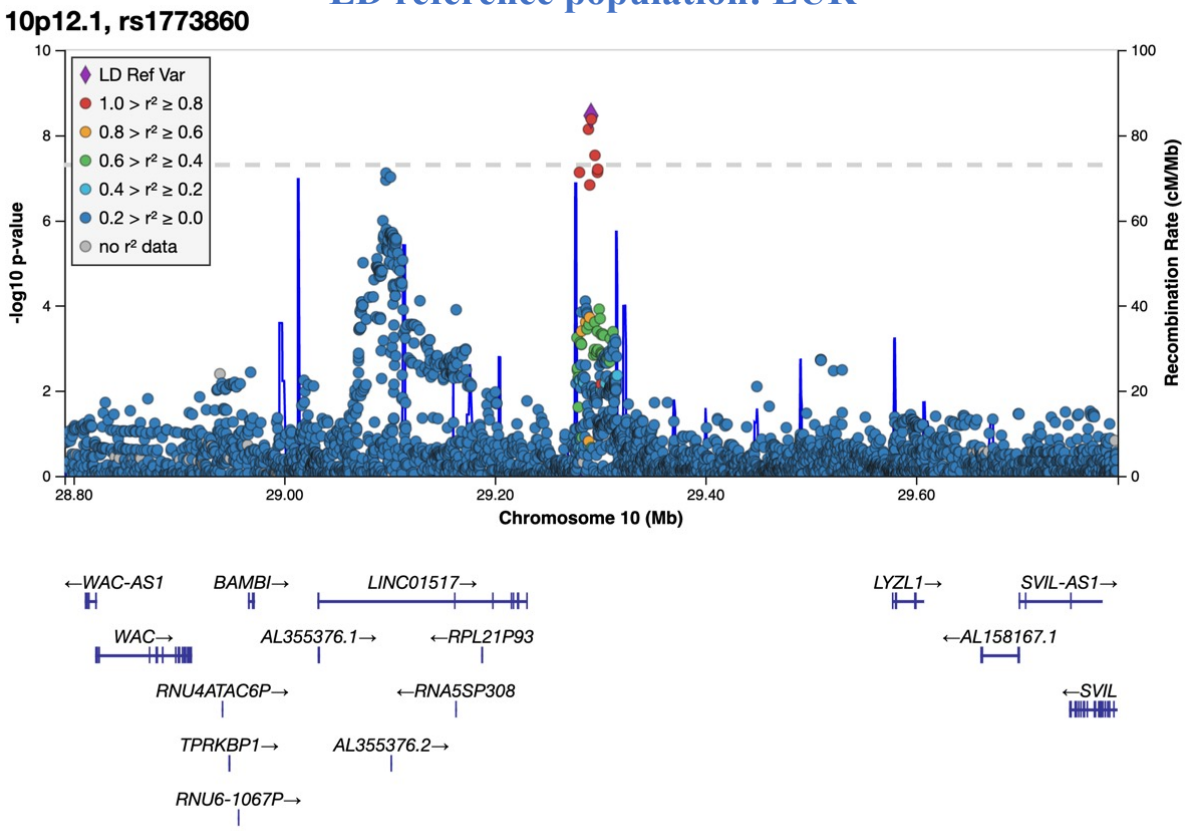



## LD reference population: EAS

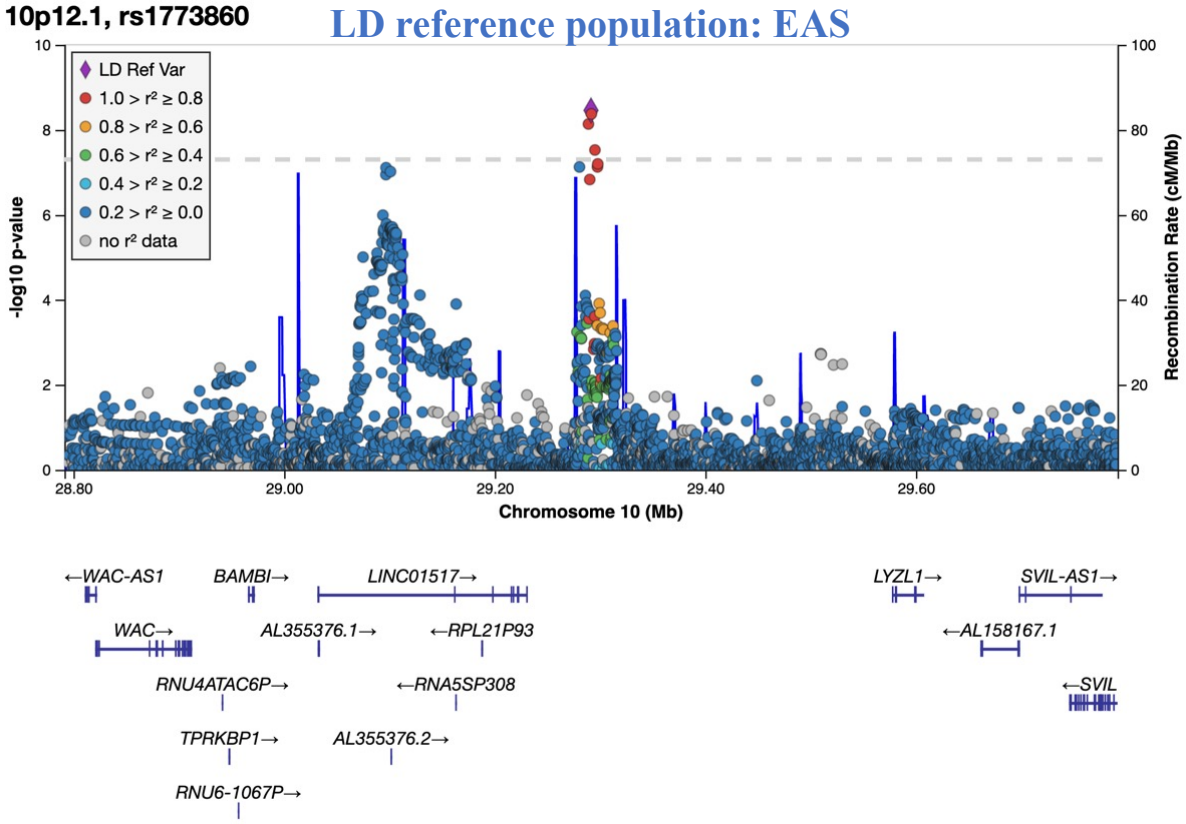
**10p12.1, rs1775910**



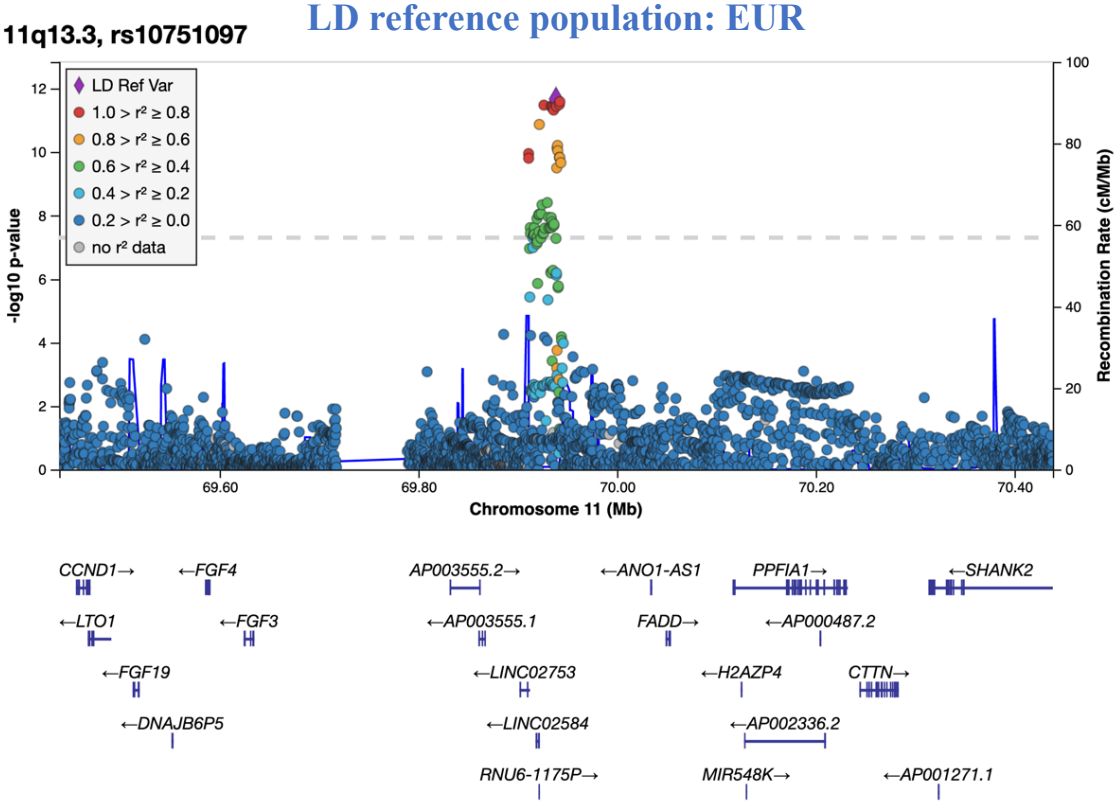
## LD reference population: EUR



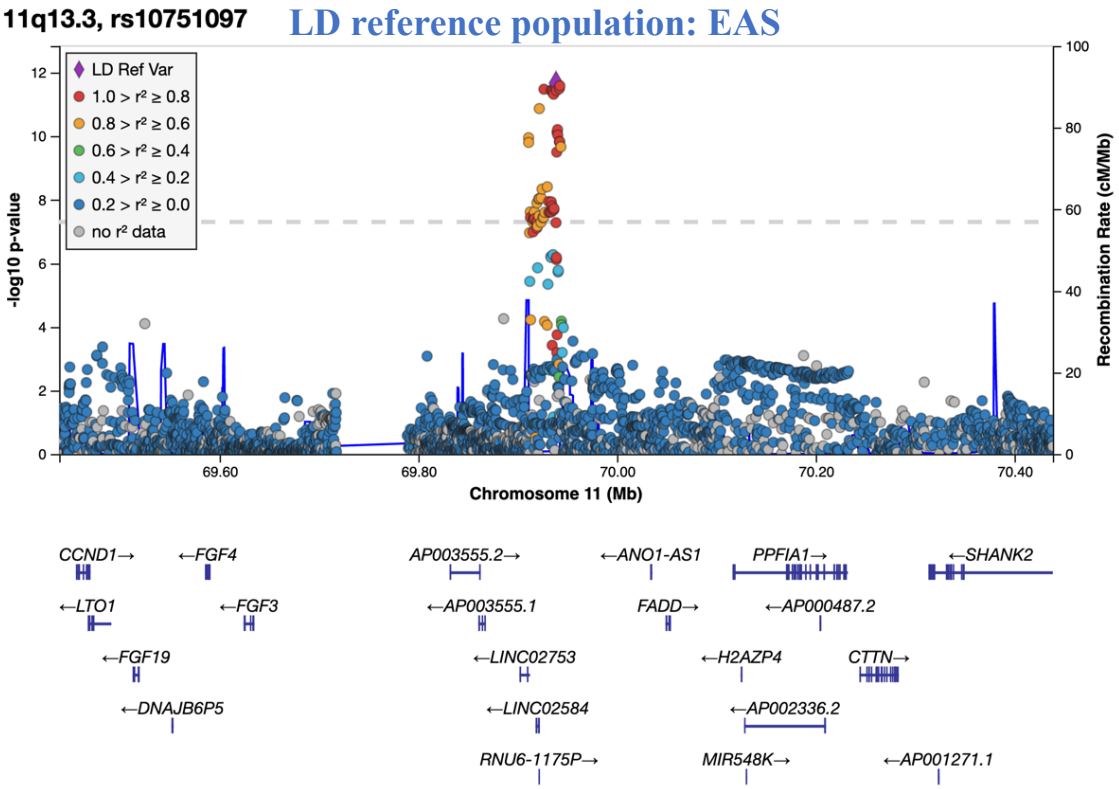

## LD reference population: EAS



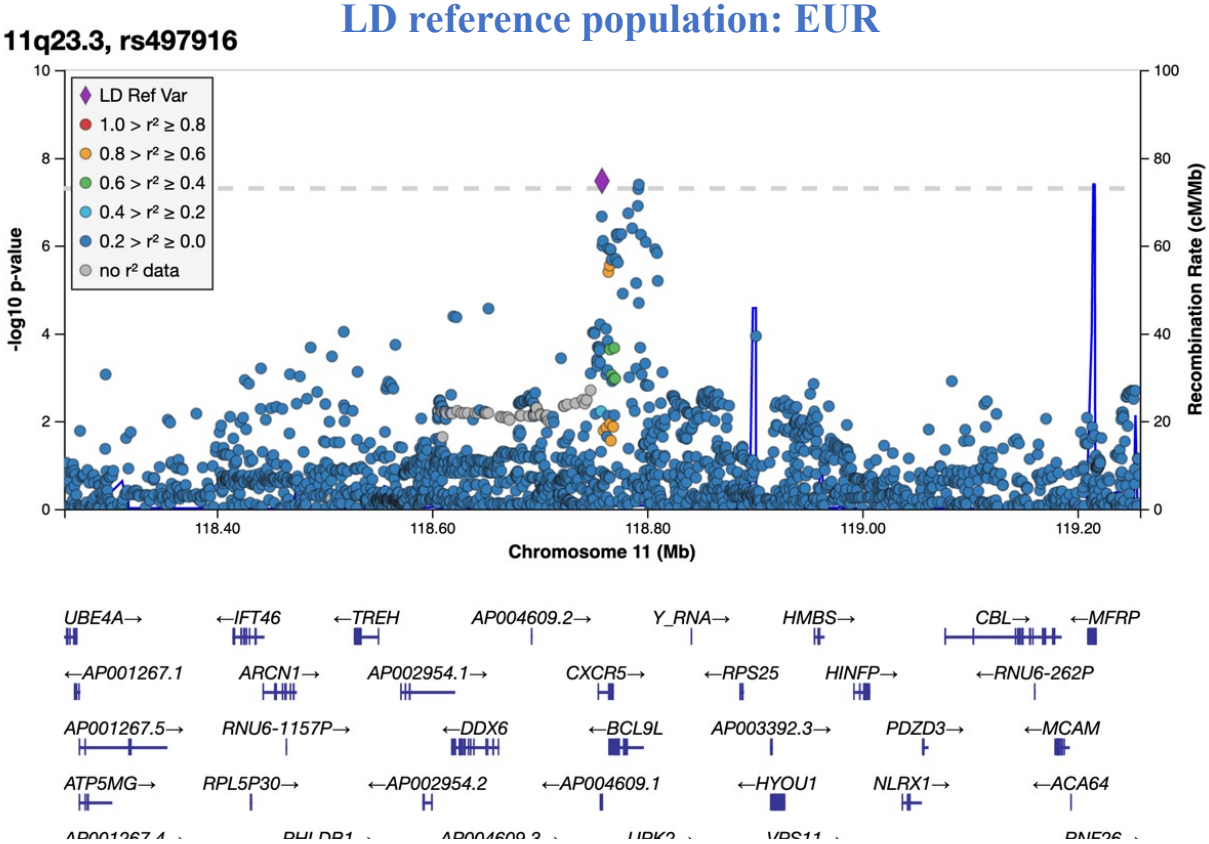
## LD reference population: EUR



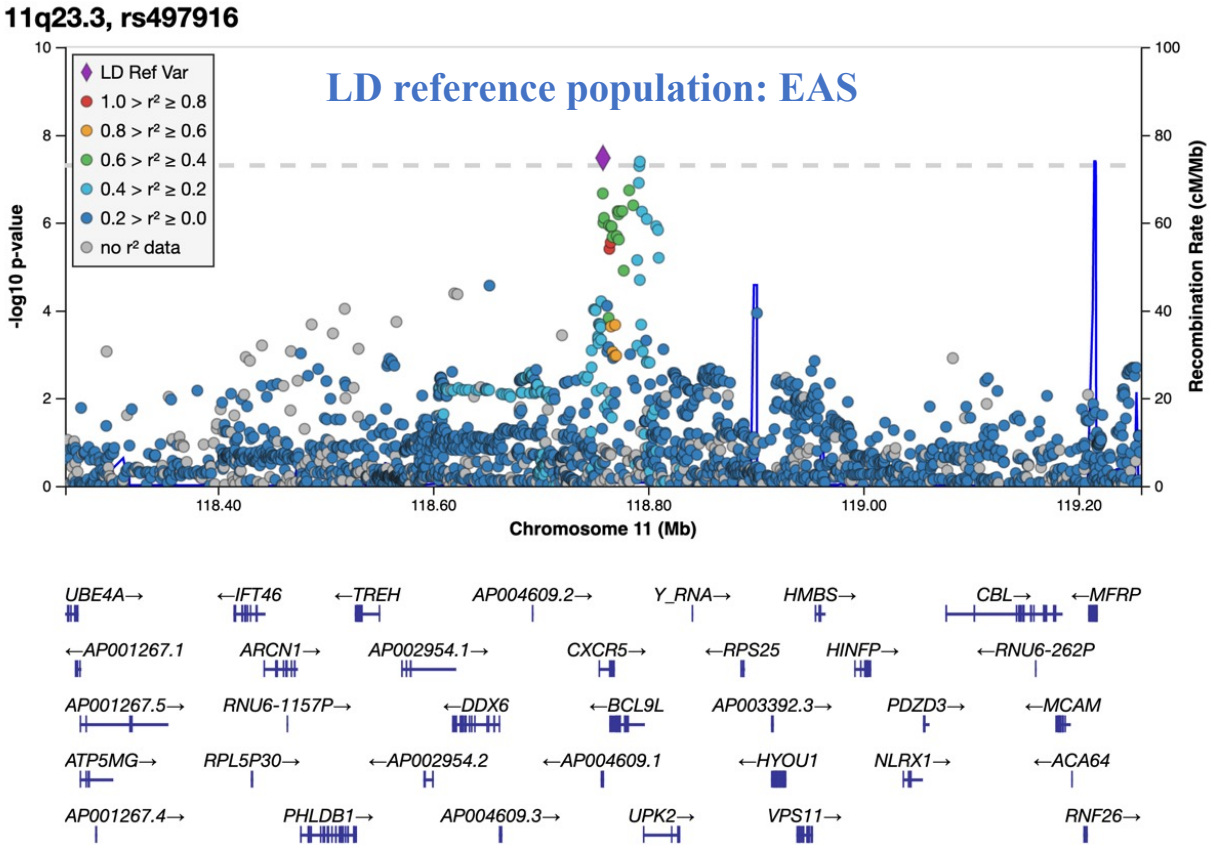

## LD reference population: EAS

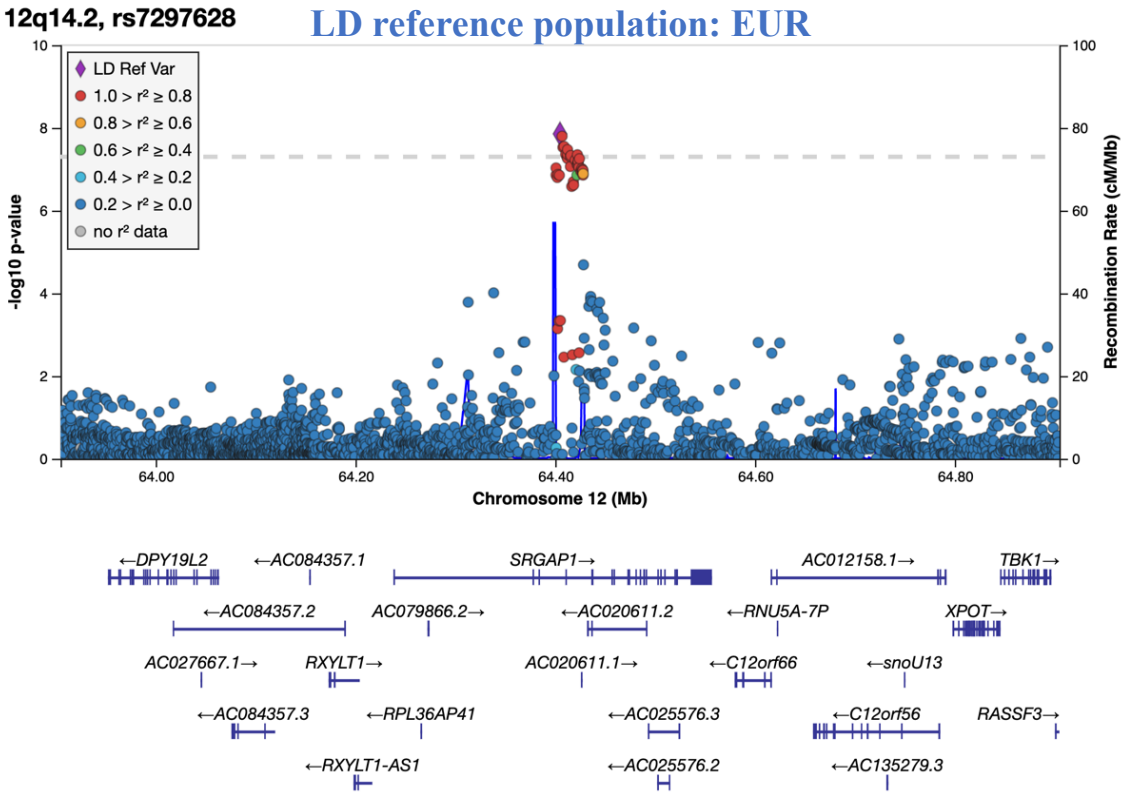


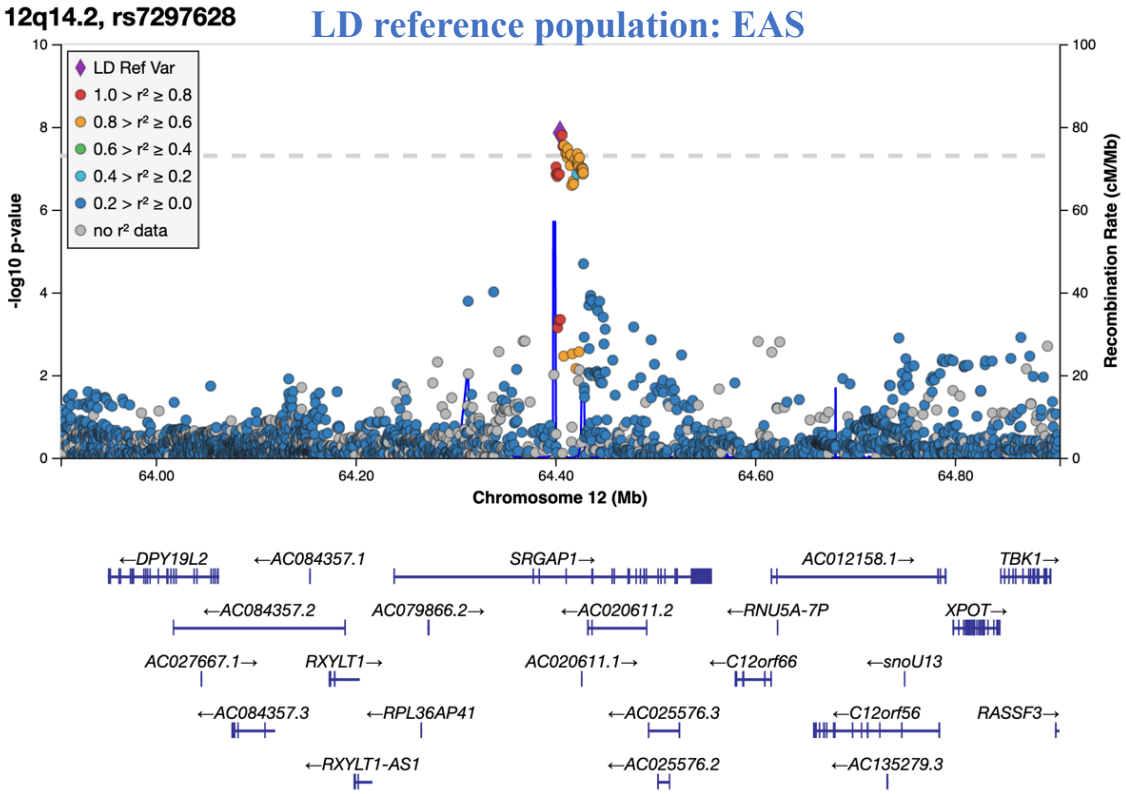
## LD reference population: EUR



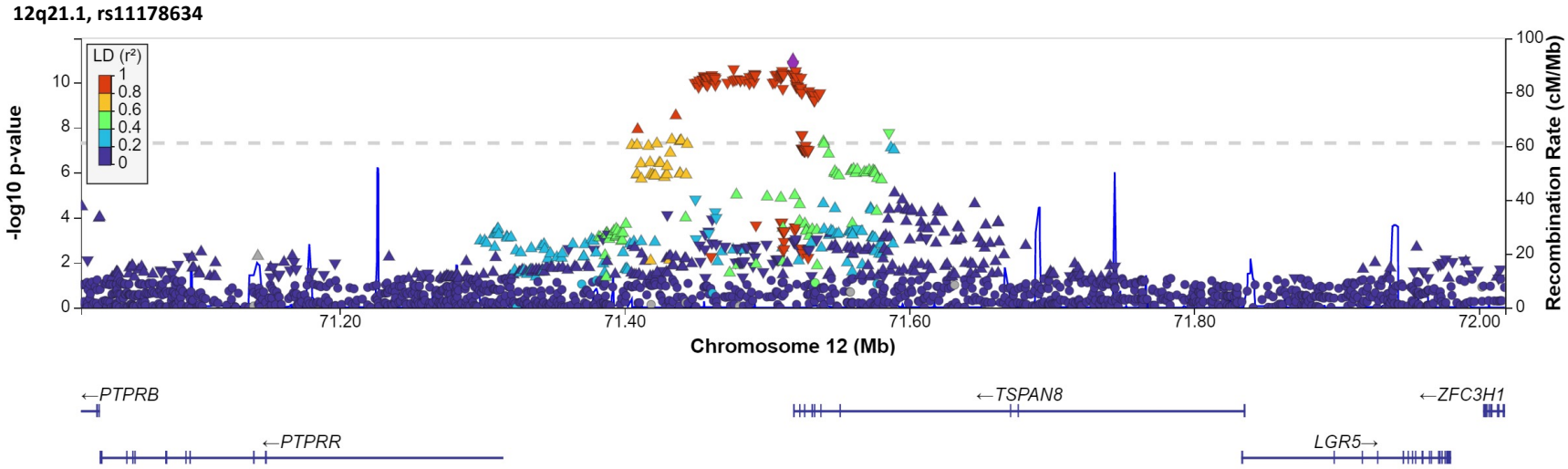

## LD reference population: EAS



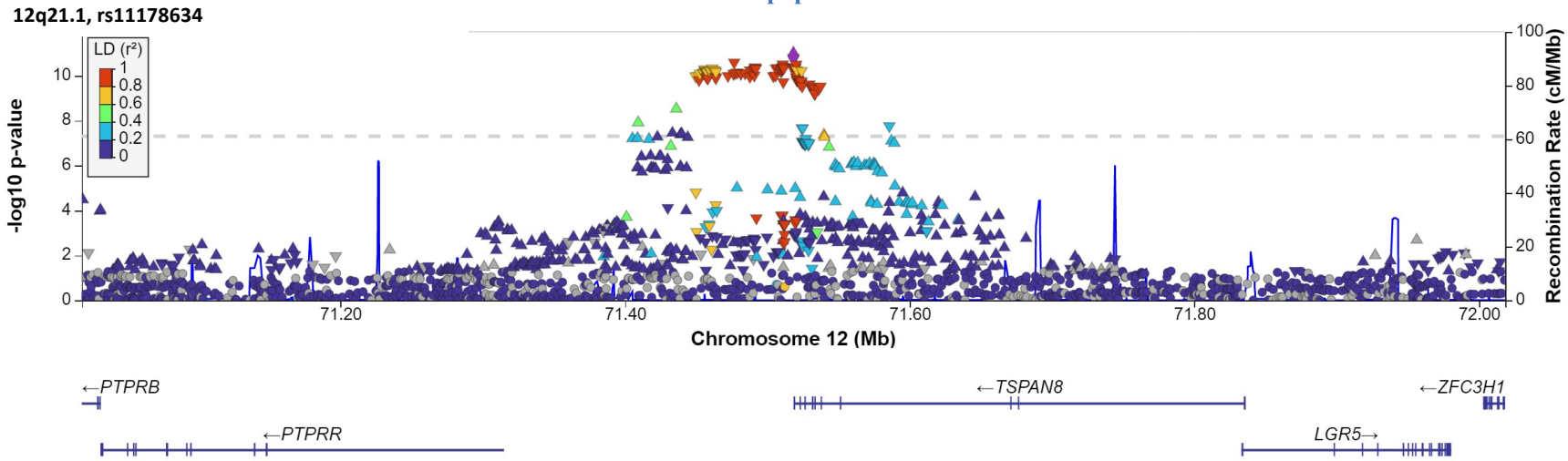

12q14.2, rs7297628
LD reference population: EUR
LD Ref Var
1.0 > r² ≥ 0.8
0.8 > r² ≥ 0.6
0.6 > r² ≥ 0.4
0.4 > r² ≥ 0.2
0.2 > r² ≥ 0.0
no r² data
-log10 p-value
Recombination Rate (cM/Mb)
Chromosome 12 (Mb)
64.00
64.20
64.40
64.60
64.80
←DPY19L2
←AC084357.1
SRGAP1→
AC012158.1→
TBK1→
←AC084357.2
AC079866.2→
←AC020611.2
←RNU5A-7P
XPOT→
AC027667.1→
RXYLT1→
AC020611.1→
←C12orf66
←snoU13
←AC084357.3
←RPL36AP41
←AC025576.3
←C12orf56
RASSF3→
←RXYLT1-AS1
←AC025576.2
←AC135279.3


12q14.2, rs7297628
LD reference population: EAS
LD Ref Var
1.0 > r² ≥ 0.8
0.8 > r² ≥ 0.6
0.6 > r² ≥ 0.4
0.4 > r² ≥ 0.2
0.2 > r² ≥ 0.0
no r² data
-log10 p-value
Recombination Rate (cM/Mb)
Chromosome 12 (Mb)
64.00
64.20
64.40
64.60
64.80
←DPY19L2
←AC084357.1
SRGAP1→
AC012158.1→
TBK1→
←AC084357.2
AC079866.2→
←AC020611.2
←RNU5A-7P
XPOT→
AC027667.1→
RXYLT1→
AC020611.1→
←C12orf66
←snoU13
←AC084357.3
←RPL36AP41
←AC025576.3
←C12orf56
RASSF3→
←RXYLT1-AS1
←AC025576.2
←AC135279.3

## LD reference population: EUR

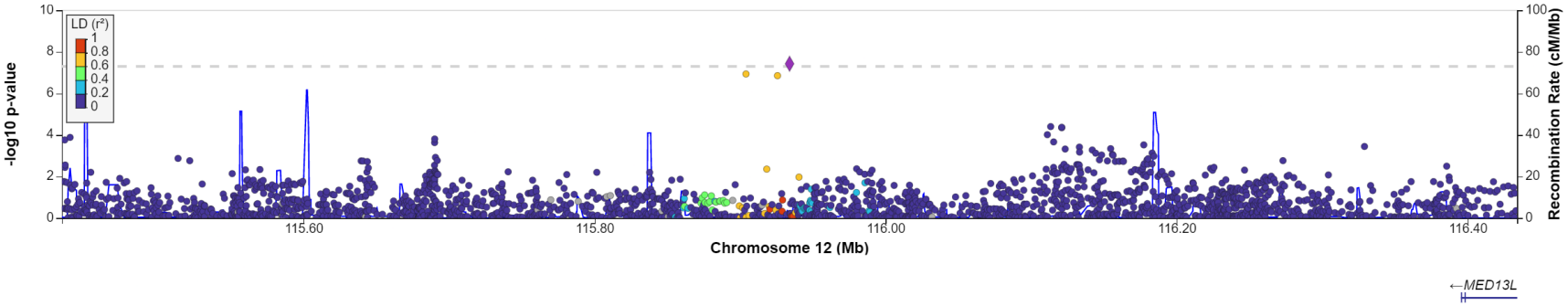



## LD reference population: EAS



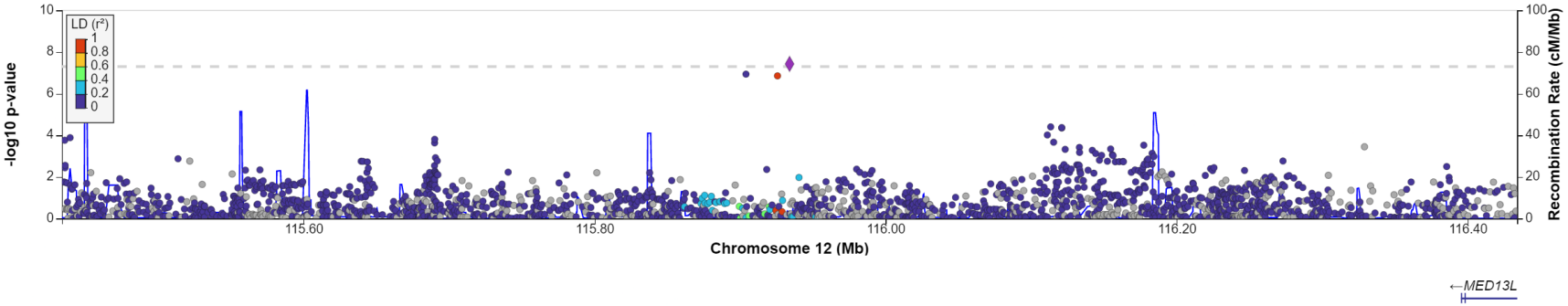
## LD reference population: EUR

**12q24.1, rs7299936**

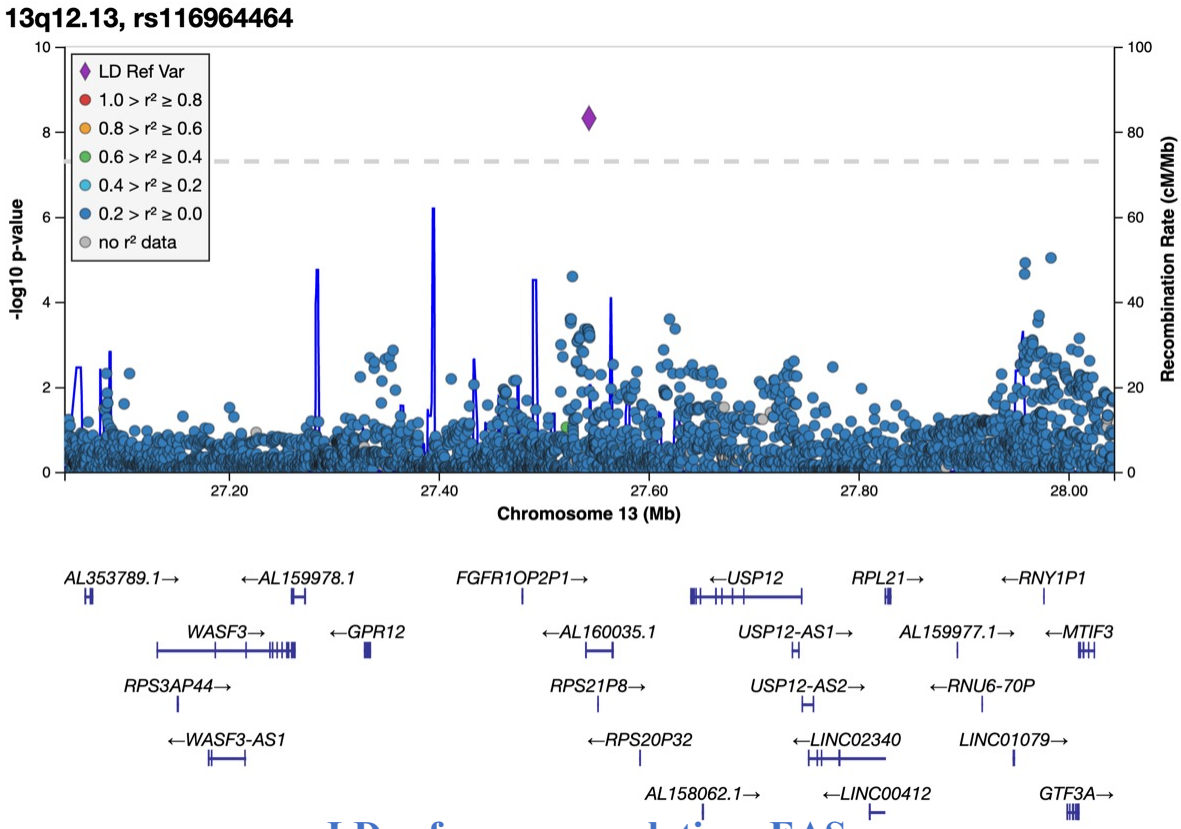



## LD reference population: EAS

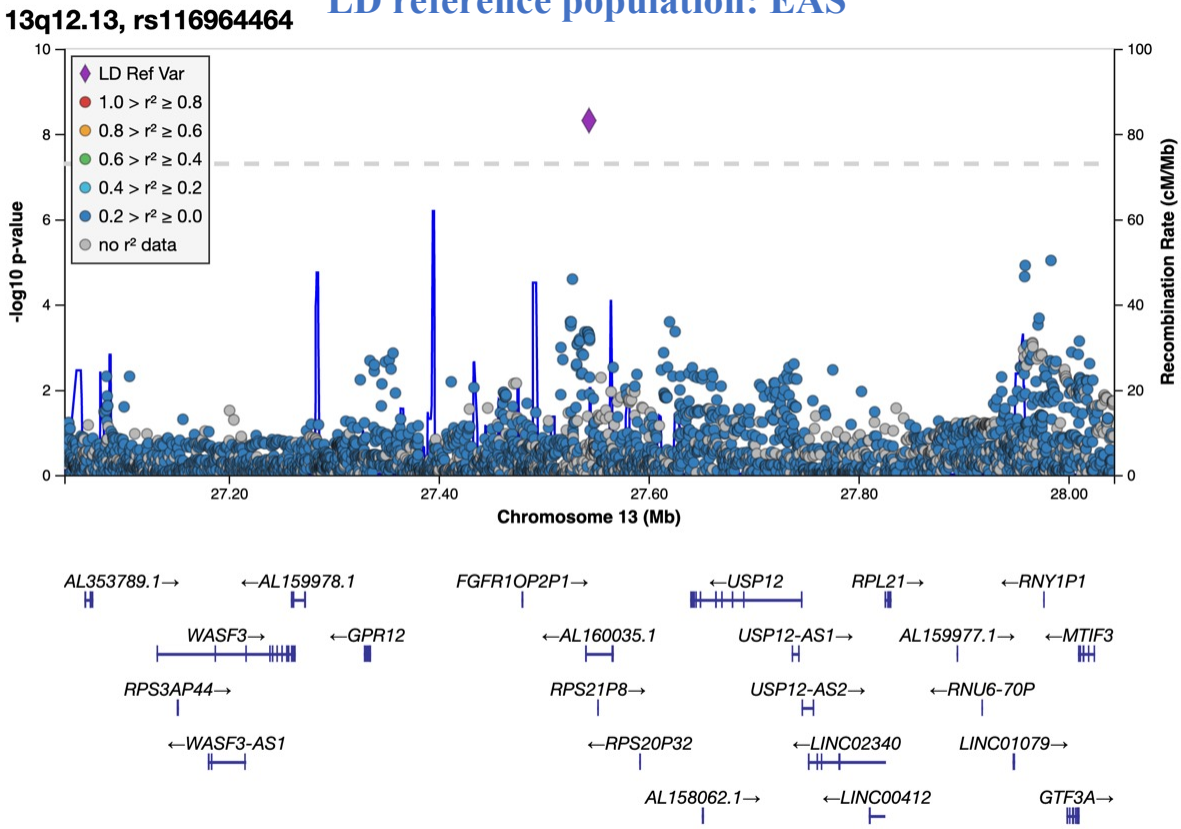
**12q24.1, rs7299936**

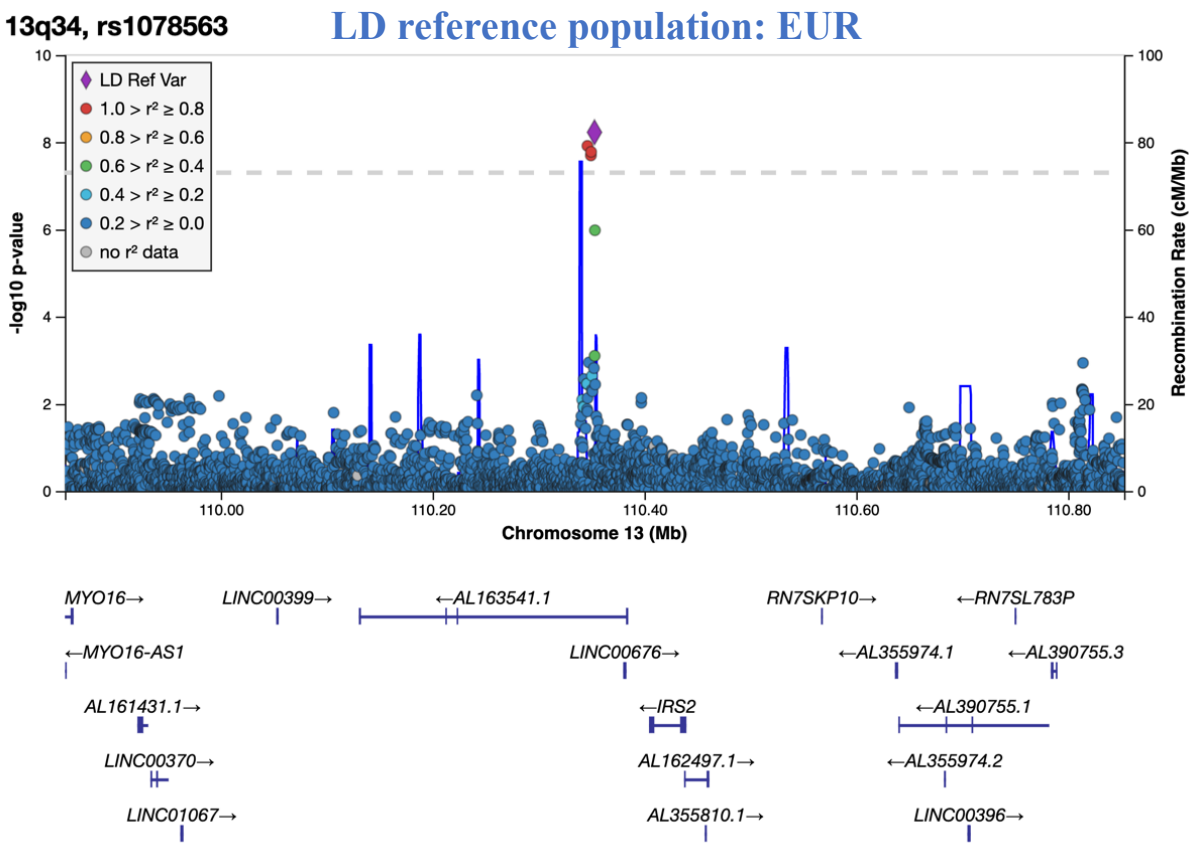


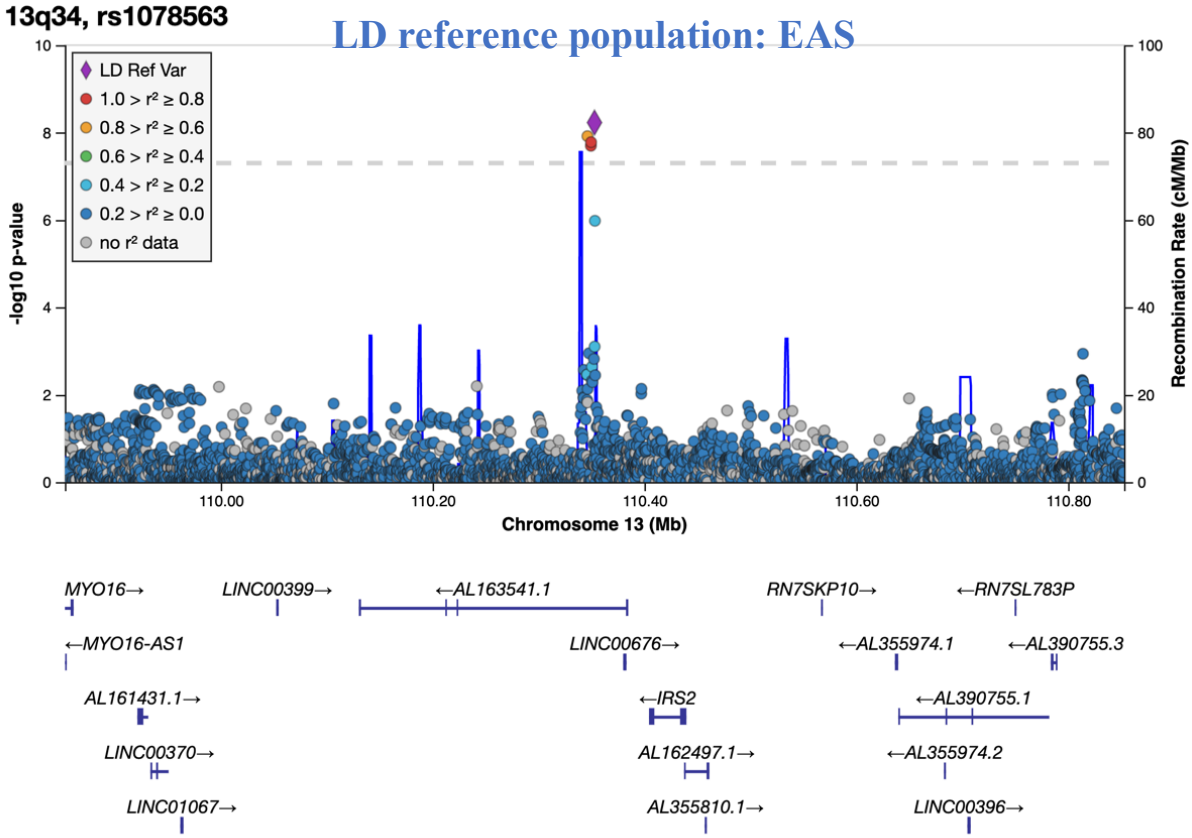
## LD reference population: EUR



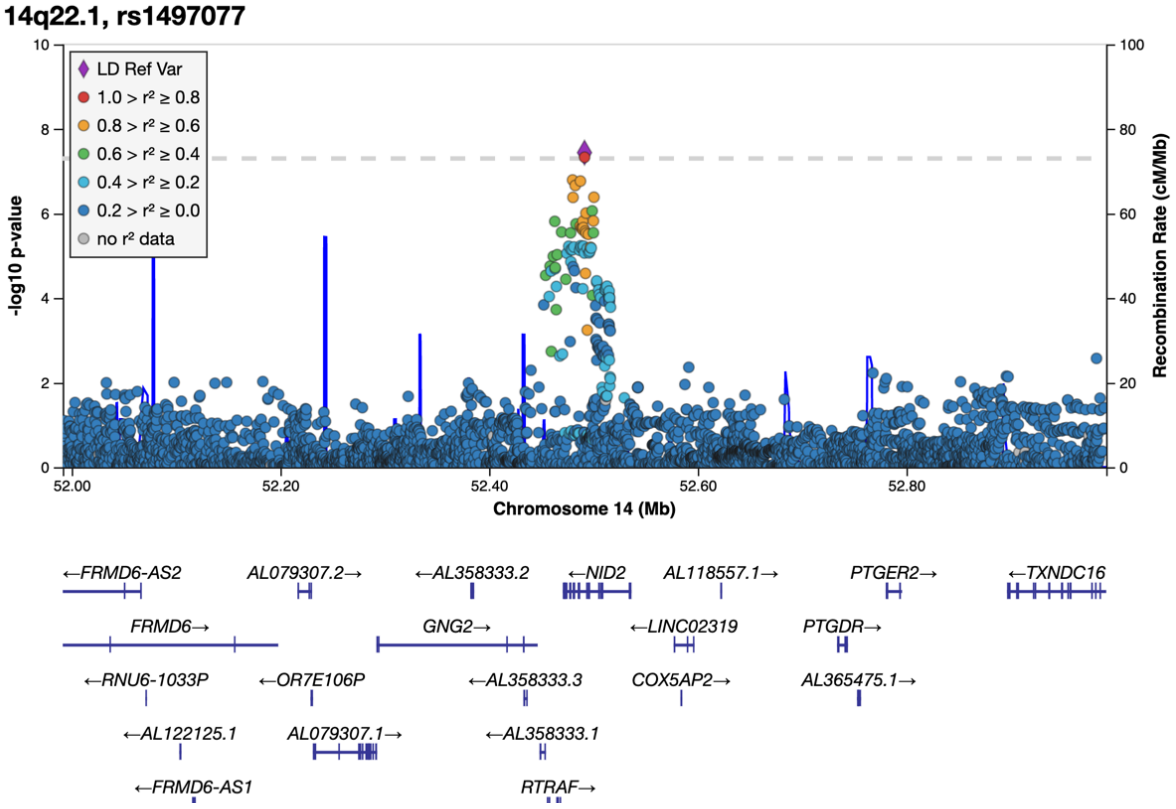

## LD reference population: EAS



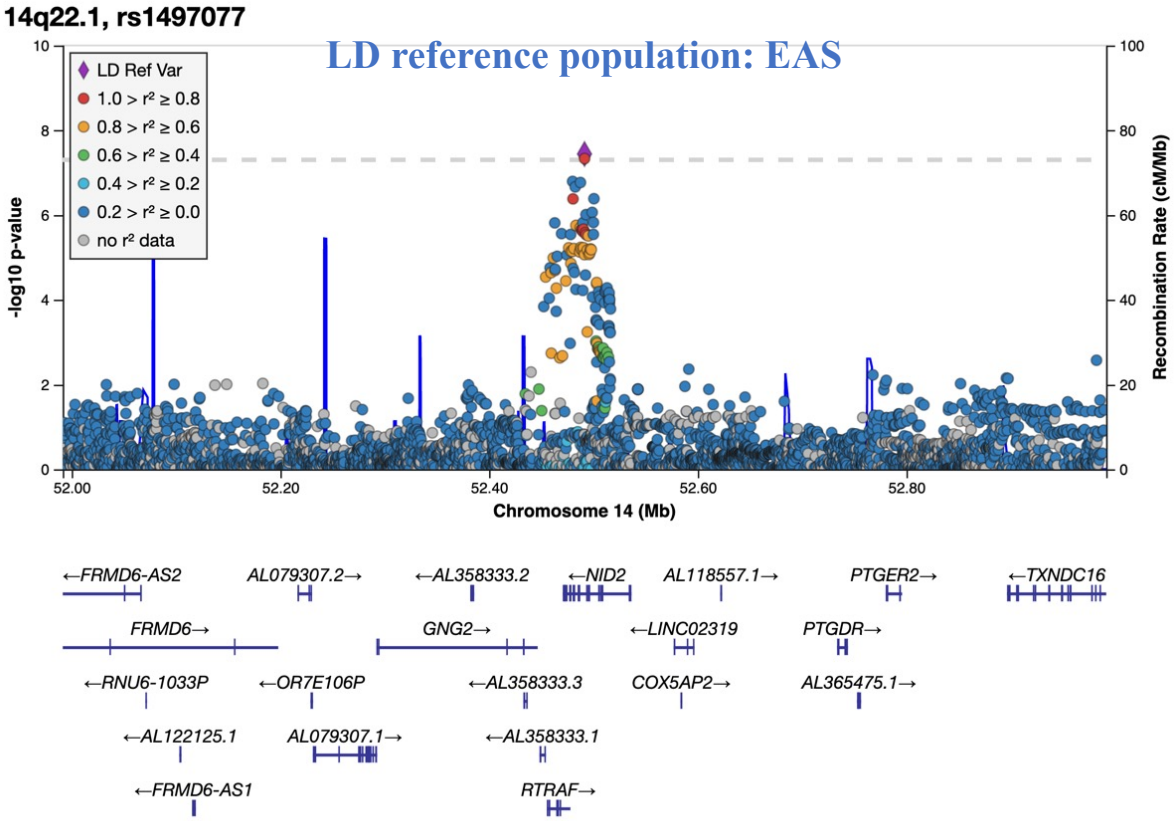

13q34, rs1078563
LD reference population: EUR
LD Ref Var
1.0 > r² ≥ 0.8
0.8 > r² ≥ 0.6
0.6 > r² ≥ 0.4
0.4 > r² ≥ 0.2
0.2 > r² ≥ 0.0
no r² data
-log10 p-value
Recombination Rate (cM/Mb)
Chromosome 13 (Mb)
110.00
110.20
110.40
110.60
110.80
MYO16→
LINC00399→
←AL163541.1
RN7SKP10→
←RN7SL783P
←MYO16-AS1
LINC00676→
←AL355974.1
←AL390755.3
AL161431.1→
←IRS2
←AL390755.1
LINC00370→
AL162497.1→
←AL355974.2
LINC01067→
AL355810.1→
LINC00396→


13q34, rs1078563
LD reference population: EAS
LD Ref Var
1.0 > r² ≥ 0.8
0.8 > r² ≥ 0.6
0.6 > r² ≥ 0.4
0.4 > r² ≥ 0.2
0.2 > r² ≥ 0.0
no r² data
-log10 p-value
Recombination Rate (cM/Mb)
Chromosome 13 (Mb)
110.00
110.20
110.40
110.60
110.80
MYO16→
LINC00399→
←AL163541.1
RN7SKP10→
←RN7SL783P
←MYO16-AS1
LINC00676→
←AL355974.1
←AL390755.3
AL161431.1→
←IRS2
←AL390755.1
LINC00370→
AL162497.1→
←AL355974.2
LINC01067→
AL355810.1→
LINC00396→

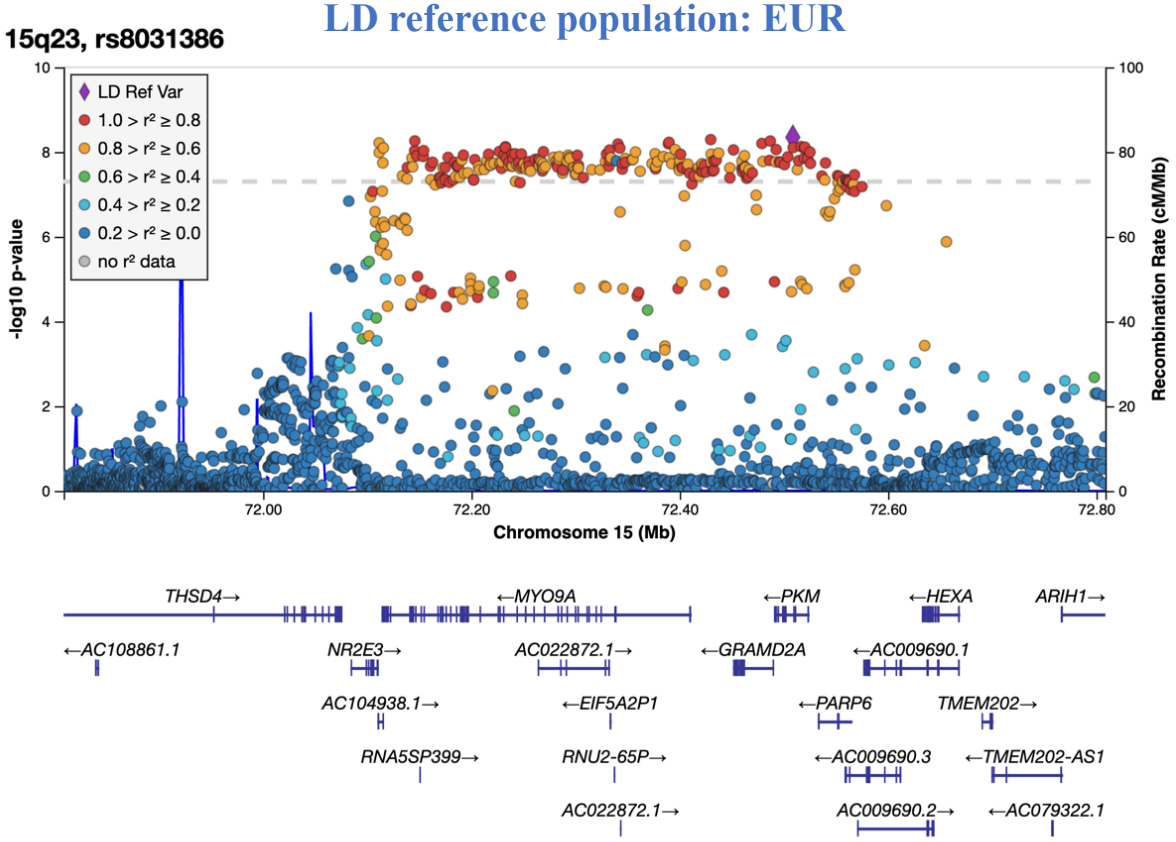
## LD reference population: EUR



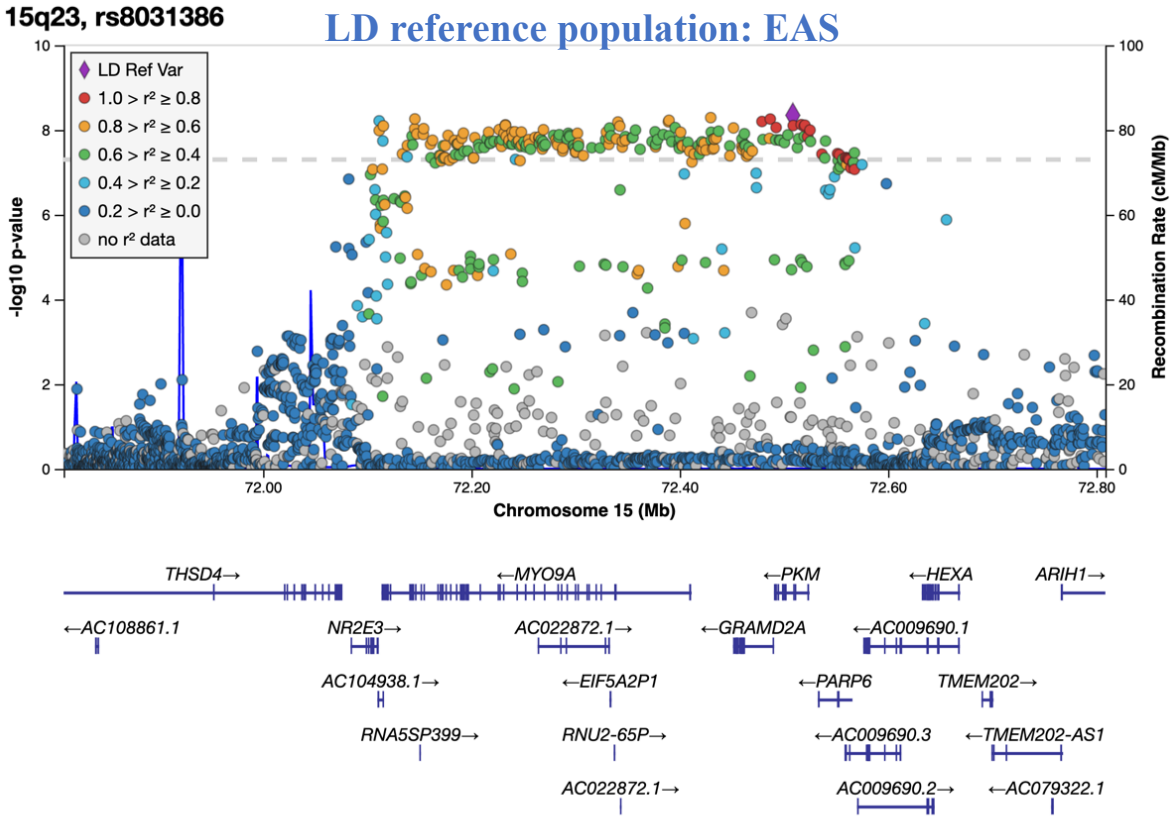

## LD reference population: EAS



## LD reference population: EUR

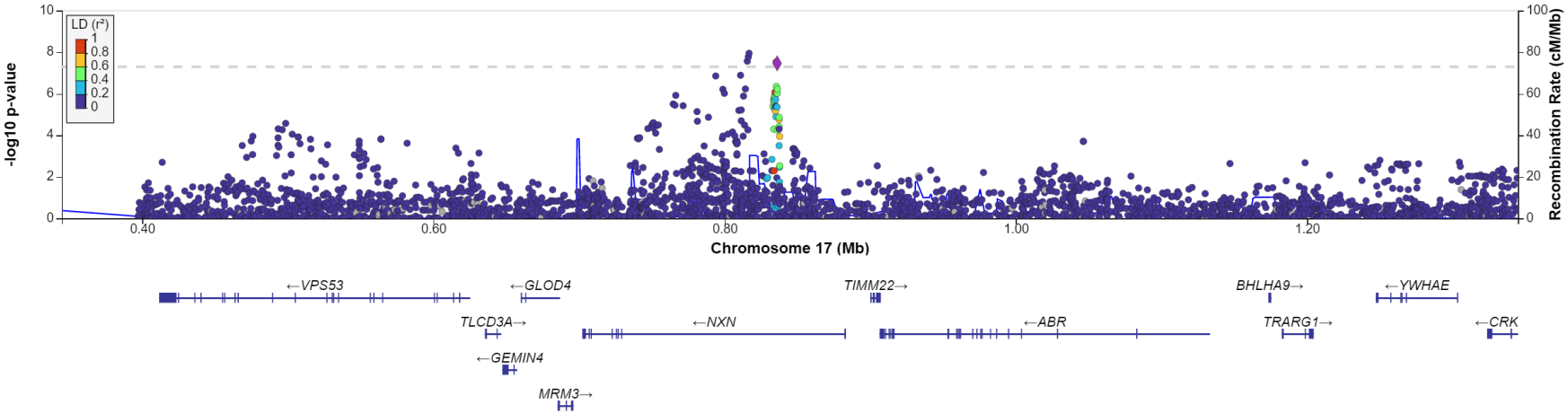



## LD reference population: EAS



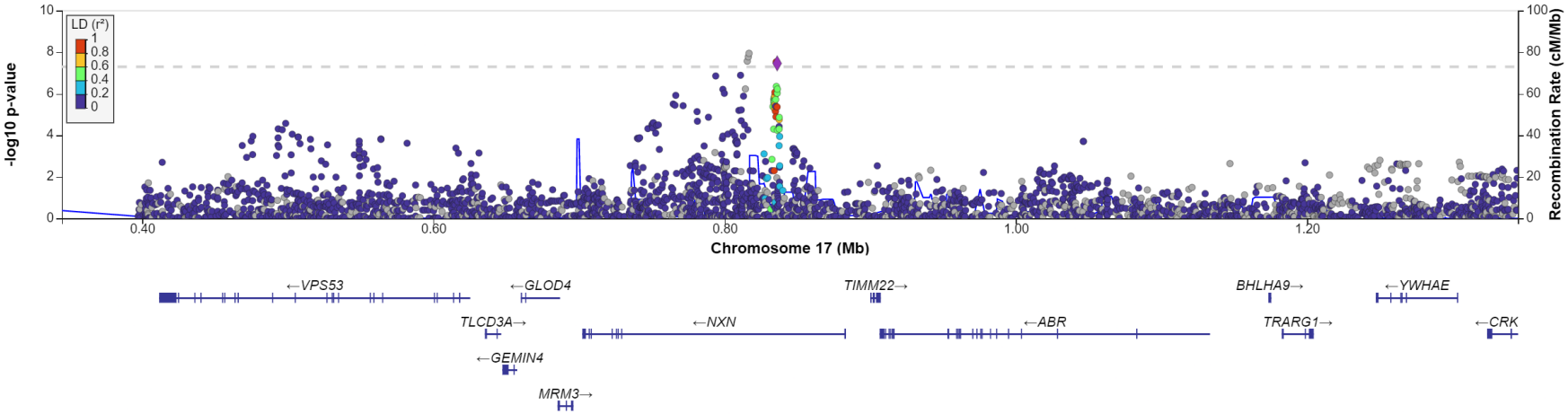
## LD reference population: EUR

**17p13.3, rs11247566**

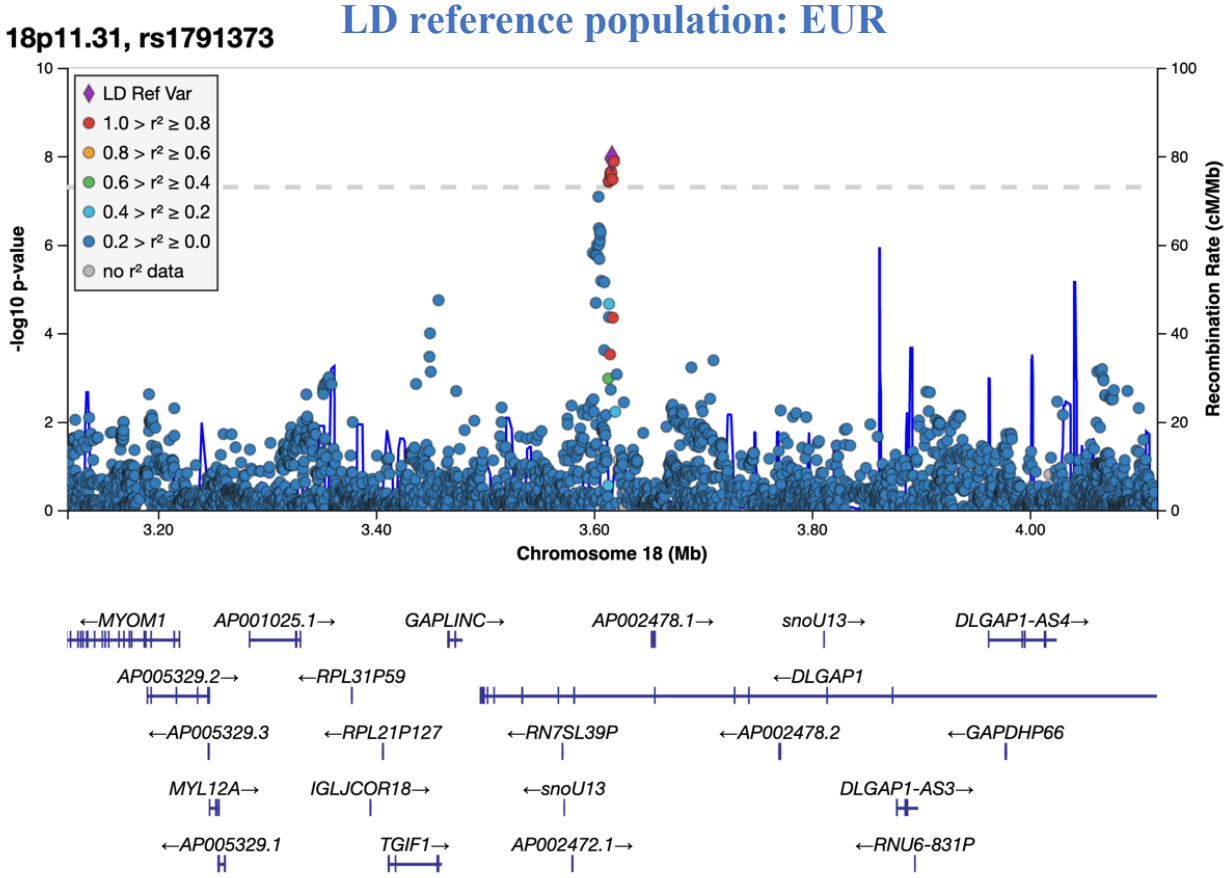



## LD reference population: EAS

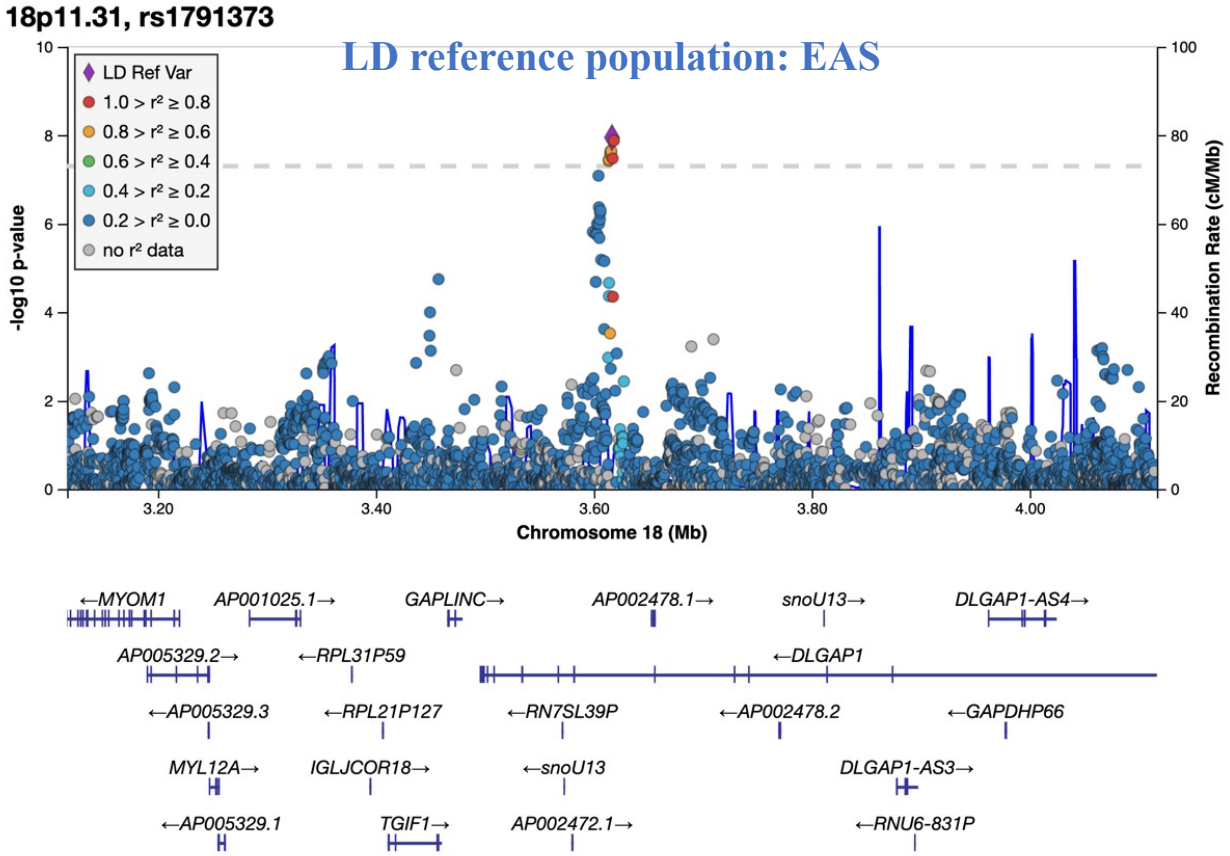
**17p13.3, rs11247566**



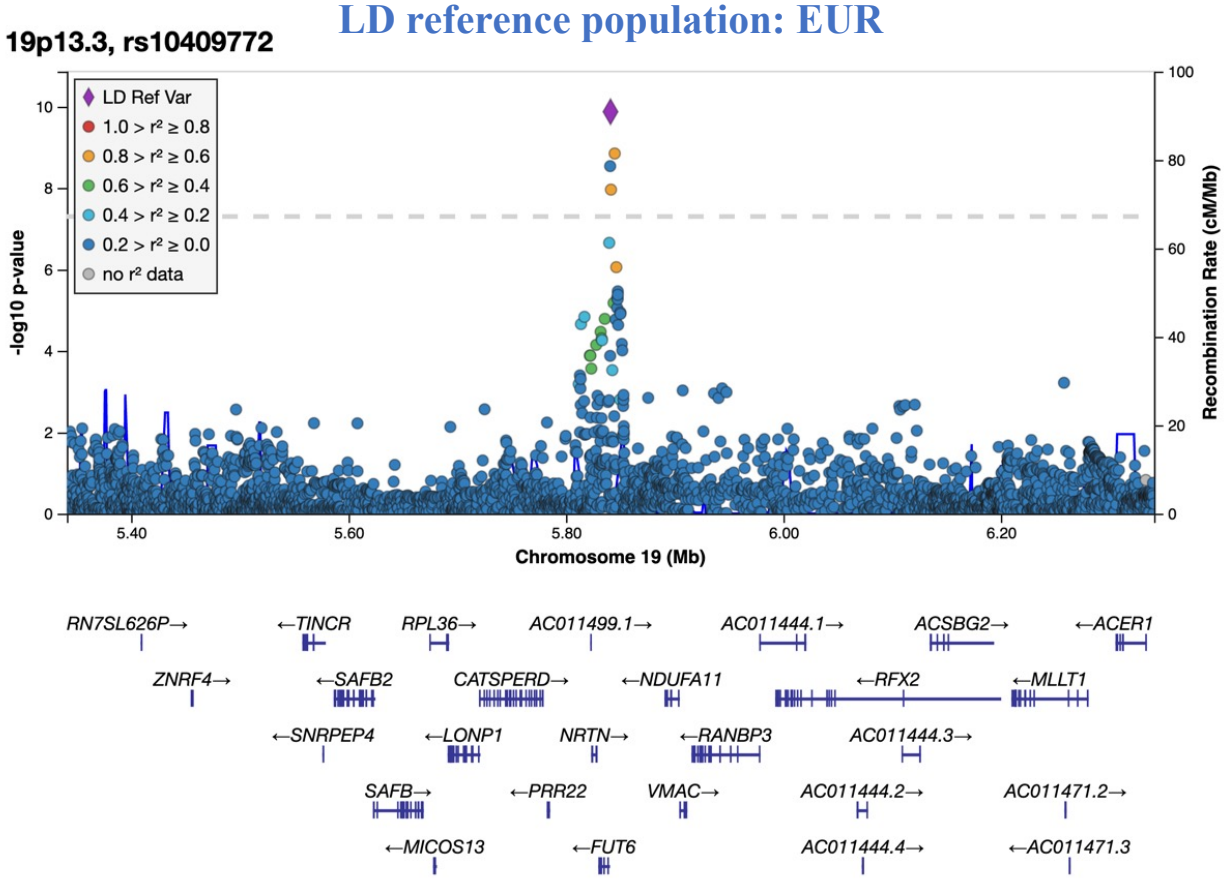
## LD reference population: EUR



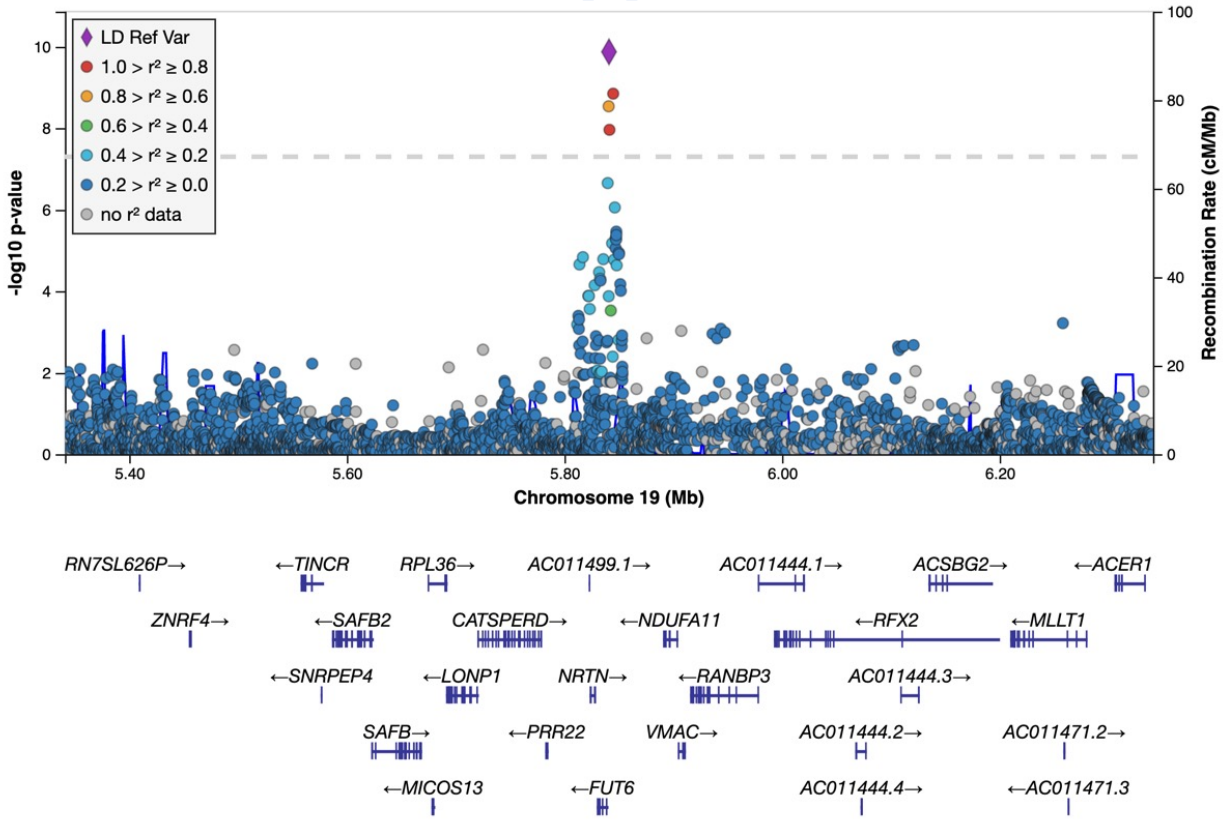

## LD reference population: EAS



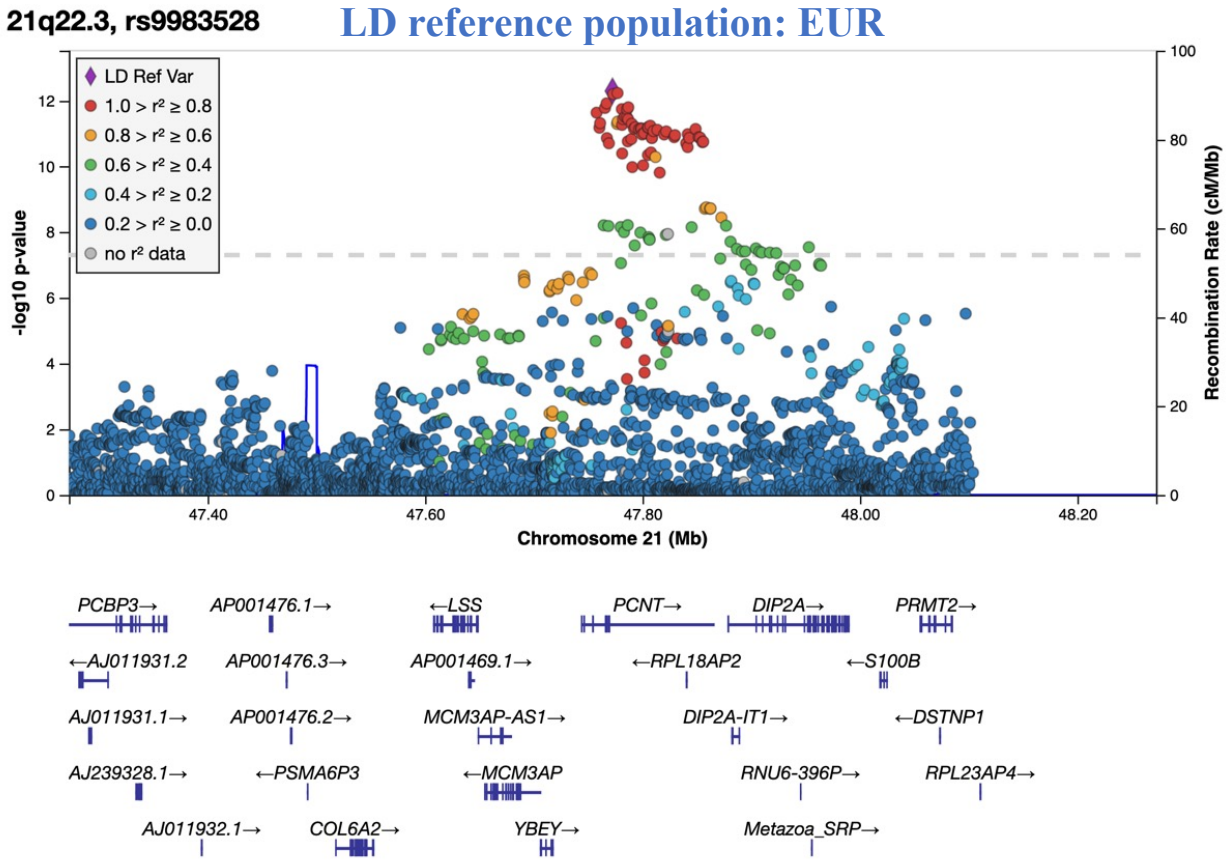
## LD reference population: EUR



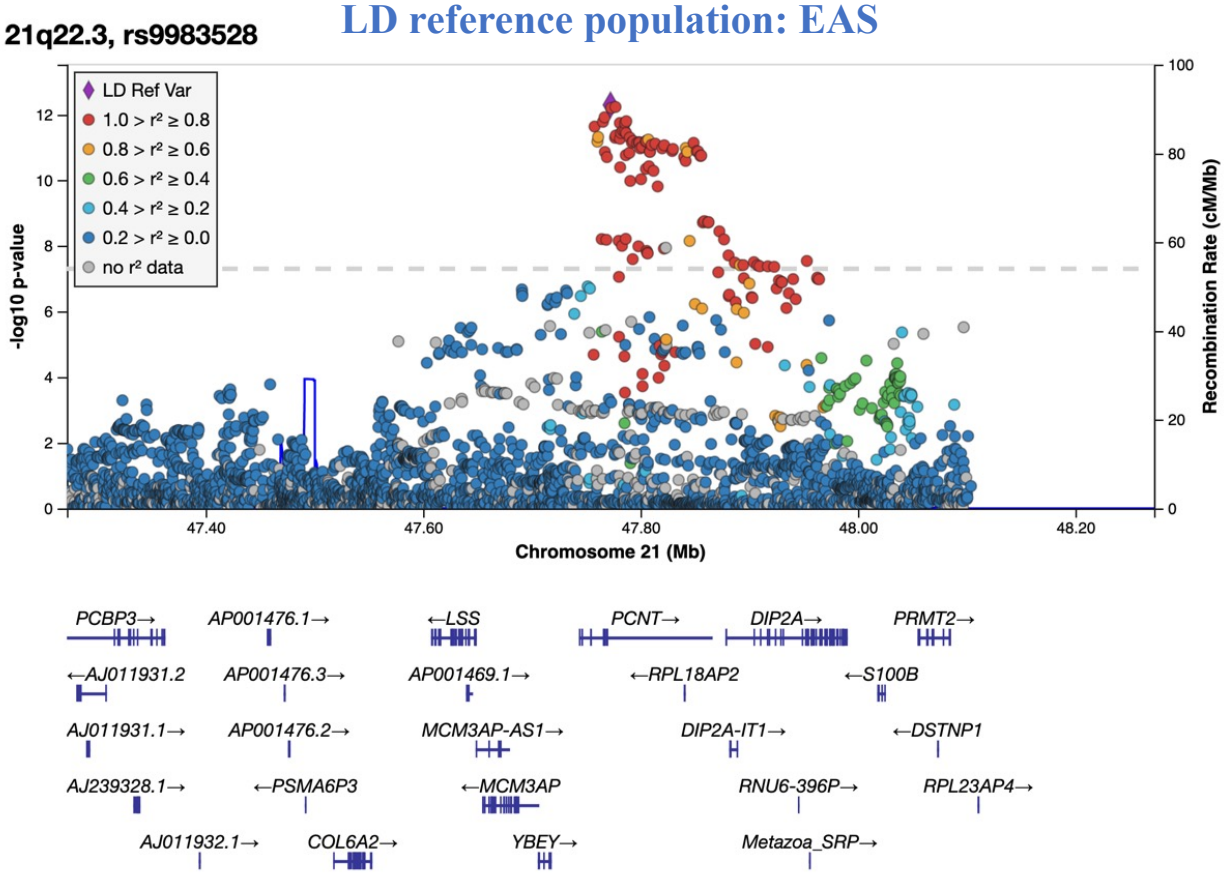

## LD reference population: EAS



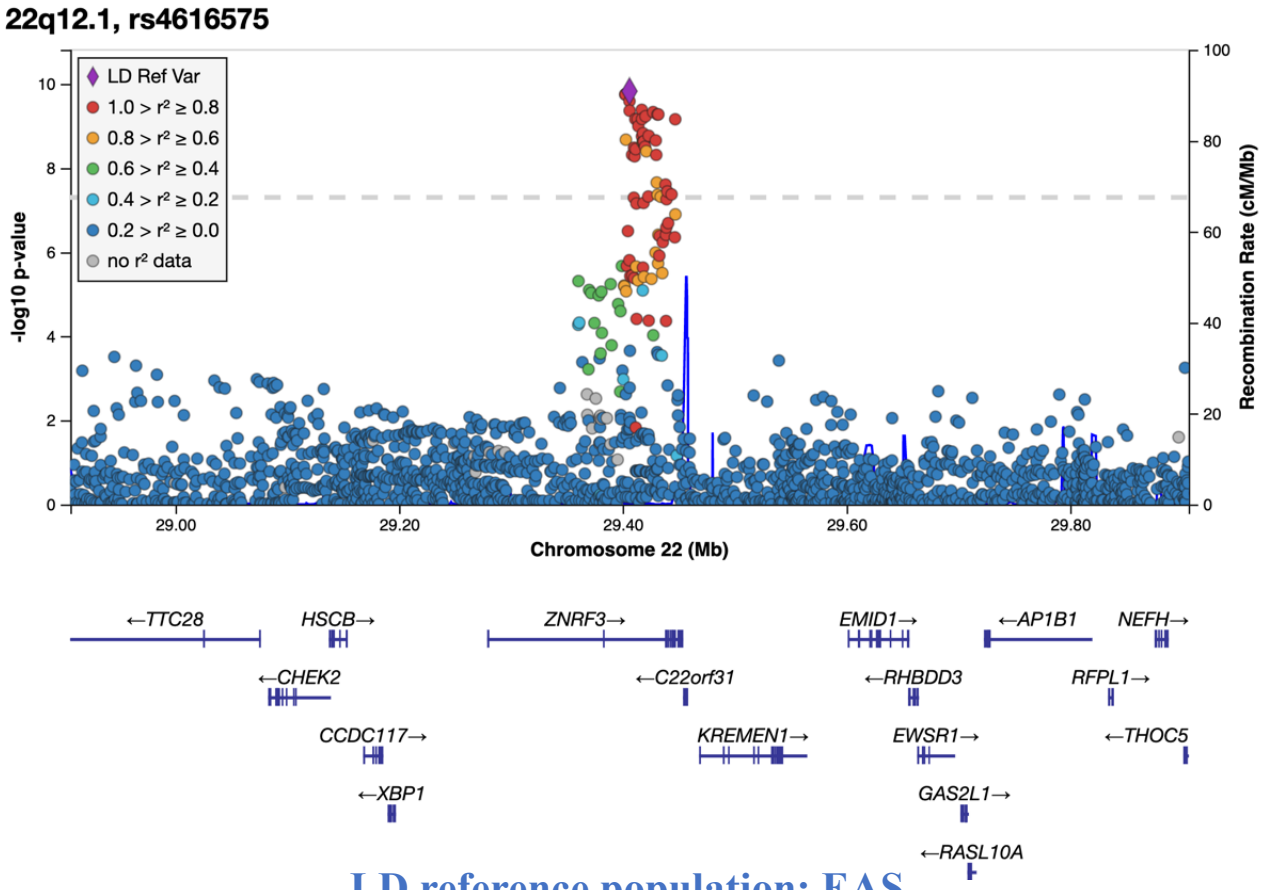
## LD reference population: EUR



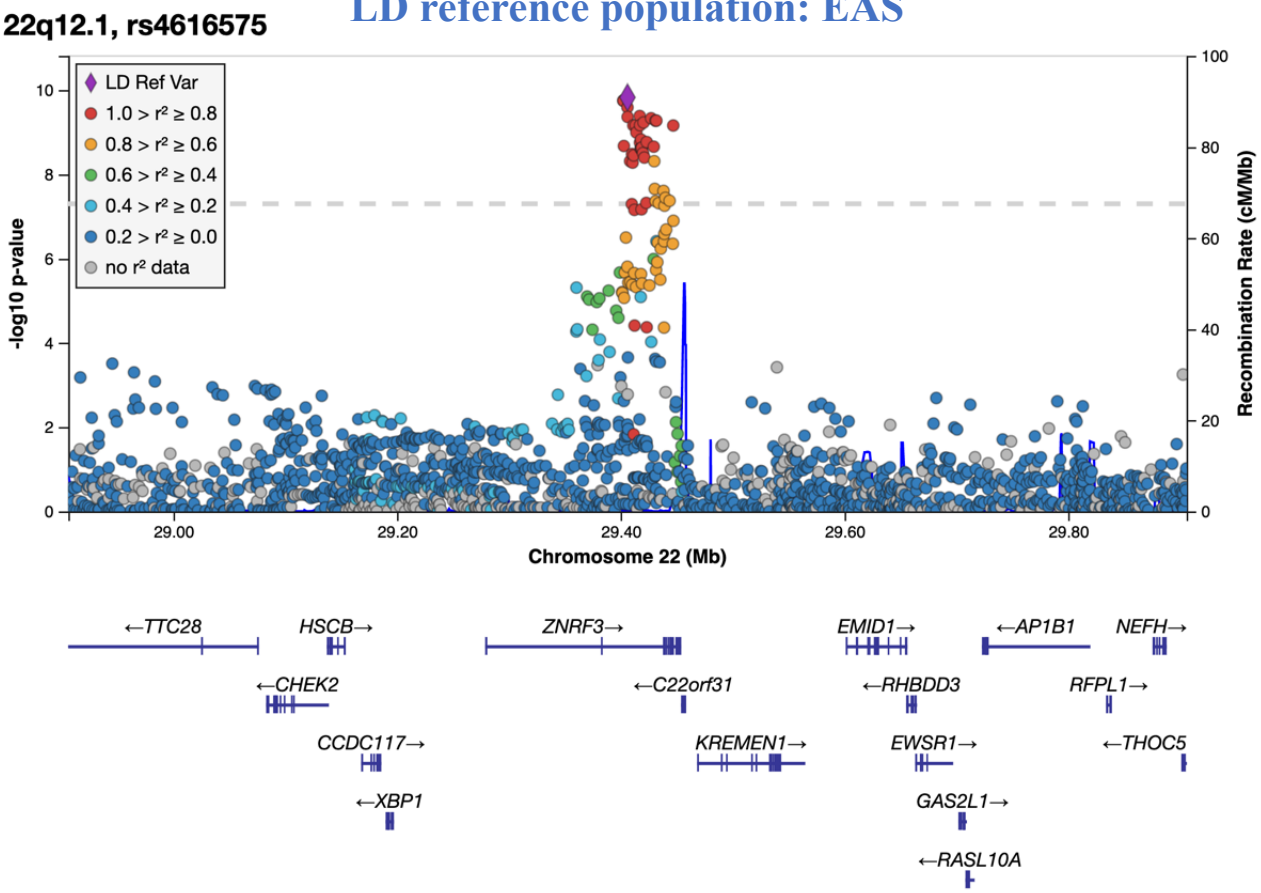

## LD reference population: EAS



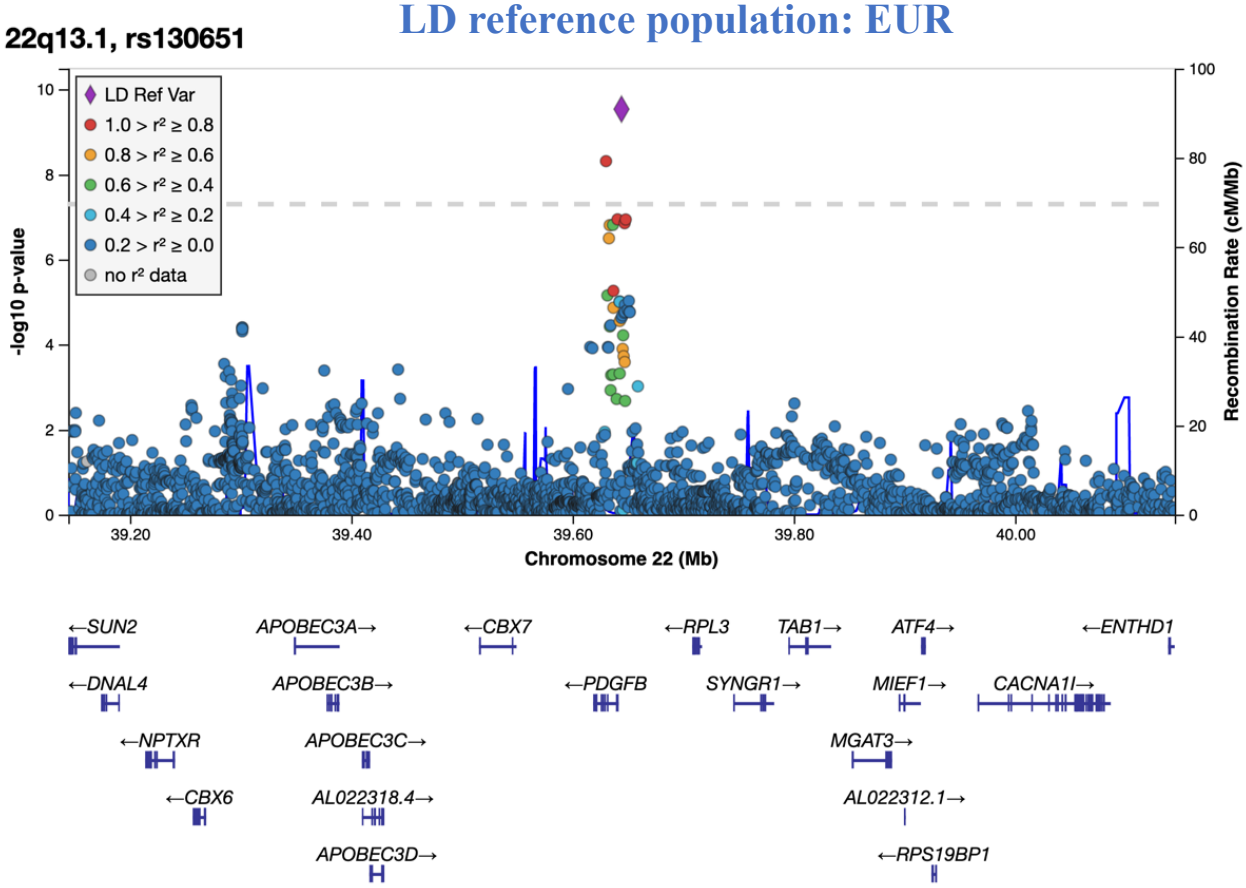
## LD reference population: EUR



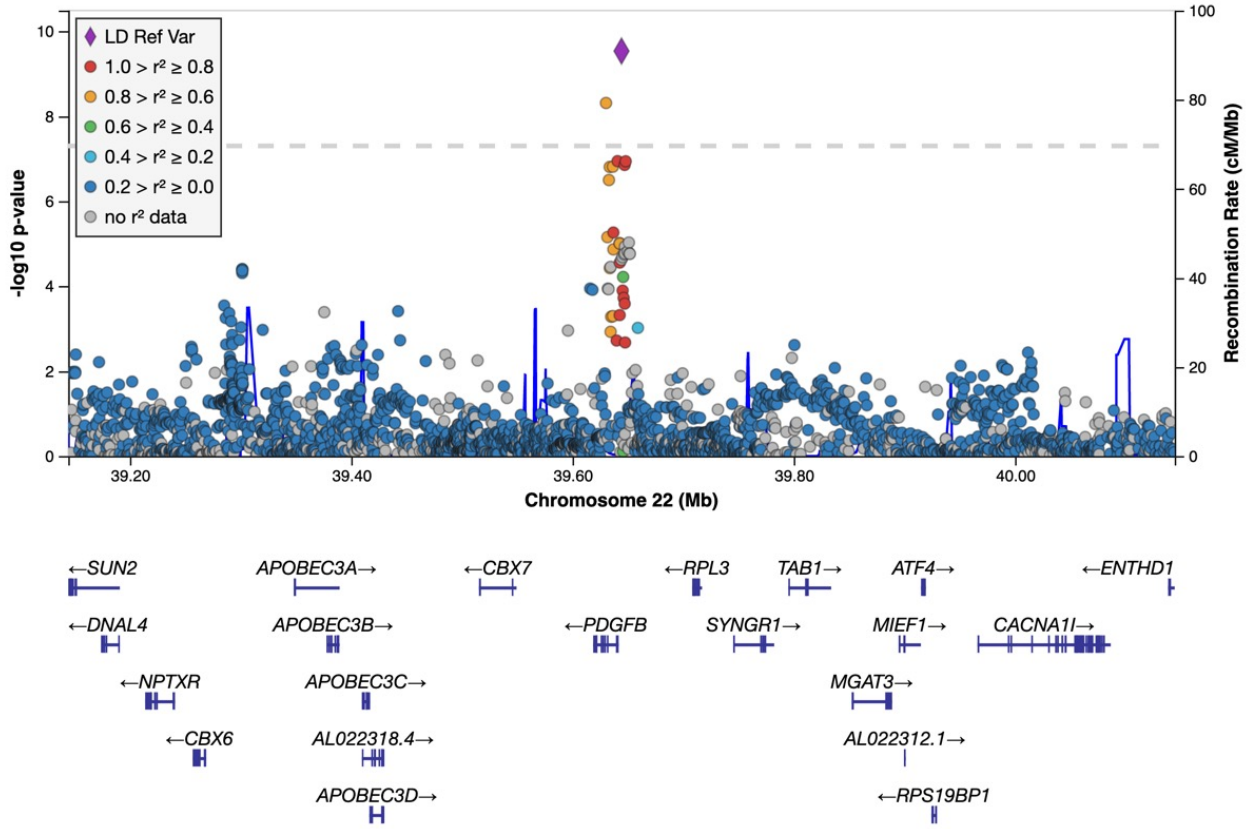

## LD reference population: EAS



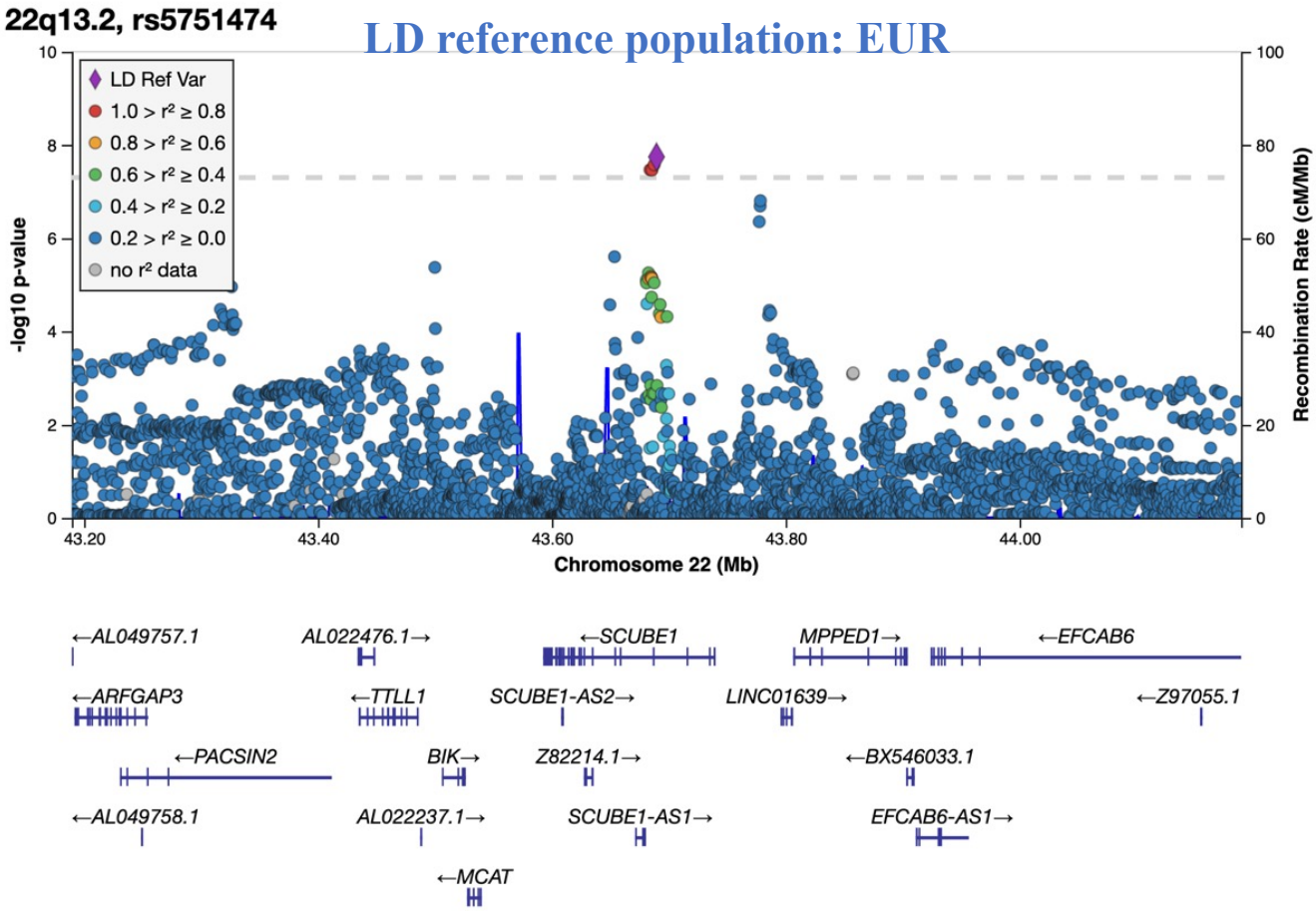
## LD reference population: EUR




## LD reference population: EAS



## LD reference population: EUR

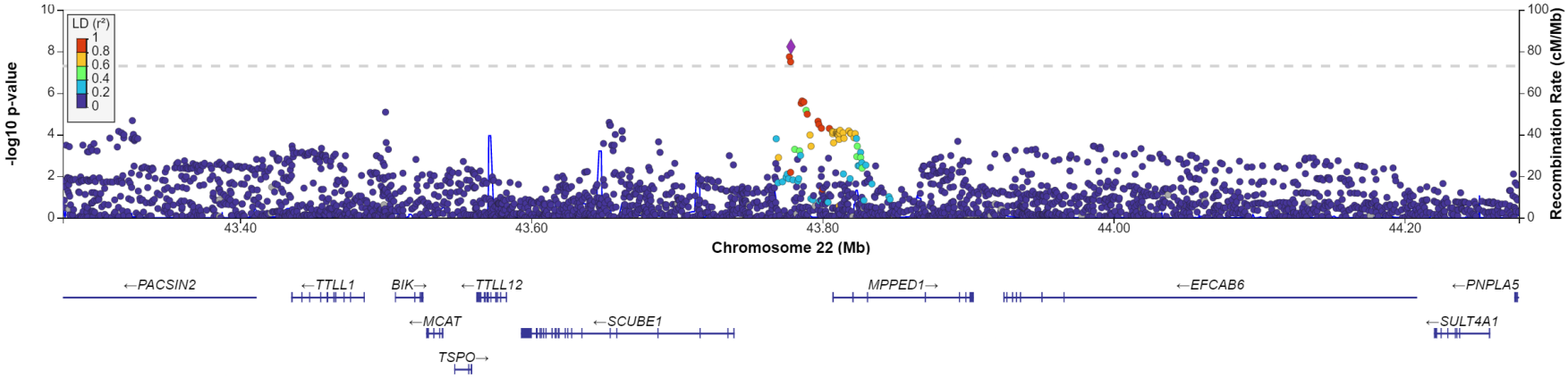



## LD reference population: EAS

rs5751474 is rare in the EAS population (MAF=0.006)

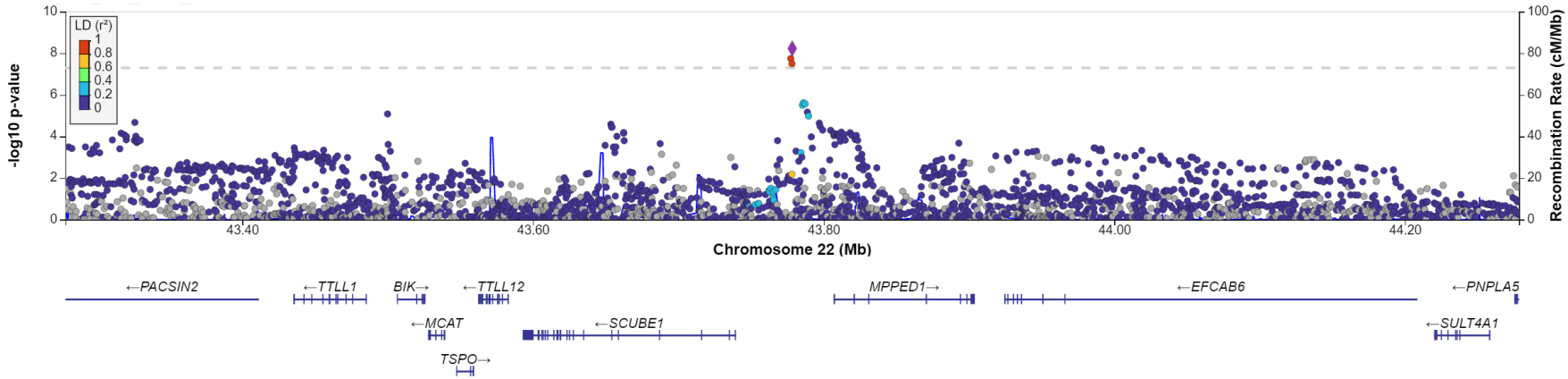
## LD reference population: EUR

**22q13.2, rs34256596**

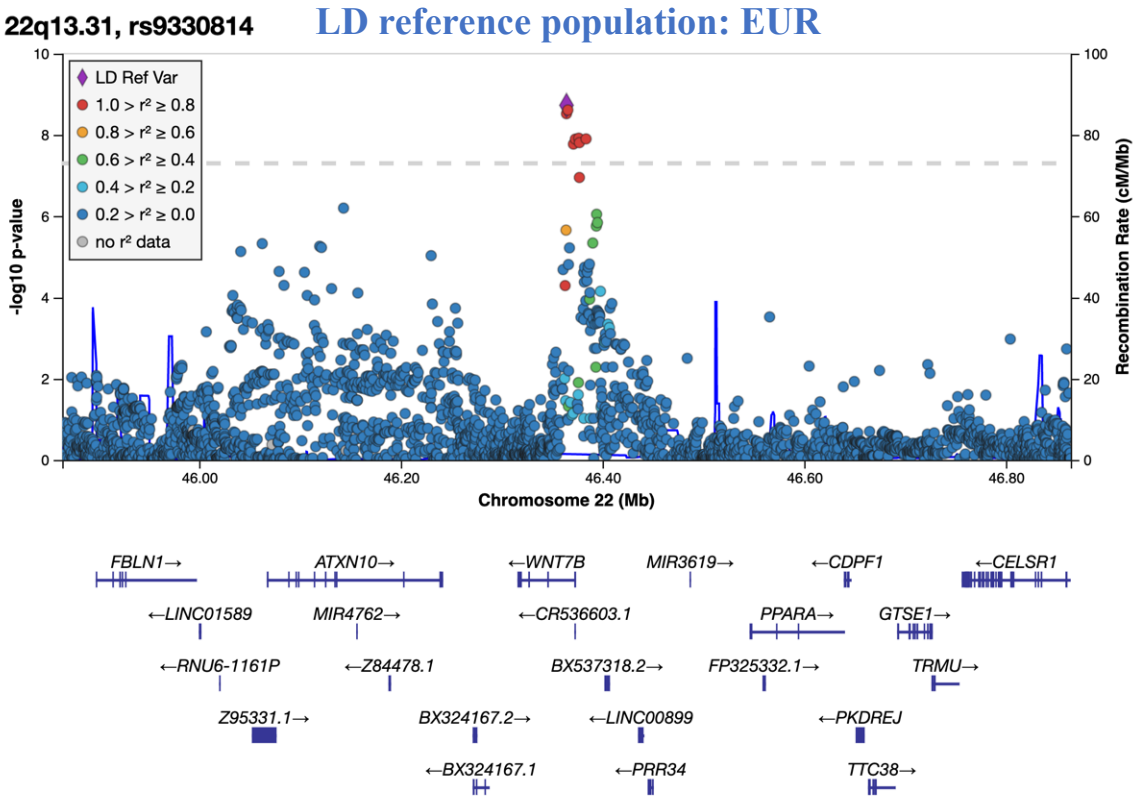



## LD reference population: EAS

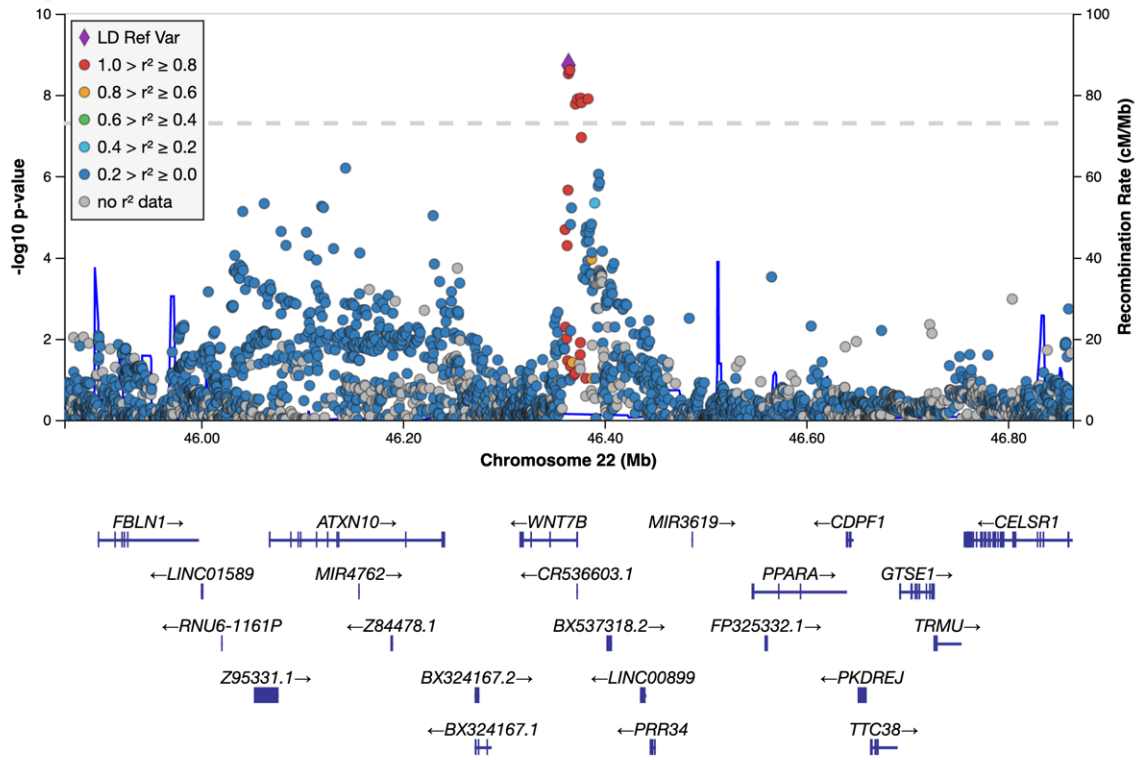
**22q13.2, rs34256596**



## LD reference population: EUR

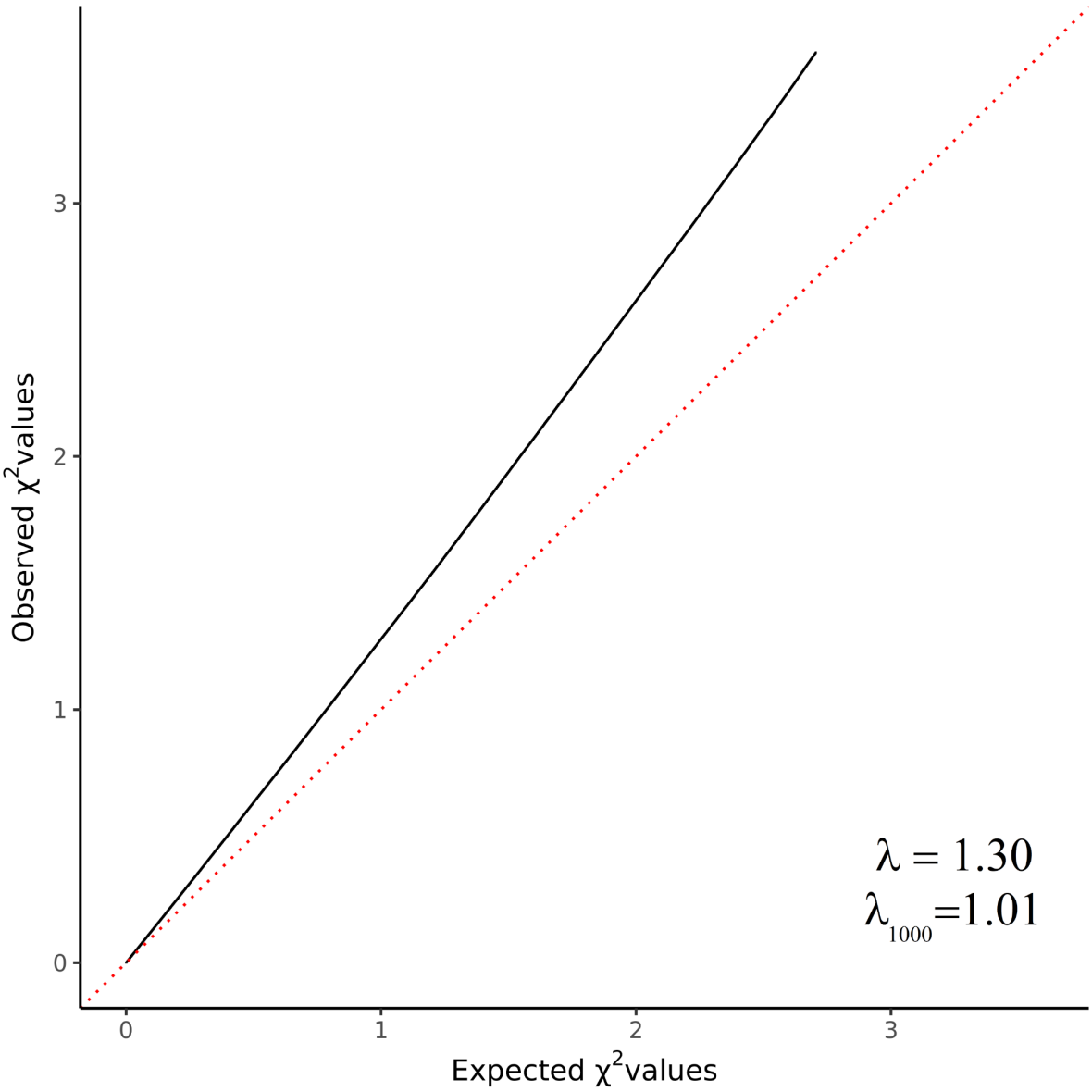



## LD reference population: EAS



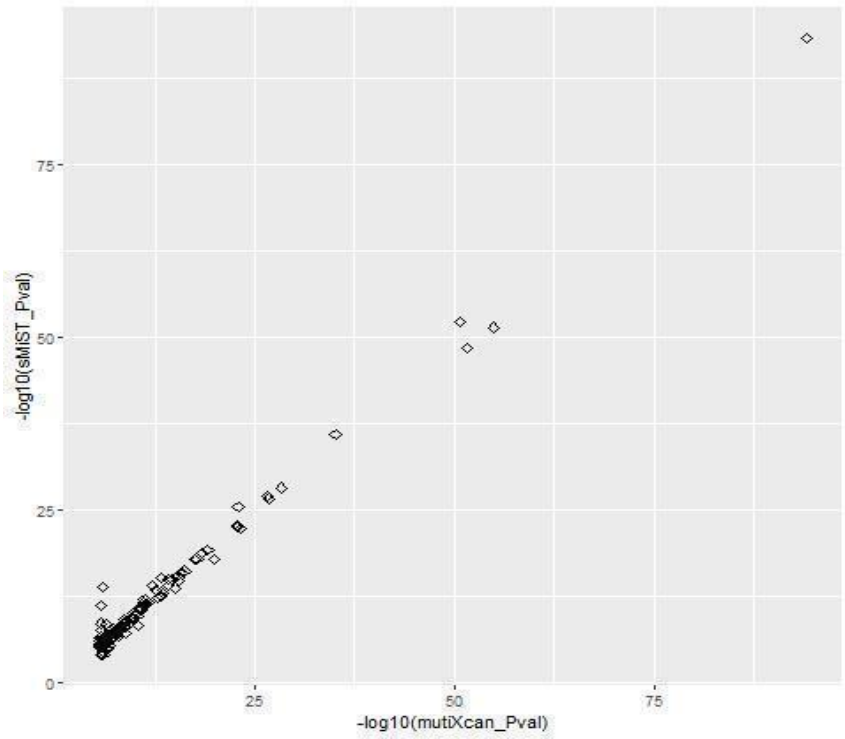

2

**Supplementary Figure 3: Quantile-Quantile (QQ) plot of observed and expected $\chi^2$ values of association between SNP genotype and colorectal cancer.** Meta $\lambda_{GC}$ = 1.30, $\lambda_{1000}$ = 1.01. The red line represents the null hypothesis of no true association.

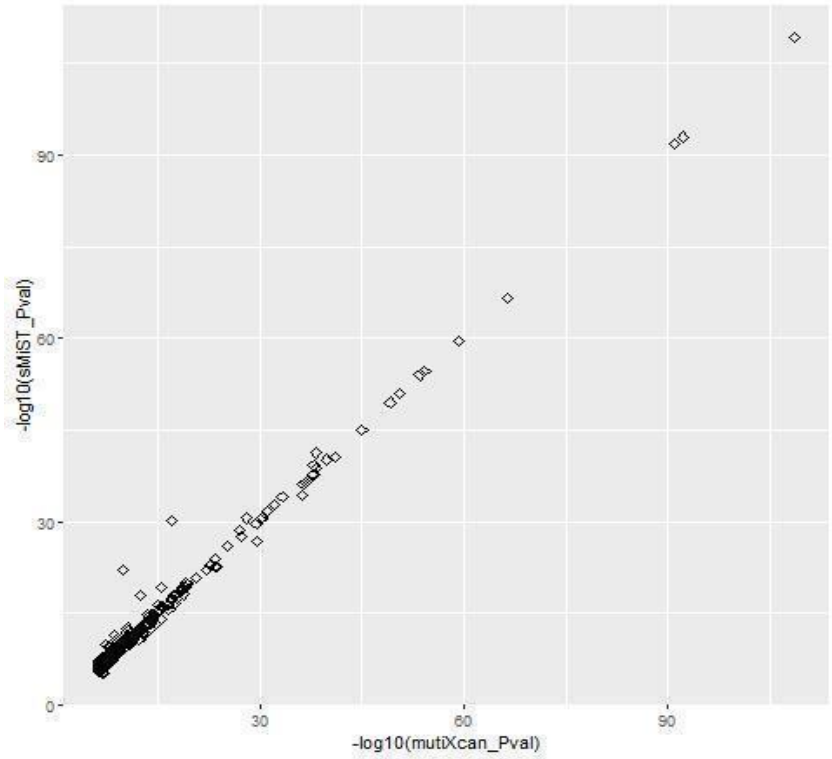


**Supplementary Figure 4. P-values from sMultiXcan and unconditional sMiST analysis of the sPrediXcan models for the 12,017 TWAS genes (A) and 88,888 MWAS CpGs (B).** All p-values are two-sided.

(A)



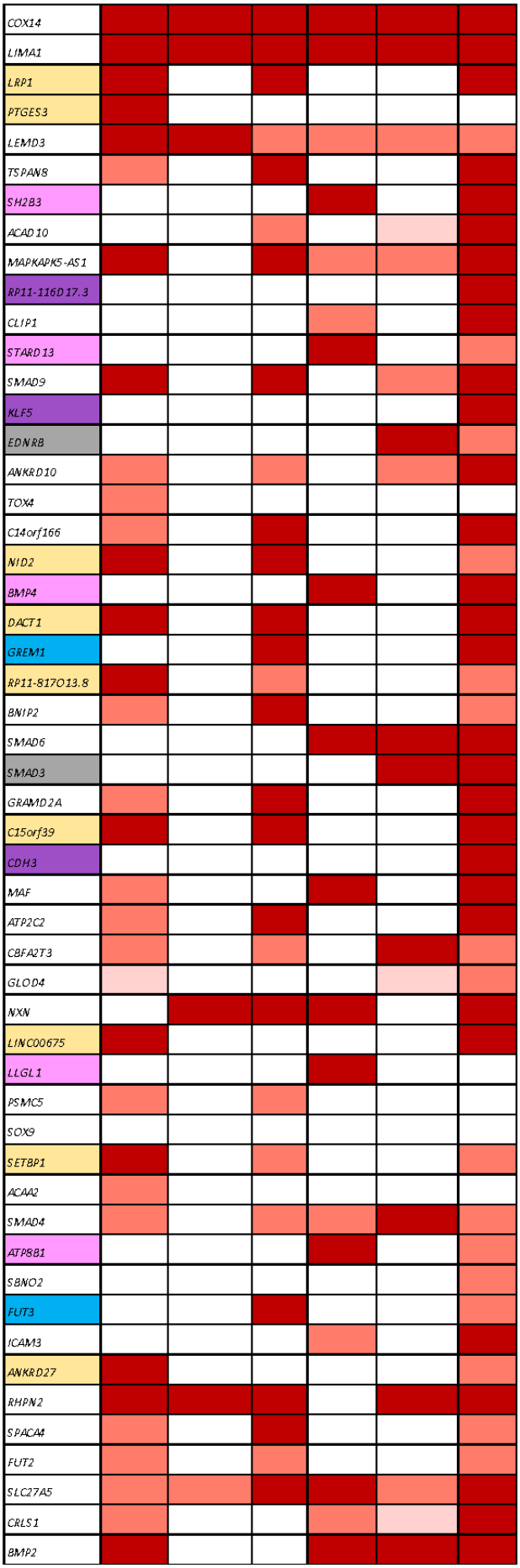

(B)

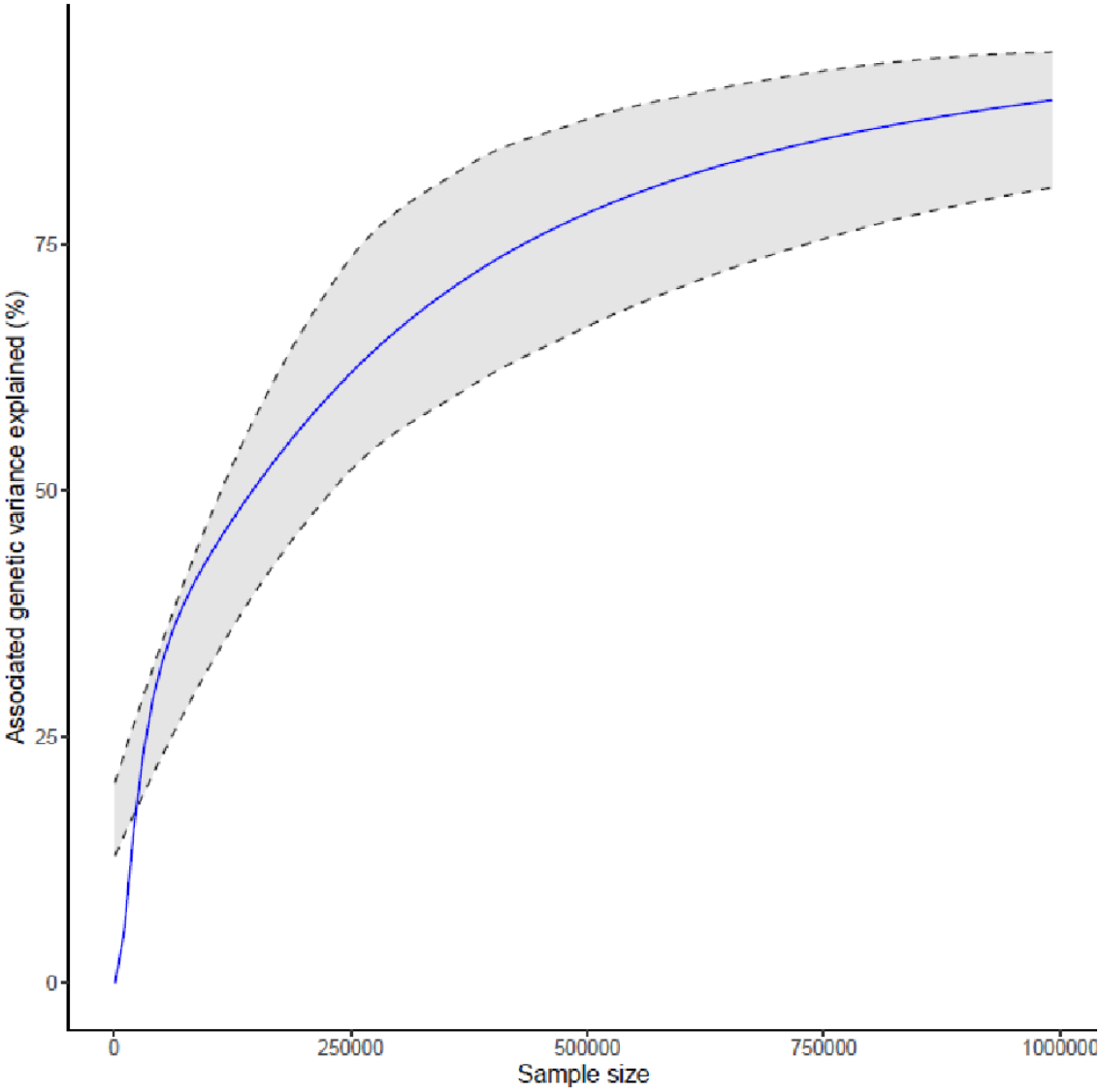


**Supplementary Figure 5. Association between effector genes and CRC risk from TWAS across tissue types.** Shading represents strength of association in each tissue in TWAS (colorectal mucosa (yellow), sigmoid colon (orange), pan-gastrointestinal "GI" (blue), immune cells (pink), mesenchyme (gray) and pan-tissue/"All" (purple). Red=Bonferroni; pink=FDR; none=no significant gene expression model or TWAs association found. The MHC region is excluded from this list, with likely disproportional effects on immune genes. We identified likely tissue-specific effects on CRC risk based on Bonferroni-significant associations in one tissue (colorectal mucosa, sigmoid colon, immune cells or mesenchyme), with no association at FDR in any of the other tissue types (including "GI" but excluding the combined "All" analysis). Gene *RREB1* was removed from the cross-tissue analysis, since it harbors a coding variant and no significant TWAS models were available for any of the tissues. Genes in yellow are associated with mucosal tissue and those in pink are associated with immune tissues.

| Single gene list | CR mucosa | Colon sigmoid | GI | Immune | Mesenchyme | All |
|---|---|---|---|---|---|---|
| SPSB1 | | | | | | |
| ARHGEF19 | | | | | | |
| WNT4 | | | | | | |
| C1QB | | | | | | |
| FHL3 | | | | | | |
| TTC22 | | | | | | |
| RPL5 | | | | | | |
| ACP6 | | | | | | |
| LINGO4 | | | | | | |
| LAMC1 | | | | | | |
| ARPC5 | | | | | | |
| LMOD1 | | | | | | |
| DSTYK | | | | | | |
| DUSP10 | | | | | | |
| FAM98A | | | | | | |
| ACTR1B | | | | | | |
| FBLN7 | | | | | | |
| ARHGEF4 | | | | | | |
| TANC1 | | | | | | |
| STK39 | | | | | | |
| SATB2 | | | | | | |
| CSRNP1 | | | | | | |
| SFMBT1 | | | | | | |
| RFT1 | | | | | | |
| ATXN7 | | | | | | |
| LRIG1 | | | | | | |
| GBE1 | | | | | | |
| BOC | | | | | | |
| WDR52 | | | | | | |
| DIRC2 | | | | | | |
| RYK | | | | | | |
| ACTRT3 | | | | | | |
| SMARCAD1 | | | | | | |
| TET2 | | | | | | |
| UGT8 | | | | | | |
| GAB1 | | | | | | |
| SMAD1 | | | | | | |
| MAB21L2 | | | | | | |
| TERT | | | | | | |
| TTC33 | | | | | | |
| CDKN2AIPNL | | | | | | |
| TXNDC15 | | | | | | |
| ERGIC1 | | | | | | |
| FBXO38 | | | | | | |
| CDX1 | | | | | | |
| RREB1 | | | | | | |
| HIVEP1 | | | | | | |
| CDKAL1 | | | | | | |
| ZKSCAN4 | | | | | | |
| TRIM27 | | | | | | |
| TULP1 | | | | | | |

| | | | | | | |
|---|---|---|---|---|---|---|
| CDKN1A | | | | | | |
| TFEB | | | | | | |
| RP1-166H4.2 | | | | | | |
| BMP5 | | | | | | |
| DCBLD1 | | | | | | |
| EPB41L2 | | | | | | |
| TCF21 | | | | | | |
| GNA12 | | | | | | |
| TBRG4 | | | | | | |
| TNS3 | | | | | | |
| RP11-114G11.5 | | | | | | |
| WBSCR27 | | | | | | |
| CDK6 | | | | | | |
| TRIM4 | | | | | | |
| LINC00513 | | | | | | |
| TOX | | | | | | |
| UTP23 | | | | | | |
| POU5F18 | | | | | | |
| RP11-384P7.7 | | | | | | |
| DCAF12 | | | | | | |
| LPAR1 | | | | | | |
| BRD3 | | | | | | |
| ABCA2 | | | | | | |
| ITIH5 | | | | | | |
| BAMBI | | | | | | |
| GPRIN2 | | | | | | |
| A1CF | | | | | | |
| SFTPA2 | | | | | | |
| LINC01475 | | | | | | |
| CUTC | | | | | | |
| CNNM2 | | | | | | |
| TCF7L2 | | | | | | |
| IFITM1 | | | | | | |
| RHOG | | | | | | |
| F2 | | | | | | |
| KBTBD4 | | | | | | |
| FADS3 | | | | | | |
| MYRF | | | | | | |
| AP000439.5 | | | | | | |
| POLD3 | | | | | | |
| CHRDL2 | | | | | | |
| ME3 | | | | | | |
| TRPC6 | | | | | | |
| COLCA1 | | | | | | |
| COLCA2 | | | | | | |
| TAGLN | | | | | | |
| BCL9L | | | | | | |
| ADAMTS15 | | | | | | |
| CCND2 | | | | | | |
| PLEKHG6 | | | | | | |
| RP1-102E24.8 | | | | | | |
| LMBR1L | | | | | | |


COX14
LIMA1
LRP1
PTGES3
LEMD3
TSPAN8
SH2B3
ACAD10
MAPKAPK5-AS1
RP11-116D17.3
CLIP1
STARD13
SMAD9
KLF5
EDNRB
ANKRD10
TOX4
C14orf166
NID2
BMP4
DACT1
GREM1
RP11-817O13.8
BNIP2
SMAD6
SMAD3
GRAMD2A
C15orf39
CDH3
MAF
ATP2C2
CBFA2T3
GLOD4
NXN
LINC00675
LLGL1
PSMC5
SOX9
SETBP1
ACAA2
SMAD4
ATP8B1
SBNO2
FUT3
ICAM3
ANKRD27
RHPN2
SPACA4
FUT2
SLC27A5
CRLS1
BMP2

| | | | | | | |
|---|---|---|---|---|---|---|
| *TMX4* | | | | | | |
| *MMP24* | | | | | | |
| *JPH2* | | | | | | |
| *PREX1* | | | | | | |
| *RP11-112L6.3* | | | | | | |
| *PARD6B* | | | | | | |
| *GNAS* | | | | | | |
| *CABLES2* | | | | | | |
| *RBBP8NL* | | | | | | |
| *YBEY* | | | | | | |
| *PCNT* | | | | | | |
| *ZNRF3* | | | | | | |
| *LIF* | | | | | | |
| *PDGFB* | | | | | | |
| *RIBC2* | | | | | | |

**Supplementary Figure 6: Projected percentage of GWAS heritability explained for a given sample size.** Results were obtained using a three-component model to estimate distribution of effect sizes. Grey shaded area represents the 95% confidence interval of the heritability estimate. The sample size indicates the total number of cases and controls, assuming a 1:1 ratio.