

Genome-wide association analyses identify distinct genetic architectures for age-related macular degeneration across ancestries

In the format provided by the
authors and unedited

Supplementary Information

Contents

- Supplementary Note 1: Summary of AMD risk factors associated in multiple regression models (**Page 2**)
- Supplementary Note 2: Summary of HLA associations in MVP EA (**Page 2**)
- Supplementary Note 3: Transcriptomic imputation models used in the multi-tissue TWAS (**Pages 2-3**)
- Supplementary Note 4: Retinal transcriptomic imputation model (**pages 3-4**)
- Supplementary Discussion (**page 4**)
- Supplementary Note 5: Million Veteran Program: Consortium Acknowledgement for Manuscripts (**pages 5-8**)
- Supplementary Fig. 1. Penetrance of the AMD PRS in females only and in all subjects. (**page 9**)
- Supplementary Fig. 2. Regional association plots of novel loci from the European-ancestry meta-analysis. (**pages 10-14**)
- Supplementary Fig. 3. QQ plots of the multi-ancestry GWAS and meta-analysis. (**page 15**)
- Supplementary Fig. 4. Genome-wide significant loci from the AA and HA GWAS. (**page 16**)
- Supplementary Fig. 5. Novel genome-wide significant loci from the multi-ancestry meta-analysis. (**page 17**)
- Supplementary Fig. 6. Co-localization of multi-ancestry GWAS and eQTL fine-mapped signals identifies candidate causal variants. (**pages 18-19**)
- Supplementary Fig. 7. Regional Miami plots of genes with novel TWAS associations. (**pages 20-25**)
- Captions for supplementary tables (**pages 26-32**)
- References (**pages 33-34**)

Supplementary tables are provided separately in an Excel file.

Supplementary Note 1: Summary of AMD risk factors associated in multiple regression models

In a multiple regression model, smoking and history of heavy drinking (score of 8 or higher on at least one AUDIT-C questionnaire) were significant risk factors in the MVP EA cohort, and were positively associated but below significance in AA or HA. Elevated HDL cholesterol was significantly associated in the EA and AA cohorts, while BMI showed a modest negative association with AMD in the EA cohort; a standard deviation increase was associated with an odds ratio (OR) of 0.96 [0.95-0.98]; $p=2\times 10^{-6}$), even in the fully adjusted model, which was replicated with similar effect size (though not with statistical significance) in the AA and HA cohorts. Reduced serum albumin was significantly associated with AMD in all three cohorts, with similar effect sizes.

Supplementary Note 2: Summary of HLA associations in MVP EA

Within MVP EA (32,567 cases and 130,444 controls), we replicated HLA allele associations¹ with DQB*02:01 ($p=1.2\times 10^{-7}$), DQB*02:02 ($p=4.7\times 10^{-10}$), and DRB*03:01 ($p=2.5\times 10^{-7}$); however, the DQB*02:01 and DRB*03:01 alleles were no longer significant in the fully adjusted model, especially after adjustment for the nearby *CFB/C2* SNP rs429608. Eight alleles were significant after Bonferroni correction, with four novel associations reaching genome-wide significance (A*11:01, B*14:01, DRB1*07:01, and DQA1*02:01).

Supplementary Note 3: Transcriptomic imputation models used in the multi-tissue TWAS

We used the following transcriptomic imputation models in the multi-tissue meta-analysis:

- Adipose tissue: subcutaneous (GTEx & STARNET) and visceral (GTEx & STARNET)
- Adrenal gland (GTEx)
- Arterial tissue: aorta (GTEx & STARNET)
- Blood (GTEx & STARNET)
- Brain (GTEx): anterior cingulate cortex, frontal pole cortex, hippocampus, and substantia nigra
- Brain (PsychENCODE) (genes and transcripts)
- Colon (GTEx): sigmoid and transverse
- Coronary artery (GTEx)
- Esophagus (GTEx): gastroesophageal junction, mucosa and muscularis
- Heart (GTEx): atrial appendage and left ventricle
- Liver (GTEx & STARNET)

- Lung (GTEx)
- Mammary artery (STARNET)
- Mammary tissue (GTEx)
- Ovary (GTEx)
- Pancreas (GTEx)
- Retina (EyeGTEx)
- Skeletal muscle (GTEx & STARNET)
- Skin (GTEx): not sun-exposed and sun-exposed
- Spleen (GTEx)
- Stomach (GTEx)
- Terminal ileum (GTEx)

All models were derived with EpiXcan, except for retina (PrediXcan).

Supplementary Note 4: Retinal transcriptomic imputation model

Genotypes and gene expression data were subjected to additional quality control steps as described below. For population classification we used individuals of known ancestry from 1000 Genomes. We excluded variants in regions of high linkage disequilibrium, variants with minor allele frequency < 0.05 , variant with high missingness (> 0.01), and variants with Hardy-Weinberg equilibrium $p < 1 \times 10^{-10}$; the remaining variants were pruned (`--indep-pairwise 1000 10 0.02` with PLINK²) and PCA was performed with PLINK³ version 2.0. We used the first (PC1), second (PC2) and third (PC3) ancestral principal components to define an ellipsoid based on 1000Gp3v5 EUR samples⁴ and samples within 3 SD from the ellipsoid center were classified as EUR; based on this definition of EUR samples, we excluded one non-European ancestry EA individual. In the remaining samples ($n = 405$), we performed additional sample-level quality control by retaining non-related samples (`--king-cutoff 0.0884` with PLINK³ version 2.0) with sample-level missingness < 0.015 for variants with variant-level missingness < 0.02 , and heterozygosity rate of $< 3SD$ away from the mean; of note, no samples were excluded by these steps. For the next step of our pipeline, we performed outlier testing in the gene expression data. After performing counts per million filtering (> 0.5 counts per million in at least 30% of samples) and voom normalization, PCA was performed, and we excluded individuals located more than 4 SD away from the mean of the ellipsoid defined by PC1 to PC3. This did not remove any individuals but assured us that our data did not contain any outliers. In this final set of individuals, we performed variant-level quality control of the genotypes by removing variants with less than 0.01 minor allele frequency (for all variants possible, we utilized minor allele frequencies reported by Allele Frequency Aggregator

European population⁵, to reduce minor allele frequency bias from the comparatively small imputation model training population), 5 minor allele counts and 0.02 missingness rate; only variants present in the reference panel of the Haplotype Reference Consortium were retained to ensure good representation of variants in the target GWAS⁶. We used this final set of quality-controlled genotypes in conjunction with our normalized expression data to discover the optimal number of PEER factors to find expression quantitative trait loci. Our analysis led to the decision to utilize 15 PEER factors, which had resulted in the discovery of 4,299 significant eQTLs⁷. This was the closest value to 90% of the maximum value of eQTLs discovered by any chosen number of PEER factors (4,844 significant eQTLs from 50 PEER factors). This allowed us to retain the maximum signal for gene expression prediction without overcorrecting our data. After residualization for 15 PEER factors, expression data were quantile normalized. Genotypes were then converted to dosages, and missing values were replaced with twice the variant's minor allele frequency before dosages were rounded to the nearest whole number.

Supplementary Discussion

A general conclusion from years of investigative efforts is that the genetic architectures at the broader *CFH* and *ARMS2/HTRA1* loci are multifaceted^{8,9}. At *CFH*, both common and rare variation¹⁰ as well as structural variation¹¹ plays a role. At *ARMS2/HTRA1*, broad linkage disequilibrium between variants has made fine-mapping and identification of causal mechanisms difficult. Current literature points to a tissue-specific cis-regulatory element within *ARMS2* that is disrupted by the risk allele, leading to downregulation of *HTRA1*¹². *HTRA1*, a serine protease with reported roles in TGF- β signaling¹³ and regulation of monocyte elimination¹⁴, has also been shown to cleave neutrophil elastase inhibitor α -1 antitrypsin¹⁵ (*SERPINA1*), producing NET inhibitory peptides, or NIPs^{16,17}. NETs activate complement^{18–20}, and *CFH* itself may both regulate NETosis²¹ and bind to NETs²², potentially connecting the biology of the two principal AMD risk loci. Supporting this pathway, we discovered that two low-frequency pathogenic mutations in *SERPINA1* that cause α -1 antitrypsin deficiency are associated with AMD, but with opposite directions of effect. Additionally, consistent with histological²³ and observational²⁴ evidence, our AMD PRS-PheWAS experiments in both EA and AA showed association with increased neutrophil counts. Thus, integration of epidemiologic and genetic data from European and non-European populations synergizes and unifies evidence and refines risk models.

Supplementary Note 5: Million Veteran Program: Consortium Acknowledgement for Manuscripts

MVP Executive Committee

- Co-Chair: J. Michael Gaziano, M.D., M.P.H.
- Co-Chair: Rachel Ramoni, D.M.D., Sc.D.
- Jim Breeling, M.D. (ex-officio)
- Kyong-Mi Chang, M.D.
- Grant Huang, Ph.D.
- Sumitra Muralidhar, Ph.D.
- Christopher J. O'Donnell, M.D., M.P.H.
- Philip S. Tsao, Ph.D.

MVP Program Office

- Sumitra Muralidhar, Ph.D.
- Jennifer Moser, Ph.D.

MVP Recruitment/Enrollment

- Recruitment/Enrollment Director/Deputy Director, Boston
- Stacey B. Whitbourne, Ph.D.; Jessica V. Brewer, M.P.H.
- MVP Coordinating Centers
 - o Clinical Epidemiology Research Center (CERC), West Haven – John Concato, M.D., M.P.H.
 - o Cooperative Studies Program Clinical Research Pharmacy Coordinating Center, Albuquerque - Stuart Warren, J.D., Pharm D.; Dean P. Argyres, M.S.
 - o Genomics Coordinating Center, Palo Alto – Philip S. Tsao, Ph.D.
 - o Massachusetts Veterans Epidemiology Research Information Center (MAVERIC), Boston - J. Michael Gaziano, M.D., M.P.H.
 - o MVP Information Center, Canandaigua – Brady Stephens, M.S.
- Core Biorepository, Boston – Mary T. Brophy M.D., M.P.H.; Donald E. Humphries, Ph.D.
- MVP Informatics, Boston – Nhan Do, M.D.; Shahpoor Shayan
- Data Operations/Analytics, Boston – Xuan-Mai T. Nguyen, Ph.D.

MVP Science

- Genomics - Christopher J. O'Donnell, M.D., M.P.H.; Saiju Pyarajan Ph.D.; Philip S.

Tsao, Ph.D.

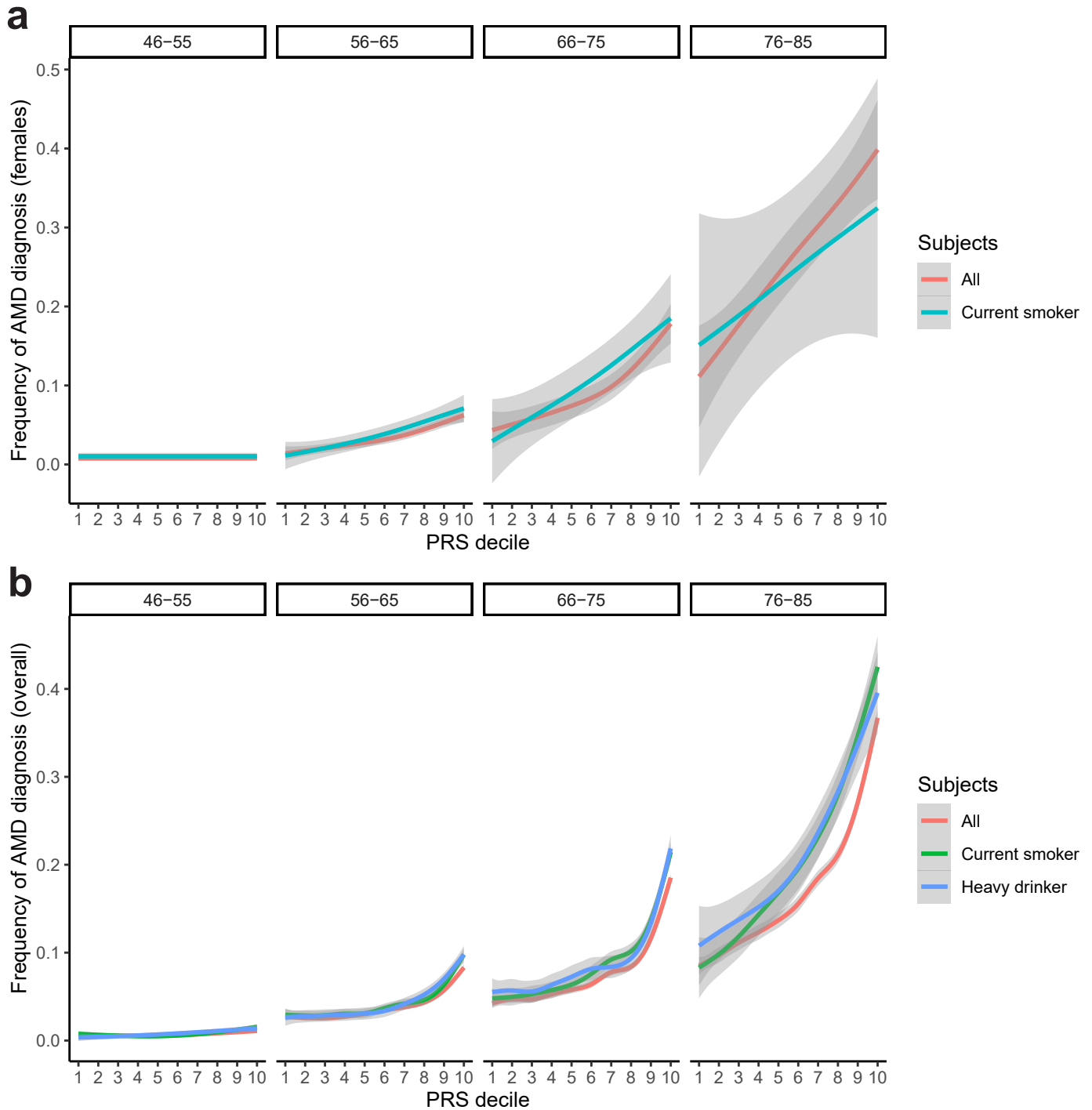
- Phenomics - Kelly Cho, M.P.H, Ph.D.
- Data and Computational Sciences – Saiju Pyarajan, Ph.D.
- Statistical Genetics – Elizabeth Hauser, Ph.D.; Yan Sun, Ph.D.; Hongyu Zhao, Ph.D.

MVP Local Site Investigators

- Atlanta VA Medical Center (Peter Wilson) - Bay Pines VA Healthcare System (Rachel McArdle)
- Birmingham VA Medical Center (Louis Dellitalia)
- Cincinnati VA Medical Center (John Harley)
- Clement J. Zablocki VA Medical Center (Jeffrey Whittle)
- Durham VA Medical Center (Jean Beckham)
- Edith Nourse Rogers Memorial Veterans Hospital (John Wells)
- Edward Hines, Jr. VA Medical Center (Salvador Gutierrez)
- Fayetteville VA Medical Center (Gretchen Gibson)
- VA Health Care Upstate New York (Laurence Kaminsky)
- New Mexico VA Health Care System (Gerardo Villareal)
- VA Boston Healthcare System (Scott Kinlay)
- VA Western New York Healthcare System (Junzhe Xu)
- Ralph H. Johnson VA Medical Center (Mark Hamner)
- Wm. Jennings Bryan Dorn VA Medical Center (Kathlyn Sue Haddock)
- VA North Texas Health Care System (Sujata Bhushan)
- Hampton VA Medical Center (Pran Iruvanti)
- Hunter Holmes McGuire VA Medical Center (Michael Godschalk)
- Iowa City VA Health Care System (Zuhair Ballas)
- Jack C. Montgomery VA Medical Center (Malcolm Buford)
- James A. Haley Veterans' Hospital (Stephen Mastorides)
- Louisville VA Medical Center (Jon Klein)
- Louis Stokes Cleveland VA Medical Center (Frank Jacono)
- Manchester VA Medical Center (Nora Ratcliffe)
- Miami VA Health Care System (Hermes Florez)

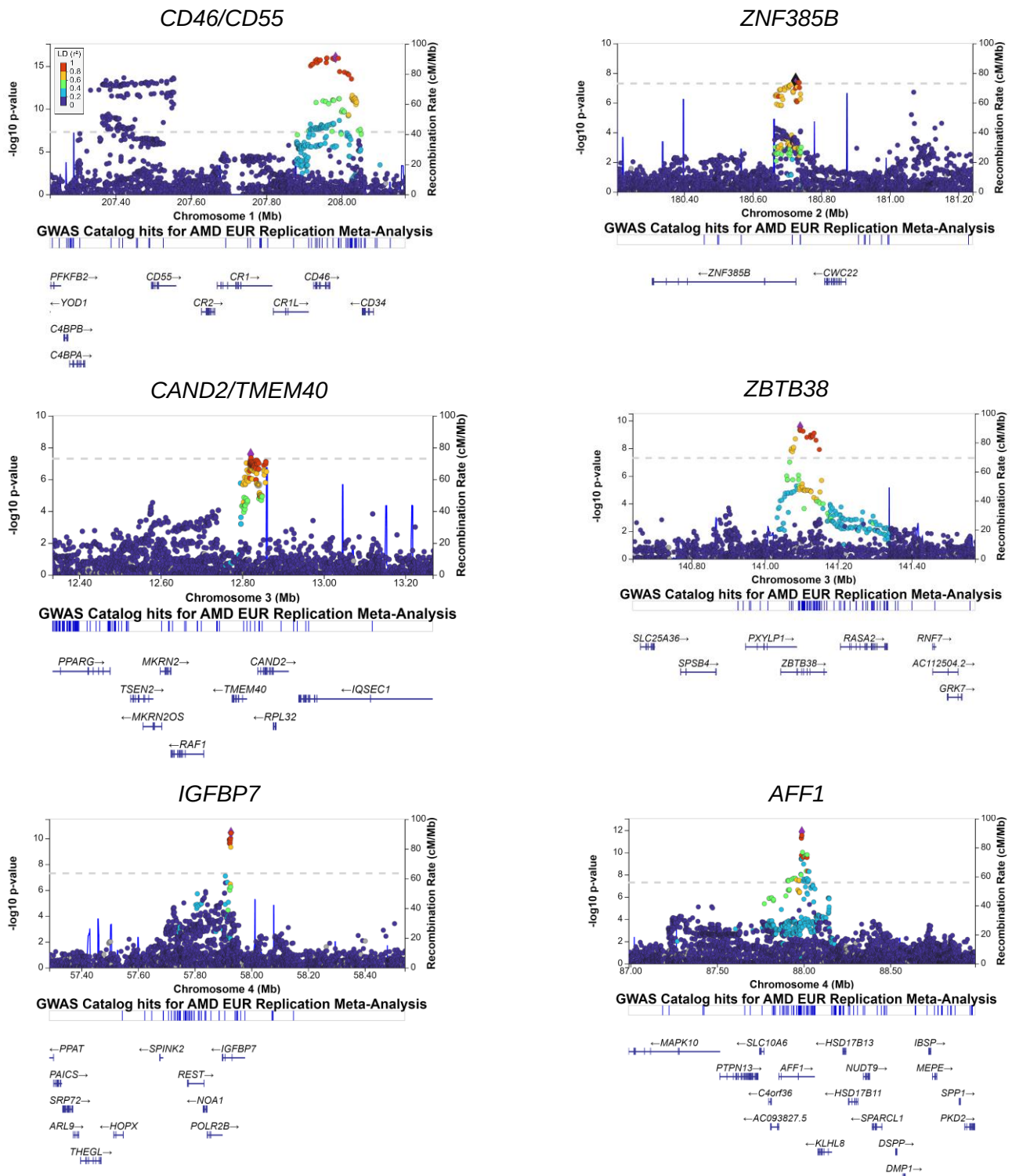
- Michael E. DeBakey VA Medical Center (Alan Swann)
- Minneapolis VA Health Care System (Maureen Murdoch)
- N. FL/S. GA Veterans Health System (Peruvemba Sriram)
- Northport VA Medical Center (Shing Shing Yeh)
- Overton Brooks VA Medical Center (Ronald Washburn)
- Philadelphia VA Medical Center (Darshana Jhala)
- Phoenix VA Health Care System (Samuel Aguayo)
- Portland VA Medical Center (David Cohen)
- Providence VA Medical Center (Satish Sharma)
- Richard Roudebush VA Medical Center (John Callaghan)
- Salem VA Medical Center (Kris Ann Oursler)
- San Francisco VA Health Care System (Mary Whooley)
- South Texas Veterans Health Care System (Sunil Ahuja)
- Southeast Louisiana Veterans Health Care System (Amparo Gutierrez)
- Southern Arizona VA Health Care System (Ronald Schiffman)
- Sioux Falls VA Health Care System (Jennifer Greco)
- St. Louis VA Health Care System (Michael Rauchman)
- Syracuse VA Medical Center (Richard Servatius)
- VA Eastern Kansas Health Care System (Mary Oehlert)
- VA Greater Los Angeles Health Care System (Agnes Wallbom)
- VA Loma Linda Healthcare System (Ronald Fernando)
- VA Long Beach Healthcare System (Timothy Morgan)
- VA Maine Healthcare System (Todd Stapley)
- VA New York Harbor Healthcare System (Scott Sherman)
- VA Pacific Islands Health Care System (Gwenevere Anderson)
- VA Palo Alto Health Care System (Philip Tsao)
- VA Pittsburgh Health Care System (Elif Sonel)
- VA Puget Sound Health Care System (Edward Boyko)
- VA Salt Lake City Health Care System (Laurence Meyer)
- VA San Diego Healthcare System (Samir Gupta)
- VA Southern Nevada Healthcare System (Joseph Fayad)

- VA Tennessee Valley Healthcare System (Adriana Hung)
- Washington DC VA Medical Center (Jack Lichy)
- W.G. (Bill) Hefner VA Medical Center (Robin Hurley)
- White River Junction VA Medical Center (Brooks Robey)
- William S. Middleton Memorial Veterans Hospital (Robert Striker)

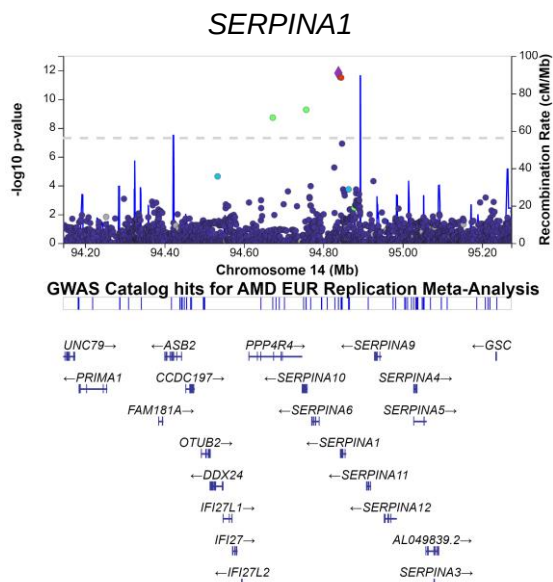
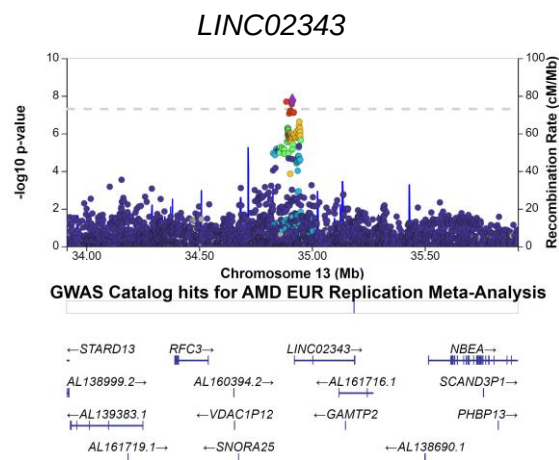
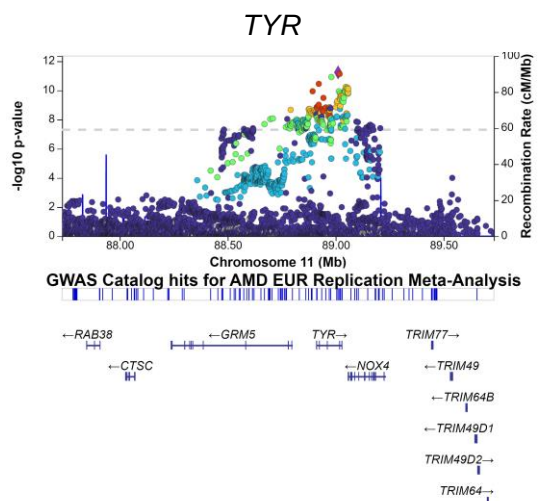
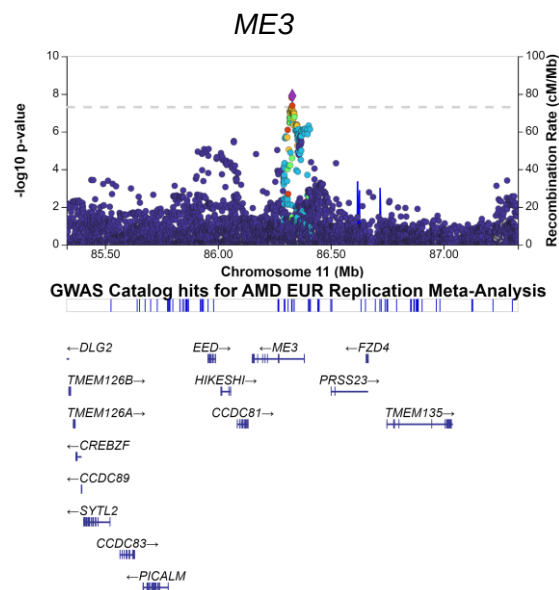
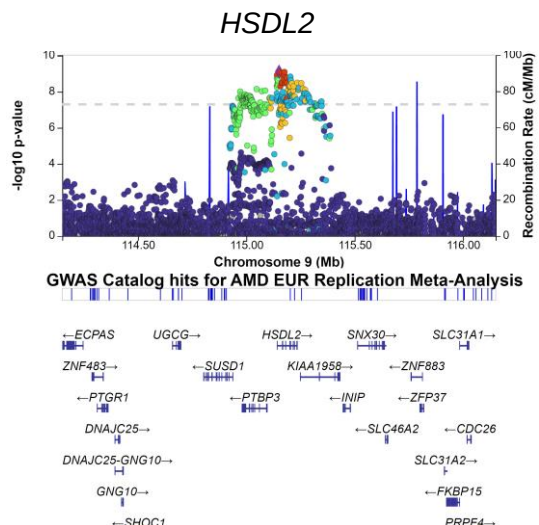
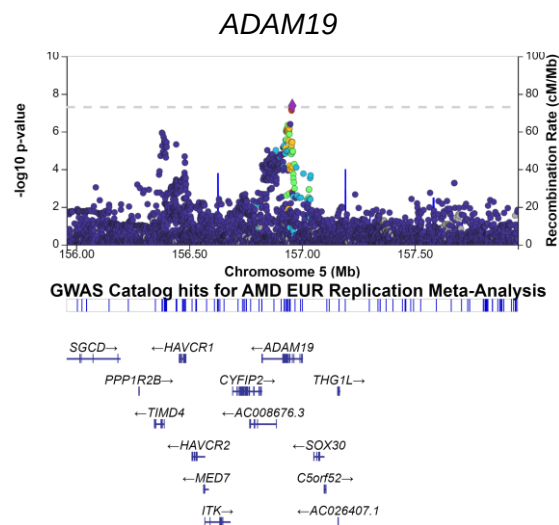


Supplementary Fig. 1. Penetrance of the AMD PRS in females only and in all subjects. (a) Prevalence of AMD in EA females in MVP as a function of PRS decile, age range, and lifestyle risk factors (N=1,002 cases; 25,198 controls). Data are presented as mean values with 95% c.i. error band. There were insufficient numbers of females with a history of heavy drinking to obtain reliable estimates. **(b)** Prevalence of AMD in both males and females in MVP as a function of PRS decile, age range, and lifestyle risk factors (N=29,728 cases; 413,463 controls). Data are presented as mean values $\pm$ 95% c.i. error band. The corresponding plot for EA males is shown in **Fig. 2a**.

Supplementary Fig. 2. Regional association plots of novel loci from the European-ancestry replication meta-analysis (57,290 cases, 324,430 controls). The 1000 Genomes EUR panel was used as the reference for plotting LD r^2 .

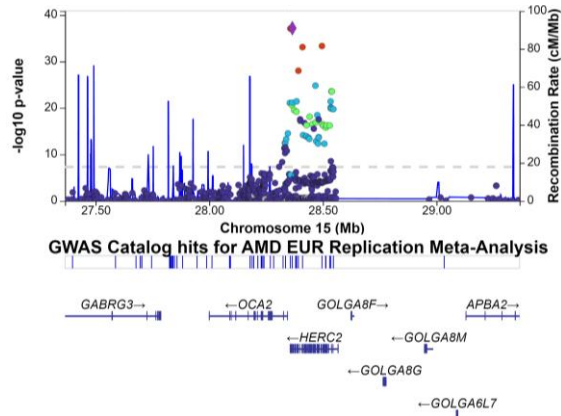


Supplementary Fig. 2

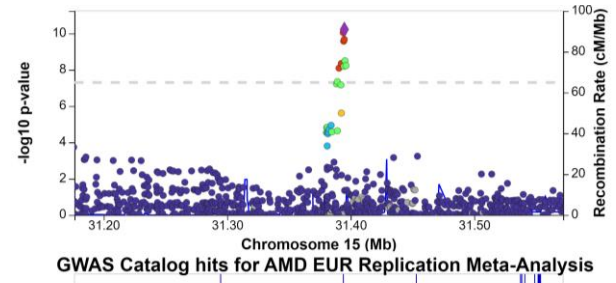


Supplementary Fig. 2

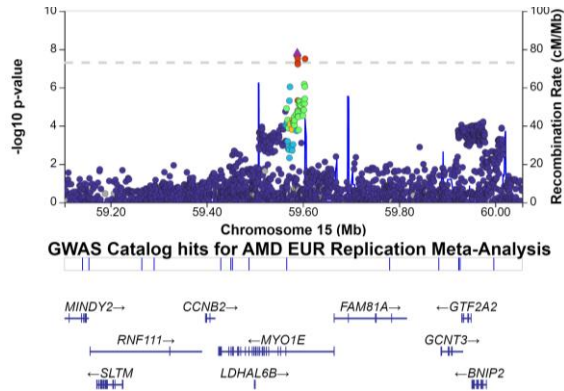
HERC2/OCA2



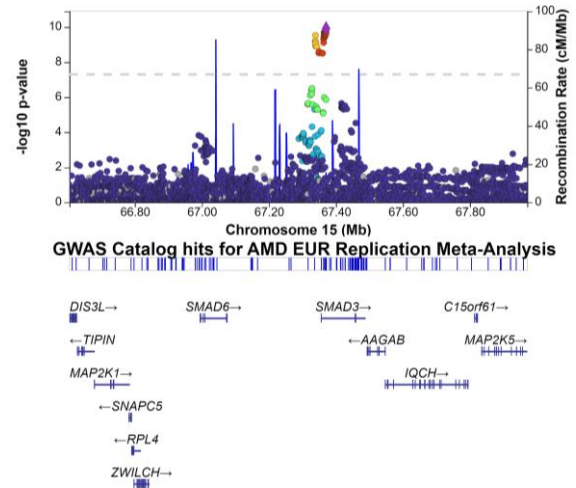
TRMP1



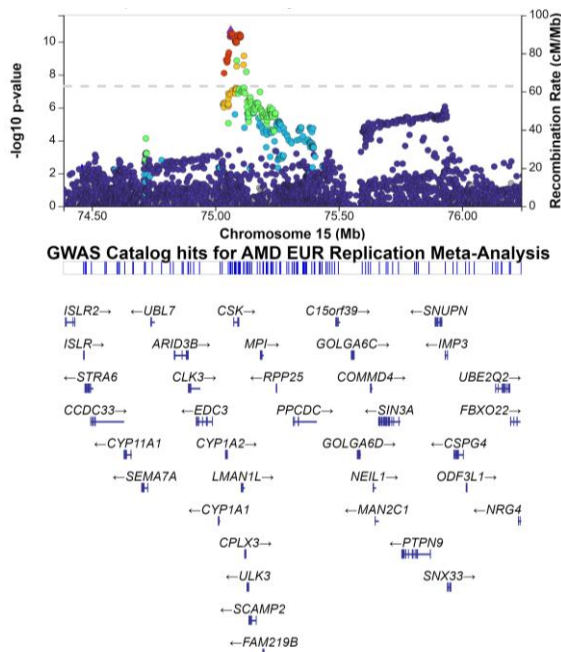
MYO1E



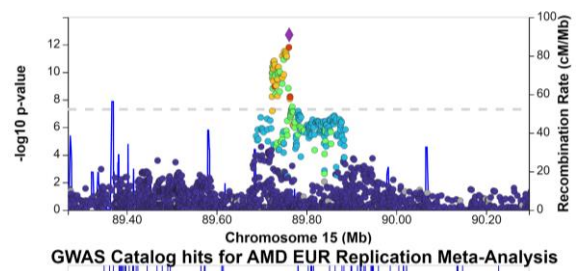
SMAD3



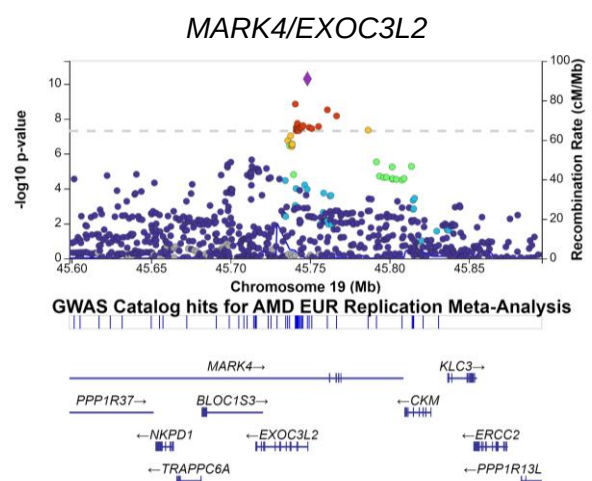
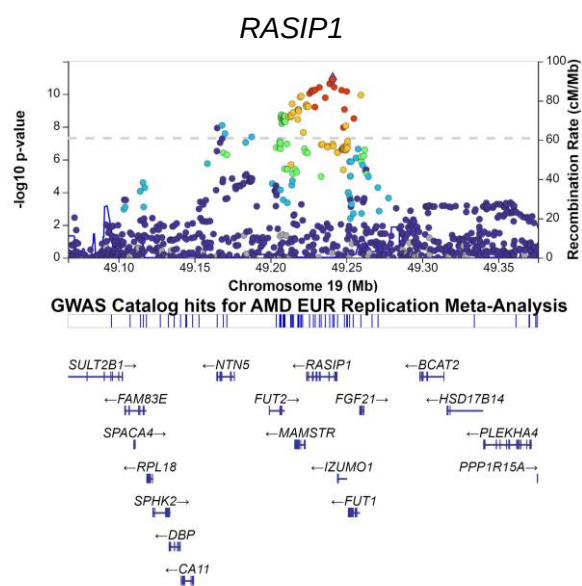
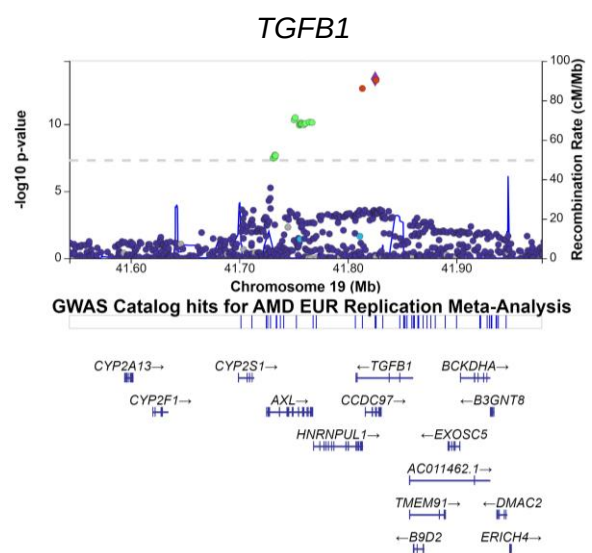
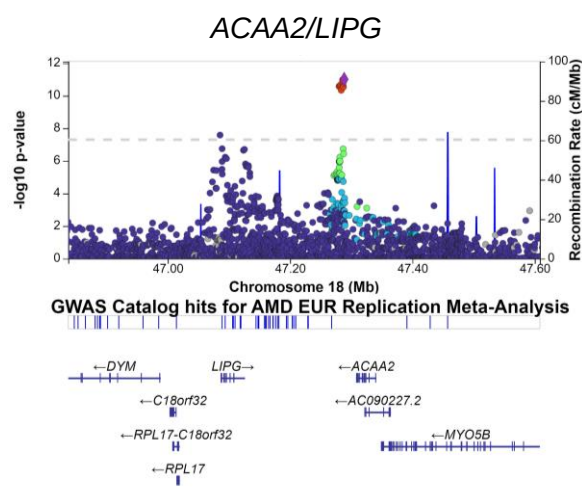
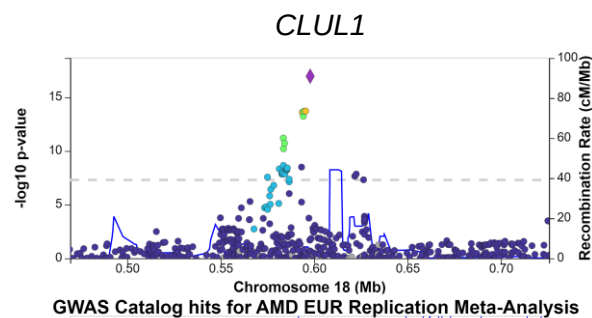
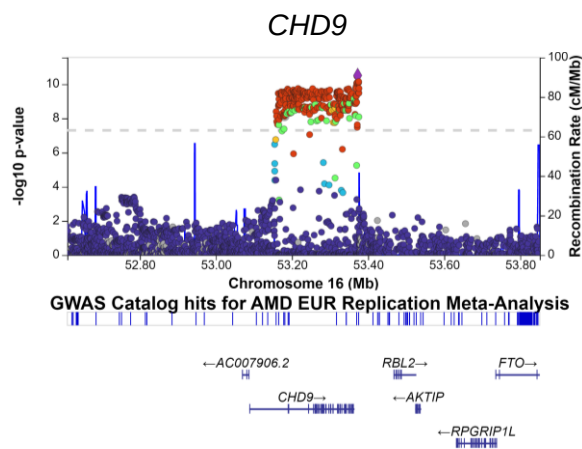
CSK



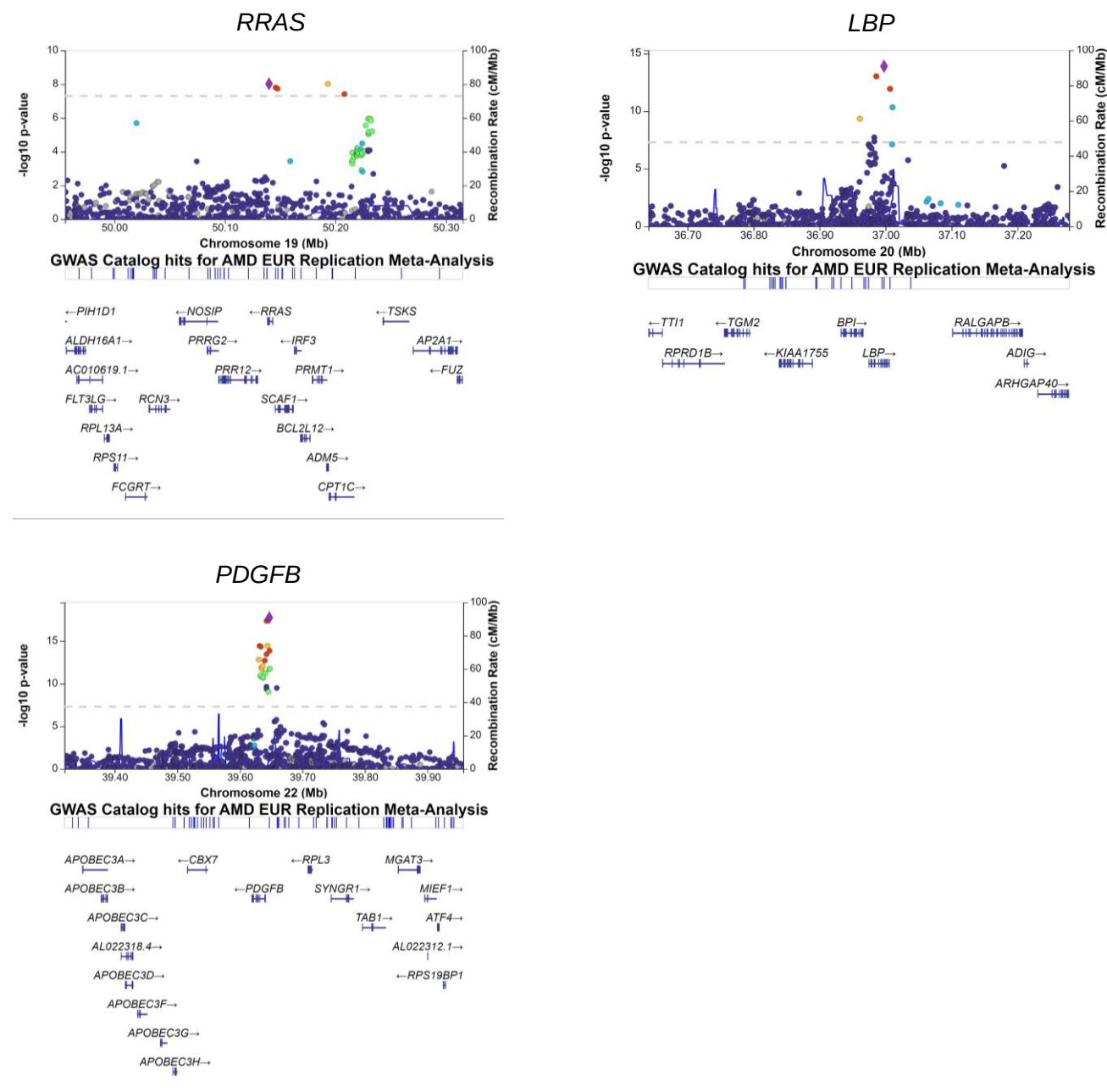
RLBP1

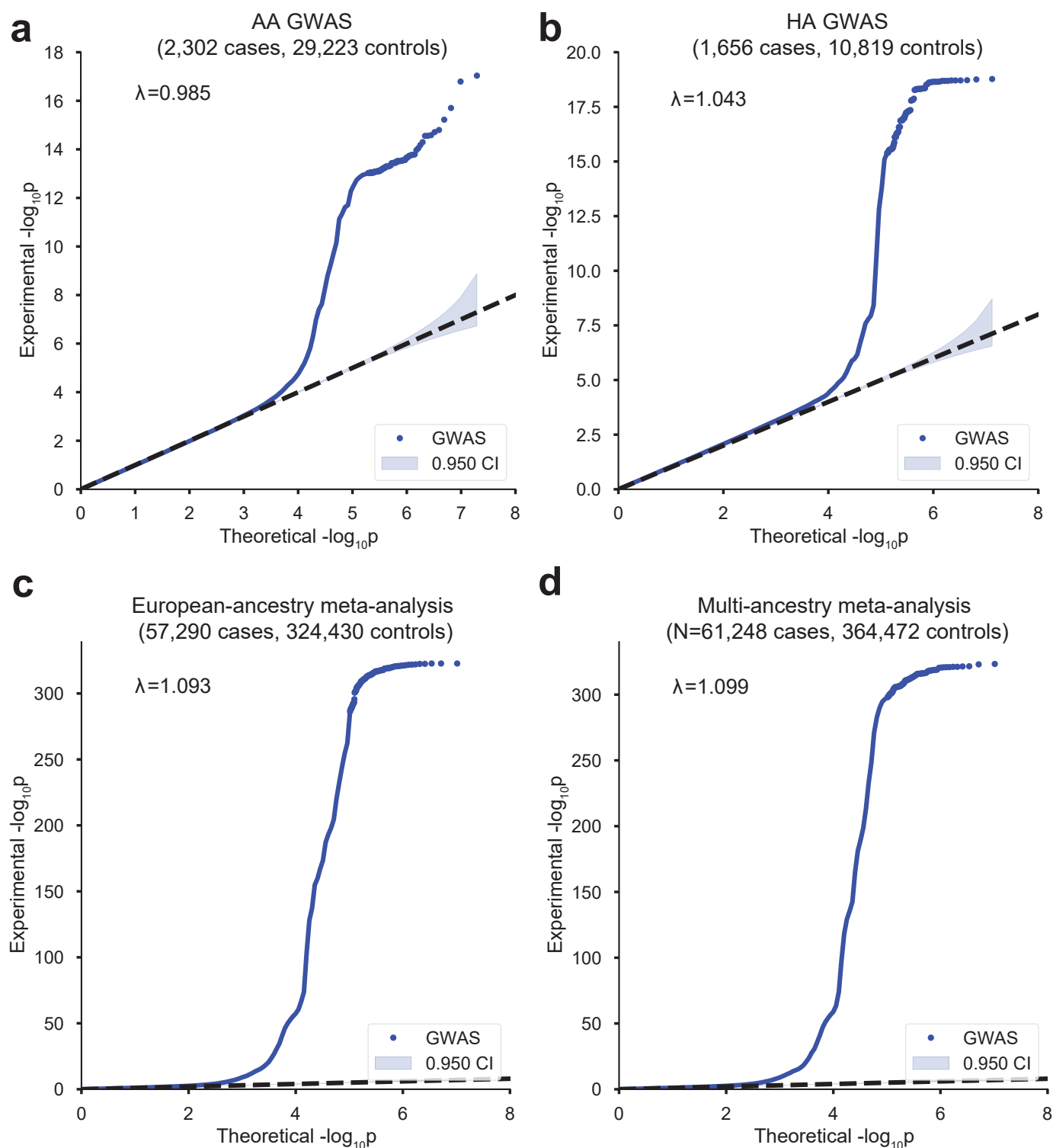


Supplementary Fig. 2



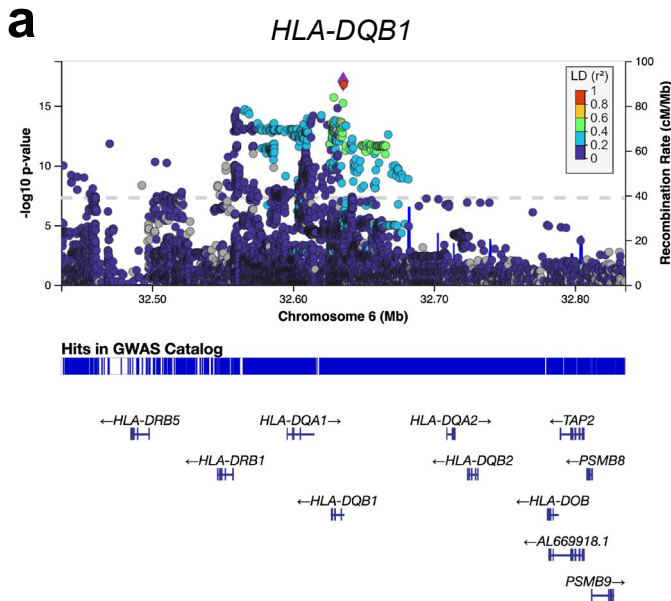
Supplementary Fig. 2



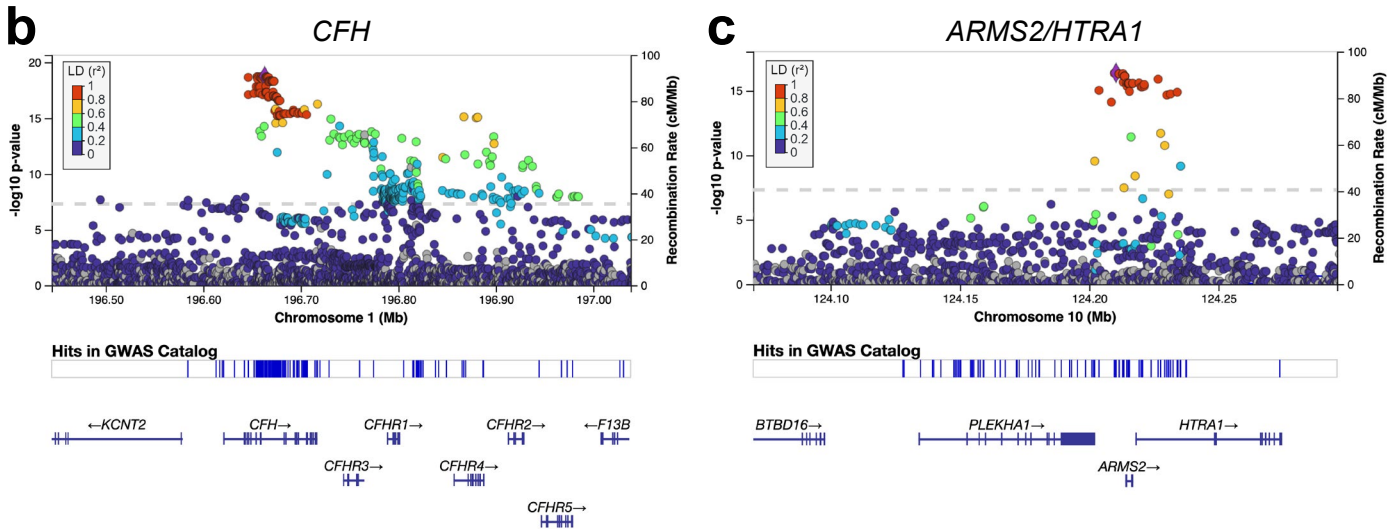


Supplementary Fig 3. QQ plots of the multi-ancestry GWAS and meta-analysis. (a) MVP AA GWAS, (b) MVP HA GWAS, (c) EA meta-analysis, and (d) multi-ancestry meta-analysis. SNPs with minor allele frequency greater than 0.5% are plotted and included in the genomic control inflation calculation. Error bands represent 95% confidence intervals around the theoretical distribution under the null.

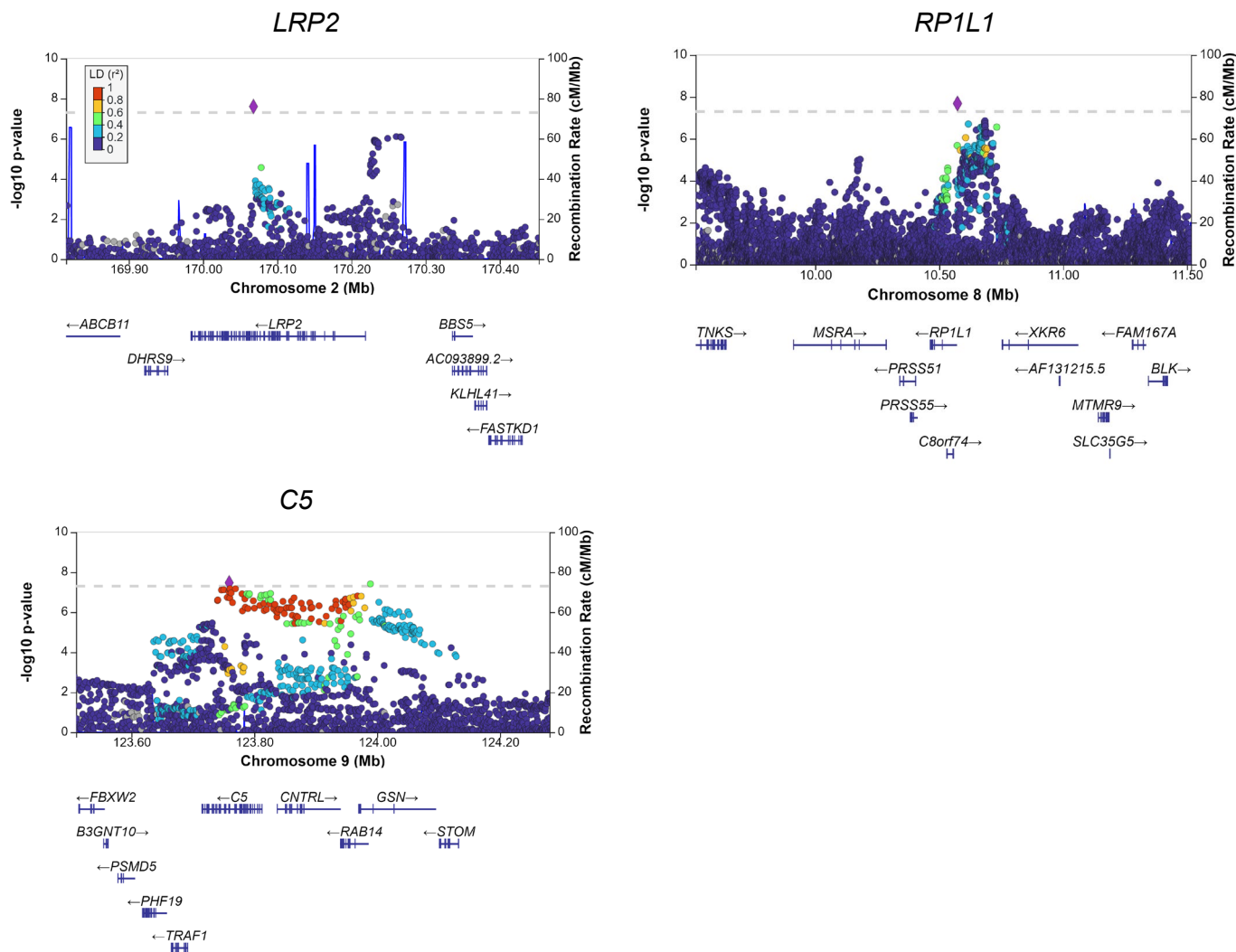
African-ancestry (AA) GWAS



Hispanic/Latino-ancestry (HA) GWAS

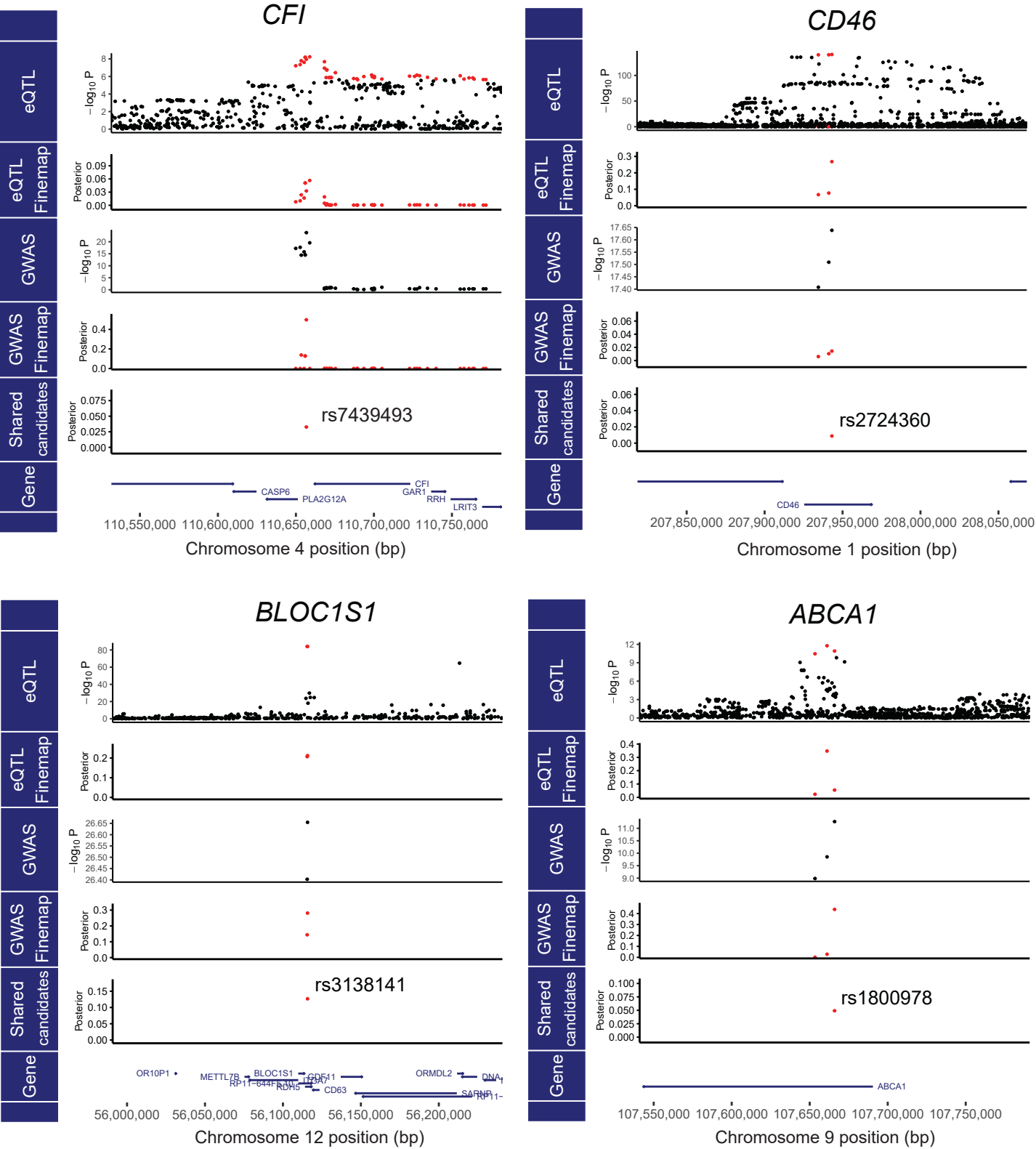


Supplementary Fig. 4. Genome-wide significant loci from the AA and HA GWAS in MVP. (a) Regional association plot of the *HLA-DQB1* locus from the AA GWAS (2,302 cases, 29,223 controls). The 1000 Genomes AFR panel was used as the reference for plotting LD r^2 . (b-c) Regional association plots of the *CFH* and *ARMS2-HTRA1* loci from the HA GWAS (1,656 cases, 10,819 controls). The 1000 Genomes AMR panel was used as the reference for plotting LD r^2 .



Supplementary Fig. 5. Novel genome-wide significant loci from the multi-ancestry meta-analysis. Regional association plots of novel loci from the multi-ancestry meta-analysis (61,258 cases, 364,472 controls) beyond that reported for the EA meta-analysis. The 1000 Genomes EUR panel was used as the reference for plotting LD r^2 .

Supplementary Fig. 6. Co-localization of multi-ancestry GWAS and eQTL fine-mapped signals identifies candidate causal variants. Shown are significant co-localized signals (co-localization posterior probability > 0.01) from the multi-ancestry discovery meta-analysis (61,258 cases, 364,472 controls) using the mmQTL method.



RDH5

eQTL

eQTL Finemap

GWAS

GWAS Finemap

Shared candidates

Gene

rs3138142

Chromosome 12 position (bp)

PMS2P1

eQTL

eQTL Finemap

GWAS

GWAS Finemap

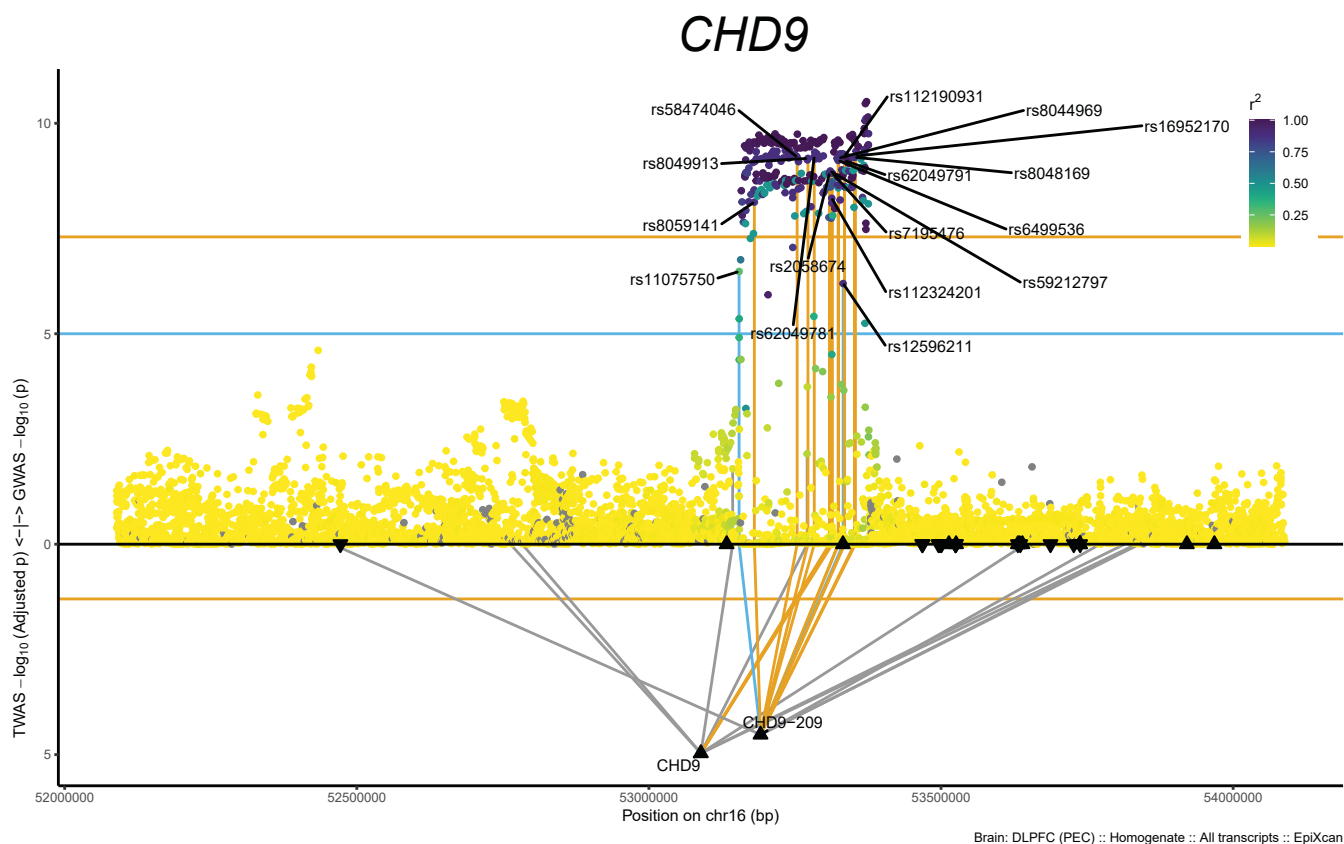
Shared candidates

Gene

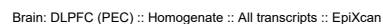
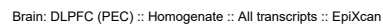
rs2108269

Chromosome 7 position (bp)

Supplementary Fig. 7. Regional Miami plots of genes with novel TWAS associations. The TWAS was based on the DLPFC brain model jointly testing genes and transcript isoforms. Results are based on the EA meta-analysis (N=57,290 cases, 324,430 controls). The GWAS p-values are shown above the axis, and the TWAS p-values are shown below it. SNPs are colored according to linkage disequilibrium r^2 with the lead SNP (EA panel). Lines connecting SNP predictors of genetically regulated gene expression are colored according to the significance level of the SNP: orange, $p < 5 \times 10^{-8}$; cyan, $5 \times 10^{-8} \leq p < 1 \times 10^{-5}$; gray, $p \geq 1 \times 10^{-5}$.

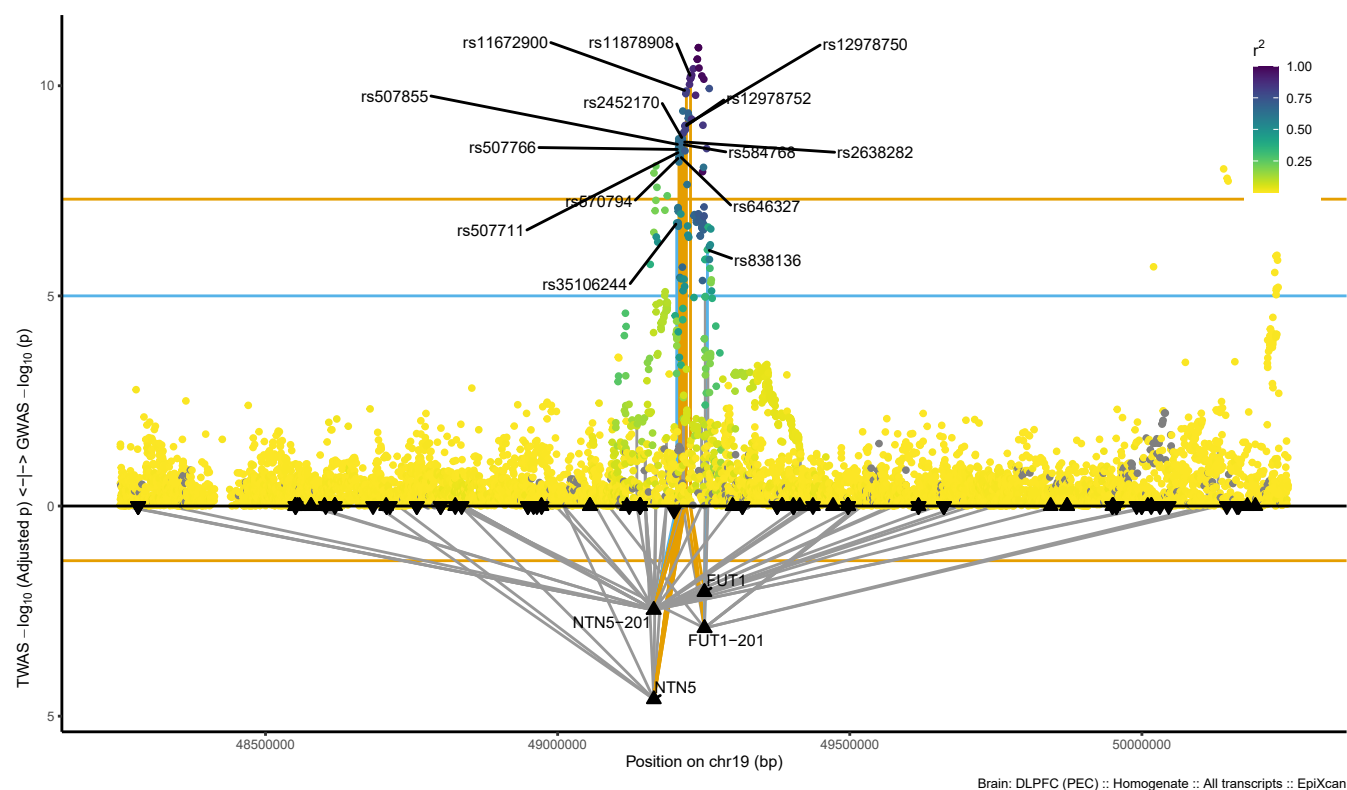


MAN2C1-201

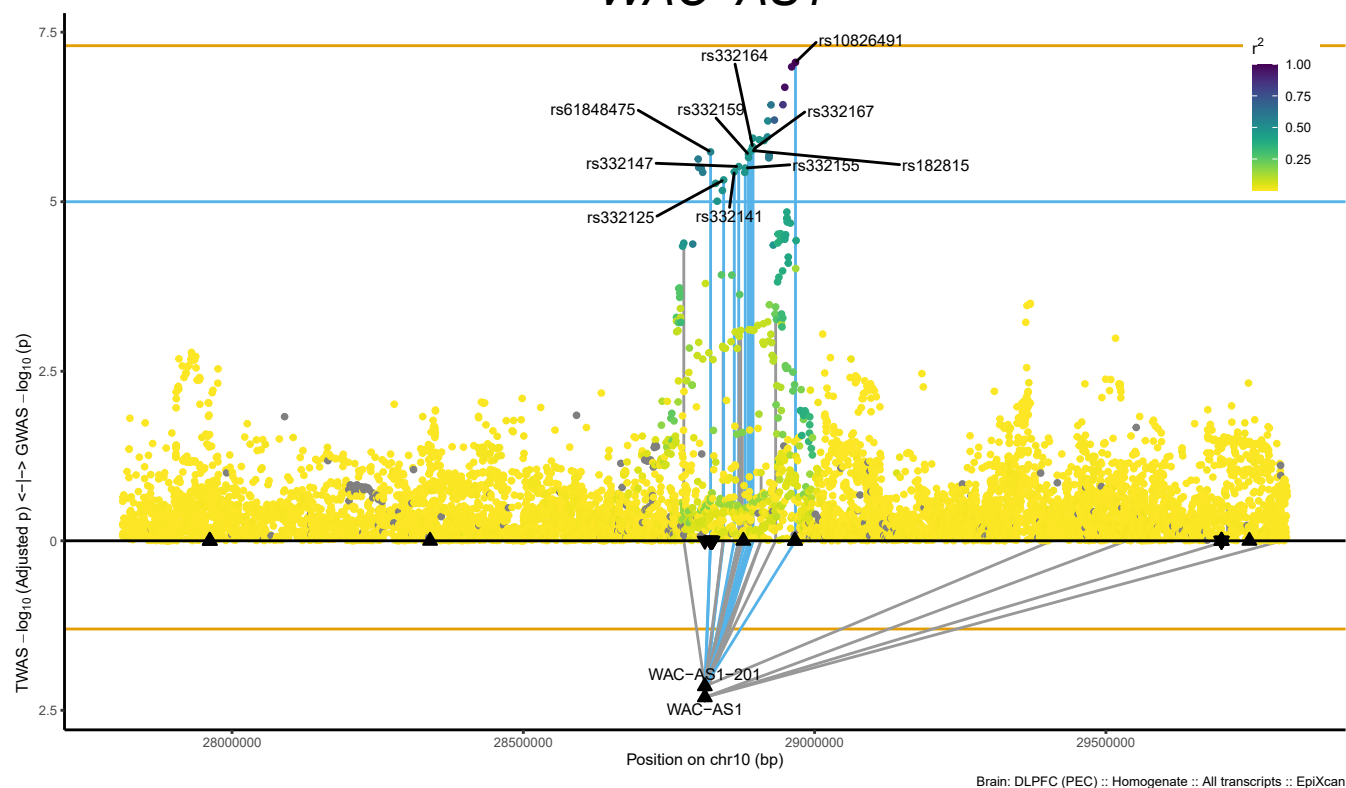


Supplementary Fig. 7

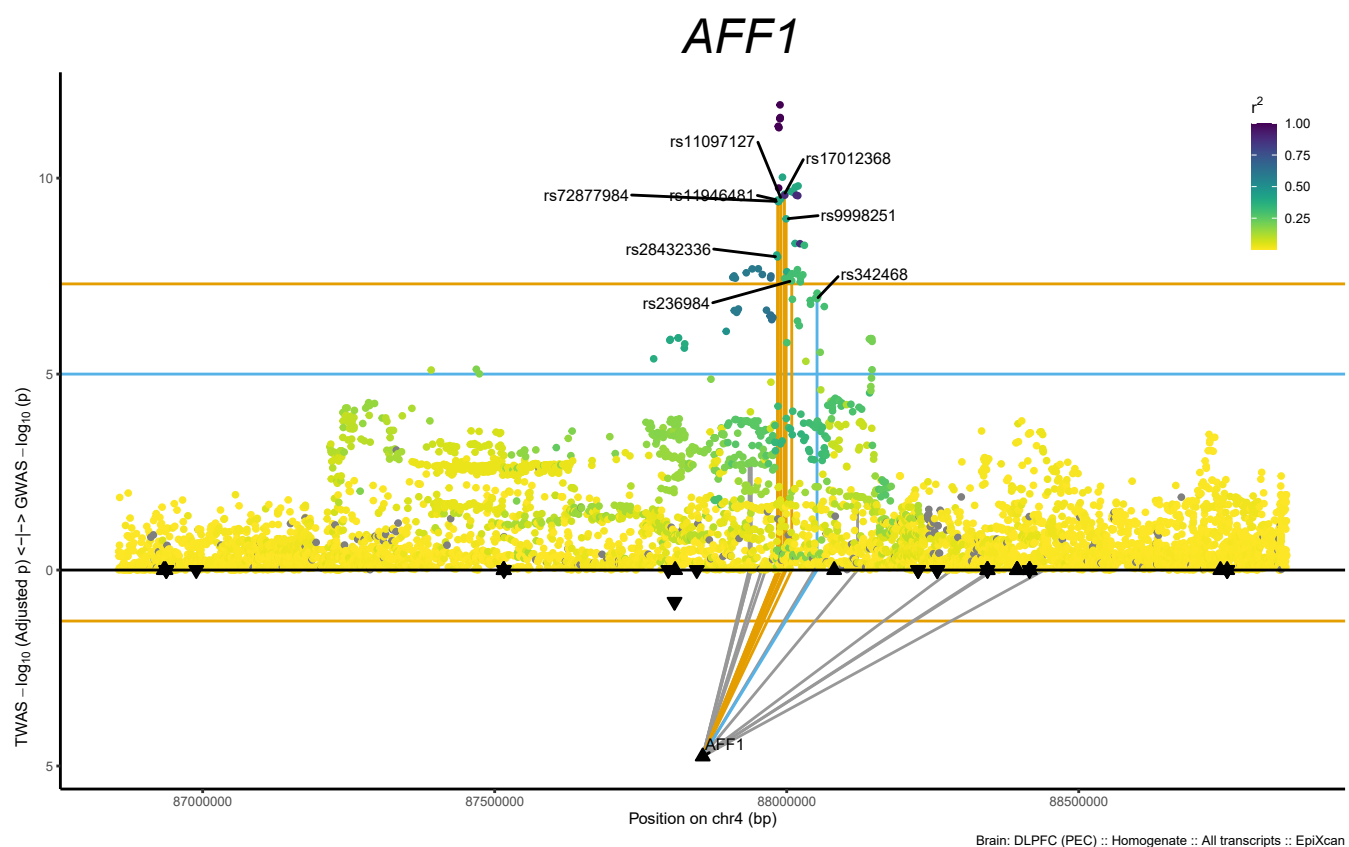
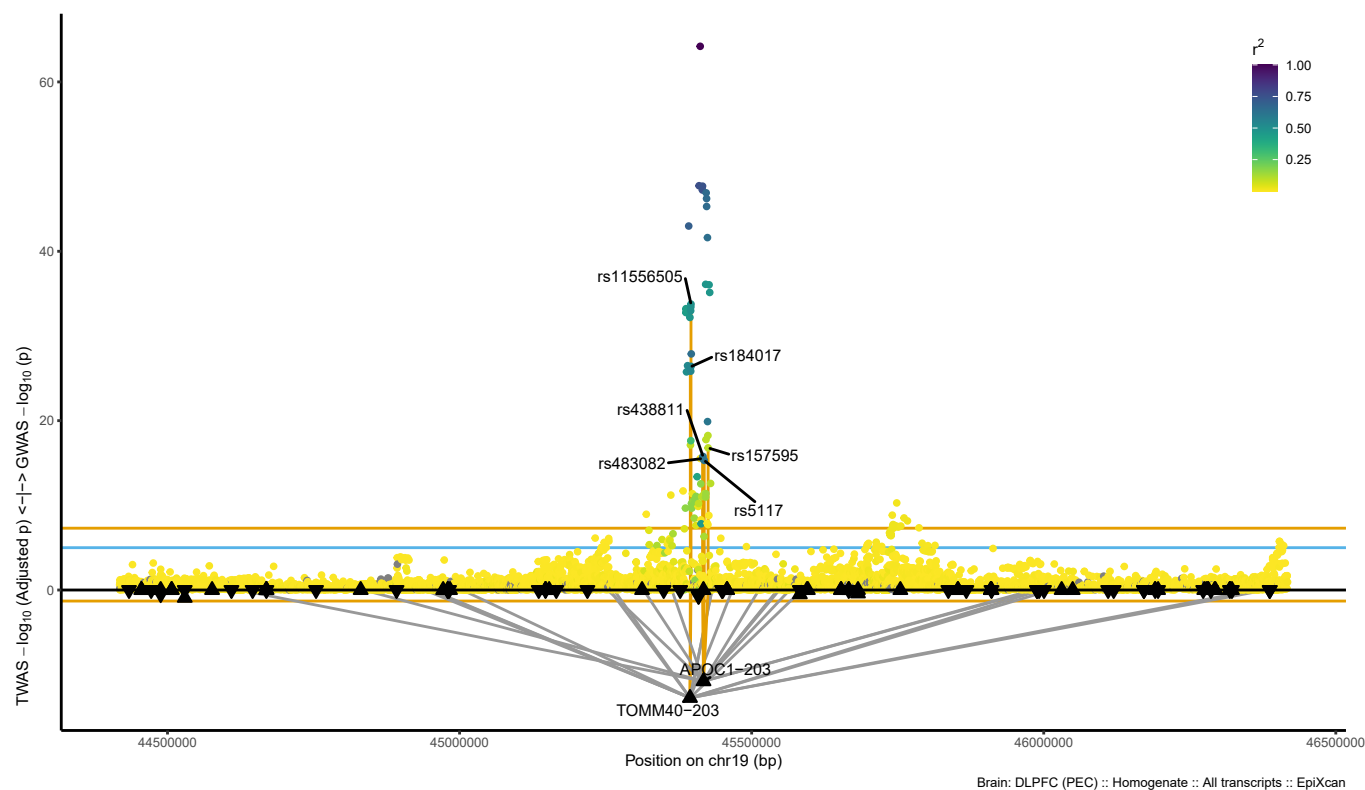
NTN5 / FUT1-201



WAC-AS1

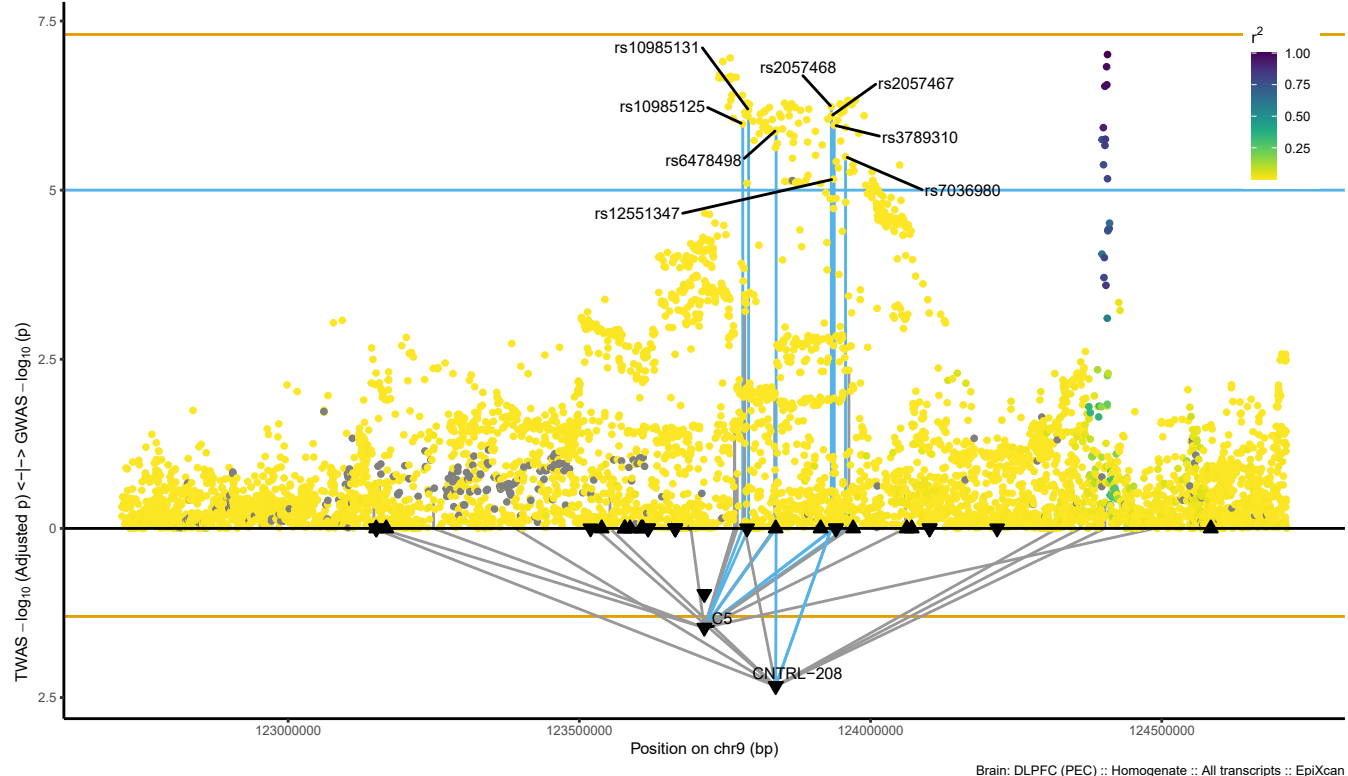


Supplementary Fig. 7 *TOMM40-203 / APOC1-203*

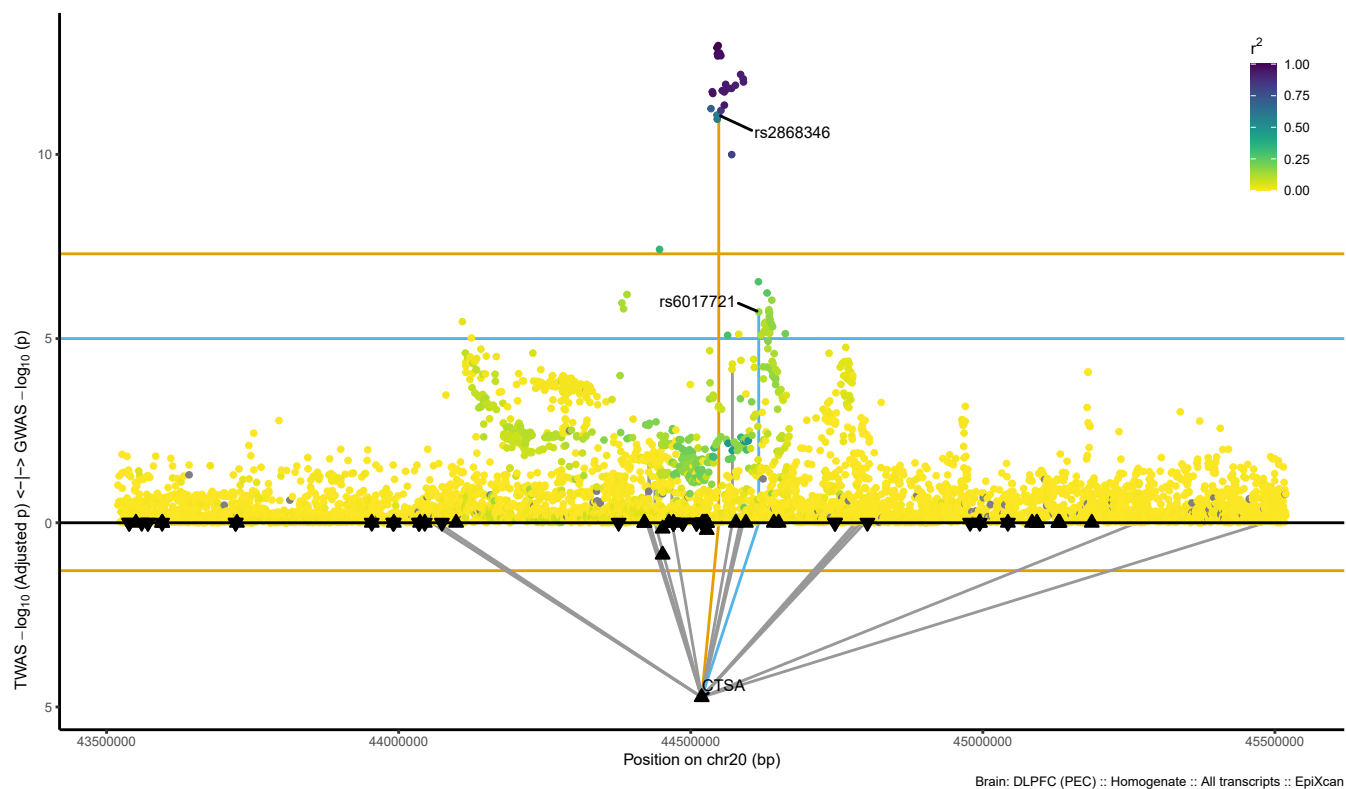


Supplementary Fig. 7

C5 / CNTRL-208

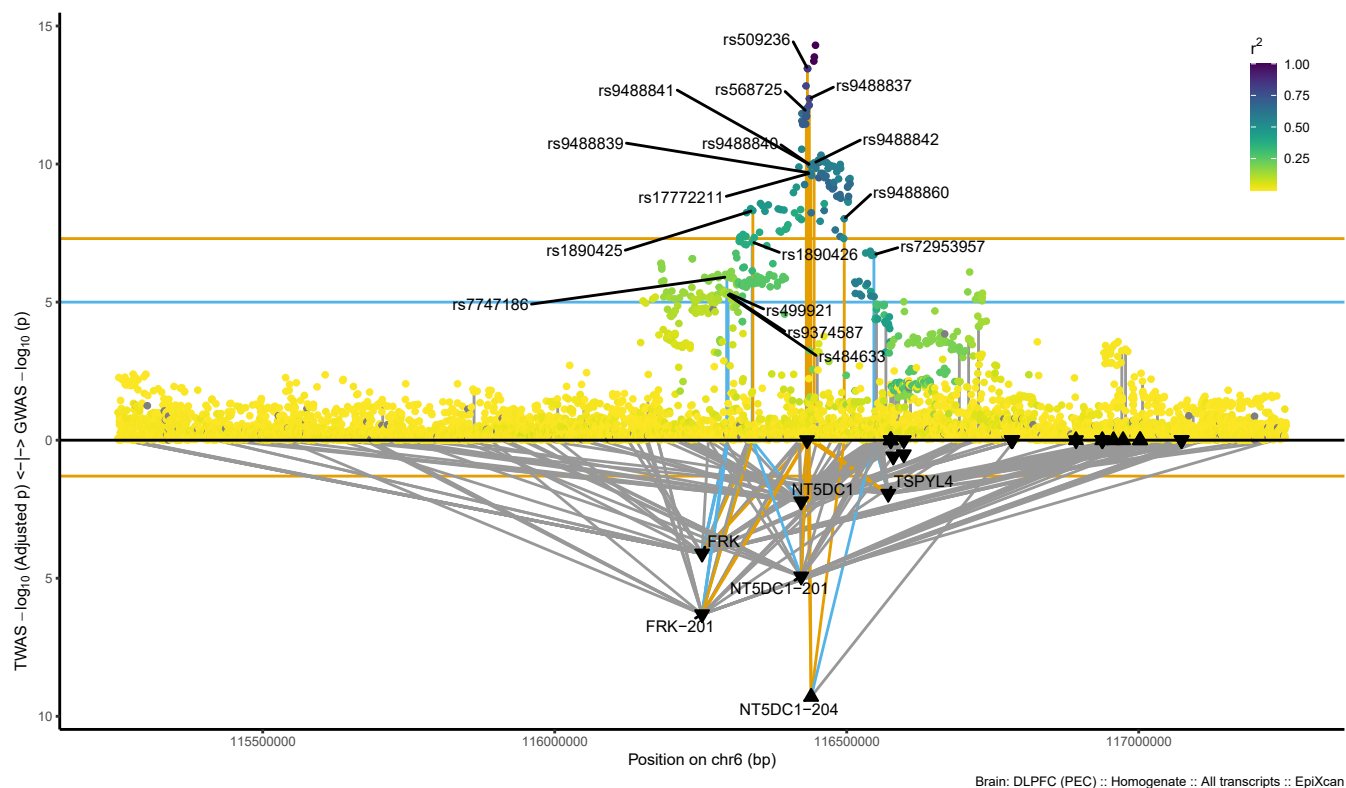


CTSA

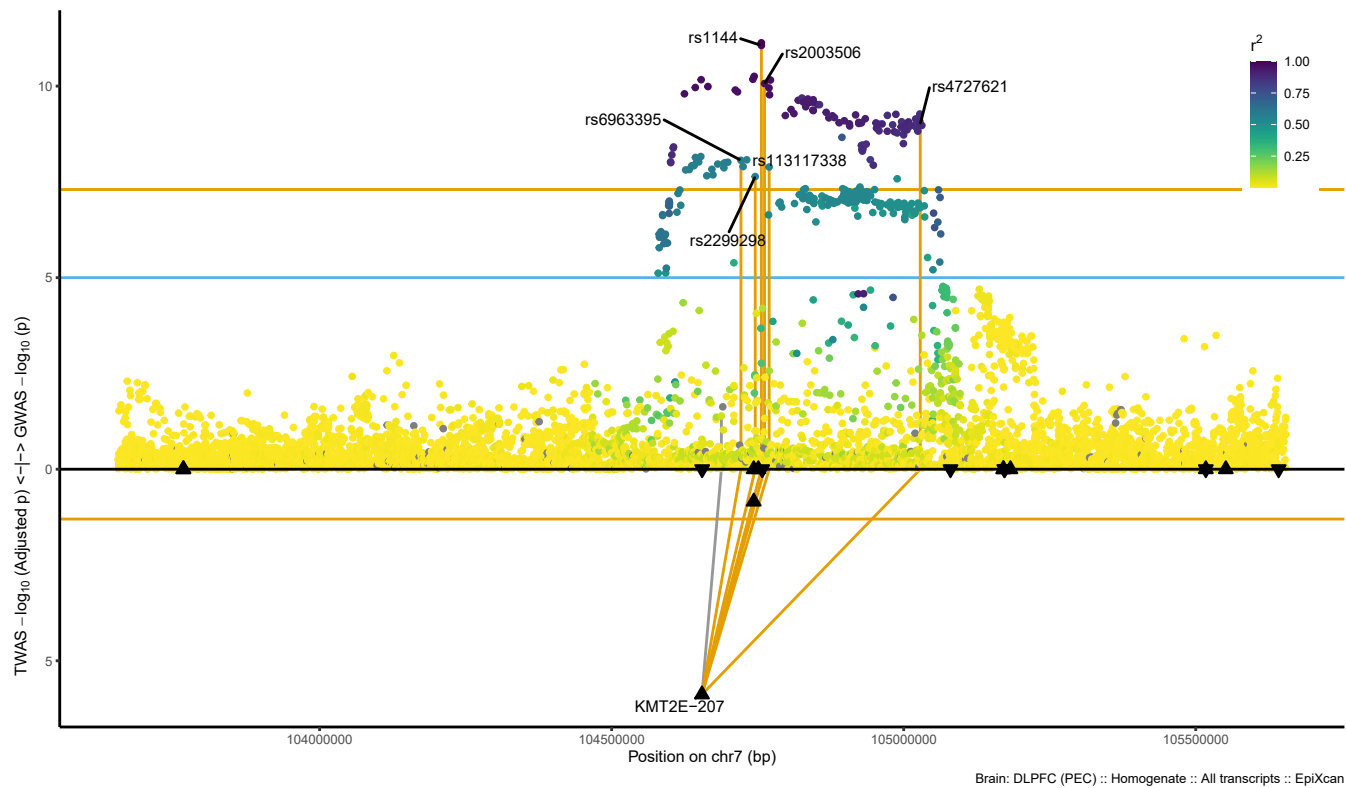


Supplementary Fig. 7

NT5DC1-204 / FRK-201



KMT2E-207



Captions for Supplementary Tables

Supplementary Table 1. Sample demographics for the GWAS cohorts comprising the European-ancestry meta-analysis and multi-ancestry meta-analysis.

Supplementary Table 2. Observational associations of AMD in the Million Veteran Program. a) Effects of demographic parameters on AMD case/control status in the MVP EA, AA and HA samples, estimated in independent single logistic regression models. b) Effects of demographic parameters on AMD case/control status in the MVP EA, AA and HA samples, estimated in multiple logistic regression models stratified by ancestry. c) Multi-ancestry time-to-AMD-diagnosis analysis using a Cox proportional-hazards model. Abbreviations: EA, European ancestry; AA, African ancestry; HA, Hispanic/Latino ancestry; OR, odds ratio; HR, hazards ratio; p, unadjusted two-tailed p-value; CI, confidence interval, PCs, principal components.

Supplementary Table 3. PheWAS summary statistics of the IAMDGC-derived PRS in European-ancestry subjects in MVP. The PRS was normalized to the unit normal distribution. Each phenotype was independently tested in a logistic regression model, adjusting for sex, age, age-squared, and 20 principal components. Additionally, top decile vs. bottom decile and top decile vs. remaining 90% contrasts were analyzed to assess risk stratification. Phenotypes with fewer than 100 cases were excluded. Abbreviations: PRS: polygenic risk score; OR: odds ratio; SE: standard error; n_total: total number of samples (cases + controls); n_cases: number of cases; p: unadjusted two-tailed p-value; p_bonferroni_adj: p-value after Bonferroni correction for the number of PheWAS phenotypes.

Supplementary Table 4. LabWAS summary statistics of the IAMDGC-derived PRS in European-ancestry subjects in MVP. Each phenotype was independently tested in a linear regression model, adjusting for sex, age, age-squared, and 20 PCs. The PRS was normalized to the unit normal distribution and labs were inverse rank normal transformed. Abbreviations: IRNT: inverse rank normal transformed; p: unadjusted two-tailed p-value; p_bonferroni_adj: p-value after Bonferroni correction for the number of LabWAS phenotypes.

Supplementary Table 5. Index SNPs from genome-wide significant loci in the European-ancestry inverse variance-weighted fixed-effects meta-analysis (N=57,290 cases, 324,430 controls).

Supplementary Table 6. Comparison of logistic regression effect sizes at the index SNPs across the IAMDGC, GERA, and MVP European-ancestry tranche 3 GWAS.

Supplementary Table 7. Conditionally independent genome-wide significant markers from the European meta-analysis. Conditionally independent signals were determined using stepwise conditional and joint analysis (COJO-slc) from the GCTA software package (Yang, 2011; PMID 21167468). Abbreviations: EA, effect allele; freq, frequency of the effect allele; beta, effect size;

se, standard error; p, p-value from GWAS; N_eff_cojo, estimated effective sample size in COJO; freq_genos, frequency of the effect allele in the reference panel; bJ, effect size from joint analysis of selected SNPs; bJ_se, standard error from joint analysis of selected SNPs; pJ, p-value from joint analysis of selected SNPs; LD_r, LD correlation between the SNP i and SNP i + 1 for the SNPs on the list.

Supplementary Table 8. Genetic correlation of AMD with selected eye-related traits, estimated from LD score regression. Abbreviations: rg, genetic correlation; se, standard error of the genetic correlation; p, unadjusted two-tailed p-value; h2_obs, observed heritability of trait 2; h2_obs_se, standard error of the observed heritability; h2_int, LD score regression intercept for trait 2; h2_int_se, standard error of the LD score regression intercept; gcov, genetic covariance; gcov_se, standard error of the genetic covariance.

Supplementary Table 9. Genetic correlation of AMD with blood-based biomarkers, hematological parameters, and other traits, estimated from LD score regression. Trait GWAS were from the Pan-UKB project (Pan-UKB team, <https://pan.ukbb.broadinstitute.org>, 2020) unless noted otherwise. Abbreviations: rg: genetic correlation; se: standard error of the genetic correlation; p: unadjusted two-tailed p-value; h2_obs: observed heritability of trait 2; h2_obs_se: standard error of the observed heritability; h2_int: LD score regression intercept for trait 2; h2_int_se: standard error of the LD score regression intercept; gcov: genetic covariance; gcov_se: standard error of the genetic covariance.

Supplementary Table 10. Mendelian randomization analyses of pigmentation traits on AMD. Traits were tested using five methods implemented in the TwoSample MR package: Inverse variance weighted, MR Egger, simple mode, weighted mode, and weighted median. Horizontal pleiotropy was tested via the MR Egger intercept. Abbreviations: se, standard error; pval, unadjusted two-tailed p-value; or, odds ratio; lo_ci and up_ci, lower and upper bounds of the 95% confidence interval.

Supplementary Table 11. Index SNPs from genome-wide significant loci in the African-ancestry GWAS (N=2,302 cases, 29,223 controls). Summary statistics are provided for the two loci reaching genome-wide significance in the AA logistic regression GWAS; the signal near HLA-DQB1 was replicated in the larger EA meta-analysis, while a second potential locus near COPS2 was not.

Supplementary Table 12. LabWAS summary statistics of the MVP AA GWAS-derived PRS in unrelated African-ancestry subjects in MVP not used in the GWAS cohort. Each phenotype was independently tested in a linear regression model, adjusting for sex, age, age-squared, and 20 PCs. The PRS was normalized to the unit normal distribution and labs were inverse rank normal

transformed (IRNT). A PheWAS was also performed, with no significant associations. Abbreviations: IRNT: inverse rank normal transformed; se: standard error; p: unadjusted two-tailed p-value.

Supplementary Table 13. Index SNPs from genome-wide significant loci in the Hispanic/Latino-ancestry GWAS (N=1,656 cases, 10,819 controls). Summary statistics are provided for the two loci reaching genome-wide significance in the HA logistic regression GWAS, CFH and ARMS2, both of which are known loci in EA.

Supplementary Table 14. Index SNPs from genome-wide significant loci in the multi-ancestry meta-analysis (N=61,248 cases, 364,472 controls). Abbreviations: RE2, Han-Eskin random effects model; RE2_STAT1, mean effect statistic; RE2_STAT2, heterogeneity statistic.

Supplementary Table 15. SNPs with posterior probability > 0.01 from fine-mapping the European-ancestry meta-analysis (N=57,290 cases, 324,430 controls). Fine-mapping was conducted at each genome-wide significant locus using FINEMAP software (Benner et al. 2016; PMID: 26773131). Abbreviations: allele1, effect allele; allele2, other allele; maf, minor allele frequency; prob, posterior probability; log10bf, logarithm of the Bayes factor; mean, marginalized shrinkage estimate for the posterior effect size mean; sd, marginalized shrinkage estimate for the posterior effect size standard deviation; mean_incl, conditional estimate for the posterior effect size mean; sd_incl, conditional estimate for the posterior effect size standard deviation; p, p-value from the GWAS; cs1-10, whether a variant belongs to credible sets 1-10 for the locus. The maximum number of credible sets in a locus was set to 10.

Supplementary Table 16. SNPs with posterior probability > 0.01 from fine-mapping the multi-ancestry meta-analysis (N=61,248 cases, 364,472 controls). Fine-mapping was conducted at each genome-wide significant locus using FINEMAP software (Benner et al. 2016; PMID: 26773131). Abbreviations: allele1, effect allele; allele2, other allele; maf, minor allele frequency; prob, posterior probability; log10bf, logarithm of the Bayes factor; mean, marginalized shrinkage estimate for the posterior effect size mean; sd, marginalized shrinkage estimate for the posterior effect size standard deviation; mean_incl, conditional estimate for the posterior effect size mean; sd_incl, conditional estimate for the posterior effect size standard deviation; p, p-value from the GWAS; cs1-10, whether a variant belongs to credible sets 1-10 for the locus. The maximum number of credible sets in a locus was set to 10.

Supplementary Table 17. Candidate causal variants from eQTL co-localization analysis using the mmQTL method (Zeng et al. 2022; PMID: 35058635). Shown are SNPs with co-localization posterior probability (CLPP) greater than 0.01. a) eQTL co-localization with the multi-ancestry fine-mapping. b) eQTL co-localization with the European-ancestry fine-mapping.

Supplementary Table 18. Local ancestry analysis of haplotypes at CFH in admixed MVP populations. a) Enrichment of European-ancestry haplotypes (EUR) in AA AMD cases. b) Enrichment of individuals with homozygous European ancestry at CFH (EUR/EUR) in AA AMD cases. c) Enrichment of EUR haplotypes, compared to non-EUR haplotypes in HA AMD cases. d) Enrichment of EUR haplotypes in HA AMD cases, compared to African (AFR) and Native American (NAT) haplotypes. For each, a two-sided Fisher's exact test was performed.

Supplementary Table 19. Haplotype analysis of the CFH locus in admixed African-ancestry (AA) subjects. Case and control frequencies were derived for all haplotypes and stratified by individuals with homozygous African local ancestry (AFR/AFR) or homozygous European local ancestry (EUR/EUR). a) Haplotype model used in Cipriani et al. (2021; PMID: 34260948). b) Haplotype model used in Pappas et al. (2021; PMID: 34563268).

Supplementary Table 20. Local ancestry analysis of haplotypes at the ARMS2 locus in admixed MVP populations. a) No significant enrichment of European-ancestry haplotypes (EUR) in AA AMD cases. b) No significant enrichment of individuals with homozygous European ancestry at ARMS2 (EUR/EUR) in AA AMD cases. c) No significant enrichment of EUR haplotypes, compared to non-EUR haplotypes in HA AMD cases. d) Slight enrichment of Native American (NAT) haplotypes in HA AMD cases, compared to African (AFR) and EUR. For each, a two-sided Fisher's exact test was performed.

Supplementary Table 21. Modeling epistatic interaction between the CFH Y402H and ARMS2 A69S risk alleles in MVP EA subjects. a) Summary statistics of the multivariable logistic regression model including the A69S*Y402H multiplicative interaction term. b) Summary statistics of the multivariable logistic regression model excluding the multiplicative interaction term. c) Likelihood ratio test comparing the two models, with p-value derived from a one-sided chi-squared test.

Supplementary Table 22. HLA fine-mapping summary statistics for unrelated EA subjects in MVP. Association with AMD was tested for each four-digit allele at HLA-A, B, C, DPB1, DQA1, DQB1, and DRB1 imputed by HIBAG software (Zheng et al. 2014; PMID: 23712092), considered additively with all other alleles as reference, in independent Firth logistic regression models. Association was tested using two models: 1) sex and ten within-ancestry PCs only, and 2) sex, ten within-ancestry PCs, and risk allele dosages for CFH (rs1061170), ARMS2 (rs10490924), CFB/C2 (rs429608), and C3 (rs2230199). p-values (unadjusted, two-tailed) and confidence intervals are based on the penalized likelihood ratio test. Alleles with minor allele count less than 50 were excluded from analysis. Additionally, summary statistics from chi-squared and Fisher's exact tests are provided.

Supplementary Table 23. HLA fine-mapping summary statistics for unrelated AA subjects in MVP. Association with AMD was tested for each four-digit allele at HLA-A, B, C, DPB1, DQA1, DQB1, and DRB1 imputed by HIBAG software (Zheng et al. 2014; PMID: 23712092), considered additively with all other alleles as reference, in independent Firth logistic regression models. Association was tested using two models: 1) sex and ten within-ancestry PCs only, and 2) sex, ten within-ancestry PCs, and risk allele dosages for CFH (rs1061170), ARMS2 (rs10490924), CFB/C2 (rs429608), and C3 (rs2230199). p-values (unadjusted, two-tailed) and confidence intervals are based on the penalized likelihood ratio test. Alleles with minor allele count less than 50 were excluded from analysis. Additionally, summary statistics from chi-squared and Fisher's exact tests are provided.

Supplementary Table 24. HLA fine-mapping summary statistics for unrelated HA subjects in MVP. Association with AMD was tested for each four-digit allele at HLA-A, B, C, DPB1, DQA1, DQB1, and DRB1 imputed by HIBAG software (Zheng et al. 2014; PMID: 23712092), considered additively with all other alleles as reference, in independent Firth logistic regression models. Association was tested using two models: 1) sex and ten within-ancestry PCs only, and 2) sex, ten within-ancestry PCs, and risk allele dosages for CFH (rs1061170), ARMS2 (rs10490924), CFB/C2 (rs429608), and C3 (rs2230199). p-values (unadjusted, two-tailed) and confidence intervals are based on the penalized likelihood ratio test. Alleles with minor allele count less than 50 were excluded from analysis. Additionally, summary statistics from chi-squared and Fisher's exact tests are provided.

Supplementary Table 25. Transcriptome-wide association scan (TWAS) summary statistics using the brain DLPFC (PsychENCODE) EpiXcan tissue model (N=57,290 cases, 364,472 controls). Abbreviations: feature, gene or transcript ENSEMBL ID as listed in the tissue transcriptome model; Feature_name, gene or transcript name as listed by the transcriptome model, typically HUGO; feature_start, start position of the feature; feature_end, end position of the feature; feature_width, width of the feature; feature_strand, strand (+/-) of the feature; zscore, S-EpiXcan association z-score for the feature; pvalue, S-EpiXcan association p-value for the feature; Bonferroni_trait_model, S-EpiXcan association Bonferroni-adjusted p-value for the feature (only considering this specific trait-model combination); fdr_trait_model, S-EpiXcan association FDR-adjusted p-value for the feature (only considering this specific trait-model combination); Bonferroni_adjusted_fdr_trait, S-EpiXcan association FDR-adjusted p-value for the feature which is then Bonferroni-adjusted study-wide; effect_size, S-EpiXcan association effect size for the gene; var_g, Variance of the gene expression, calculated as $W' * G * W$ (where W is the vector of SNP weights in a gene's model, W' is its transpose, and G is the covariance matrix);

pred_perf_r2, Cross-validated R2 of tissue model's correlation to gene's measured transcriptome (prediction performance); pred_perf_pval, p-value of tissue model's correlation to gene's measured transcriptome (prediction performance); pred_perf_qval, q-value of tissue model's correlation to gene's measured transcriptome (prediction performance); n_snps_used, Number of SNPs from GWAS that were used in S-EpiXcan analysis; n_snps_in_cov, Number of SNPs in the covariance matrix; n_snps_in_model, Number of SNPs in the imputation model; fdr_all, FDR-adjusted p-value considering all trait-model combinations; Bonferroni_all, Bonferroni-adjusted p-value considering all trait-model combinations.

Supplementary Table 26. Transcriptome-wide association scan (TWAS) summary statistics using the retina (EyeGEx) PrediXcan tissue model (N=57,290 cases, 364,472 controls). Abbreviations: feature, gene or transcript ENSEMBL ID as listed in the tissue transcriptome model; Feature_name, gene or transcript name as listed by the transcriptome model, typically HUGO; feature_start, start position of the feature; feature_end, end position of the feature; feature_width, width of the feature; feature_strand, strand (+/-) of the feature; zscore, S-EpiXcan association z-score for the feature; pvalue, S-EpiXcan association p-value for the feature; Bonferroni_trait_model, PrediXcan association Bonferroni-adjusted p-value for the feature (only considering this specific trait-model combination); fdr_trait_model, PrediXcan association FDR-adjusted p-value for the feature (only considering this specific trait-model combination); Bonferroni_adjusted_fdr_trait, PrediXcan association FDR-adjusted p-value for the feature which is then Bonferroni-adjusted study-wide; effect_size, S-EpiXcan association effect size for the gene; var_g, Variance of the gene expression, calculated as $W' * G * W$ (where W is the vector of SNP weights in a gene's model, W' is its transpose, and G is the covariance matrix); pred_perf_r2, Cross-validated R2 of tissue model's correlation to gene's measured transcriptome (prediction performance); pred_perf_pval, p-value of tissue model's correlation to gene's measured transcriptome (prediction performance); pred_perf_qval, q-value of tissue model's correlation to gene's measured transcriptome (prediction performance); n_snps_used, Number of SNPs from GWAS that were used in PrediXcan analysis; n_snps_in_cov, Number of SNPs in the covariance matrix; n_snps_in_model, Number of SNPs in the imputation model; fdr_all, FDR-adjusted p-value considering all trait-model combinations; Bonferroni_all, Bonferroni-adjusted p-value considering all trait-model combinations.

Supplementary Table 27. TWAS summary statistics from the ACAT (aggregated Cauchy association test) meta-analysis of 38 tissue models (N=57,290 cases, 364,472 controls). Abbreviations: pvalue, unadjusted significance of gene feature in ACAT test; Bonferroni_trait_model, Bonferroni-adjusted p-value for the gene feature; fdr_trait_model, FDR-

adjusted p-value for the gene feature; Bonferroni_adjusted_fdr_trait, FDR-adjusted p-value for the feature which is then Bonferroni-adjusted study-wide (equals fdr_trait_model for the meta-analysis).

Supplementary Table 28. Comparison of the AMD TWAS results to previous studies (Strunz et al. 2020; PMID: 32005911; Ratnapriya et al. 2019; PMID: 30742112).

Supplementary Table 29. Summary-based Mendelian randomization (SMR) summary statistics for retina gene expression instruments (EyeGEx). Analysis was performed in the European-ancestry meta-analysis (N=57,290 cases, 364,472 controls) using the SMR function in the GCTA software package. Abbreviations: A1, effect allele; A2, other allele; Freq, effect allele frequency; b, se, p: beta, standard error, and unadjusted p-value from the GWAS, eQTL, and SMR experiment, respectively; nsnp_HEIDI and p_HEIDI, number of SNPs used in HEIDI (heterogeneity of independent instruments) outlier testing and corresponding p-value.

Supplementary Table 30. Gene-set enrichment analysis (GSEA) summary statistics based on the DLPFC TWAS. Abbreviations: pval, unadjusted gene set enrichment p-value; Jaccard, Jaccard similarity coefficient ranges from 0 (no overlap) to 1 (complete overlap); FDR_AdjP, FDR-adjusted p-value; BH_AdjP, Benjamini-Hochberg-adjusted p-value; Bonf_AdjP, Bonferroni-adjusted p-value; OR (95% C.I.), odds ratio and 95% confidence interval; qvalue, minimum false discovery rate at which a test result is considered statistically significant; ** = FDR < 0.01; * = FDR < 0.05.

Supplementary Table 31. Gene-set enrichment analysis (GSEA) summary statistics based on the TWAS tissue meta-analysis. Abbreviations: pval, unadjusted gene set enrichment p-value; Jaccard, Jaccard similarity coefficient ranges from 0 (no overlap) to 1 (complete overlap); FDR_AdjP, FDR-adjusted p-value; BH_AdjP, Benjamini-Hochberg-adjusted p-value; Bonf_AdjP, Bonferroni-adjusted p-value; OR (95% C.I.), odds ratio and 95% confidence interval; qvalue, minimum false discovery rate at which a test result is considered statistically significant; ** = FDR < 0.01; * = FDR < 0.05.

Supplementary Table 32. Summary statistics from the gene-based rare variant burden test. All SNPs were ultra-rare (frequency less than 0.25%) and genotyped on the MVP 1.0 array (Hunter-Zinck et al. 2020; PMID: 32243820). Three masks were defined: mask1, containing only frameshift, nonsense, stop-gained, and stop-loss loss of function (LoF) SNPs; mask2 containing LoF SNPs plus high-impact missense variants; and mask3, containing LoF + missense + high-impact splice-site mutations. Logistic regression burden tests over the masks were conducted in REGENIE (Mbatchou et al. 2021; PMID: 34017140).

References

1. Jorgenson, E. *et al.* Common coding variants in the HLA-DQB1 region confer susceptibility to age-related macular degeneration. *Eur. J. Hum. Genet.* **24**, 1049–1055 (2016).
2. Purcell, S. *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).
3. Chang, C. C. *et al.* Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* **4**, 7 (2015).
4. 1000 Genomes Project Consortium *et al.* A global reference for human genetic variation. *Nature* **526**, 68–74 (2015).
5. Phan, L. *et al.* ALFA: Allele frequency aggregator. *National Center for Biotechnology Information, US National Library of Medicine* (2020).
6. McCarthy, S. *et al.* A reference panel of 64,976 haplotypes for genotype imputation. *Nat. Genet.* **48**, 1279–1283 (2016).
7. Ongen, H., Buil, A., Brown, A. A., Dermitzakis, E. T. & Delaneau, O. Fast and efficient QTL mapper for thousands of molecular phenotypes. *Bioinformatics* **32**, 1479–1485 (2016).
8. Cantsilieris, S. *et al.* Recurrent structural variation, clustered sites of selection, and disease risk for the complement factor H (CFH) gene family. *Proc. Natl. Acad. Sci. U. S. A.* **115**, E4433–E4442 (2018).
9. Cipriani, V. *et al.* Beyond factor H: The impact of genetic-risk variants for age-related macular degeneration on circulating factor-H-like 1 and factor-H-related protein concentrations. *Am. J. Hum. Genet.* **108**, 1385–1400 (2021).
10. Fritsche, L. G. *et al.* A large genome-wide association study of age-related macular degeneration highlights contributions of rare and common variants. *Nat. Genet.* **48**, 134–143 (2016).
11. Hughes, A. E. *et al.* A common CFH haplotype, with deletion of CFHR1 and CFHR3, is associated with lower risk of age-related macular degeneration. *Nat. Genet.* **38**, 1173–1177 (2006).
12. Williams, B. L. *et al.* Chromosome 10q26-driven age-related macular degeneration is associated with reduced levels of HTRA1 in human retinal pigment epithelium. *Proc. Natl. Acad. Sci. U. S. A.* **118**, (2021).
13. Friedrich, U. *et al.* Synonymous variants in HTRA1 implicated in AMD susceptibility impair its capacity to regulate TGF- β signaling. *Hum. Mol. Genet.* **24**, 6361–6373 (2015).
14. Beguier, F. *et al.* The 10q26 risk haplotype of age-related macular degeneration aggravates subretinal inflammation by impairing monocyte elimination. *Immunity* **53**, 429–441.e8 (2020).

15. Frochaux, V., Hildebrand, D., Talke, A., Linscheid, M. W. & Schlüter, H. Alpha-1-antitrypsin: a novel human high temperature requirement protease A1 (HTRA1) substrate in human placental tissue. *PLoS One* **9**, e109483 (2014).
16. Campbell, R. A. *et al.* Placental HTRA1 cleaves α 1-antitrypsin to generate a NET-inhibitory peptide. *Blood* **138**, 977–988 (2021).
17. Oka, C. & Kawaichi, M. HtrA1 meets NETs: does it open a new field? *Blood* vol. 138 920–921 (2021).
18. Wang, H., Wang, C., Zhao, M.-H. & Chen, M. Neutrophil extracellular traps can activate alternative complement pathways. *Clin. Exp. Immunol.* **181**, 518–527 (2015).
19. Yuen, J. *et al.* NETosing neutrophils activate complement both on their own NETs and bacteria via alternative and non-alternative pathways. *Front. Immunol.* **7**, 137 (2016).
20. de Bont, C. M., Boelens, W. C. & Pruijn, G. J. M. NETosis, complement, and coagulation: a triangular relationship. *Cell. Mol. Immunol.* **16**, 19–27 (2019).
21. Schneider, A. E., Sándor, N., Kárpáti, É. & Józsi, M. Complement factor H modulates the activation of human neutrophil granulocytes and the generation of neutrophil extracellular traps. *Mol. Immunol.* **72**, 37–48 (2016).
22. Halder, L. D. *et al.* Factor H binds to extracellular DNA traps released from human blood monocytes in response to *Candida albicans*. *Front. Immunol.* **7**, 671 (2016).
23. Ghosh, S. *et al.* Neutrophils homing into the retina trigger pathology in early age-related macular degeneration. *Commun Biol* **2**, 348 (2019).
24. Niazi, S., Krogh Nielsen, M., Sørensen, T. L. & Subhi, Y. Neutrophil-to-lymphocyte ratio in age-related macular degeneration: a systematic review and meta-analysis. *Acta Ophthalmol.* **97**, 558–566 (2019).