medRxiv preprint doi: https://doi.org/10.1101/2022.08.16.22278855; this version posted August 21, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. This article is a US Government work. It is not subject to copyright under 17 USC 105 and is also made available for use under a CC0 license.

Distinctive cross-ancestry genetic architecture for age-related macular degeneration

Authors: Bryan R. Gorman[1,2,†], Georgios Voloudakis[3,4,†], Robert P. Igo, Jr[5,6,†,‡], Tyler Kinzy[6], Christopher W. Halladay[7], Tim B. Bigdeli[8,9], Biao Zeng[10], Sanan Venkatesh[3], Jessica N. Cooke Bailey[5,6,11], Dana C. Crawford[5,6,12,11], Kyriacos Markianos[1], Frederick Dong[1,2], Patrick Schreiner[1,2], Wen Zhang[3], VA Million Veteran Program[13], International AMD Genomics Consortium (IAMDGC)[14], Tamer Hadi[5,15], Matthew D. Anger[16,17], Amy D. Stockwell[18], Ronald B. Melles[19], Jie Yin[20], Hélène Choquet[20], Rebecca Kaye[21], Karina Patasova[21], Praveen J. Patel[21], Brian L. Yaspan[18], Eric Jorgenson[22], Pirro G. Hysi[21,23,24], Andrew J. Lotery[25,26], J. Michael. Gaziano[27,28], Philip S. Tsao[29,30,31], Steven J. Fliesler[16,17], Jack M. Sullivan[16,17], Paul B. Greenberg[32,33], Wen-Chih Wu[34,35], Themistocles L. Assimes[29,30], Saiju Pyarajan[1,36], Panos Roussos[3,4,§], Neal S. Peachey[5,37,38,§], Sudha K. Iyengar[5,6,15,12,11,§,*]

Affiliations:

[1]Center for Data and Computational Sciences (C-DACS), VA Cooperative Studies Program, VA Boston Healthcare System; Boston, MA, USA.

[2]Booz Allen Hamilton; McLean, VA, USA.

[3]Center for Disease Neurogenomics, Department of Psychiatry, Friedman Brain Institute, Department of Genetics and Genomic Science and Institute for Multiscale Biology, Icahn School of Medicine at Mount Sinai; New York, NY, USA.

[4]Mental Illness Research, Education, and Clinical Center, James J. Peters VA Medical Center; Bronx, NY, USA.

[5]Louis Stokes Cleveland VA Medical Center; Cleveland, OH, USA.

[6]Department of Population and Quantitative Health Sciences, Case Western Reserve University; Cleveland, OH, USA.

[7]Center for Innovation in Long Term Services and Supports, Providence VA Medical Center; Providence, RI, USA.

[8]VA New York Harbor Healthcare System; Brooklyn, NY, USA.

[9]Department of Psychiatry and Behavioral Sciences, SUNY Downstate Health Sciences University; Brooklyn, NY, USA.

[10]Department of Genetics and Genomic Science and Institute for Multiscale Biology, Icahn School of Medicine at Mount Sinai; New York, NY, USA.

[11]Cleveland Institute for Computational Biology, Case Western Reserve University; Cleveland, OH, USA.

[12]Department of Genetics & Genome Sciences, Case Western Reserve University; Cleveland, OH, USA.

[13]VA Boston Healthcare System; Boston, MA, USA.

[14]Institute for Human Genomics, University of Miami Miller School of Medicine; Miami, FL, USA.

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

medRxiv preprint doi: https://doi.org/10.1101/2022.08.16.22278855; this version posted August 21, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. This article is a US Government work. It is not subject to copyright under 17 USC 105 and is also made available for use under a CC0 license.

[15]Department of Ophthalmology and Visual Sciences, University Hospitals Eye Institute; Cleveland, OH, USA.
[16]Research Service, VA Western NY Healthcare System; Buffalo, NY, USA.
[17]Department of Ophthalmology and Neuroscience Graduate Program, SUNY-University at Buffalo; Buffalo, NY, USA.
[18]Department of Human Genetics, Genentech; South San Francisco, CA, USA.
[19]Department of Ophthalmology, Kaiser Permanente Northern California; Redwood City, CA, USA.
[20]Division of Research, Kaiser Permanente Northern California; Oakland, CA, USA.
[21]Section of Ophthalmology, School of Life Course Sciences, King's College London; London, UK.
[22]Regeneron Pharmaceuticals; Tarrytown, NY, USA.
[23]Department of Twin Research and Genetic Epidemiology; King's College London, London, UK.
[24]UCL Great Ormond Street Institute of Child Health, King's College London; London, UK.
[25]Southampton Eye Unit, University Hospital Southampton National Health Services Trust; Southampton, UK.
[26]Clinical and Experimental Sciences, Faculty of Medicine, University of Southampton; Southampton, UK.
[27]Massachusetts Veterans Epidemiology Research and Information Center (MAVERIC), VA Boston Healthcare System; Boston, MA, USA.
[28]Division of Aging, Department of Medicine, Brigham and Women's Hospital, Harvard Medical School; Boston, MA, USA.
[29]VA Palo Alto Epidemiology Research and Information Center for Genomics, VA Palo Alto Health Care System; Palo Alto, CA, USA.
[30]Department of Medicine, Stanford University School of Medicine; Stanford, CA, USA.
[31]Stanford Cardiovascular Institute, Stanford University School of Medicine; Stanford, CA, USA.
[32]Section of Ophthalmology, Providence VA Medical Center; Providence, RI, 02908, USA.
[33]Division of Ophthalmology, Alpert Medical School, Brown University; Providence, RI, USA.
[34]Section of Cardiology, Medical Service, Providence VA Medical Center; Providence, RI, USA.
[35]Division of Cardiology, Department of Medicine, Alpert Medical School, Brown University; Providence, RI, USA.
[36]Department of Medicine, Brigham and Women's Hospital, Harvard Medical School; Boston, MA, USA.
[37]Department of Ophthalmology, Cleveland Clinic Lerner College of Medicine of Case Western Reserve University; Cleveland, OH, USA.
[38]Cole Eye Institute, Cleveland Clinic; Cleveland, OH, USA.

† denotes equal contribution

‡ deceased

§ denotes equal contribution

*Corresponding author: Email: ski@case.edu

medRxiv preprint doi: https://doi.org/10.1101/2022.08.16.22278855; this version posted August 21, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. This article is a US Government work. It is not subject to copyright under 17 USC 105 and is also made available for use under a CC0 license.

Abstract:

To effectively reduce vision loss due to age-related macular generation (AMD) on a global scale, knowledge of its genetic architecture in diverse populations is necessary. A critical element, AMD risk profiles in African and Hispanic/Latino ancestries, remains largely unknown due to lower lifetime prevalence. We combined genetic and clinical data in the Million Veteran Program with five other cohorts to conduct the first multi-ancestry genome-wide association study of AMD and discovered 63 loci (30 novel). We observe marked cross-ancestry heterogeneity at major risk loci, especially in African-ancestry populations which demonstrate a primary signal in a Major Histocompatibility Complex Class II haplotype and reduced risk at the established *CFH* and *ARMS2/HTRA1* loci. Broadening efforts to include ancestrally-distinct populations helped uncover genes and pathways which boost risk in an ancestry-dependent manner, and are potential targets for corrective therapies.

One Sentence Summary: Probing electronic health record data with genomics unearths novel paths to age-related macular degeneration.

medRxiv preprint doi: https://doi.org/10.1101/2022.08.16.22278855; this version posted August 21, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. This article is a US Government work. It is not subject to copyright under 17 USC 105 and is also made available for use under a CC0 license.

Main Text:

Introduction

Blindness is one of the most feared health problems, reflecting the critical role that vision plays in daily living. Age-related macular degeneration (AMD) is a common disorder that impacts the central retina region or macula which is specialized for high acuity vision. Because AMD spares the peripheral retina, patients retain useful side vision, but the loss of central vision interferes with important daily activities such as reading, close work, driving, and recognizing faces.

AMD is a progressive disease that alters various histological layers of the retina, although progression rates vary by age, sex, and ancestry (*1*). Lipoproteinaceous deposits called drusen formed between the Bruchs' membrane and the retinal pigment epithelium (RPE) are the earliest indications. The clinical presentation of dry AMD typically involves drusen, pigmentary abnormalities and geographic atrophy occupying the central fovea. Decreased choroidal perfusion and retinal tissue oxygenation may stimulate choroidal neovascularization, resulting in wet AMD.

AMD shows differential risk across diverse populations (*2*, *3*) with Black subjects being 74% less likely to receive an AMD diagnosis compared to non-Hispanic white subjects after adjustment for age and sex (*3*). In the same study, Hispanic/Latino subjects showed a 44% reduction, while Asian subjects showed a 19% reduction in AMD diagnosis. The cause of the reduced risks in these populations is not well understood, although genetic or environmental exposure differences, or disparities in life expectancy are likely factors (*4*). While genome-wide association studies (GWAS) for AMD have been published for European ancestry (EA) and Asian ancestry (AS) populations, comparable analyses have not been conducted for African ancestry (AA) or Hispanic/Latino-ancestry (HA) populations in part because of lower AMD prevalence.

medRxiv preprint doi: https://doi.org/10.1101/2022.08.16.22278855; this version posted August 21, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. This article is a US Government work. It is not subject to copyright under 17 USC 105 and is also made available for use under a CC0 license.

GWAS have been highly successful in exploring the etiology of AMD; *CFH* is widely regarded as the first disease susceptibility locus discovered through a GWAS (*5*–*7*). The largest genetic analysis for AMD focused primarily on EA populations (*8*), leaving the question of risk in other ancestral groups unanswered. A central feature of the current investigation is the multi-ancestry nature of the biobank-scale sample available from the Department of Veterans Affairs (VA) Million Veteran Program (MVP) (*9*). Examining 61,248 cases and 364,472 controls from multiple cohorts, we discovered 30 new genetic loci, and validated 33 known loci. In this study, we synthesize several lines of evidence to show that decreased AMD prevalence in AA is at least partly mediated by reduced risk at *CFH*, and other alternative complement pathway genes, allowing effects of other loci to be isolated. This work underscores the value of inclusive population studies to advance our mechanistic understanding of AMD and likely other diseases as well.

Results

Epidemiology of AMD in MVP

An overview of the present study is depicted in **Fig. 1**, with additional details in **Fig. S1. Table S1** summarizes the AMD cohorts used for our analyses. Among demographic factors **(Tables S2a, S2b)**, not surprisingly, increased age was significantly associated with AMD status. Female sex was also significantly associated with AMD status, even in MVP, where 91% of participants are male. 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]; $p=2\times10^{-6}$), even in the fully adjusted model, which was replicated with similar effect size (though not with statistical significance) in the AA and HA

medRxiv preprint doi: https://doi.org/10.1101/2022.08.16.22278855; this version posted August 21, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. This article is a US Government work. It is not subject to copyright under 17 USC 105 and is also made available for use under a CC0 license.

cohorts. Reduced serum albumin was significantly associated with AMD in all three cohorts, with similar effect sizes.

Using summary statistics from a prior GWAS conducted by the International AMD Genomics Consortium (IAMDGC) (*8*), we constructed polygenic risk scores (PRS) in MVP using PRS-CS software (*10*). We observed strong evidence of increasing penetrance of AMD with genetic risk and age in MVP males **(Fig. 2a)**, ranging from a prevalence of approximately 35% in the top PRS decile within the 75-85 age range, compared to less than 10% in the bottom decile. History of heavy alcohol use and smoking are significant lifestyle risk factors **(Table S2b)**, yet they demonstrate smaller changes in penetrance with age compared to the PRS.

To quantify if there were an overall reduction in risk for AMD in other ancestry groups across the entire MVP cohort, we performed a Cox proportional-hazards model with time to AMD diagnosis as the outcome and found that AA veterans had a 57% reduction in risk and HA veterans had a 26% reduction in risk, relative to EA veterans **(Table S2c)**. Supporting this overall result, the PRS demonstrated weak transferability to non-European ancestries in our selected case-control sample **(Fig. S2)**: in our EA sample, a standard deviation increase in the PRS was associated with an AMD OR=1.76 [95% confidence interval (CI) 1.73-1.78] in a fully adjusted model **(Table S2b)**, but only OR=1.13 [1.08-1.18; $p=1.2\times10^{-7}$] in our AA sample and OR=1.47 [1.39-1.56; $p=1.6\times10^{-41}$] in our HA sample.

We next interrogated pleiotropy of the AMD PRS using diagnostic codes (i.e., phenome-wide association study, or PheWAS) and laboratory measures (i.e., laboratory-wide association study, or LabWAS) recorded in the EHR. We evaluated PheWAS associations of 1,665 ICD code-based phenotypes with the normalized PRS and contrasted the top vs. bottom deciles and top decile vs. remaining 90% to evaluate risk stratification. Fifty-four phenotypes were significantly associated **(Fig. 2b; Table S3)**. As expected, retinal codes were highly significant -- compared to the bottom PRS decile, the top decile had OR=17.8 [15.1-20.9] for a diagnosis of wet AMD. Unsurprisingly, alcohol and tobacco exposure also showed strong associations. Other notable

medRxiv preprint doi: https://doi.org/10.1101/2022.08.16.22278855; this version posted August 21, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. This article is a US Government work. It is not subject to copyright under 17 USC 105 and is also made available for use under a CC0 license.

positive associations included cataracts, refractive error, and mood disorders. Interestingly, the AMD PRS was also associated with lowered risks of Alzheimer's disease and other dementias, which may reflect opposing roles of HDL cholesterol. Of the 69 lab measurements, 16 were significantly associated, including positive associations with HDL cholesterol, red cell distribution width, platelets, serum magnesium, monocytes, and neutrophils, compared with negative associations with triglycerides, serum albumin and serum glucose **(Fig 2c; Table S4)**.

European meta-analysis for AMD identifies 27 novel loci

GWAS of AMD in two MVP tranches and meta-analysis with the summary statistics from five other cohorts (IAMDGC, GERA, UK Biobank, Genentech geographic atrophy, and Genentech choroidal neovascularization) was followed by replication, i.e., meta-analysis with a third MVP tranche that later became available to us, for a total EA analysis comprising 57,290 AMD cases and 324,430 controls **(Table S1)**. This analysis **(Fig. 3a)** revealed 60 index variants: 33 previously reported and 27 novel significant loci **(Table S5; Figs. S3-S4)**. Due to oversampling for severity in the IAMDGC, effect sizes of the index variants at major association peaks are lower overall in the MVP EA cohorts, similar to other EHR-based cohorts **(Table S6)**. However, genetic correlation (*11*) between the MVP and IAMDGC GWAS was very high (r_G=0.95; se=0.04). Of the 34 loci reported by the IAMDGC (*8*), 31 are genome-wide significant in the EA meta-analysis; the exceptions are *PRLS-SPEF2*, *TRPM3*, and *VTN*.

We observed novel loci in or near genes related to complement cascade regulation (*CD46/CD55*)**,** TGF-β signaling (*TGFB1*, *SMAD3*, *ADAM19*), pigmentation (*TYR, HERC2/OCA2, TRPM1*), angiogenesis and vascular homeostasis (*RRAS*, *RASIP1, IGFBP7*, *PDGFB*, *MYO1E*, *EXOC3L2*), inflammation (*CSK/ULK3*), proteolysis (*SERPINA1*), cell proliferation and apoptosis (*ZBTB38*, *ZNF385B*), ubiquitin-proteasome system (*CAND2/TEMEM40*), lipid metabolism and biogenesis (*CHD9*, *LBP*, *HSDL2*, *AFF1*, *ACAA2/LIPG*, *ME3*), and photoreceptor function (*RLBP1, CLUL1*). Finally, we observed a novel locus in an intergenic region near *LINC02343*, a non-coding

medRxiv preprint doi: https://doi.org/10.1101/2022.08.16.22278855; this version posted August 21, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. This article is a US Government work. It is not subject to copyright under 17 USC 105 and is also made available for use under a CC0 license.

RNA of uncharacterized function. A conditional analysis revealed 176 mutually significant variants in or near the 60 loci **(Table S7)**.

At the 20q13 locus, the previously reported *MMP9* index SNP was not replicated in our meta-analysis, and a new signal at nearby *PLTP* (phospholipid transfer protein), a *CETP* and *LBP* paralog, emerged. Although partially in LD ($r^2 = 0.2$), conditioning on the *MMP9* SNP (rs1888235) in our meta-analysis, and on the *PLTP* SNP (rs17447545) in IAMDGC, weakened but did not eliminate either signal, demonstrating independence **(Fig. S5)**. This result is consistent with the reported specificity of the *MMP9* signal for wet AMD, which is enriched in IAMDGC (*8*).

Rare protective missense mutations in novel genes *CFD* and *RRAS* emerged, independent of any common variation. The *CFD* missense mutation (p.Glu69Lys, rs35186399), which was not genotyped in the IAMDGC study, has OR=0.73 [0.66-0.81; $p=8.0\times10^{-10}$], with a frequency just under 1%, and is a pQTL for decreased complement C8 levels (*12*). Although rs35186399 is not predicted by most variant effect classifiers to have a significant impact on protein function (*13*, *14*), it is within a CTCF binding site, which could potentially affect the insulation of topologically associated domains and enhancer-promoter regulatory interactions. We also identified a novel protective mutation in *RRAS*, a Ras-family GTPase with reported regulatory functions in angiogenesis and vascular homeostasis (*15*–*17*). The putatively protein-altering mutation, rs61760904 (p.Asp133Asn), has a frequency of 0.6% in MVP EA, with OR=0.75 [0.68-0.83; $p=9.5\times10^{-9}$], and also associates with increased blood pressure (*18*).

At *SERPINA1*, which encodes α-1 antitrypsin (A1AT), we observed associations with the pathogenic PiZ (p.Glu366Lys; rs28929474) and PiS (p.Glu288Val; rs17580) alleles. Intriguingly, PiZ is protective for AMD (OR=0.81 [0.77-0.86; $p=3.2\times10^{-12}$]), while PiS increases risk (OR=1.10 [1.06-1.14; $p=1.2\times10^{-7}$]). A1AT is a serine protease inhibitor primarily expressed in the liver which regulates proinflammatory neutrophil elastase. Notably, A1AT is cleaved by HTRA1 to form NET (neutrophil extracellular trap) inhibitory peptides (NIPs) (*19*). Z and S allele heterozygosity are

medRxiv preprint doi: https://doi.org/10.1101/2022.08.16.22278855; this version posted August 21, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. This article is a US Government work. It is not subject to copyright under 17 USC 105 and is also made available for use under a CC0 license.

associated with numerous inflammatory biomarkers, including serum albumin (*20*), C reactive protein, and glycoprotein acetyls (*21*).

Genetic correlations with a number of eye traits and related phenotypes revealed some of the strongest correlations with pigment dispersion syndrome and glaucoma (*22*), as well as eye color (*23*) and hair color (*24*) **(Table S8)**. Similarly, genetic risk to intraocular pressure was inversely correlated with that of AMD. We further scanned genetic correlations with blood-based biomarkers (Pan-UKB team, 2020) and found the strongest correlations with serum albumin (r_G=-0.13; p=4.2×10^{-8}), albumin/globulin ratio, and kidney biomarkers phosphate, calcium, and cystatin C **(Table S9)**.

We further explored the causality of pigmentation associations using Mendelian randomization (MR) **(Fig. 4; Table S10)**. Drawing instrumental variables from some of the largest studies of pigmentation, we report for the first time a causal effect of pigmentation on AMD in EA populations. Comparing AMD data with those from a study (*23*) in which eye color was measured on a quantitative scale (range 0-6) from blue to dark brown, we find that each step towards darker eye color decreases the risk of AMD by about 5% (OR=0.95 [0.94-0.97]). Selecting only the instruments with known associations in other pigmentary traits, which likely represent the best proxies for melanin metabolism in the retina (*23*), we obtain a similar effect estimate (OR=0.95 [0.93-0.98]). To further confirm the involvement of melanin metabolic pathways in AMD, we examined other pigmentary traits. We found that each step up from blond to black hair color (range 0-4) (*24*) decreases the risk of AMD by 12% (OR=0.88 [0.85-0.92]). The more modest protective effect of eye color relative to hair color may be due to the higher phenotypic contribution of light diffraction genes (*23*). Finally, we find also that proxies of skin color, such as non-melanoma skin cancers (*25*) (OR=1.07 [1.02-1.12]) and low tanning response (*26*) (OR=1.07 [1.02-1.13]) are significantly associated in MR models with higher AMD risk. We strongly believe that these examples are simply proxies for fundamental pigmentation processes and approximate the presence of melanin in the RPE and choroid.

medRxiv preprint doi: https://doi.org/10.1101/2022.08.16.22278855; this version posted August 21, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. This article is a US Government work. It is not subject to copyright under 17 USC 105 and is also made available for use under a CC0 license.

Multi-ancestry GWAS for AMD identifies the first loci in admixed Black and Hispanic-Latino populations

We conducted the first GWAS on AMD in individuals of African and Hispanic/Latino ancestries. In the AA GWAS (2,302 cases and 29,223 controls), a primary locus centered at *HLA-DQB1* (index SNP rs3844313; OR=1.41 [1.30-1.52]; $p=9.1\times10^{-18}$) reached genome-wide significance **(Figs. 3b and S6)**. Supporting the AA result, the *HLA-DQB1* locus was significant in the EA GWAS at a direction-consistent but lower OR **(Fig. 5a; Table S11)**. In contrast to the EA population, much smaller effect sizes were observed in the AA GWAS at major AMD risk loci *CFH*, *ARMS2*, and *CFB/C2* (**Fig. 5a**). However, EA index markers have similar allele frequencies in the MVP AA sample.

We constructed a new AA-specific PRS based on the AA GWAS and used it to score unrelated AA subjects in MVP that were not included in our GWAS cohort. We then performed PheWAS and LabWAS on the AA PRS (N up to 80,345 depending on the phenotype) to better understand the pleiotropy of genetic risk in AA populations. We observed significant increases in median neutrophil count, white blood cell count, neutrophil fraction, and HDL cholesterol, and a decrease in lymphocyte fraction **(Table S12)**. This result demonstrates that lipid metabolism and immune pathways continue to play a role in AMD pathogenesis in AA despite divergent genetic architecture.

In the HA GWAS (1,545 cases and 10,930 controls), *CFH* and *ARMS2* reached genome-wide significance **(Fig. 3c and S6; Table S13)**, despite smaller sample size than AA. The HA GWAS showed effect sizes more comparable to the EA estimates, though still generally smaller at the major EA loci **(Fig. 5b)**. In summary, AMD risk is attenuated in AA and HA which we ascribe at least partly to genetic precursors, with AA showing the most dissimilar profiles to EA.

medRxiv preprint doi: https://doi.org/10.1101/2022.08.16.22278855; this version posted August 21, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. This article is a US Government work. It is not subject to copyright under 17 USC 105 and is also made available for use under a CC0 license.

Multi-ancestry meta-analysis and fine-mapping identifies three additional novel loci and candidate causal variants

To capture cross-ancestry effects and to increase discovery power, we conducted a multi-ancestry meta-analysis, combining our EA meta-analysis, EA replication GWAS, and AA and HA GWAS reaching a total of 61,248 cases and 364,472 controls. A total of 62 loci were genome-wide significant, including three novel loci beyond that reported above for the EA meta-analysis: *LRP2*, *RP1L1*, and complement *C5* **(Table S14; Fig. S7)**, and all significant EA loci except for *ADAM19*. *LRP2* encodes the transmembrane low-density lipoprotein receptor megalin, which is expressed in the RPE (*27*, *28*), and *RP1L1* encodes a retina-specific protein involved in photoreceptor differentiation and function (*29*).

We leveraged differential LD structure across ancestries, along with multi-model expression data, to identify candidate causal variants. We fine-mapped genome-wide significant loci in the multi-ancestry and EA discovery meta-analyses using the program FINEMAP (*30*) **(Tables S15-S16)**. To co-localize eQTLs, multivariate multiple QTL (mmQTL) (*31*) was first applied to detect eQTL, and eCAVIAR (*32*) was then used to obtain a co-localization posterior probability (CLPP) for each GWAS credible set variant. Based on co-localization with the multi-ancestry meta-analysis, we identified seven candidate causal variants modifying the gene expression of *CD46, RDH5*, *BLOC1S1, FAM227A*, *CFI*, *TNFRSF10A*, and *ABCA1* **(Fig. S8; Table S17)**. Expression co-localization of the variant with the highest CLPP, *TNFRSF10A*, is shown in **Fig. S8a**. *TNFRSF10A* encodes for a receptor, which is activated by tumor necrosis factor and induces cell apoptosis.

medRxiv preprint doi: https://doi.org/10.1101/2022.08.16.22278855; this version posted August 21, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. This article is a US Government work. It is not subject to copyright under 17 USC 105 and is also made available for use under a CC0 license.

Genetic risk at *CFH* in admixed Black and Hispanic/Latino populations comes from European haplotypes

We conducted local ancestry analyses to investigate the contribution of genetic risk at *CFH* in admixed individuals. We inferred local ancestry (*33*), and extracted ancestry-specific imputed dosages and haplotype counts. On average, the MVP AA cohort had 81% African (AFR) and 19% European (EUR) admixture, and the MVP HA cohort had 31% Native American (NAT), 61% EUR, and 8% AFR admixture. We used Fisher's Exact Test to test for significant differences in the proportion of cases and controls with EUR haplotypes at *CFH* **(Table S18)**. We conducted a 2x2 test on haplotypes in AA subjects, which was statistically significant (OR=1.12 [1.04-1.20]; p=0.0033) and a 3x2 test on haplotypes in admixed HA subjects (AFR/EUR/NAT x case/control) which was also statistically significant ($p=1.7\times10^{-5}$).

The *CFH* locus was tested using the Tractor method (*34*), a local ancestry-aware GWAS model which partitions the contributions to genetic risk according to the ancestry of origin of each haplotype to obtain ancestry-specific marginal effect size estimates **(Fig. 6)**. In AA subjects, the conventional GWAS found a modest effect at the established risk polymorphism *CFH* Y402H (rs1061170) with OR=1.13 [1.06-1.20, $p=1.8\times10^{-4}$]. However, the Tractor GWAS found differential effect across haplotype ancestry, with AFR haplotypes ("AFR tract") having OR=1.07 [1.00-1.15; p=0.045] and EUR haplotypes ("EUR tract") having OR=1.40 [1.23-1.60; $p=7.1\times10^{-7}$] **(Fig. 6c)**. Moreover, the genetic architecture of the EUR tract was similar to the analysis conducted within EA individuals **(Fig. 6a)**. Within HA subjects **(Fig. 6b-c)**, however, the conventional GWAS demonstrated a large effect (OR=1.49 [1.37-1.63]; $p=2.4\times10^{-19}$) similar to that in EA only. Examining the risk by local ancestry, the EUR and the native American (NAT) tracts demonstrated a large effect size, whereas the AFR tract did not (OR=1.04 [0.78-1.39; p = 0.79]), consistent with the AFR tract in the AA population. However, the NAT tract had much smaller allele frequency

medRxiv preprint doi: https://doi.org/10.1101/2022.08.16.22278855; this version posted August 21, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. This article is a US Government work. It is not subject to copyright under 17 USC 105 and is also made available for use under a CC0 license.

(3% in NAT vs 38% in EUR haplotypes). Thus, most of the risk at *CFH* in the AA and HA GWAS arises from EUR haplotypes.

Smaller *CFH* Y402H effect size observed in AFR could be explained through haplotype analysis. In particular, a protective deletion of *CFHR3* and *CFHR1* (*35*) has a higher frequency in AFR haplotypes (40% compared to 20% in EUR haplotypes) (*36*) and has been proposed as an explanation for lower rates of AMD observed in AA (*37*). As no SNPs tag the *CFHR3-CFHR1* deletion with $r^2 > 0.8$ in AFR haplotypes, we called the deletion directly from genotype intensities and phased the deletion with the genotypes. We found that the deletion is largely out of phase with the *CFH* Y402H risk allele (D'=0.87) in AFR haplotypes; thus, it cannot fully explain the reduced effect size. Stratified by local ancestry, haplotypes in AA were analyzed using two models (*38*, *39*), excluding SNPs that are rare in AFR haplotypes **(Table S19)**. While both models suggest that the putative risk haplotype may have a lower frequency in AFR compared to EUR, neither model demonstrated enrichment in cases with homozygous AFR local ancestry; they did, however, find haplotypes with enrichment in cases with homozygous EUR local ancestry. Thus, our analysis suggests a smaller effect size in AFR haplotypes, which may contribute to the lower incidence of AMD in AA; we acknowledge that the *CFH* locus comprises a complex genetic architecture (*40*).

Genetic risk at *ARMS2* is lower in admixed Black subjects, but does not vary by haplotype ancestry

Similarly, we dissected the local ancestry contribution of genetic risk at the *ARMS2-HTRA1* locus, which also demonstrated marked cross-ancestry heterogeneity. In contrast to *CFH*, we did not find significant differences in the proportion of European haplotypes at *ARMS2* in admixed African or Hispanic subjects using Fisher's Exact Test **(Table S20)**. We tested the *ARMS2* locus using the Tractor method **(Fig. 6d)**. Within AA subjects, OR at the *ARMS2* A69S

medRxiv preprint doi: https://doi.org/10.1101/2022.08.16.22278855; this version posted August 21, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. This article is a US Government work. It is not subject to copyright under 17 USC 105 and is also made available for use under a CC0 license.

polymorphism were close to 1 and not significant in both the AFR and EUR tracts. However, within admixed HA subjects, EUR, NAT, and AFR local ancestry tracts surprisingly all showed similar statistically significant OR. We hypothesized that risk at *ARMS2* was partially dependent on genetic architecture at *CFH*, and isolated admixed AA subjects with EUR/EUR local ancestry at *CFH*. We observed OR in AFR and EUR tracts similar to the EA and HA GWAS, with the AFR tract reaching statistical significance (p=0.0026). In contrast, OR were close to 1 in both AFR and EUR tracts in subjects with homozygous AFR local ancestry at *CFH*. Thus, our results suggest that the causal variant at *ARMS2* has similar effect size across ancestries, and the lack of association in AA may be due to epistasis or pathway effects related to the conspicuous lack of complement loci in the AA GWAS. Supporting this hypothesis, we modeled epistatic interaction between the *CFH* and *ARMS2* risk alleles in MVP EA **(Table S21)** and observed a significant association of AMD with the interaction term (OR=1.11 [1.08-1.14]; $p=1.6\times10^{-13}$).

An MHC Class II haplotype is associated with AMD in African and European ancestries

As the AA GWAS demonstrated a primary association with the MHC and not *CFH* and *ARMS2* as observed in other ancestries, we dissected the MHC signal by testing HLA alleles imputed with HIBAG (*41*). Full summary statistics are provided in **Tables S22-S24**. Within MVP EA (32,567 cases and 130,444 controls), we replicated HLA allele associations (*42*) with DQB*02:01 ($p=1.2\times10^{-7}$), DQB*02:02 ($p=4.7\times10^{-10}$), and DRB*03:01 ($p=2.5\times10^{-7}$); 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). Within AA subjects (2,296 cases and 28,478 controls), three Class II alleles were highly significant: DRB1*07:01 (1.52 [1.37-1.68]; p=5.2×10-

medRxiv preprint doi: https://doi.org/10.1101/2022.08.16.22278855; this version posted August 21, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. This article is a US Government work. It is not subject to copyright under 17 USC 105 and is also made available for use under a CC0 license.

^{15}), DQA1*02:01 (1.39 [1.26-1.52]; p=2.4×10^{-11}), and DQB1*02:02 (1.39 [1.27-1.52]; p=2.1×10^{-12}). These three alleles are known to form a haplotype (HLA-DRB1*07:01-DQA1*02:01-DQB1*02:02) that is associated with autoimmune conditions such as celiac disease (*43*) and asparaginase hypersensitivity (*44*, *45*). In an additive model, the risk haplotype was highly associated with AMD, with OR=1.51 [1.37-1.66; p=1.8×10^{-16}] in AA and replicated with a smaller effect size, OR=1.16 [1.12-1.20; p=2.1×10^{-15}], in EA, with all other haplotypes as reference. A PheWAS of the risk haplotype in AA led to significant enrichment of diagnostic signals for drusen (p=1.2×10^{-26}) and dry AMD (p=4.7×10^{-9}), which were corroborated in the EA analysis (p=6.7×10^{-4} and p=2.4×10^{-3}, respectively). Additionally, we replicated the celiac disease association in the EA cohort (p=1.0×10^{-12}). Antibodies directed against HLA-DR showed immunoreactivity in drusen (*46*), supporting biological plausibility.

Multi-tissue TWAS of AMD identifies novel AMD-associated genes

To pair the GWAS signals with functional gene units, predictive models of gene expression (*47*) were used to perform a multi-tissue transcriptome-wide association study (TWAS), based on our EA discovery meta-analysis (N=45,045 cases, 275,806 controls). We performed TWAS using the high-powered dorsolateral prefrontal cortex (DLPFC) dataset from the PsychENCODE Consortium (PEC) (*48*) **(Table S25)**, as well as retina (*49*) **(Table S26)**, and 35 other tissue models built from GTEx v8 (*50*) and STARNET (*51*). We then meta-analyzed tissue p-values using ACAT (*52*) **(Table S27)**. AMD-related gene expression was highly correlated across tissues **(Fig. S9a)**. Consistent with the high protein and gene expression similarity among retina and brain (Human Protein Atlas, proteinatlas.org) (*53*), we found that the retina TWAS clustered together with the TWAS in DLPFC and other brain tissues **(Fig. S9b)**.

Applying the DLPFC model, 156 genes and 326 transcripts were significant at the FDR-corrected p<0.05 association threshold **(Fig. 3a)**. Novel TWAS genes include regulatory genes for protein glycosylation (*MAN2C1*, *MPI*, *TMEM199*, and *CTSA*), lipid metabolism (*CHD9*), cell

medRxiv preprint doi: https://doi.org/10.1101/2022.08.16.22278855; this version posted August 21, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. This article is a US Government work. It is not subject to copyright under 17 USC 105 and is also made available for use under a CC0 license.

cycle (*FRK*), and neurogenesis and neural differentiation (*NTN5* and *NIF3L1*). Significant genes not at a genome-wide significant GWAS locus include *CAB39*, *CUL1* antisense transcript *AC005229.4* (ENSG00000273314), and *WAC-AS1* **(Fig. S10)**. Along with the novel GWAS locus near *CAND2*, *CUL1* and *WAC* are components of the ubiquitin-proteasome system, suggesting a larger role for that pathway in AMD than previously appreciated.

At *CD46/CD55* we observed significant downregulation of the *CD55*-203 transcript and upregulation of *CD46* overall, with differential regulation of some *CD46* transcripts **(Fig. 5c)**, likely corresponding to a strong splice QTL at the index SNP (rs2724360) observed in many tissues in GTEx ($p=1.3\times10^{-25}$ in DFPLC). We further performed retina summary-based MR (SMR) experiments **(Table S28)**, which provide support for a causal link between AMD risk variants and expression of *CD46*, but not *CD55*.

We then conducted gene set enrichment analysis (GSEA) on significant TWAS genes in the DLPFC-based model **(Table S29; Fig. S11a)** and the tissue meta-analysis **(Table S30; Fig. S11b)**. In addition to complement cascade, humoral immunity, and HDL cholesterol regulation pathways, we observed significant enrichment in genes belonging to pathways related to regulation of immune and inflammatory responses, and death receptor-mediated apoptosis. We further performed semantic clustering of significant GO terms in the tissue meta-analysis GSEA, which identified two major themes associated with AMD pathology: immune functions and lipid homeostasis **(Fig. S12)**. Finally, we conducted gene-based rare variant burden analyses of the MVP EA cohort **(Table S31)**, which confirmed the risk-increasing effect of rare *CFH* and *CFI* mutations, and performed enrichment using our TWAS results **(Fig. S13)**. In addition to *CFH* and *CFI,* genes prioritized by this analysis include *CETP, ABCA7, ELP5* and *B3GLCT*.

medRxiv preprint doi: https://doi.org/10.1101/2022.08.16.22278855; this version posted August 21, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. This article is a US Government work. It is not subject to copyright under 17 USC 105 and is also made available for use under a CC0 license.

Discussion

We performed the largest GWAS meta-analysis to date and the first GWAS of AMD in AA and HA populations using data from veterans participating in MVP, nearly quadrupling the number of cases and doubling associated loci. Importantly, we uncovered marked differences in risk between major AMD loci across populations even at loci with known large effect sizes, especially *CFH* and *ARMS2/HTRA1*, which may partially explain lower rates of AMD in these groups. We submit that the overall reduction in risk in AA is due to tempered involvement of the complement pathway, substantiated through overall weak PRS transferability. In contrast to other ancestries, our AA GWAS had a primary peak at *HLA-DQB1*, corresponding to an MHC Class II risk haplotype which was replicated with a smaller effect size in EA subjects. This heterogeneity in genetic architecture may reflect the role of natural selection in shaping immune responses across ancestries (*54*, *55*). However, our PRS-PheWAS experiment using the AA GWAS as a base, which demonstrated associations between the AA AMD PRS and increases in both HDL cholesterol and neutrophil counts, points to shared disease mechanisms.

We dissected cross-ancestry heterogeneity at the *CFH* and *ARMS2* loci using haplotype-based local ancestry analyses, “painting” the ancestral origin of each chromosome. These analyses revealed that AFR haplotypes have a smaller marginal effect size at the *CFH* risk allele, compared to EUR and NAT haplotypes in the same individuals, in both AA and HA populations. In contrast with *CFH*, the lack of association at *ARMS2* in AA is not due to ancestry-based haplotype-specific effects. In AA individuals, both AFR and EUR haplotypes demonstrated similarly small effect sizes. However, in HA individuals, which show high risk at *ARMS2*, AFR, EUR, and NAT haplotypes all showed similarly large marginal effect sizes. We infer from these observations a diminished role for the *ARMS2/HTRA1* locus in the absence of genetic risk at *CFH*. We further leveraged the statistical power offered by MVP to confirm a robust interaction effect between the *CFH* and *ARMS2* risk alleles in EA.

medRxiv preprint doi: https://doi.org/10.1101/2022.08.16.22278855; this version posted August 21, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. This article is a US Government work. It is not subject to copyright under 17 USC 105 and is also made available for use under a CC0 license.

A general conclusion from years of investigative efforts is that the genetic architectures at the broader *CFH* and *ARMS2/HTRA1* loci are multifaceted (*40*, *56*). At *CFH*, both common and rare variation (*8*) as well as structural variation (*35*) 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* (*57*). HTRA1, a serine protease with reported roles in TGF-β signaling (*58*) and regulation of monocyte elimination (*59*), has also been shown to cleave neutrophil elastase inhibitor α-1 antitrypsin (*SERPINA1*) (*60*), producing NET inhibitory peptides, or NIPs (*19*, *61*). NETs activate complement (*62*–*64*), and CFH itself may both regulate NETosis (*65*) and bind to NETs (*66*), 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 histological (*67*) and observational (*68*) 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.

Other loci to emerge in our GWAS meta-analysis continue to expand on known AMD themes, including numerous novel loci related to the complement cascade, lipid metabolism, angiogenesis, TGF-β signaling, photoreceptor function, apoptosis, and inflammation. We also identified the first pigmentation loci associated with AMD: *HERC2/OCA2*, *TYR*, and *TRPM1*. We explored this theme further by demonstrating for the first time a protective effect of pigmentation on AMD in EA using MR, supporting a causal relationship where long-standing observational evidence has been inconsistent (*69*).

To pinpoint causal changes in genetically regulated gene expression, we performed multi-tissue TWAS and SMR experiments. Our analysis adds transcriptomic imputation models from

medRxiv preprint doi: https://doi.org/10.1101/2022.08.16.22278855; this version posted August 21, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. This article is a US Government work. It is not subject to copyright under 17 USC 105 and is also made available for use under a CC0 license.

new cohorts (PsychENCODE and STARNET), expands sample sizes in GTEx (GTEx v8 vs. GTEx v7), leverages a higher-powered GWAS, and introduces expression information at the transcript level in the DLPFC model. A comparison with previous TWAS studies (*49*, *70*) is summarized in **Table S32**. TWAS enabled us to refine several loci, such as at *CD46/CD55*, where we identified specific *CD46* transcripts associated with AMD. Our TWAS analysis also enabled us to identify numerous additional novel AMD-associated genes across several pathways, including ubiquitin-proteasome system, protein glycosylation, cell cycle regulation, and neural differentiation.

While our analysis demonstrated a very high genetic correlation with IAMDGC (95%), one limitation is that our phenotyping approach used only ICD code-based diagnoses and demographic parameters (*71*), which limited our ability for deep phenotyping. An important future direction will be to review ophthalmologic imaging data to assess disease severity and pathology in patients diagnosed with AMD and associate findings with specific genetic risk factors, such as the HLA risk haplotype in AA. To date, most large-scale imaging studies of AMD have focused on EA populations, and as our results demonstrate, increasing diversity is important.

In conclusion, our study expands the number of genetic loci associated with AMD and fills in important gaps in the literature regarding the genetics of AMD in non-European ancestries. Moreover, it illustrates the importance of considering diverse admixed genomes and using local ancestry-aware analyses (*34*), which can leverage differences in effect size, frequency, and LD to tease apart mechanisms of genetic risk and improve risk prediction.

medRxiv preprint doi: https://doi.org/10.1101/2022.08.16.22278855; this version posted August 21, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. This article is a US Government work. It is not subject to copyright under 17 USC 105 and is also made available for use under a CC0 license.

Materials and Methods

Samples

A general description of the MVP sample may be found in Gaziano et al. (*9*). Composition of the cohorts studied here is available in **Table S1.** Informed consent was obtained from all participants, and all studies were performed with approval from the Institutional Review Boards at participating centers, in accordance with the Declaration of Helsinki.

MVP

We used ICD9/ICD10 codes in VA electronic health records to define AMD case/control status, using Algorithm 4 in Halladay et al. (*71*) in which AMD cases and controls were required to be 50 and 65 years of age, respectively. Briefly, AMD cases had ICD9/ICD10 codes indicating AMD from at least two separate eye clinic visits; controls had either no eye clinic visits or no AMD diagnoses. The three EA cohorts represent three separate releases of the MVP genotype data.

A description of the custom MVP Affymetrix Biobank chip, and general quality control (QC) procedures for genetic data, have been published (*72*). Briefly, the MVP sample was genotyped for a custom Affymetrix Biobank chip, containing 725,000 variants. Samples were imputed to the 1000 Genomes Phase 3 panel. AA and HA samples were additionally imputed to the African Genome Resources imputation panel from the Sanger Institute (*73*) to improve imputation of African haplotypes. Imputed genotypes were available for a total of 658,000 Veterans across all ethnic groups. We used HARE categories (*74*) AFR (African American), EUR (European American) and HIS (Hispanic American) to harmonize genetic ancestry with self-identified race/ethnicity. Principal components for ancestry were determined within HARE groups.

Genentech

Study Design and Populations. We performed whole-genome sequencing of DNA derived from blood samples obtained from patients with choroidal neovascularization (CNV) participating

medRxiv preprint doi: https://doi.org/10.1101/2022.08.16.22278855; this version posted August 21, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. This article is a US Government work. It is not subject to copyright under 17 USC 105 and is also made available for use under a CC0 license.

in clinical trials for Ranibizumab (NCT00891735 [HARBOR], NCT00061594 [ANCHOR] and NCT00056836 [MARINA]) and from patients with geographic atrophy (GA) participating in clinical trials for Lampalizumab (NCT02247479 [CHROMA], NCT02247531 [SPECTRI], NCT01229215 [MAHALO]) and an observational study (NCT02479386 [PROXIMA]). Patients had to consent for genetic analysis for inclusion eligibility and these cohorts were selected for inclusion based on available phenotypic information and DNA availability for whole-genome sequencing.

Two independent sets of non-AMD controls were used for the GA risk analysis and the CNV risk analysis. Samples and data for non-AMD controls without CNV were obtained from clinical trial cohorts of asthma and RA. Samples and data for non-AMD controls without GA were obtained from clinical trial cohorts of asthma, colorectal cancer, COPD, inflammatory bowel disease, IPF and RA. No healthy controls were used because no whole-genome sequencing data was available. All non-AMD controls were aged 50 or older and of European ancestry.

All patients (AMD and non-AMD controls) included in this study provided written informed consent for whole-genome sequencing of their DNA. Ethical approval was provided as per the original clinical trials.

DNA analysis. The whole-genome sequencing data was generated to a read depth of 30X using the HiSeq platform (Illumina X10, San Diego, CA, USA) processed using the Burrows-Wheeler Aligner (BWA) / Genome Analysis Toolkit (GATK) best practices pipeline. Whole genome sequencing short reads were mapped to hg38 / GRCh38 (GCA_000001405.15), including alternate assemblies, using BWA version 0.7.9a-r786 to generate BAM files. All sequencing data was subject to quality control and was checked for concordance with SNP fingerprint data collected before sequencing. After filtering for genotypes with a GATK genotype quality greater than 90, samples with heterozygote concordance with SNP chip data of less than 75% were removed. Sample contamination was determined with VerifyBamID software. And samples with a freemix parameter of more than 0.03 were excluded. Joint variant calling was done using the GATK best practices joint genotyping pipeline to generate a single variant call format (VCF) file.

medRxiv preprint doi: https://doi.org/10.1101/2022.08.16.22278855; this version posted August 21, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. This article is a US Government work. It is not subject to copyright under 17 USC 105 and is also made available for use under a CC0 license.

The called variants were then processed using ASDPEx to filter out spurious variant calls in the alternate regions.

Quality Control. Samples were then excluded if the call rate was less than 90%. Identity by descent analysis was used to detect and filter out relatedness in the dataset; samples were excluded if PI_HAT was 0.4 or higher. Samples were removed if they showed excess heterozygosity with more than three standard deviations of the mean. This resulted in 1,703 GA patients with 2,611 non-GA controls, and 1,175 CNV patients with 3,225 non-CNV controls.

Sample genotypes were set to missing if the Genotype Quality score was less than 20 and SNPs were removed if the missingness was higher than 5%. SNPs were filtered if the significance level for the Hardy-Weinberg equilibrium test was less than 5×10^{-8}. The allele depth balance test was performed to test for equal allele depth at heterozygote carriers using a binomial test; SNPs were excluded if the p-value was less than 1×10^{-5}.

GWAS

MVP EA. GWAS on the first two EA tranches of the MVP cohort were conducted in PLINK 2.0, adjusting for sex and six ancestry-specific principal components.

MVP AA and HA. Individuals were classified by genetic ancestry using the HARE method (*74*) which harmonizes self-reported ethnicity with genetic ancestry. We conducted GWAS in MVP AA and HA populations with REGENIE v1.6.7 (*75*) using sex and 10 ancestry-specific PCs as covariates. We used the approximate Firth mode (--firth --approx) with a p-value threshold of 0.05 to control type I error rate in rare SNPs. We additionally used option --firth-se to generate standard errors computed from the Firth effect size and likelihood ratio test p-value for meta-analysis purposes.

MVP EA replication GWAS. We conducted a GWAS in a third tranche of MVP data that later became available to us. To ensure that novel associations in the EA meta-analysis were robust to population structure, we conducted the EA replication GWAS in REGENIE v2.2.4 (*75*)

medRxiv preprint doi: https://doi.org/10.1101/2022.08.16.22278855; this version posted August 21, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. This article is a US Government work. It is not subject to copyright under 17 USC 105 and is also made available for use under a CC0 license.

using sex and ten ancestry-specific PCs as covariates. We used the approximate Firth mode (--firth --approx) with a p-value threshold of 0.05 to control type I error rate in rare SNPs. We additionally used option --firth-se to generate standard errors computed from the Firth effect size and likelihood ratio test p-value for meta-analysis purposes.

IAMDGC. Phenotyping and GWAS analysis in IAMDGC is described in detail elsewhere (*8*). Briefly, European-ancestry subjects were genotyped on a custom Illumina HumanCoreExome array and imputed to the 1000 Genomes Phase 1 panel. GWAS analyses were conducted using EPACTS using DNA type (whether whole genome-amplified or not) and two PCs as covariates.

GERA. Phenotyping and sample collection in GERA are described in detail elsewhere (*42*, *76*). Cases and controls were selected from European-ancestry subjects. Samples were genotyped on an Affymetrix Axiom array and imputed to the 1000 Genomes panel. GWAS were conducted in PLINK 1.07 using age, gender, PCs 1-10 (calculated within European-ancestry subjects), and percentage of Ashkenazi ancestry as covariates.

UK Biobank. Cases and controls were selected from subjects of at least 65 years of age. Samples were genotyped on the UK BiLEVE Axiom Array or the UK Biobank Axiom Array and imputed to the Haplotype Reference Consortium (HRC) reference panel. GWAS analyses were conducted in PLINK 1.9 using sex and the first two PCs as covariates.

Genentech GA and CNV. Common variant (MAF>=1%) genome-wide association studies (GWAS) were conducted to separately assess GA risk and CNV risk as compared to independent non-AMD control cohorts. PLINK was used to perform logistic regression using an additive model, adjusting for age, sex and the first three principal components.

Statistical Analysis

Epidemiological analyses. Observational association analyses were conducted using the statsmodels (version 0.12.2) Python module. Case-control analyses were performed with logistic regression. Time-to-AMD-diagnosis analysis was performed using a Cox proportional hazards

medRxiv preprint doi: https://doi.org/10.1101/2022.08.16.22278855; this version posted August 21, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. This article is a US Government work. It is not subject to copyright under 17 USC 105 and is also made available for use under a CC0 license.

model. The index date was defined as the date of first recorded visit to the VA healthcare system. An event was defined as a subject receiving AMD diagnoses on two unique days. The time-to-event was defined as the time from the first recorded visit to the first recorded AMD diagnosis. Subjects with fewer than two unique days with AMD diagnoses were right-censored by the date of the last recorded VA healthcare system visit.

GWAS Meta-analysis. GWAS meta-analyses were conducted using the inverse variance-weighted fixed effects scheme as implemented in METAL (*77*).

Additional QC. We imposed the following criteria for the GWAS and meta-analysis: imputation quality score ≥ 0.5, minor allele frequency ≥ 0.005, and $p \geq 1\times10^{-50}$ from an exact test for Hardy-Weinberg proportions.

Locus definition. Association loci were defined using the FUMA GWAS online tool (*78*) with default options (merging together LD blocks closer than 250 kb into a locus).

IAMDGC-based polygenic risk score analysis. We derived AMD PRS scores for subjects in MVP across all ancestries using the IAMDGC GWAS (*8*) as a base. Weights were derived from the base GWAS using PRS-CS (*10*) with a global shrinkage prior of 1×10^{-4} (the default recommended for less-polygenic traits), using the UK Biobank European-ancestry LD reference. Imputed SNPs meeting $R^2 > 0.8$ and minor allele frequency > 1% QC criteria were included in the model. Scores were then normalized to zero mean and unit variance within each ancestry group.

AA GWAS-based polygenic risk score analysis. To better understand the pleiotropy of AMD genetic risk in African ancestries, we used the AA summary statistics as the base for scoring polygenic risk in unrelated MVP AA subjects that were not included as cases or controls in the GWAS cohort. As our phenotyping approach required subjects to have visited a VA eye clinic and been evaluated by an eye specialist, the PRS was evaluated in the remaining unrelated AA subjects with no such history. Weights were derived from the base MVP AA GWAS using PRS-CS (*10*) with a global shrinkage prior of 1×10^{-4} (the default recommended for less-polygenic traits),

medRxiv preprint doi: https://doi.org/10.1101/2022.08.16.22278855; this version posted August 21, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. This article is a US Government work. It is not subject to copyright under 17 USC 105 and is also made available for use under a CC0 license.

using the UK Biobank African-ancestry LD reference. Scores were then normalized to zero mean and unit variance.

Conditional association analysis. We determined a set of mutually independent genome-wide significant variants using the conditional analysis procedure implemented in GCTA-COJO (*79*). A subset of 100,000 MVP EA individuals (as classified by the HARE method) formed the reference population for LD structure. Independent significant variants were chosen over all loci simultaneously by means of the --cojo-slct procedure for stepwise regression. Selected variants had association $p < 5\times10^{-8}$ while adjusting for all other variants and were not necessarily pairwise independent.

HLA analysis. Four-digit HLA alleles were imputed in all subjects in MVP using HIBAG (*41*) with the Axiom UK Biobank Array multi-ethnic model as reference. Allele calls with posterior probability less than 50% were filtered out. Alleles were tested additively relative to all other alleles at each locus using the logistf R package for Firth logistic regression (*80*). Two models were applied: the 'standard' model with adjustment for sex and PCs 1-10, and the 'fully adjusted' model with adjustment for sex, PCs 1-10, and major AMD risk alleles at *CFH*, *ARMS2*, *CFB/C2*, and *C3*. Alleles with a minor allele count less than 50 were not tested. P-values and confidence intervals are based on the penalized likelihood ratio test and profile penalized likelihood, while the standard error and z-score are based on the Wald method.

*HLA-DRB1*07:01-DQA1*02:01-DQB1*02:02 haplotype analysis*. Two SNPs (rs28383172 and rs7775228) were recently reported to tag the haplotype with high sensitivity and specificity (*81*), both of which are genotyped on the MVP array. We phased genotypes using SHAPEIT4 (*82*) and designated the risk haplotypes as those with G and C alleles, respectively.

Local ancestry analysis. We inferred local ancestry within our AA sample assuming two-way (AFR/EUR) admixture, and within our HA sample assuming three-way (AFR/EUR/NAT) admixture. The 1000 Genomes YRI (N=108) and CEU (N=99) populations, were used as the AFR and EUR reference, respectively, and 43 Native American samples from Mao et al. (*83*, *84*) were

medRxiv preprint doi: https://doi.org/10.1101/2022.08.16.22278855; this version posted August 21, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. This article is a US Government work. It is not subject to copyright under 17 USC 105 and is also made available for use under a CC0 license.

used as the NAT reference. We used RFMIX version 2 (*33*) to generate local ancestry calls for phased genotypes. We then extracted ancestry-specific dosages from the imputed data into PLINK 2.0-compatible files (*85*) using custom scripts based on the Tractor workflow (*34*). For the AA analysis, EUR-specific dosages were put into a PGEN file, and AFR-specific dosages and EUR haplotype counts were interlaced in a zstandard-compressed table. For the HA analysis, EUR-specific dosages were put into a PGEN file, with AFR and NAT-specific dosages and EUR and AFR haplotype counts interlaced into a zstandard-compressed table. We used these files to conduct a local ancestry-aware GWAS using the PLINK 2.0 local covariates feature, obtaining ancestry-specific marginal effect size estimates.

CFHR3-CFHR1 deletion calling. We calculated log-R ratio intensities for 12 probesets spanning the *CFHR3* and *CFHR1* deletion. We then clustered intensities by constructing a two-dimensional UMAP embedding (*86*), which separated samples into 0, 1, and 2-copy number clusters. In order to conduct haplotype analyses, the deletion was phased with the genotype data using SHAPEIT 4.2 (*82*), and the locus was re-imputed using Minimac4 (*87*) to obtain accurate deletion calls alongside imputed genotypes.

Rare variant gene-based analysis. We conducted gene-based rare variant burden association analyses of the entire MVP EA cohort (N=34,046 cases, 135,775 controls). We considered only variants genotyped on the MVP 1.0 array (*72*), which is enriched in protein-altering rare variants, and applied a recent technical advance (*88*) that improves the positive predictive value of rare genotype calls. We further restricted the included markers to ultra-rare variants (MAF < 0.25% in controls) classified as “high-impact” (*89*). We then defined a series of three “masks” over which the burden test was performed: mask 1 including frameshift and nonsense loss of function (LoF) variants only, mask 2 including LoF + missense variants, and mask 3 including LoF + missense + splice site variants. Burden test association analyses were conducted in REGENIE v2.02 with Firth logistic regression (*75*).

medRxiv preprint doi: https://doi.org/10.1101/2022.08.16.22278855; this version posted August 21, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. This article is a US Government work. It is not subject to copyright under 17 USC 105 and is also made available for use under a CC0 license.

Relatedness. For experiments in unrelated samples, we used KING version 2.2.7 (*90*) to generate a maximal set of samples that are unrelated within two degrees.

Summary-based MR. We performed summary-based MR experiments on retina eQTL data from EyeGEx using the Summary-based Mendelian Randomization (SMR) software tool (*91*).

Two-sample MR. We performed MR on selected pigmentation traits and AMD using the TwoSampleMR R package (*92*). Instruments were derived from genome-wide significant loci reported in the exposure trait GWASs and clumped using the TwoSampleMR clump_data tool with default options ($r^2 < 0.001$) to ensure no correlation between instruments.

Genetic correlation. Genetic correlation analyses were performed using LDSC 1.01 (*11*) using the provided European-ancestry LD scores derived from the 1000 Genomes project.

Fine-mapping. We performed Bayesian fine-mapping of each genome-wide significant locus in the European-ancestry meta-analysis and trans-ancestry meta-analysis using FINEMAP 1.4 (*30*). Pairwise SNP correlations were calculated directly from imputed dosages on 320,831 European-ancestry samples in MVP using LDSTORE 2.0. The maximum number of allowed causal SNPs at each locus was set to 10 (the default used in the FinnGen fine-mapping pipeline: https://github.com/FINNGEN/finemapping-pipeline). Finally, we performed multivariate multiple QTL (mmQTL) analysis as previously described (*31*), leveraging the multi-ancestry RNA-seq-based brain gene expression reference dataset derived from 2,119 donors (*31*) and the GWAS summary statistics from the AMD multi-ancestry discovery meta-analysis.

PheWAS. We performed PheWAS and LabWAS in unrelated subjects as implemented in the R PheWAS package version 0.12 (*93*). For each ICD code, cases were defined as at least 2 counts of the code on separate days, and controls were 0 counts. Subjects with 1 count were excluded. Additionally, we applied phenotype-based control exclusion (i.e., exclusion of controls that were cases in similar phenotypes) and we excluded females in male-specific phenotypes and vice versa. After this process, only phenotypes with 100 or more cases were considered. For labs,

medRxiv preprint doi: https://doi.org/10.1101/2022.08.16.22278855; this version posted August 21, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. This article is a US Government work. It is not subject to copyright under 17 USC 105 and is also made available for use under a CC0 license.

we used the median value in regression analyses, and required that each individual have at least two lab measurements to be included. Labs with measurements in at least 50 individuals were considered. Lab measurements were then normalized using inverse rank normal transformation (IRNT). All ICD code-based phenotypes were tested with logistic regression, at the Bonferroni-corrected p-value of 3×10^{-5}. All 69 lab-based phenotypes were tested with linear regression, at Bonferroni-corrected p-value of 7×10^{-4}. Both analyses adjusted for age, age-squared, sex (for non-sex specific phenotypes), and 20 ancestry-specific PCs.

TWAS

Transcriptomic imputation model construction. Transcriptomic imputation models are constructed as previously described (*47*, *94*) for tissues of the GTEx v8 (*50*), STARNET (*51*) and PsychENCODE (*48*, *95*) cohorts. For GTEx and STARNET cohorts, we consider adipose tissue: subcutaneous (GTEx & STARNET) and visceral (GTEx & STARNET); arterial tissue: aorta (GTEx & STARNET), coronary artery (GTEx) and mammary artery (STARNET); Blood (GTEx & STARNET); brain (GTEx): anterior cingulate cortex, frontal pole cortex, hippocampus and substantia nigra; adrenal gland (GTEx); colon (GTEx): sigmoid and trasverse; esophagus (GTEx): gastroesophageal junction, mucosa and mascularis; pancreas (GTEx); stomach (GTEx); terminal ileum (GTEx); heart (GTEx): atrial appendage and left ventricle; liver (GTEx & STARNET), skeletal muscle (GTEx & STARNET); mammary tissue (GTEx); ovary (GTEx); lung (GTEx); skin (GTEx): not sun and sun exposed; spleen (GTEx). The genetic datasets of the GTEx (*50*), STARNET (*51*) and PsychENCODE (*95*) cohorts are uniformly processed for quality control (QC) steps before genotype imputation as previously described (*47*, *94*). We restrict our analysis to samples with European ancestry as previously described (*47*). Genotypes are imputed using the University of Michigan server (*87*) with the Haplotype Reference Consortium (HRC) reference panel (*73*). Gene expression information is derived from RNA-seq gene level counts which are adjusted for known and hidden confounders, followed by quantile normalization. For GTEx, we

medRxiv preprint doi: https://doi.org/10.1101/2022.08.16.22278855; this version posted August 21, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. This article is a US Government work. It is not subject to copyright under 17 USC 105 and is also made available for use under a CC0 license.

use publicly available, quality-controlled, gene expression datasets from the GTEx consortium (http://www.gtexportal.org/). RNA-seq data for STARNET were obtained in the form of residualized gene counts from a previously published study (*51*). For the dorsolateral prefrontal cortex from PsychENCODE we used post-quality-control RNA-seq data that were fully processed, filtered, normalized, and extensively corrected for all known biological and technical covariates except the diagnosis status (*48*) as previously described (*94*). Finally, we constructed a retinal transcriptomic imputation model based on reference data from Ratnapriya et al. (*49*) comprising imputed genotypes and expression data for 406 individuals. 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 $p < 1\times10^{-10}$; the remaining variants were pruned (--indep-pairwise 1000 10 0.02 with PLINK (*96*)) and PCA was performed with PLINK 2 (*85*). We used the first (PC1), second (PC2) and third (PC3) ancestral principal components to define an ellipsoid based on 1000Gp3v5 EUR samples (*97*) 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 individual. In the remaining samples (n = 405), we performed additional sample-level quality control by retaining non-related samples (--king-cutoff 0.0884 with PLINK 2 (*85*)) 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; we performed PCA on raw gene counts and excluded individuals located more than 4 SD away from the mean of the ellipsoid defined by PC1 to PC3. The latter led to the exclusion of twelve samples, giving us a final sample size of 393 individuals for the training of the model. In this final set of individuals, we performed variant-level quality control of the genotypes by removing variants with greater than 0.05 minor allele frequency and 0.02

medRxiv preprint doi: https://doi.org/10.1101/2022.08.16.22278855; this version posted August 21, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. This article is a US Government work. It is not subject to copyright under 17 USC 105 and is also made available for use under a CC0 license.

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 GWAS (*73*). After conversion of genotypes to dosages, missing values were replaced with twice the variant's minor allele frequency. Expression data corresponding to the retained individuals were prepared as follows: genes with > 0.5 counts per million in at least 30% of samples were retained, expression values were normalized by voom (*98*), residualized for 10 PEER factors (*99*) and quantile normalized. For training, we used PrediXcan (*100*) for the construction of the retinal transcriptomic imputation model due to a lack of SNP epigenetic annotation information; for all other models, we used EpiXcan (*47*).

Multi-tissue transcriptome-wide association study (TWAS). We performed the gene-trait association analysis as previously described (*47*). Briefly, we applied the S-PrediXcan method (*101*) to integrate the summary statistics and the transcriptomic imputation models constructed above to obtain gene-level association results. Results were corrected for multiple testing with the Benjamini & Hochberg (FDR) method (*102*). P-values across tissues were meta-analyzed using ACAT (*52*) ≤ 0.05 and predictive r^2 > 0.01 (to control both for significance and variance explained).

Gene set enrichment analysis for TWAS results. To investigate whether the genes associated with AMD exhibit enrichment for biological pathways, we use gene sets from MsigDB 5.1 (*103*) and filter out non-protein coding genes, genes located at MHC as well as genes whose expression cannot be reliably imputed. In addition, we assay enrichment for genesets deriving from the rare variant analysis as above. Statistical significance is evaluated with one-sided Fisher's exact test and the corrected p-values are obtained by the Benjamini-Hochberg (FDR) method (*102*). Semantic enrichment analysis of the TWAS results was performed as follows: a) gene ontology (GO) data (*104*, *105*) were retrieved on 2021-07-10; b) we run Fisher-based gene

medRxiv preprint doi: https://doi.org/10.1101/2022.08.16.22278855; this version posted August 21, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. This article is a US Government work. It is not subject to copyright under 17 USC 105 and is also made available for use under a CC0 license.

set enrichment analysis as above and kept enrichments with FDR-corrected p-value ≤ 0.1; c) we used GOGO (*106*) to generate the directed acyclic graphs (DAGs) and estimated all possible pairwise semantic similarities when both GO terms of each pair fell within one of the three DAG domains: molecular function ontology (MFO), biological process ontology (BPO), and cellular component ontology (CCO); d) to characterize/illustrate the semantic clusters deriving from the significantly enriched gene sets, we converted semantic similarity between GO terms to distance and determined the optimal number of clusters with NbClust (*107*) using the ward.D2 method (*108*) according to the Silhouette criterion (*109*); e) finally, we built a word cloud for each semantic cluster by mining text terms (*110*) from its members' GO term titles.

medRxiv preprint doi: https://doi.org/10.1101/2022.08.16.22278855; this version posted August 21, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. This article is a US Government work. It is not subject to copyright under 17 USC 105 and is also made available for use under a CC0 license.

References:

1. M. Fleckenstein, T. D. L. Keenan, R. H. Guymer, U. Chakravarthy, S. Schmitz-Valckenberg, C. C. Klaver, W. T. Wong, E. Y. Chew, Age-related macular degeneration. *Nat Rev Dis Primers*. **7**, 31 (2021).

2. B. L. Vanderbeek, D. N. Zacks, N. Talwar, B. Nan, D. C. Musch, J. D. Stein, Racial differences in age-related macular degeneration rates in the United States: a longitudinal analysis of a managed care network. *Am. J. Ophthalmol.* **152**, 273–282.e3 (2011).

3. M. A. Mahr, D. O. Hodge, J. C. Erie, Racial Differences in Age-Related Macular Degeneration and Associated Anti-Vascular Endothelial Growth Factor Intravitreal Injections among Medicare Beneficiaries. *Ophthalmol Retina*. **2**, 1188–1195 (2018).

4. D. Wang, Y. Jiang, M. He, J. Scheetz, W. Wang, Disparities in the Global Burden of Age-Related Macular Degeneration: An Analysis of Trends from 1990 to 2015. *Curr. Eye Res.* **44**, 657–663 (2019).

5. J. L. Haines, M. A. Hauser, S. Schmidt, W. K. Scott, L. M. Olson, P. Gallins, K. L. Spencer, S. Y. Kwan, M. Noureddine, J. R. Gilbert, N. Schnetz-Boutaud, A. Agarwal, E. A. Postel, M. A. Pericak-Vance, Complement factor H variant increases the risk of age-related macular degeneration. *Science*. **308**, 419–421 (2005).

6. R. J. Klein, C. Zeiss, E. Y. Chew, J.-Y. Tsai, R. S. Sackler, C. Haynes, A. K. Henning, J. P. SanGiovanni, S. M. Mane, S. T. Mayne, M. B. Bracken, F. L. Ferris, J. Ott, C. Barnstable, J. Hoh, Complement factor H polymorphism in age-related macular degeneration. *Science*. **308**, 385–389 (2005).

7. A. O. Edwards, R. Ritter 3rd, K. J. Abel, A. Manning, C. Panhuysen, L. A. Farrer, Complement factor H polymorphism and age-related macular degeneration. *Science*. **308**, 421–424 (2005).

8. L. G. Fritsche, W. Igl, J. N. C. Bailey, F. Grassmann, S. Sengupta, J. L. Bragg-Gresham, K. P. Burdon, S. J. Hebbring, C. Wen, M. Gorski, I. K. Kim, D. Cho, D. Zack, E. Souied, H. P. N. Scholl, E. Bala, K. E. Lee, D. J. Hunter, R. J. Sardell, P. Mitchell, J. E. Merriam, V. Cipriani, J. D. Hoffman, T. Schick, Y. T. E. Lechanteur, R. H. Guymer, M. P. Johnson, Y. Jiang, C. M. Stanton, G. H. S. Buitendijk, X. Zhan, A. M. Kwong, A. Boleda, M. Brooks, L. Gieser, R. Ratnapriya, K. E. Branham, J. R. Foerster, J. R. Heckenlively, M. I. Othman, B. J. Vote, H. H. Liang, E. Souzeau, I. L. McAllister, T. Isaacs, J. Hall, S. Lake, D. A. Mackey, I. J. Constable, J. E. Craig, T. E. Kitchner, Z. Yang, Z. Su, H. Luo, D. Chen, H. Ouyang, K. Flagg, D. Lin, G. Mao, H. Ferreyra, K. Stark, C. N. von Strachwitz, A. Wolf, C. Brandl, G. Rudolph, M. Olden, M. A. Morrison, D. J. Morgan, M. Schu, J. Ahn, G. Silvestri, E. E. Tsironi, K. H. Park, L. A. Farrer, A. Orlin, A. Brucker, M. Li, C. A. Curcio, S. Mohand-Saïd, J.-A. Sahel, I. Audo, M. Benchaboune, A. J. Cree, C. A. Rennie, S. V. Goverdhan, M. Grunin, S. Hagbi-Levi, P. Campochiaro, N. Katsanis, F. G. Holz, F. Blond, H. Blanché, J.-F. Deleuze, R. P. Igo Jr, B. Truitt, N. S. Peachey, S. M. Meuer, C. E. Myers, E. L. Moore, R. Klein, M. A. Hauser, E. A. Postel, M. D. Courtenay, S. G. Schwartz, J. L. Kovach, W. K. Scott, G. Liew, A. G. Tan, B. Gopinath, J. C. Merriam, R. T. Smith, J. C. Khan, H. Shahid, A. T. Moore, J. A. McGrath, R. Laux, M. A. Brantley Jr, A. Agarwal, L. Ersoy, A. Caramoy, T. Langmann, N. T. M. Saksens, E. K. de Jong, C. B. Hoyng, M. S. Cain, A. J. Richardson, T. M. Martin, J. Blangero, D. E. Weeks, B. Dhillon, C. M. van Duijn, K. F. Doheny, J. Romm, C. C. W. Klaver, C.

medRxiv preprint doi: https://doi.org/10.1101/2022.08.16.22278855; this version posted August 21, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. This article is a US Government work. It is not subject to copyright under 17 USC 105 and is also made available for use under a CC0 license.

Hayward, M. B. Gorin, M. L. Klein, P. N. Baird, A. I. den Hollander, S. Fauser, J. R. W. Yates, R. Allikmets, J. J. Wang, D. A. Schaumberg, B. E. K. Klein, S. A. Hagstrom, I. Chowers, A. J. Lotery, T. Léveillard, K. Zhang, M. H. Brilliant, A. W. Hewitt, A. Swaroop, E. Y. Chew, M. A. Pericak-Vance, M. DeAngelis, D. Stambolian, J. L. Haines, S. K. Iyengar, B. H. F. Weber, G. R. Abecasis, I. M. Heid, A large genome-wide association study of age-related macular degeneration highlights contributions of rare and common variants. *Nat. Genet.* **48**, 134–143 (2016).

9. J. M. Gaziano, J. Concato, M. Brophy, L. Fiore, S. Pyarajan, J. Breeling, S. Whitbourne, J. Deen, C. Shannon, D. Humphries, P. Guarino, M. Aslan, D. Anderson, R. LaFleur, T. Hammond, K. Schaa, J. Moser, G. Huang, S. Muralidhar, R. Przygodzki, T. J. O'Leary, Million Veteran Program: A mega-biobank to study genetic influences on health and disease. *J. Clin. Epidemiol.* **70**, 214–223 (2016).

10. T. Ge, C.-Y. Chen, Y. Ni, Y.-C. A. Feng, J. W. Smoller, Polygenic prediction via Bayesian regression and continuous shrinkage priors. *Nat. Commun.* **10**, 1776 (2019).

11. B. K. Bulik-Sullivan, P.-R. Loh, H. K. Finucane, S. Ripke, J. Yang, Schizophrenia Working Group of the Psychiatric Genomics Consortium, N. Patterson, M. J. Daly, A. L. Price, B. M. Neale, LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat. Genet.* **47**, 291–295 (2015).

12. M. Pietzner, E. Wheeler, J. Carrasco-Zanini, J. Raffler, N. D. Kerrison, E. Oerton, V. P. W. Auyeung, J. 'an Luan, C. Finan, J. P. Casas, R. Ostroff, S. A. Williams, G. Kastenmüller, M. Ralser, E. R. Gamazon, N. J. Wareham, A. D. Hingorani, C. Langenberg, Genetic architecture of host proteins involved in SARS-CoV-2 infection. *Nat. Commun.* **11**, 6397 (2020).

13. P. Kumar, S. Henikoff, P. C. Ng, Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat. Protoc.* **4**, 1073–1081 (2009).

14. I. A. Adzhubei, S. Schmidt, L. Peshkin, V. E. Ramensky, A. Gerasimova, P. Bork, A. S. Kondrashov, S. R. Sunyaev, A method and server for predicting damaging missense mutations. *Nat. Methods*. **7**, 248–249 (2010).

15. M. Komatsu, E. Ruoslahti, R-Ras is a global regulator of vascular regeneration that suppresses intimal hyperplasia and tumor angiogenesis. *Nat. Med.* **11**, 1346–1350 (2005).

16. J. Sawada, F. Li, M. Komatsu, R-Ras protein inhibits autophosphorylation of vascular endothelial growth factor receptor 2 in endothelial cells and suppresses receptor activation in tumor vasculature. *J. Biol. Chem.* **290**, 8133–8145 (2015).

17. M. Vähätupa, S. Prince, S. Vataja, T. Mertimo, M. Kataja, K. Kinnunen, V. Marjomäki, H. Uusitalo, M. Komatsu, T. A. H. Järvinen, H. Uusitalo-Järvinen, Lack of R-Ras Leads to Increased Vascular Permeability in Ischemic Retinopathy. *Invest. Ophthalmol. Vis. Sci.* **57**, 4898–4909 (2016).

18. P. Surendran, F. Drenos, R. Young, H. Warren, J. P. Cook, A. K. Manning, N. Grarup, X. Sim, D. R. Barnes, K. Witkowska, J. R. Staley, V. Tragante, T. Tukiainen, H. Yaghootkar, N. Masca, D. F. Freitag, T. Ferreira, O. Giannakopoulou, A. Tinker, M. Harakalova, E. Mihailov, C. Liu, A. T. Kraja, S. Fallgaard Nielsen, A. Rasheed, M. Samuel, W. Zhao, L. L. Bonnycastle, A. U. Jackson, N. Narisu, A. J. Swift, L. Southam, J. Marten, J. R. Huyghe, A. Stančáková,

medRxiv preprint doi: https://doi.org/10.1101/2022.08.16.22278855; this version posted August 21, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. This article is a US Government work. It is not subject to copyright under 17 USC 105 and is also made available for use under a CC0 license.

C. Fava, T. Ohlsson, A. Matchan, K. E. Stirrups, J. Bork-Jensen, A. P. Gjesing, J. Kontto, M. Perola, S. Shaw-Hawkins, A. S. Havulinna, H. Zhang, L. A. Donnelly, C. J. Groves, N. W. Rayner, M. J. Neville, N. R. Robertson, A. M. Yiorkas, K.-H. Herzig, E. Kajantie, W. Zhang, S. M. Willems, L. Lannfelt, G. Malerba, N. Soranzo, E. Trabetti, N. Verweij, E. Evangelou, A. Moayyeri, A.-C. Vergnaud, C. P. Nelson, A. Poveda, T. V. Varga, M. Caslake, A. J. de Craen, S. Trompet, J. 'an Luan, R. A. Scott, S. E. Harris, D. C. Liewald, R. Marioni, C. Menni, A.-E. Farmaki, G. Hallmans, F. Renström, J. E. Huffman, M. Hassinen, S. Burgess, R. S. Vasan, J. F. Felix, CHARGE-Heart Failure Consortium, M. Uria-Nickelsen, A. Malarstig, D. F. Reily, M. Hoek, T. Vogt, H. Lin, W. Lieb, EchoGen Consortium, M. Traylor, H. F. Markus, METASTROKE Consortium, H. M. Highland, A. E. Justice, E. Marouli, GIANT Consortium, J. Lindström, M. Uusitupa, P. Komulainen, T. A. Lakka, R. Rauramaa, O. Polasek, I. Rudan, O. Rolandsson, P. W. Franks, G. Dedoussis, T. D. Spector, EPIC-InterAct Consortium, P. Jousilahti, S. Männistö, I. J. Deary, J. M. Starr, C. Langenberg, N. J. Wareham, M. J. Brown, A. F. Dominiczak, J. M. Connell, J. W. Jukema, N. Sattar, I. Ford, C. J. Packard, T. Esko, R. Mägi, A. Metspalu, R. A. de Boer, P. van der Meer, P. van der Harst, Lifelines Cohort Study, G. Gambaro, E. Ingelsson, L. Lind, P. I. de Bakker, M. E. Numans, I. Brandslund, C. Christensen, E. R. Petersen, E. Korpi-Hyövälti, H. Oksa, J. C. Chambers, J. S. Kooner, A. I. Blakemore, S. Franks, M.-R. Jarvelin, L. L. Husemoen, A. Linneberg, T. Skaaby, B. Thuesen, F. Karpe, J. Tuomilehto, A. S. Doney, A. D. Morris, C. N. Palmer, O. L. Holmen, K. Hveem, C. J. Willer, T. Tuomi, L. Groop, A. Käräjämäki, A. Palotie, S. Ripatti, V. Salomaa, D. S. Alam, A. A. Shafi Majumder, E. Di Angelantonio, R. Chowdhury, M. I. McCarthy, N. Poulter, A. V. Stanton, P. Sever, P. Amouyel, D. Arveiler, S. Blankenberg, J. Ferrières, F. Kee, K. Kuulasmaa, M. Müller-Nurasyid, G. Veronesi, J. Virtamo, P. Deloukas, Wellcome Trust Case Control Consortium, P. Elliott, Understanding Society Scientific Group, E. Zeggini, S. Kathiresan, O. Melander, J. Kuusisto, M. Laakso, S. Padmanabhan, D. Porteous, C. Hayward, G. Scotland, F. S. Collins, K. L. Mohlke, T. Hansen, O. Pedersen, M. Boehnke, H. M. Stringham, EPIC-CVD Consortium, P. Frossard, C. Newton-Cheh, CHARGE+ Exome Chip Blood Pressure Consortium, M. D. Tobin, B. G. Nordestgaard, T2D-GENES Consortium, GoT2DGenes Consortium, ExomeBP Consortium, CHD Exome+ Consortium, M. J. Caulfield, A. Mahajan, A. P. Morris, M. Tomaszewski, N. J. Samani, D. Saleheen, F. W. Asselbergs, C. M. Lindgren, J. Danesh, L. V. Wain, A. S. Butterworth, J. M. Howson, P. B. Munroe, Trans-ancestry meta-analyses identify rare and common variants associated with blood pressure and hypertension. *Nat. Genet.* **48**, 1151–1161 (2016).

19. R. A. Campbell, H. D. Campbell, J. S. Bircher, C. V. de Araujo, F. Denorme, J. L. Crandell, J. L. Rustad, J. Monts, M. J. Cody, Y. Kosaka, C. C. Yost, Placental HTRA1 cleaves α1-antitrypsin to generate a NET-inhibitory peptide. *Blood*. **138**, 977–988 (2021).

20. K. A. Fawcett, K. Song, G. Qian, A.-E. Farmaki, R. Packer, C. John, N. Shrine, R. Granell, S. Ring, N. J. Timpson, L. M. Yerges-Armstrong, R. Eastell, L. V. Wain, R. A. Scott, M. D. Tobin, I. P. Hall, Pleiotropic associations of heterozygosity for the SERPINA1 Z allele in the UK Biobank. *ERJ Open Res.* **7** (2021), doi:10.1183/23120541.00049-2021.

21. X. Yin, L. S. Chan, D. Bose, A. U. Jackson, P. VandeHaar, A. E. Locke, C. Fuchsberger, H. M. Stringham, R. Welch, K. Yu, L. Fernandes Silva, S. K. Service, D. Zhang, E. C. Hector, E. Young, L. Ganel, I. Das, H. Abel, M. R. Erdos, L. L. Bonnycastle, J. Kuusisto, N. O. Stitziel, I. M. Hall, G. R. Wagner, FinnGen, J. Kang, J. Morrison, C. F. Burant, F. S. Collins, S. Ripatti, A. Palotie, N. B. Freimer, K. L. Mohlke, L. J. Scott, X. Wen, E. B. Fauman, M. Laakso, M. Boehnke, Genome-wide association studies of metabolites in Finnish men identify disease-relevant loci. *Nat. Commun.* **13**, 1644 (2022).

medRxiv preprint doi: https://doi.org/10.1101/2022.08.16.22278855; this version posted August 21, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. This article is a US Government work. It is not subject to copyright under 17 USC 105 and is also made available for use under a CC0 license.

22. M. J. Simcoe, A. Shah, B. Fan, H. Choquet, N. Weisschuh, N. H. Waseem, C. Jiang, R. B. Melles, R. Ritch, O. A. Mahroo, B. Wissinger, E. Jorgenson, J. L. Wiggs, D. F. Garway-Heath, P. G. Hysi, C. J. Hammond, Genome-Wide Association Study Identifies Two Common Loci Associated with Pigment Dispersion Syndrome/Pigmentary Glaucoma and Implicates Myopia in its Development. *Ophthalmology* (2022), doi:10.1016/j.ophtha.2022.01.005.

23. M. Simcoe, A. Valdes, F. Liu, N. A. Furlotte, D. M. Evans, G. Hemani, S. M. Ring, G. D. Smith, D. L. Duffy, G. Zhu, S. D. Gordon, S. E. Medland, D. Vuckovic, G. Girotto, C. Sala, E. Catamo, M. P. Concas, M. Brumat, P. Gasparini, D. Toniolo, M. Cocca, A. Robino, S. Yazar, A. Hewitt, W. Wu, P. Kraft, C. J. Hammond, Y. Shi, Y. Chen, C. Zeng, C. C. W. Klaver, A. G. Uitterlinden, M. A. Ikram, M. A. Hamer, C. M. van Duijn, T. Nijsten, J. Han, D. A. Mackey, N. G. Martin, C.-Y. Cheng, 23andMe Research Team, International Visible Trait Genetics Consortium, D. A. Hinds, T. D. Spector, M. Kayser, P. G. Hysi, Genome-wide association study in almost 195,000 individuals identifies 50 previously unidentified genetic loci for eye color. *Sci Adv.* **7** (2021), doi:10.1126/sciadv.abd1239.

24. P. G. Hysi, A. M. Valdes, F. Liu, N. A. Furlotte, D. M. Evans, V. Bataille, A. Visconti, G. Hemani, G. McMahon, S. M. Ring, G. D. Smith, D. L. Duffy, G. Zhu, S. D. Gordon, S. E. Medland, B. D. Lin, G. Willemsen, J. Jan Hottenga, D. Vuckovic, G. Girotto, I. Gandin, C. Sala, M. P. Concas, M. Brumat, P. Gasparini, D. Toniolo, M. Cocca, A. Robino, S. Yazar, A. W. Hewitt, Y. Chen, C. Zeng, A. G. Uitterlinden, M. A. Ikram, M. A. Hamer, C. M. van Duijn, T. Nijsten, D. A. Mackey, M. Falchi, D. I. Boomsma, N. G. Martin, International Visible Trait Genetics Consortium, D. A. Hinds, M. Kayser, T. D. Spector, Genome-wide association meta-analysis of individuals of European ancestry identifies new loci explaining a substantial fraction of hair color variation and heritability. *Nat. Genet.* **50**, 652–656 (2018).

25. U. E. Liyanage, M. H. Law, X. Han, J. An, J.-S. Ong, P. Gharahkhani, S. Gordon, R. E. Neale, C. M. Olsen, 23andMe Research Team, S. MacGregor, D. C. Whiteman, Combined analysis of keratinocyte cancers identifies novel genome-wide loci. *Hum. Mol. Genet.* **28**, 3148–3160 (2019).

26. A. Visconti, D. L. Duffy, F. Liu, G. Zhu, W. Wu, Y. Chen, P. G. Hysi, C. Zeng, M. Sanna, M. M. Iles, P. A. Kanetsky, F. Demenais, M. A. Hamer, A. G. Uitterlinden, M. A. Ikram, T. Nijsten, N. G. Martin, M. Kayser, T. D. Spector, J. Han, V. Bataille, M. Falchi, Genome-wide association study in 176,678 Europeans reveals genetic loci for tanning response to sun exposure. *Nat. Commun.* **9**, 1684 (2018).

27. S. Lundgren, T. Carling, G. Hjälm, C. Juhlin, J. Rastad, U. Pihlgren, L. Rask, G. Akerström, P. Hellman, Tissue distribution of human gp330/megalin, a putative Ca(2+)-sensing protein. *J. Histochem. Cytochem.* **45**, 383–392 (1997).

28. T. Storm, S. Heegaard, E. I. Christensen, R. Nielsen, Megalin-deficiency causes high myopia, retinal pigment epithelium-macromelanosomes and abnormal development of the ciliary body in mice. *Cell Tissue Res.* **358**, 99–107 (2014).

29. N. C. L. Noel, I. M. MacDonald, RP1L1 and inherited photoreceptor disease: A review. *Surv. Ophthalmol.* **65**, 725–739 (2020).

30. C. Benner, C. C. A. Spencer, A. S. Havulinna, V. Salomaa, S. Ripatti, M. Pirinen, FINEMAP: efficient variable selection using summary data from genome-wide association studies. *Bioinformatics*. **32**, 1493–1501 (2016).

medRxiv preprint doi: https://doi.org/10.1101/2022.08.16.22278855; this version posted August 21, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. This article is a US Government work. It is not subject to copyright under 17 USC 105 and is also made available for use under a CC0 license.

31. B. Zeng, J. Bendl, R. Kosoy, J. F. Fullard, G. E. Hoffman, P. Roussos, Multi-ancestry eQTL meta-analysis of human brain identifies candidate causal variants for brain-related traits. *Nat. Genet.* **54**, 161–169 (2022).

32. F. Hormozdiari, M. van de Bunt, A. V. Segrè, X. Li, J. W. J. Joo, M. Bilow, J. H. Sul, S. Sankararaman, B. Pasaniuc, E. Eskin, Colocalization of GWAS and eQTL Signals Detects Target Genes. *Am. J. Hum. Genet.* **99**, 1245–1260 (2016).

33. B. K. Maples, S. Gravel, E. E. Kenny, C. D. Bustamante, RFMix: a discriminative modeling approach for rapid and robust local-ancestry inference. *Am. J. Hum. Genet.* **93**, 278–288 (2013).

34. E. G. Atkinson, A. X. Maihofer, M. Kanai, A. R. Martin, K. J. Karczewski, M. L. Santoro, J. C. Ulirsch, Y. Kamatani, Y. Okada, H. K. Finucane, K. C. Koenen, C. M. Nievergelt, M. J. Daly, B. M. Neale, Tractor uses local ancestry to enable the inclusion of admixed individuals in GWAS and to boost power. *Nat. Genet.* (2021), doi:10.1038/s41588-020-00766-y.

35. A. E. Hughes, N. Orr, H. Esfandiary, M. Diaz-Torres, T. Goodship, U. Chakravarthy, 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).

36. T. A. Sivakumaran, R. P. Igo Jr, J. M. Kidd, A. Itsara, L. J. Kopplin, W. Chen, S. A. Hagstrom, N. S. Peachey, P. J. Francis, M. L. Klein, E. Y. Chew, V. L. Ramprasad, W.-T. Tay, P. Mitchell, M. Seielstad, D. E. Stambolian, A. O. Edwards, K. E. Lee, D. V. Leontiev, G. Jun, Y. Wang, L. Tian, F. Qiu, A. K. Henning, T. LaFramboise, P. Sen, M. Aarthi, R. George, R. Raman, M. K. Das, L. Vijaya, G. Kumaramanickavel, T. Y. Wong, A. Swaroop, G. R. Abecasis, R. Klein, B. E. K. Klein, D. A. Nickerson, E. E. Eichler, S. K. Iyengar, A 32 kb critical region excluding Y402H in CFH mediates risk for age-related macular degeneration. *PLoS One*. **6**, e25598 (2011).

37. L. V. Holmes, L. Strain, S. J. Staniforth, I. Moore, K. Marchbank, D. Kavanagh, J. A. Goodship, H. J. Cordell, T. H. J. Goodship, Determining the population frequency of the CFHR3/CFHR1 deletion at 1q32. *PLoS One*. **8**, e60352 (2013).

38. V. Cipriani, L. Lorés-Motta, F. He, D. Fathalla, V. Tilakaratna, S. McHarg, N. Bayatti, İ. E. Acar, C. B. Hoyng, S. Fauser, A. T. Moore, J. R. W. Yates, E. K. de Jong, B. P. Morgan, A. I. den Hollander, P. N. Bishop, S. J. Clark, Increased circulating levels of Factor H-Related Protein 4 are strongly associated with age-related macular degeneration. *Nat. Commun.* **11**, 778 (2020).

39. C. M. Pappas, M. A. Zouache, S. Matthews, C. D. Faust, J. L. Hageman, B. L. Williams, B. T. Richards, G. S. Hageman, Protective chromosome 1q32 haplotypes mitigate risk for age-related macular degeneration associated with the CFH-CFHR5 and ARMS2/HTRA1 loci. *Hum. Genomics*. **15**, 60 (2021).

40. V. Cipriani, A. Tierney, J. R. Griffiths, V. Zuber, P. I. Sergouniotis, J. R. W. Yates, A. T. Moore, P. N. Bishop, S. J. Clark, R. D. Unwin, 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).

41. X. Zheng, J. Shen, C. Cox, J. C. Wakefield, M. G. Ehm, M. R. Nelson, B. S. Weir, HIBAG--HLA genotype imputation with attribute bagging. *Pharmacogenomics J.* **14**, 192–200 (2014).

medRxiv preprint doi: https://doi.org/10.1101/2022.08.16.22278855; this version posted August 21, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. This article is a US Government work. It is not subject to copyright under 17 USC 105 and is also made available for use under a CC0 license.

42. E. Jorgenson, R. B. Melles, T. J. Hoffmann, X. Jia, L. C. Sakoda, M. N. Kvale, Y. Banda, C. Schaefer, N. Risch, L. Shen, Common coding variants in the HLA-DQB1 region confer susceptibility to age-related macular degeneration. *Eur. J. Hum. Genet.* **24**, 1049–1055 (2016).

43. F. Megiorni, A. Pizzuti, HLA-DQA1 and HLA-DQB1 in Celiac disease predisposition: practical implications of the HLA molecular typing. *J. Biomed. Sci.* **19**, 88 (2012).

44. N. Kutszegi, X. Yang, A. Gézsi, G. Schermann, D. J. Erdélyi, Á. F. Semsei, K. M. Gábor, J. C. Sági, G. T. Kovács, A. Falus, H. Zhang, C. Szalai, HLA-DRB1*07:01-HLA-DQA1*02:01-HLA-DQB1*02:02 haplotype is associated with a high risk of asparaginase hypersensitivity in acute lymphoblastic leukemia. *Haematologica*. **102**, 1578–1586 (2017).

45. C. A. Fernandez, C. Smith, W. Yang, M. Daté, D. Bashford, E. Larsen, W. P. Bowman, C. Liu, L. B. Ramsey, T. Chang, V. Turner, M. L. Loh, E. A. Raetz, N. J. Winick, S. P. Hunger, W. L. Carroll, S. Onengut-Gumuscu, W.-M. Chen, P. Concannon, S. S. Rich, P. Scheet, S. Jeha, C.-H. Pui, W. E. Evans, M. Devidas, M. V. Relling, HLA-DRB1*07:01 is associated with a higher risk of asparaginase allergies. *Blood*. **124**, 1266–1276 (2014).

46. R. F. Mullins, S. R. Russell, D. H. Anderson, G. S. Hageman, Drusen associated with aging and age-related macular degeneration contain proteins common to extracellular deposits associated with atherosclerosis, elastosis, amyloidosis, and dense deposit disease. *FASEB J.* **14**, 835–846 (2000).

47. W. Zhang, G. Voloudakis, V. M. Rajagopal, B. Readhead, J. T. Dudley, E. E. Schadt, J. L. M. Björkegren, Y. Kim, J. F. Fullard, G. E. Hoffman, P. Roussos, Integrative transcriptome imputation reveals tissue-specific and shared biological mechanisms mediating susceptibility to complex traits. *Nat. Commun.* **10**, 3834 (2019).

48. M. J. Gandal, P. Zhang, E. Hadjimichael, R. L. Walker, C. Chen, S. Liu, H. Won, H. van Bakel, M. Varghese, Y. Wang, A. W. Shieh, J. Haney, S. Parhami, J. Belmont, M. Kim, P. Moran Losada, Z. Khan, J. Mleczko, Y. Xia, R. Dai, D. Wang, Y. T. Yang, M. Xu, K. Fish, P. R. Hof, J. Warrell, D. Fitzgerald, K. White, A. E. Jaffe, PsychENCODE Consortium, M. A. Peters, M. Gerstein, C. Liu, L. M. Iakoucheva, D. Pinto, D. H. Geschwind, Transcriptome-wide isoform-level dysregulation in ASD, schizophrenia, and bipolar disorder. *Science*. **362** (2018), doi:10.1126/science.aat8127.

49. R. Ratnapriya, O. A. Sosina, M. R. Starostik, M. Kwicklis, R. J. Kapphahn, L. G. Fritsche, A. Walton, M. Arvanitis, L. Gieser, A. Pietraszkiewicz, S. R. Montezuma, E. Y. Chew, A. Battle, G. R. Abecasis, D. A. Ferrington, N. Chatterjee, A. Swaroop, Retinal transcriptome and eQTL analyses identify genes associated with age-related macular degeneration. *Nat. Genet.* **51**, 606–610 (2019).

50. GTEx Consortium, The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science*. **369**, 1318–1330 (2020).

51. O. Franzén, R. Ermel, A. Cohain, N. K. Akers, A. Di Narzo, H. A. Talukdar, H. Foroughi-Asl, C. Giambartolomei, J. F. Fullard, K. Sukhavasi, S. Köks, L.-M. Gan, C. Giannarelli, J. C. Kovacic, C. Betsholtz, B. Losic, T. Michoel, K. Hao, P. Roussos, J. Skogsberg, A. Ruusalepp, E. E. Schadt, J. L. M. Björkegren, Cardiometabolic risk loci share downstream cis- and trans-gene regulation across tissues and diseases. *Science*. **353**, 827–830 (2016).

medRxiv preprint doi: https://doi.org/10.1101/2022.08.16.22278855; this version posted August 21, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. This article is a US Government work. It is not subject to copyright under 17 USC 105 and is also made available for use under a CC0 license.

52. Y. Liu, S. Chen, Z. Li, A. C. Morrison, E. Boerwinkle, X. Lin, ACAT: A Fast and Powerful p Value Combination Method for Rare-Variant Analysis in Sequencing Studies. *Am. J. Hum. Genet.* **104**, 410–421 (2019).

53. E. Sjöstedt, W. Zhong, L. Fagerberg, M. Karlsson, N. Mitsios, C. Adori, P. Oksvold, F. Edfors, A. Limiszewska, F. Hikmet, J. Huang, Y. Du, L. Lin, Z. Dong, L. Yang, X. Liu, H. Jiang, X. Xu, J. Wang, H. Yang, L. Bolund, A. Mardinoglu, C. Zhang, K. von Feilitzen, C. Lindskog, F. Pontén, Y. Luo, T. Hökfelt, M. Uhlén, J. Mulder, An atlas of the protein-coding genes in the human, pig, and mouse brain. *Science*. **367** (2020), doi:10.1126/science.aay5947.

54. Y. Nédélec, J. Sanz, G. Baharian, Z. A. Szpiech, A. Pacis, A. Dumaine, J.-C. Grenier, A. Freiman, A. J. Sams, S. Hebert, A. Pagé Sabourin, F. Luca, R. Blekhman, R. D. Hernandez, R. Pique-Regi, J. Tung, V. Yotova, L. B. Barreiro, Genetic Ancestry and Natural Selection Drive Population Differences in Immune Responses to Pathogens. *Cell*. **167**, 657–669.e21 (2016).

55. H. E. Randolph, J. K. Fiege, B. K. Thielen, C. K. Mickelson, M. Shiratori, J. Barroso-Batista, R. A. Langlois, L. B. Barreiro, Genetic ancestry effects on the response to viral infection are pervasive but cell type specific. *Science*. **374**, 1127–1133 (2021).

56. S. Cantsilieris, B. J. Nelson, J. Huddleston, C. Baker, L. Harshman, K. Penewit, K. M. Munson, M. Sorensen, A. E. Welch, V. Dang, F. Grassmann, A. J. Richardson, R. H. Guymer, T. A. Graves-Lindsay, R. K. Wilson, B. H. F. Weber, P. N. Baird, R. Allikmets, E. E. Eichler, 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).

57. B. L. Williams, N. A. Seager, J. D. Gardiner, C. M. Pappas, M. C. Cronin, C. Amat di San Filippo, R. A. Anstadt, J. Liu, M. A. Toso, L. Nichols, T. J. Parnell, J. R. Eve, P. L. Bartel, M. A. Zouache, B. T. Richards, G. S. Hageman, 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), doi:10.1073/pnas.2103617118.

58. U. Friedrich, S. Datta, T. Schubert, K. Plössl, M. Schneider, F. Grassmann, R. Fuchshofer, K.-J. Tiefenbach, G. Längst, B. H. F. Weber, Synonymous variants in HTRA1 implicated in AMD susceptibility impair its capacity to regulate TGF-β signaling. *Hum. Mol. Genet.* **24**, 6361–6373 (2015).

59. F. Beguier, M. Housset, C. Roubeix, S. Augustin, Y. Zagar, C. Nous, T. Mathis, C. Eandi, M. Benchaboune, A. Drame-Maigné, W. Carpentier, S. Chardonnet, S. Touhami, G. Blot, J. B. Conart, H. Charles-Messance, A. Potey, J.-F. Girmens, M. Paques, F. Blond, T. Leveillard, E. Koertvely, J. E. Roger, J.-A. Sahel, P. Sapieha, C. Delarasse, X. Guillonneau, F. Sennlaub, The 10q26 Risk Haplotype of Age-Related Macular Degeneration Aggravates Subretinal Inflammation by Impairing Monocyte Elimination. *Immunity*. **53**, 429–441.e8 (2020).

60. V. Frochaux, D. Hildebrand, A. Talke, M. W. Linscheid, H. Schlüter, Alpha-1-antitrypsin: a novel human high temperature requirement protease A1 (HTRA1) substrate in human placental tissue. *PLoS One*. **9**, e109483 (2014).

61. C. Oka, M. Kawaichi, HtrA1 meets NETs: does it open a new field? *Blood*. **138** (2021), pp. 920–921.

medRxiv preprint doi: https://doi.org/10.1101/2022.08.16.22278855; this version posted August 21, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. This article is a US Government work. It is not subject to copyright under 17 USC 105 and is also made available for use under a CC0 license.

62. H. Wang, C. Wang, M.-H. Zhao, M. Chen, Neutrophil extracellular traps can activate alternative complement pathways. *Clin. Exp. Immunol.* **181**, 518–527 (2015).

63. J. Yuen, F. G. Pluthero, D. N. Douda, M. Riedl, A. Cherry, M. Ulanova, W. H. A. Kahr, N. Palaniyar, C. Licht, NETosing Neutrophils Activate Complement Both on Their Own NETs and Bacteria via Alternative and Non-alternative Pathways. *Front. Immunol.* **7**, 137 (2016).

64. C. M. de Bont, W. C. Boelens, G. J. M. Pruijn, NETosis, complement, and coagulation: a triangular relationship. *Cell. Mol. Immunol.* **16**, 19–27 (2019).

65. A. E. Schneider, N. Sándor, É. Kárpáti, M. Józsi, Complement factor H modulates the activation of human neutrophil granulocytes and the generation of neutrophil extracellular traps. *Mol. Immunol.* **72**, 37–48 (2016).

66. L. D. Halder, M. A. Abdelfatah, E. A. H. Jo, I. D. Jacobsen, M. Westermann, N. Beyersdorf, S. Lorkowski, P. F. Zipfel, C. Skerka, Factor H Binds to Extracellular DNA Traps Released from Human Blood Monocytes in Response to Candida albicans. *Front. Immunol.* **7**, 671 (2016).

67. S. Ghosh, A. Padmanabhan, T. Vaidya, A. M. Watson, I. A. Bhutto, S. Hose, P. Shang, N. Stepicheva, M. Yazdankhah, J. Weiss, M. Das, S. Gopikrishna, Aishwarya, N. Yadav, T. Berger, T. W. Mak, S. Xia, J. Qian, G. A. Lutty, A. Jayagopal, J. S. Zigler Jr, S. Sethu, J. T. Handa, S. C. Watkins, A. Ghosh, D. Sinha, Neutrophils homing into the retina trigger pathology in early age-related macular degeneration. *Commun Biol*. **2**, 348 (2019).

68. S. Niazi, M. Krogh Nielsen, T. L. Sørensen, Y. Subhi, Neutrophil-to-lymphocyte ratio in age-related macular degeneration: a systematic review and meta-analysis. *Acta Ophthalmol.* **97**, 558–566 (2019).

69. B. E. K. Klein, K. P. Howard, S. K. Iyengar, T. A. Sivakumaran, K. J. Meyers, K. J. Cruickshanks, R. Klein, Sunlight exposure, pigmentation, and incident age-related macular degeneration. *Invest. Ophthalmol. Vis. Sci.* **55**, 5855–5861 (2014).

70. T. Strunz, S. Lauwen, C. Kiel, International AMD Genomics Consortium (IAMDGC), A. den Hollander, B. H. F. Weber, A transcriptome-wide association study based on 27 tissues identifies 106 genes potentially relevant for disease pathology in age-related macular degeneration. *Sci. Rep.* **10**, 1584 (2020).

71. C. W. Halladay, T. Hadi, M. D. Anger, P. B. Greenberg, J. M. Sullivan, P. E. Konicki, N. S. Peachey, R. P. Igo Jr, S. K. Iyengar, W.-C. Wu, D. C. Crawford, VA Million Veteran Program, Genetically-guided algorithm development and sample size optimization for age-related macular degeneration cases and controls in electronic health records from the VA Million Veteran Program. *AMIA Jt Summits Transl Sci Proc*. **2019**, 153–162 (2019).

72. H. Hunter-Zinck, Y. Shi, M. Li, B. R. Gorman, S.-G. Ji, N. Sun, T. Webster, A. Liem, P. Hsieh, P. Devineni, P. Karnam, X. Gong, L. Radhakrishnan, J. Schmidt, T. L. Assimes, J. Huang, C. Pan, D. Humphries, M. Brophy, J. Moser, S. Muralidhar, G. D. Huang, R. Przygodzki, J. Concato, J. M. Gaziano, J. Gelernter, C. J. O'Donnell, E. R. Hauser, H. Zhao, T. J. O'Leary, P. S. Tsao, S. Pyarajan, Genotyping Array Design and Data Quality Control in the Million Veteran Program. *Am. J. Hum. Genet.* **106**, 535–548 (2020).

73. S. McCarthy, S. Das, W. Kretzschmar, O. Delaneau, A. R. Wood, A. Teumer, H. M. Kang, C.

medRxiv preprint doi: https://doi.org/10.1101/2022.08.16.22278855; this version posted August 21, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. This article is a US Government work. It is not subject to copyright under 17 USC 105 and is also made available for use under a CC0 license.

Fuchsberger, P. Danecek, K. Sharp, Y. Luo, C. Sidore, A. Kwong, N. Timpson, S. Koskinen, S. Vrieze, L. J. Scott, H. Zhang, A. Mahajan, J. Veldink, U. Peters, C. Pato, C. M. van Duijn, C. E. Gillies, I. Gandin, M. Mezzavilla, A. Gilly, M. Cocca, M. Traglia, A. Angius, J. C. Barrett, D. Boomsma, K. Branham, G. Breen, C. M. Brummett, F. Busonero, H. Campbell, A. Chan, S. Chen, E. Chew, F. S. Collins, L. J. Corbin, G. D. Smith, G. Dedoussis, M. Dorr, A.-E. Farmaki, L. Ferrucci, L. Forer, R. M. Fraser, S. Gabriel, S. Levy, L. Groop, T. Harrison, A. Hattersley, O. L. Holmen, K. Hveem, M. Kretzler, J. C. Lee, M. McGue, T. Meitinger, D. Melzer, J. L. Min, K. L. Mohlke, J. B. Vincent, M. Nauck, D. Nickerson, A. Palotie, M. Pato, N. Pirastu, M. McInnis, J. B. Richards, C. Sala, V. Salomaa, D. Schlessinger, S. Schoenherr, P. E. Slagboom, K. Small, T. Spector, D. Stambolian, M. Tuke, J. Tuomilehto, L. H. Van den Berg, W. Van Rheenen, U. Volker, C. Wijmenga, D. Toniolo, E. Zeggini, P. Gasparini, M. G. Sampson, J. F. Wilson, T. Frayling, P. I. W. de Bakker, M. A. Swertz, S. McCarroll, C. Kooperberg, A. Dekker, D. Altshuler, C. Willer, W. Iacono, S. Ripatti, N. Soranzo, K. Walter, A. Swaroop, F. Cucca, C. A. Anderson, R. M. Myers, M. Boehnke, M. I. McCarthy, R. Durbin, Haplotype Reference Consortium, A reference panel of 64,976 haplotypes for genotype imputation. *Nat. Genet.* **48**, 1279–1283 (2016).

74. H. Fang, Q. Hui, J. Lynch, J. Honerlaw, T. L. Assimes, J. Huang, M. Vujkovic, S. M. Damrauer, S. Pyarajan, J. M. Gaziano, S. L. DuVall, C. J. O'Donnell, K. Cho, K.-M. Chang, P. W. F. Wilson, P. S. Tsao, VA Million Veteran Program, Y. V. Sun, H. Tang, Harmonizing Genetic Ancestry and Self-identified Race/Ethnicity in Genome-wide Association Studies. *Am. J. Hum. Genet.* **105**, 763–772 (2019).

75. J. Mbatchou, L. Barnard, J. Backman, A. Marcketta, J. A. Kosmicki, A. Ziyatdinov, C. Benner, C. O'Dushlaine, M. Barber, B. Boutkov, L. Habegger, M. Ferreira, A. Baras, J. Reid, G. Abecasis, E. Maxwell, J. Marchini, Computationally efficient whole-genome regression for quantitative and binary traits. *Nat. Genet.* **53**, 1097–1103 (2021).

76. L. Shen, T. J. Hoffmann, R. B. Melles, L. C. Sakoda, M. N. Kvale, Y. Banda, C. Schaefer, N. Risch, E. Jorgenson, Differences in the Genetic Susceptibility to Age-Related Macular Degeneration Clinical Subtypes. *Invest. Ophthalmol. Vis. Sci.* **56**, 4290–4299 (2015).

77. C. J. Willer, Y. Li, G. R. Abecasis, METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics*. **26**, 2190–2191 (2010).

78. K. Watanabe, E. Taskesen, A. van Bochoven, D. Posthuma, Functional mapping and annotation of genetic associations with FUMA. *Nat. Commun.* **8**, 1826 (2017).

79. J. Yang, S. H. Lee, M. E. Goddard, P. M. Visscher, GCTA: a tool for genome-wide complex trait analysis. *Am. J. Hum. Genet.* **88**, 76–82 (2011).

80. G. Heinze, M. Schemper, A solution to the problem of separation in logistic regression. *Stat. Med.* **21**, 2409–2419 (2002).

81. N. Kutszegi, A. Gézsi, Á. F Semsei, J. Müller, R. Simon, E. R. Kovács, K. Hegedüs, G. T. Kovács, C. Szalai, D. J. Erdélyi, Two tagging single-nucleotide polymorphisms to capture HLA-DRB1*07:01-DQA1*02:01-DQB1*02:02 haplotype associated with asparaginase hypersensitivity. *Br. J. Clin. Pharmacol.* **87**, 2542–2548 (2021).

82. O. Delaneau, J.-F. Zagury, M. R. Robinson, J. L. Marchini, E. T. Dermitzakis, Accurate, scalable and integrative haplotype estimation. *Nat. Commun.* **10**, 5436 (2019).

medRxiv preprint doi: https://doi.org/10.1101/2022.08.16.22278855; this version posted August 21, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. This article is a US Government work. It is not subject to copyright under 17 USC 105 and is also made available for use under a CC0 license.

83. X. Mao, A. W. Bigham, R. Mei, G. Gutierrez, K. M. Weiss, T. D. Brutsaert, F. Leon-Velarde, L. G. Moore, E. Vargas, P. M. McKeigue, M. D. Shriver, E. J. Parra, A genomewide admixture mapping panel for Hispanic/Latino populations. *Am. J. Hum. Genet.* **80**, 1171–1178 (2007).

84. A. R. Martin, C. R. Gignoux, R. K. Walters, G. L. Wojcik, B. M. Neale, S. Gravel, M. J. Daly, C. D. Bustamante, E. E. Kenny, Human Demographic History Impacts Genetic Risk Prediction across Diverse Populations. *Am. J. Hum. Genet.* **100**, 635–649 (2017).

85. C. C. Chang, C. C. Chow, L. C. Tellier, S. Vattikuti, S. M. Purcell, J. J. Lee, Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience*. **4**, 7 (2015).

86. L. McInnes, J. Healy, J. Melville, UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. *arXiv [stat.ML]* (2018), (available at http://arxiv.org/abs/1802.03426).

87. S. Das, L. Forer, S. Schönherr, C. Sidore, A. E. Locke, A. Kwong, S. I. Vrieze, E. Y. Chew, S. Levy, M. McGue, D. Schlessinger, D. Stambolian, P.-R. Loh, W. G. Iacono, A. Swaroop, L. J. Scott, F. Cucca, F. Kronenberg, M. Boehnke, G. R. Abecasis, C. Fuchsberger, Next-generation genotype imputation service and methods. *Nat. Genet.* **48**, 1284–1287 (2016).

88. O. M. Man, M. H. Woehrmann, T. A. Webster, J. Gollub, A. Bivol, S. M. Keeble, K. H. Aull, A. Mittal, A. H. Roter, B. A. Wong, J. P. Schmidt, Novel genotyping algorithms for rare variants significantly improve the accuracy of Applied Biosystems™ Axiom™ array genotyping calls. *bioRxiv* (2021), p. 2021.09.13.459984.

89. E. T. Cirulli, S. White, R. W. Read, G. Elhanan, W. J. Metcalf, F. Tanudjaja, D. M. Fath, E. Sandoval, M. Isaksson, K. A. Schlauch, J. J. Grzymski, J. T. Lu, N. L. Washington, Genome-wide rare variant analysis for thousands of phenotypes in over 70,000 exomes from two cohorts. *Nat. Commun.* **11**, 542 (2020).

90. A. Manichaikul, J. C. Mychaleckyj, S. S. Rich, K. Daly, M. Sale, W.-M. Chen, Robust relationship inference in genome-wide association studies. *Bioinformatics*. **26**, 2867–2873 (2010).

91. Z. Zhu, F. Zhang, H. Hu, A. Bakshi, M. R. Robinson, J. E. Powell, G. W. Montgomery, M. E. Goddard, N. R. Wray, P. M. Visscher, J. Yang, Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. *Nat. Genet.* **48**, 481–487 (2016).

92. G. Hemani, J. Zheng, B. Elsworth, K. H. Wade, V. Haberland, D. Baird, C. Laurin, S. Burgess, J. Bowden, R. Langdon, V. Y. Tan, J. Yarmolinsky, H. A. Shihab, N. J. Timpson, D. M. Evans, C. Relton, R. M. Martin, G. Davey Smith, T. R. Gaunt, P. C. Haycock, The MR-Base platform supports systematic causal inference across the human phenome. *Elife*. **7** (2018), doi:10.7554/eLife.34408.

93. R. J. Carroll, L. Bastarache, J. C. Denny, R PheWAS: data analysis and plotting tools for phenome-wide association studies in the R environment. *Bioinformatics*. **30**, 2375–2376 (2014).

94. J. F. Fullard, H.-C. Lee, G. Voloudakis, S. Suo, B. Javidfar, Z. Shao, C. Peter, W. Zhang, S. Jiang, A. Corvelo, H. Wargnier, E. Woodoff-Leith, D. P. Purohit, S. Ahuja, N. M. Tsankova, N. Jette, G. E. Hoffman, S. Akbarian, M. Fowkes, J. F. Crary, G.-C. Yuan, P. Roussos, Single-nucleus transcriptome analysis of human brain immune response in patients with severe

medRxiv preprint doi: https://doi.org/10.1101/2022.08.16.22278855; this version posted August 21, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. This article is a US Government work. It is not subject to copyright under 17 USC 105 and is also made available for use under a CC0 license.

COVID-19. *Genome Med.* **13**, 118 (2021).

95. D. Wang, S. Liu, J. Warrell, H. Won, X. Shi, F. C. P. Navarro, D. Clarke, M. Gu, P. Emani, Y. T. Yang, M. Xu, M. J. Gandal, S. Lou, J. Zhang, J. J. Park, C. Yan, S. K. Rhie, K. Manakongtreecheep, H. Zhou, A. Nathan, M. Peters, E. Mattei, D. Fitzgerald, T. Brunetti, J. Moore, Y. Jiang, K. Girdhar, G. E. Hoffman, S. Kalayci, Z. H. Gümüş, G. E. Crawford, PsychENCODE Consortium, P. Roussos, S. Akbarian, A. E. Jaffe, K. P. White, Z. Weng, N. Sestan, D. H. Geschwind, J. A. Knowles, M. B. Gerstein, Comprehensive functional genomic resource and integrative model for the human brain. *Science*. **362** (2018), doi:10.1126/science.aat8464.

96. S. Purcell, B. Neale, K. Todd-Brown, L. Thomas, M. A. R. Ferreira, D. Bender, J. Maller, P. Sklar, P. I. W. de Bakker, M. J. Daly, P. C. Sham, PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).

97. 1000 Genomes Project Consortium, A. Auton, L. D. Brooks, R. M. Durbin, E. P. Garrison, H. M. Kang, J. O. Korbel, J. L. Marchini, S. McCarthy, G. A. McVean, G. R. Abecasis, A global reference for human genetic variation. *Nature*. **526**, 68–74 (2015).

98. C. W. Law, Y. Chen, W. Shi, G. K. Smyth, voom: Precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome Biol.* **15**, R29 (2014).

99. O. Stegle, L. Parts, M. Piipari, J. Winn, R. Durbin, Using probabilistic estimation of expression residuals (PEER) to obtain increased power and interpretability of gene expression analyses. *Nat. Protoc.* **7**, 500–507 (2012).

100. E. R. Gamazon, H. E. Wheeler, K. P. Shah, S. V. Mozaffari, K. Aquino-Michaels, R. J. Carroll, A. E. Eyler, J. C. Denny, GTEx Consortium, D. L. Nicolae, N. J. Cox, H. K. Im, A gene-based association method for mapping traits using reference transcriptome data. *Nat. Genet.* **47**, 1091–1098 (2015).

101. A. N. Barbeira, S. P. Dickinson, R. Bonazzola, J. Zheng, H. E. Wheeler, J. M. Torres, E. S. Torstenson, K. P. Shah, T. Garcia, T. L. Edwards, E. A. Stahl, L. M. Huckins, GTEx Consortium, D. L. Nicolae, N. J. Cox, H. K. Im, Exploring the phenotypic consequences of tissue specific gene expression variation inferred from GWAS summary statistics. *Nat. Commun.* **9**, 1825 (2018).

102. Y. Benjamini, Y. Hochberg, Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J. R. Stat. Soc.* **57**, 289–300 (1995).

103. A. Subramanian, P. Tamayo, V. K. Mootha, S. Mukherjee, B. L. Ebert, M. A. Gillette, A. Paulovich, S. L. Pomeroy, T. R. Golub, E. S. Lander, J. P. Mesirov, Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 15545–15550 (2005).

104. M. Ashburner, C. A. Ball, J. A. Blake, D. Botstein, H. Butler, J. M. Cherry, A. P. Davis, K. Dolinski, S. S. Dwight, J. T. Eppig, M. A. Harris, D. P. Hill, L. Issel-Tarver, A. Kasarskis, S. Lewis, J. C. Matese, J. E. Richardson, M. Ringwald, G. M. Rubin, G. Sherlock, Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat. Genet.* **25**, 25–29 (2000).

105. Gene Ontology Consortium, The Gene Ontology resource: enriching a GOld mine. *Nucleic*

medRxiv preprint doi: https://doi.org/10.1101/2022.08.16.22278855; this version posted August 21, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. This article is a US Government work. It is not subject to copyright under 17 USC 105 and is also made available for use under a CC0 license.

Acids Res. **49**, D325–D334 (2021).

106. C. Zhao, Z. Wang, GOGO: An improved algorithm to measure the semantic similarity between gene ontology terms. *Sci. Rep.* **8**, 15107 (2018).

107. M. Charrad, N. Ghazzali, V. Boiteau, A. Niknafs, NbClust: an R package for determining the relevant number of clusters in a data set. *J. Stat. Softw.* **61**, 1–36 (2014).

108. F. Murtagh, P. Legendre, Ward's hierarchical agglomerative clustering method: Which algorithms implement ward's criterion? *J. Classification*. **31**, 274–295 (2014).

109. P. J. Rousseeuw, Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *J. Comput. Appl. Math.* **20**, 53–65 (1987).

110. I. Feinerer, K. Hornik, D. Meyer, Text mining infrastructure in R. *J. Stat. Softw.* **25**, 1–54 (2008).

medRxiv preprint doi: https://doi.org/10.1101/2022.08.16.22278855; this version posted August 21, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. This article is a US Government work. It is not subject to copyright under 17 USC 105 and is also made available for use under a CC0 license.

Acknowledgements:

The unsung heroes of this publication are the U.S. Veterans who participated in this study, and who contributed time and effort to improve the lives of all Veterans and of populations worldwide. Key individuals who collected, curated or cleaned data for the Million Veteran Program are listed in the Supplementary Information. Dr. Igo passed away during the drafting of this manuscript and is remembered for his tireless contributions. We are further grateful to Catherine Tcheandjieu and Elizabeth Atkinson for advice on local ancestry inference and Tractor implementation, and to Jonathan Haines for comments on an earlier version of the manuscript. Finally, we wish to acknowledge the UK Biobank Eye and Vision Consortium who helped collect the eye phenotypes in the UK Biobank. The UK Biobank Eye and Vision Consortium is supported by grants from Moorfields Eye Charity, The NIHR Biomedical Research Centre at Moorfields Eye Hospital NHS Foundation Trust and UCL Institute of Ophthalmology, the Alcon Research Institute and the International Glaucoma Association (UK). This research has been conducted using the UK Biobank Resource under Application Number 2112. This project was supported by the Clinical and Translational Science Collaborative of Cleveland, which is funded by the National Institutes of Health, National Center for Advancing Translational Science, Clinical and Translational Science Award grant UL1TR002548. The content of this paper is solely the responsibility of the authors and does not necessarily represent the official views of the Department of Veterans Affairs or any agency of the United States federal government.

medRxiv preprint doi: https://doi.org/10.1101/2022.08.16.22278855; this version posted August 21, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. This article is a US Government work. It is not subject to copyright under 17 USC 105 and is also made available for use under a CC0 license.

Figure Legends:

Fig. 1. Overview of AMD GWAS meta-analysis primary and secondary analyses. Abbreviations: EA, European ancestry; AA, African ancestry; HA, Hispanic/Latino ancestry; TWAS, transcriptome-wide association study; PRS, polygenic risk score; eQTL, expression quantitative trait loci; PheWAS, phenome-wide association study; LabWAS, lab measurement-wide association study; HLA, human leukocyte antigen; GSEA, gene set enrichment analysis; MR, Mendelian randomization.

Fig. 2. Penetrance and pleiotropy of the AMD PRS. (**A**) Prevalence of AMD in EA males in MVP as a function of PRS decile, age range, and lifestyle risk factors. (**B**) PheWAS of the AMD PRS with 1,665 ICD code-based binary phenotypes in MVP EA. Phenotypes with association p-values less than 1×10^{-5} are shown. Odds ratios contrasting the top decile of PRS vs. the bottom decile (blue) and the top decile vs. the remaining 90% (red) are presented, with the x-axis in log scale. Phenotypes not directly related to vision are italicized. (**C**) LabWAS of the AMD PRS with 69 quantitative clinical lab measurements in MVP EA. Labs significant after Bonferroni adjustment for the number of labs are shown.

Fig. 3. GWAS analyses identify 27 novel loci in EA and the first loci in AA and HA. (**A**) Miami plot of the EA GWAS replication meta-analysis (57,290 cases, 324,430 controls) and DLPFC TWAS (45,045 cases, 275,806 controls). Red lines: genome-wide significance in the GWAS and TWAS analyses; blue line: FDR-corrected association p-value < 0.05 in the DLPFC TWAS; dashed gray line: transition from linear to log-scale on the y-axis. (**B**) Manhattan plot of the AA GWAS (2,302 cases, 29,223 controls). (A second potential locus near *COPS2* could not be confirmed in other ancestries.) (**C**) Manhattan plot of the HA GWAS (1,656 cases, 10,819 controls).

Fig. 4. MR of pigmentation traits on AMD risk in European ancestries. Forest plot of OR and 95% confidence intervals from the inverse variance-weighted MR analyses are shown. Eye color (*23*) was analyzed using all genome-wide significant instruments and using a selection of 14 instruments with known pleiotropic associations with other pigmentation traits. Other traits representing proxies for pigmentation, such as hair color (*24*), non-melanoma skin cancer (*25*), and tanning response (*26*) were also analyzed.

Fig. 5. Highlighted results from the multi-ancestry GWAS and TWAS. (**A**) Comparison of effect sizes (log-OR) of EA and AA index SNPs in the MVP EA GWAS (x-axis) and the MVP AA

medRxiv preprint doi: https://doi.org/10.1101/2022.08.16.22278855; this version posted August 21, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. This article is a US Government work. It is not subject to copyright under 17 USC 105 and is also made available for use under a CC0 license.

GWAS (y-axis). (**B**) Comparison of effect sizes in the MVP EA GWAS (x-axis) and the MVP HA GWAS (y-axis). Highlighted highly heterogeneous loci are shown in red. Error bars correspond to 95% confidence intervals. (**C**) Regional Miami plot of genetically regulated gene expression at the 1q32 (*CD46/CD55*) locus. GWAS results are above the axis and TWAS results from the DFPLC brain model below the axis. GWAS points are colored according to linkage disequilibrium r^2 with the top SNP.

Fig. 6. Local ancestry analysis of the *CFH* and *ARMS2/HTRA1* loci. (**A**) Regional association plots of AMD risk at *CFH* in a standard GWAS in AA individuals (top), and the EUR (middle) and AFR (bottom) tracts from the Tractor analysis. (**B**) Regional association plots of AMD risk at *CFH* in a standard GWAS in HA individuals (top), the EUR ancestry tract from the tractor analysis (middle) and the NAT ancestry tract (NAT) (bottom). (**C**) Forest plot comparing effect sizes and confidence intervals of the *CFH* Y402H risk allele in EA, AA, and HA populations, with Tractor analyses for AA (AFR and EUR tracts), and HA (EUR, NAT, and AFR tracts). (**D**) Forest plot comparing effect sizes and confidence intervals of the *ARMS2* A69S risk allele in EA, AA, and HA populations, with Tractor analyses for AA (AFR and EUR tracts), and HA (EUR, NAT, and AFR tracts). Additionally, Tractor effect size estimates in AA populations with either homozygous EUR ancestry at *CFH* ("EUR *CFH*") or homozygous AFR ancestry at *CFH* ("AFR *CFH*") are provided. Error bars are colored according to whether they correspond to a standard GWAS (black), or a Tractor GWAS, with blue for AFR tracts, red for EUR tracts, and green for NAT tracts.

medRxiv preprint doi: https://doi.org/10.1101/2022.08.16.22278855; this version posted August 21, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. This article is a US Government work. It is not subject to copyright under 17 USC 105 and is also made available for use under a CC0 license.

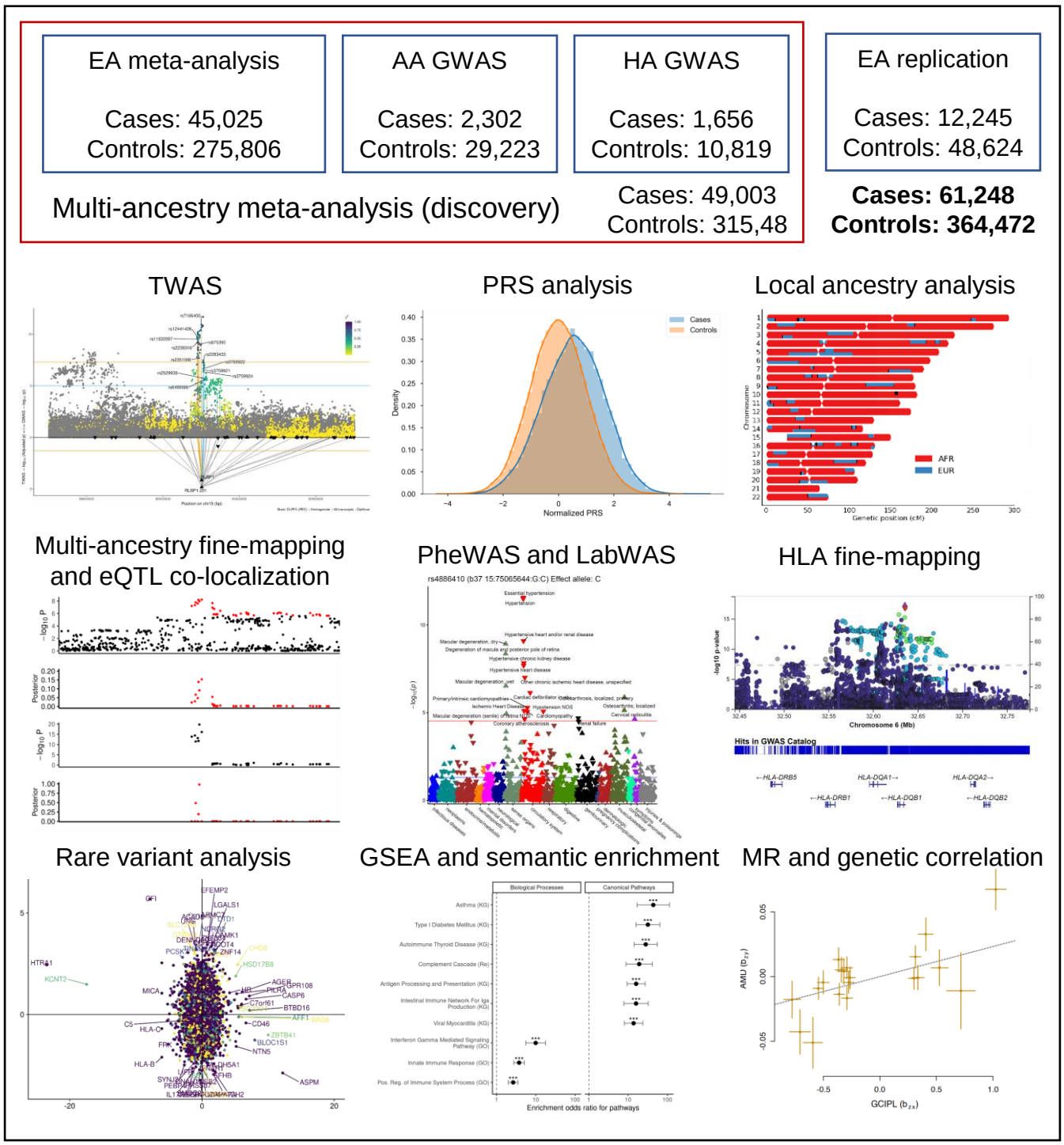


Fig. 1. Overview of AMD GWAS meta-analysis primary and secondary analyses. Abbreviations: EA, European ancestry; AA, African ancestry; HA, Hispanic/Latino ancestry; TWAS, transcriptome-wide association study; PRS, polygenic risk score; eQTL, expression quantitative trait loci; PheWAS, phenome-wide association study; LabWAS, lab measurement-wide association study; HLA, human leukocyte antigen; GSEA, gene set enrichment analysis; MR, Mendelian randomization.

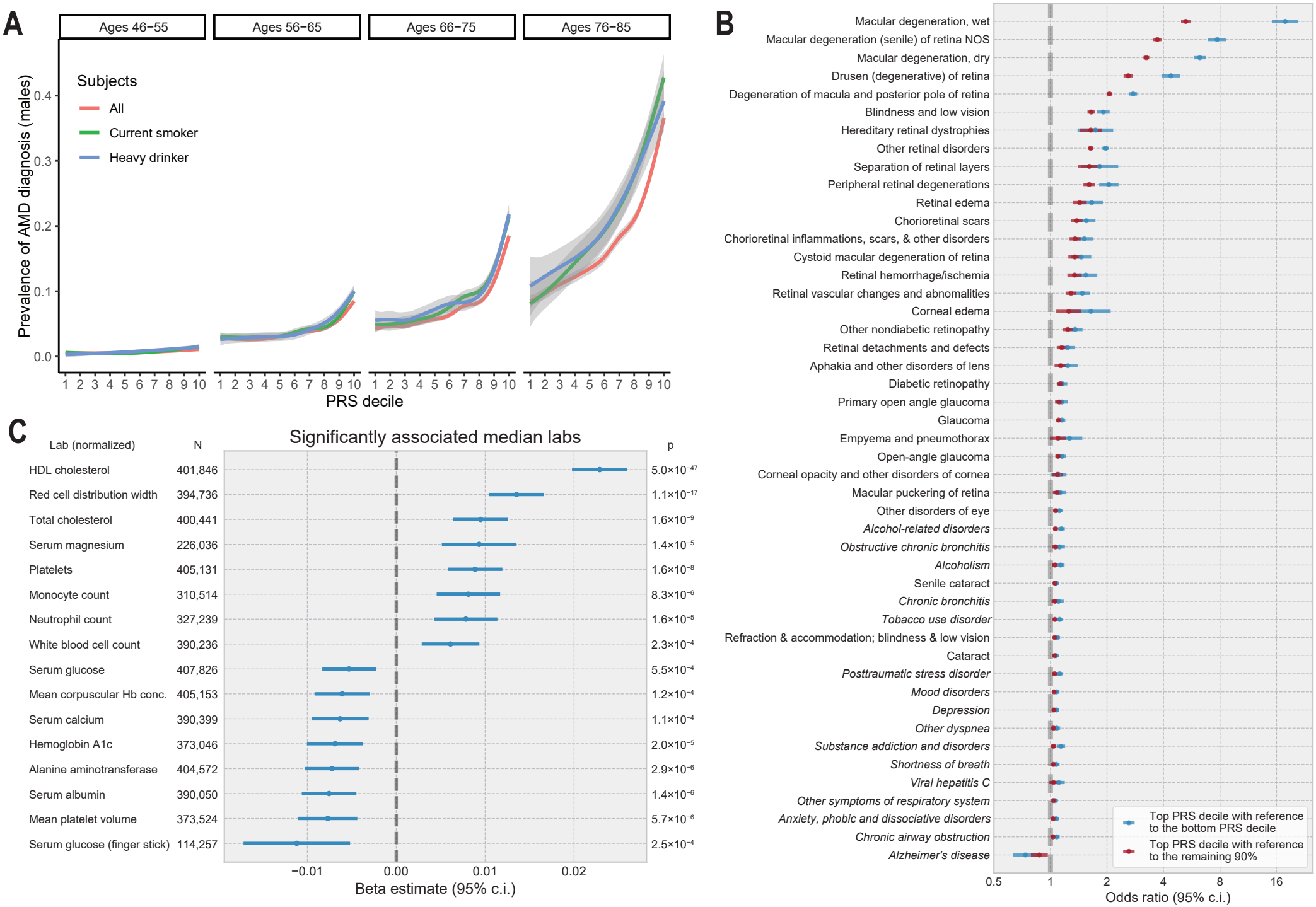
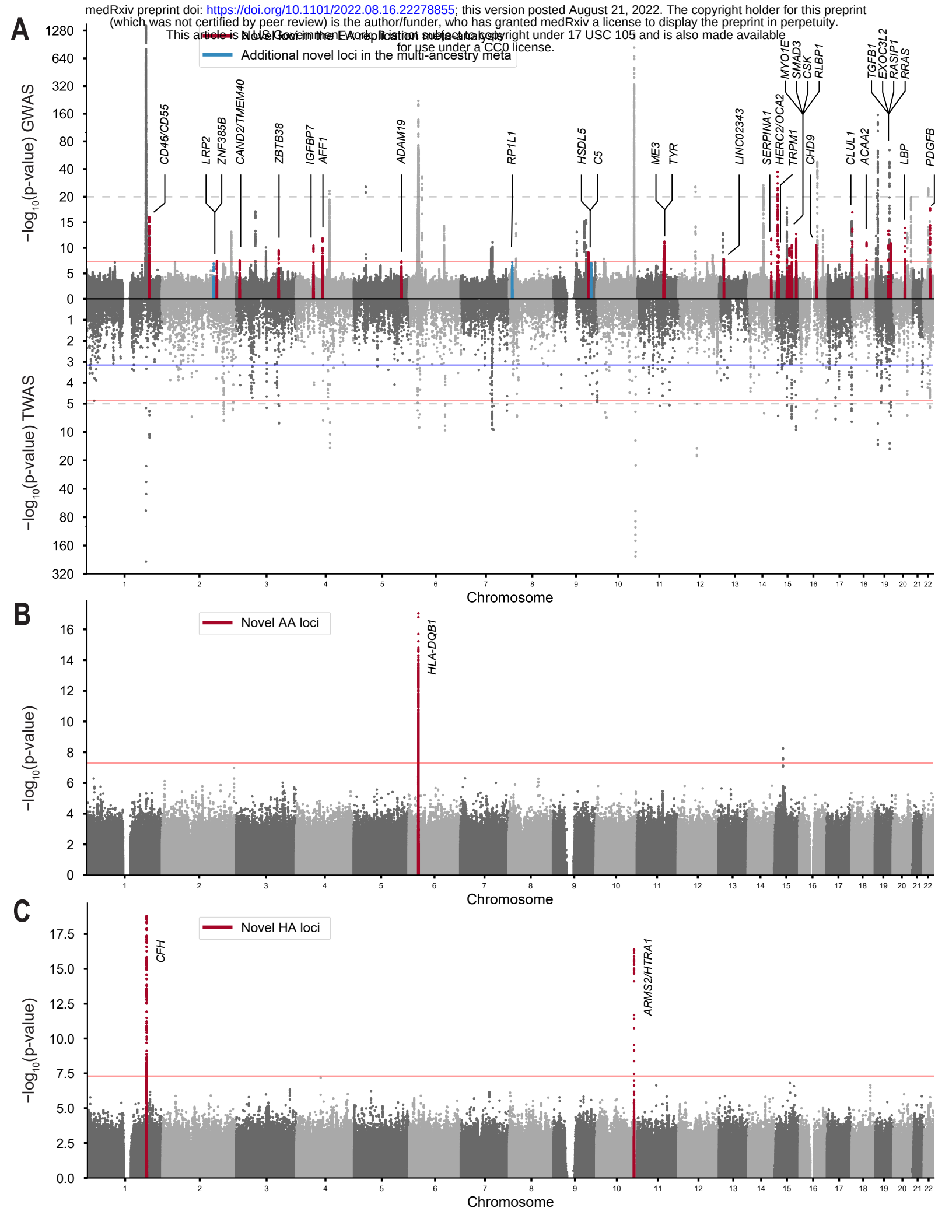


Fig. 2. Penetrance and pleiotropy of the AMD PRS. (**A**) Prevalence of AMD in EA males in MVP as a function of PRS decile, age range, and lifestyle risk factors. (**B**) PheWAS of the AMD PRS with 1,665 ICD code-based binary phenotypes in MVP EA. Phenotypes with association p-values less than 1×10^{-5} are shown. Odds ratios contrasting the top decile of PRS vs. the bottom decile (blue) and the top decile vs. the remaining 90% (red) are presented, with the x-axis in log scale. Phenotypes not directly related to vision are italicized. (**C**) LabWAS of the AMD PRS with 69 quantitative clinical lab measurements in MVP EA. Labs significant after Bonferroni adjustment for the number of labs are shown.

medRxiv preprint doi: https://doi.org/10.1101/2022.08.16.22278855; this version posted August 21, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
This article is a US Government work. It is not subject to copyright under 17 USC 105 and is also made available for use under a CC0 license.



medRxiv preprint doi: https://doi.org/10.1101/2022.08.16.22278855; this version posted August 21, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. This article is a US Government work. It is not subject to copyright under 17 USC 105 and is also made available for use under a CC0 license.

Fig. 3. GWAS analyses identify 27 novel loci in EA and the first loci in AA and HA. (**A**) Miami plot of the EA GWAS replication meta-analysis (57,290 cases, 324,430 controls) and DLPFC TWAS (45,045 cases, 275,806 controls). Red lines: genome-wide significance in the GWAS and TWAS analyses; blue line: FDR-corrected association p-value < 0.05 in the DLPFC TWAS; dashed gray line: transition from linear to log-scale on the y-axis. (**B**) Manhattan plot of the AA GWAS (2,302 cases, 29,223 controls). (A second potential locus near *COPS2* could not be confirmed in other ancestries.) (**C**) Manhattan plot of the HA GWAS (1,656 cases, 10,819 controls).

medRxiv preprint doi: https://doi.org/10.1101/2022.08.16.22278855; this version posted August 21, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. This article is a US Government work. It is not subject to copyright under 17 USC 105 and is also made available for use under a CC0 license.

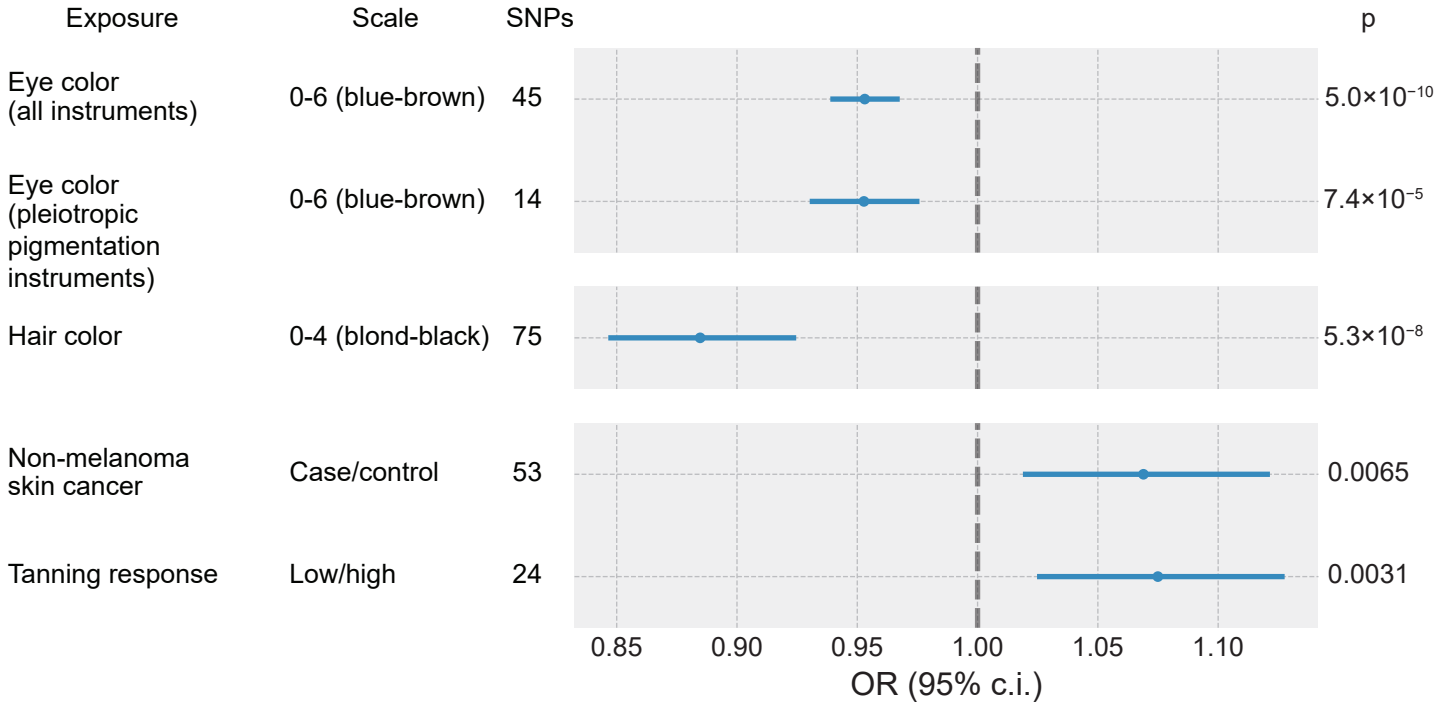


Fig. 4. MR of pigmentation traits on AMD risk in European ancestries. Forest plot of OR and 95% confidence intervals from the inverse variance-weighted MR analyses are shown. Eye color (*23*) was analyzed using all genome-wide significant instruments and using a selection of 14 instruments with known pleiotropic associations with other pigmentation traits. Other traits representing proxies for pigmentation, such as hair color (*24*), non-melanoma skin cancer (*25*), and tanning response (*26*) were also analyzed.

medRxiv preprint doi: https://doi.org/10.1101/2022.08.16.22278855; this version posted August 21, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. This article is a US Government work. It is not subject to copyright under 17 USC 105 and is also made available for use under a CC0 license.

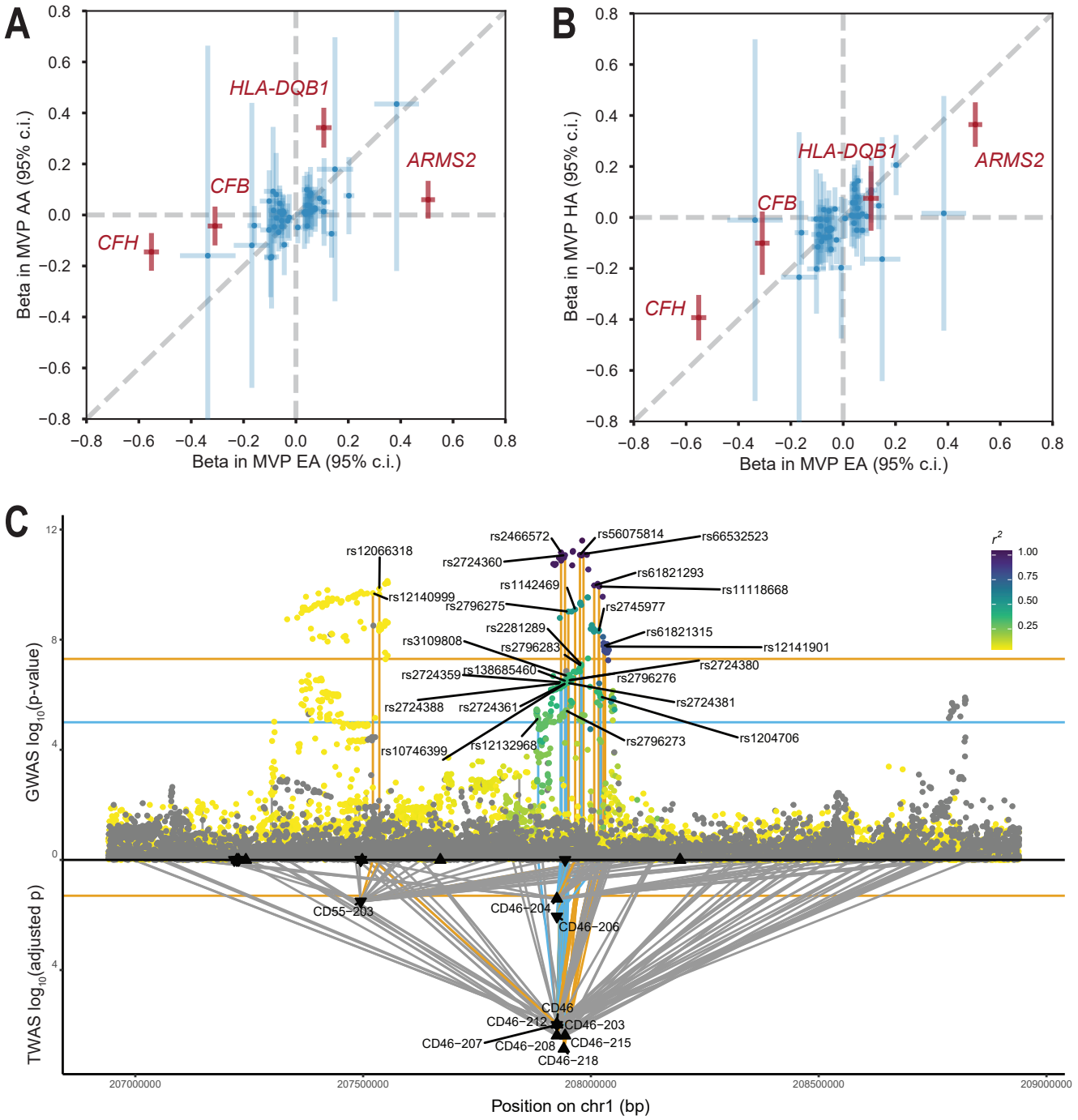
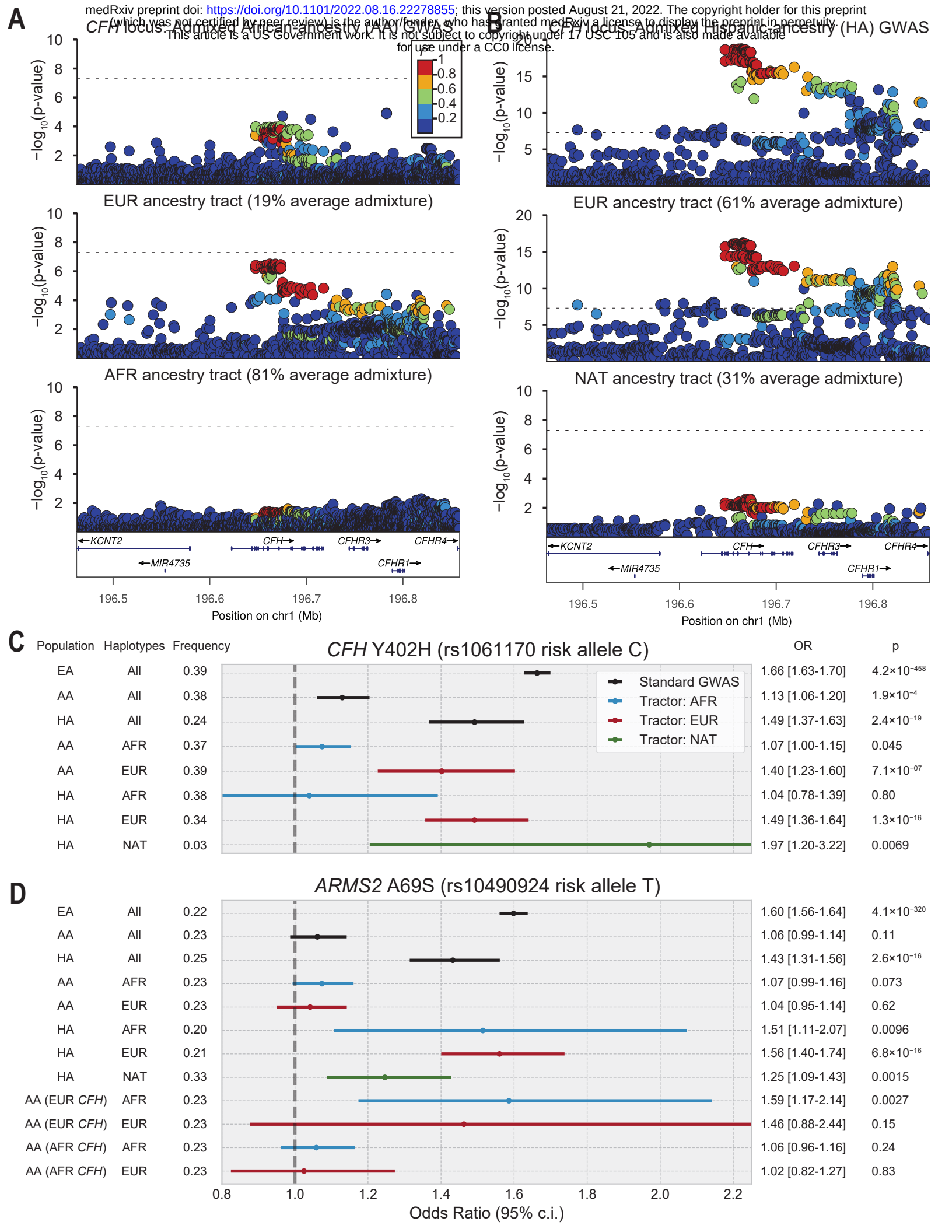


Fig. 5. Highlighted results from the multi-ancestry GWAS and TWAS. (**A**) Comparison of effect sizes (log-OR) of EA and AA index SNPs in the MVP EA GWAS (x-axis) and the MVP AA GWAS (y-axis). (**B**) Comparison of effect sizes in the MVP EA GWAS (x-axis) and the MVP HA GWAS (y-axis). Highlighted highly heterogeneous loci are shown in red. Error bars correspond to 95% confidence intervals. (**C**) Regional Miami plot of genetically regulated gene expression at the 1q32 (*CD46/CD55*) locus. GWAS results are above the axis and TWAS results from the DFPLC brain model below the axis. GWAS points are colored according to linkage disequilibrium r^2 with the top SNP.

medRxiv preprint doi: https://doi.org/10.1101/2022.08.16.22278855; this version posted August 21, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. This article is a US Government work. It is not subject to copyright under 17 USC 105 and is also made available for use under a CC0 license.



medRxiv preprint doi: https://doi.org/10.1101/2022.08.16.22278855; this version posted August 21, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. This article is a US Government work. It is not subject to copyright under 17 USC 105 and is also made available for use under a CC0 license.

Fig. 6. Local ancestry analysis of the *CFH* and *ARMS2/HTRA1* loci. (**A**) Regional association plots of AMD risk at *CFH* in a standard GWAS in AA individuals (top), and the EUR (middle) and AFR (bottom) tracts from the Tractor analysis. (**B**) Regional association plots of AMD risk at *CFH* in a standard GWAS in HA individuals (top), the EUR ancestry tract from the tractor analysis (middle) and the NAT ancestry tract (NAT) (bottom). (**C**) Forest plot comparing effect sizes and confidence intervals of the *CFH* Y402H risk allele in EA, AA, and HA populations, with Tractor analyses for AA (AFR and EUR tracts), and HA (EUR, NAT, and AFR tracts). (**D**) Forest plot comparing effect sizes and confidence intervals of the *ARMS2* A69S risk allele in EA, AA, and HA populations, with Tractor analyses for AA (AFR and EUR tracts), and HA (EUR, NAT, and AFR tracts). Additionally, Tractor effect size estimates in AA populations with either homozygous EUR ancestry at *CFH* ("EUR *CFH*") or homozygous AFR ancestry at *CFH* ("AFR *CFH*") are provided. Error bars are colored according to whether they correspond to a standard GWAS (black), or a Tractor GWAS, with blue for AFR tracts, red for EUR tracts, and green for NAT tracts.