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

Bryan R. Gorman[1,2,43], Georgios Voloudakis[3,4,43], Robert P. Igo Jr.[5,6,43,45], Tyler Kinzy[6], Christopher W. Halladay[7], Tim B. Bigdeli[8,9], Biao Zeng[3], Sanan Venkatesh[3,4], Jessica N. Cooke Bailey[5,6,10,11], Dana C. Crawford[5,6,10,11], Kyriacos Markianos[1], Frederick Dong[1,2], Patrick A. Schreiner[1,2], Wen Zhang[3], VA Million Veteran Program*, International AMD Genomics Consortium (IAMDGC)*, Tamer Hadi[12,13], Matthew D. Anger[14,15], Amy Stockwell[16], Ronald B. Melles[17], Jie Yin[18], Hélène Choquet[18], Rebecca Kaye[19,20], Karina Patasova[21,22], Praveen J. Patel[23,24], Brian L. Yaspan[16], Eric Jorgenson[25], Pirro G. Hysi[21,22,26,27], Andrew J. Lotery[19,20], J. Michael Gaziano[28,29], Philip S. Tsao[30,31,32], Steven J. Fliesler[15,33,34,35], Jack M. Sullivan[15,33,35], Paul B. Greenberg[36,37], Wen-Chih Wu[38,39], Themistocles L. Assimes[30,31], Saiju Pyarajan[1,40], Panagiotis Roussos[3,4,44✉], Neal S. Peachey[5,41,42,44✉], Sudha K. Iyengar[5,6,10,11,13,44✉],.,

[1]Center for Data and Computational Sciences (C-DACS), VA Cooperative Studies Program, VA Boston Healthcare System, Boston, MA, 02130, USA, [2]Booz Allen Hamilton, McLean, VA, 22102, USA, [3]Center for Disease Neurogenomics, Department of Psychiatry; Friedman Brain Institute; Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY, 10029, USA, [4]VISN 2 Mental Illness Research, Education, and Clinical Center (MIRECC), James J. Peters Veteran Affairs Medical Center, New York/New Jersey VA Health Care Network, Bronx, NY, 10468, USA, [5]Research Service, VA Northeast Ohio Healthcare System, Cleveland, OH, 44106, USA, [6]Department of Population and Quantitative Health Sciences, Case Western Reserve University, Cleveland, OH, 44106, USA, [7]Center for Innovation in Long Term Services and Supports, VA Providence Healthcare System, Providence, RI, 02908, USA, [8]Research Service, VA New York Harbor Healthcare System, Brooklyn, NY, 11209, USA, [9]Department of Psychiatry and Behavioral Sciences, SUNY Downstate Health Sciences University, Brooklyn, NY, 11203, USA, [10]Cleveland Institute for Computational Biology, Case Western Reserve University, Cleveland, OH, 44106, USA, [11]Department of Genetics & Genome Sciences, Case Western Reserve University, Cleveland, OH, 44106, USA, [12]Eye Clinic, VA Northeast Ohio Healthcare System, Cleveland, OH, 44106, USA, [13]Department of Ophthalmology and Visual Sciences, University Hospitals Eye Institute, Cleveland, OH, 44106, USA, [14]Eye Clinic, VA Western NY Healthcare System, Buffalo, NY, 14215, USA, [15]Ophthalmology, Jacobs School of Medicine and Biomedical Sciences, SUNY-University at Buffalo, Buffalo, NY, 14209, USA, [16]Department of Human Genetics, Genentech, South San Francisco, CA, 94080, USA, [17]Department of Ophthalmology, Kaiser Permanente Northern California, Redwood City, CA, 94063, USA, [18]Division of Research, Kaiser Permanente Northern California, Oakland, CA, 94612, USA, [19]Southampton Eye Unit, University Hospital Southampton National Health Services Trust, Southampton, U.K., [20]Clinical and Experimental Sciences, Faculty of Medicine, University of Southampton, Southampton, U.K.,

[21]Section of Ophthalmology, School of Life Course Sciences, King's College London, London, U.K., [22]Department of Twin Research and Genetic Epidemiology, King's College London, London, U.K., [23]National Institute for Health and Care Research Biomedical Research Centre, Moorfields Eye Hospital National Health Service Foundation Trust, London, U.K., [24]Institute of Ophthalmology, University College London, London, U.K., [25]Regeneron Genetics center, Tarrytown, NY, 10591, USA, [26]UCL Great Ormond Street Institute of Child Health, King's College London, London, U.K., [27]Sørlandet Sykehus Arendal, Arendal Hospital, Arendal, Norway, [28]Million Veteran Program Coordinating Center, VA Boston Healthcare System, Boston, MA, 02130, USA, [29]Division of Aging, Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, 02115, USA, [30]VA Palo Alto Epidemiology Research and Information Center for Genomics, VA Palo Alto Health Care System, Palo Alto, CA, 94304, USA, [31]Department of Medicine, Stanford University School of Medicine, Stanford, CA, 94305, USA, [32]Stanford Cardiovascular Institute, Stanford University School of Medicine, Stanford, CA, 94305, USA, [33]Research Service, VA Western NY Healthcare System, Buffalo, NY, 14215, USA, [34]Biochemistry, Jacobs School of Medicine and Biomedical Sciences, SUNY-University at Buffalo, Buffalo, NY, 14209, USA, [35]Graduate Program in Neurosciences, Jacobs School of Medicine and Biomedical Sciences, SUNY-University at Buffalo, Buffalo, NY, 14209, USA, [36]Section of Ophthalmology, VA Providence Healthcare System, Providence, RI, 02908, USA, [37]Division of Ophthalmology, Alpert Medical School, Brown University, Providence, RI, 02912, USA, [38]Section of Cardiology, Medical Service, VA Providence Healthcare System, Providence, RI, 02908, USA, [39]Division of Cardiology, Department of Medicine, Alpert Medical School, Brown University, Providence, RI, 02912, USA, [40]Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, 02115, USA, [41]Cole Eye Institute, Cleveland Clinic, Cleveland, OH, 44195, USA, [42]Department of Ophthalmology, Cleveland Clinic Lerner College of Medicine of Case Western Reserve University, Cleveland, OH, 44195, USA

*Lists of authors and their affiliations appear at the end of the paper.

[43]These authors contributed equally to this work.
[44]These authors jointly supervised this work.
[45]Deceased: Robert P. Igo, Jr.
✉e-mail: ski@case.edu; neal.peachey@va.gov; panagiotis.roussos@mssm.edu

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. Dissecting local ancestry in admixed individuals, we find significantly smaller marginal effect sizes for *CFH* risk alleles in African-ancestry haplotypes. 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.

Introduction

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, and driving.

AMD is a progressive disease, although progression rates vary by age, sex, and ancestry[1]. Lipoproteinaceous deposits called drusen formed between 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 macula. Choroidal neovascularization is the hallmark of 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]. Even greater reductions in risk are reported for wet AMD[2]. 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.

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 GWAS 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

Risk models for AMD in MVP

An overview of the present study is depicted in Fig. 1. Supplementary Table 1 summarizes the AMD cohorts used for our analyses. Among demographic factors (Supplementary Tables 2a, 2b), not surprisingly, increased age was significantly associated with AMD status. Female sex was significantly associated with AMD status, even in MVP, where 91% of participants are male; this may be partially mediated by the longer lifespan of females. In multiple regression models, we also observed associations with smoking, alcohol, high-density lipoprotein (HDL) cholesterol, body mass index, and serum albumin (Supplementary Note 1).

Using summary statistics from a prior GWAS conducted by the International AMD Genomics Consortium[8] (IAMDGC), we constructed polygenic risk scores (PRS) in MVP using PRS-CS software[10]. We observed 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, to less than 10% in the bottom decile, with similar results observed in females (Supplementary Fig. 1). History of heavy alcohol use and smoking are significant lifestyle risk factors (Supplementary Table 2b), 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 from entry into the VA healthcare system 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 (Supplementary Table 2c). Moreover, the PRS demonstrated weak transferability to non-European ancestries in our selected case-control sample (Extended Data Fig. 1): 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 (Supplementary Table 2b), 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; PheWAS) and laboratory measures (i.e., laboratory-wide association study; 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; Supplementary Table 3). 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 and OR=6.2 [5.8-6.7] for dry AMD. The difference likely reflects the high proportion of wet AMD cases in the IAMDGC base GWAS (77%) and greater severity of disease captured by the wet phecode compared to the dry phecode. Alcohol and tobacco exposure showed strong associations. Other positive associations included cataracts, refractive error, and mood disorders. 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, and negative associations with triglycerides, serum albumin and serum glucose (Fig. 2c; Supplementary Table 4).

European-ancestry meta-analysis of AMD

We performed GWAS of AMD in three MVP tranches and meta-analysis with the summary statistics from five other cohorts (IAMDGC, GERA, UK Biobank, Genentech geographic atrophy, and Genentech choroidal neovascularization) for a total EA analysis comprising 57,290 AMD cases and 324,430 controls (Supplementary Table 1). This analysis (Fig. 3a) revealed 60 index variants: 33 previously reported and 27 novel significant loci (Supplementary Table 5; Supplementary Fig. 2). The genomic control parameter (λ_{GC}) was 1.09, indicating minimal systemic inflation of the test statistic (Supplementary Fig. 3). Due to oversampling for severity in the IAMDGC, effect sizes of the index variants at major association peaks were lower overall in the MVP EA cohorts, similar to other EHR-based cohorts (Supplementary Table 6). 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* (rs114092250, $p=4.0\times10^{-5}$), *TRPM3* (rs11142636, $p=3.6\times10^{-6}$), and *VTN* (rs11080055, $p=1.9\times10^{-6}$).

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 RNA of uncharacterized function. A conditional analysis revealed 178 conditionally independent significant variants in or near the 60 loci (Supplementary Table 7).

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 (Extended Data Fig. 2). 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. Both variants were directly genotyped on the MVP array. 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 is also associated 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 (r^2 = 0.9 with the lead SNP) is protective for AMD (OR=0.81 [0.77-0.86; $p=3.2\times10^{-12}$]), while PiS (r^2 < 0.01 with the lead SNP) 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 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] (Supplementary Table 8). Similarly, genetic risk to intraocular pressure was inversely correlated with that of AMD. We scanned genetic correlations with blood-based biomarkers (Pan-UKB team, 2020) and found the strongest correlations with serum albumin (r_G=-0.12; $p=4.0\times10^{-4}$), albumin/globulin ratio, and kidney biomarkers phosphate, calcium, and cystatin C (Supplementary Table 9).

We explored the causality of pigmentation associations in EA populations using two-sample Mendelian randomization (MR) (Fig. 4; Supplementary Table 10). Comparing AMD data with those from a study[23] in which eye color was measured on a quantitative scale (range 0-6), we find that each step towards darker eye color decreases the risk of AMD by about 5% (OR=0.95 [0.94-0.97]). 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[24] (range 0-4) 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 pigmentation, 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. All traits had non-significant MR-Egger intercepts (Supplementary Table 10), consistent with most of the effect occurring through vertical (causal) pleiotropy[27]. Finally, we selected only eye color instruments with known associations across multiple pigmentary traits, which likely represent the best proxies for melanin metabolism in the retina[23] and obtained a similar effect estimate (OR=0.95 [0.93-0.98]). We strongly believe that these examples are simply proxies for fundamental pigmentation processes and approximate the presence of melanin in the RPE and choroid.

Discovery of AMD risk loci in African and Hispanic ancestries

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 (Fig. 3b; Supplementary Fig. 4). Supporting the AA result, the *HLA-DQB1* locus was significant in the EA GWAS at a direction-consistent but lower OR (Fig. 5a; Supplementary Table 11). 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 (Supplementary Table 12). 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; Supplementary Fig. 5; Supplementary Table 13), 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.

Multi-ancestry AMD meta-analysis and fine-mapping

To capture cross-ancestry effects and increase discovery power, we conducted a multi-ancestry meta-analysis, combining our EA meta-analysis and AA and HA GWAS reaching a total of 61,248 cases and 364,472 controls. 62 loci were genome-wide significant, including three novel loci beyond those reported above for the EA meta-analysis: *LRP2*, *RP1L1*, and complement *C5* (Supplementary Table 14; Supplementary Fig. 5), and all significant EA loci except for *ADAM19*. *LRP2* encodes the transmembrane low-density lipoprotein receptor megalin, which is expressed in the RPE[28,29], and *RP1L1* encodes a retina-specific protein involved in photoreceptor differentiation and function[30]. As a sensitivity analysis, we further performed random effects meta-analyses using the Han-Eskin method[31] (RE2), which achieved similar results (Supplementary Table 14).

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 EA and multi-ancestry meta-analyses using FINEMAP[32] (Supplementary Tables 15-16). To co-localize eQTLs, multivariate multiple QTL[33] (mmQTL) was first applied to detect eQTLs; eCAVIAR[34] 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 six candidate causal variants modifying the gene expression of *CD46, RDH5*, *BLOC1S1*, *PMS2P1*, *CFI*, and *ABCA1* (Supplementary Table 17). Expression co-localization of the gene-variant pair with the highest CLPP, BLOC1S1 (rs3138141), is shown in Supplementary Fig. 6.

Haplotype ancestry at CFH alters risk in admixed populations

We conducted local ancestry analyses to investigate the contribution of genetic risk at *CFH* in admixed individuals. We performed local ancestry inference using RFMIX2[35], 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* (Supplementary Table 18). 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[36], 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), the conventional GWAS demonstrated a large effect (OR=1.49 [1.37-1.63]; $p=2.4\times10^{-19}$) like 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 (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.

We explored whether smaller *CFH* Y402H effect size observed in AFR could be explained through haplotype analysis. In particular, a protective deletion of *CFHR3* and *CFHR1*[37] has a higher frequency in AFR haplotypes (40% compared to 20% in EUR haplotypes[38]) and has been proposed as an explanation for lower rates of AMD observed in AA[39]. 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[40,41], excluding SNPs that are rare in AFR haplotypes (Supplementary Table 19). 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[42].

ARMS2 genetic risk does not vary by haplotype ancestry

We similarly 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 EUR haplotypes at *ARMS2* in admixed AA or HA subjects using Fisher's Exact Test (Supplementary Table 20). We tested the *ARMS2* locus using the Tractor method (Fig. 6d; Extended Data Fig. 3). Within AA subjects, ORs at the *ARMS2* A69S 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 ORs. 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 ORs in AFR and EUR tracts similar to the EA and HA GWAS, with the AFR tract reaching statistical significance (p=0.0026). In contrast, ORs 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 (Supplementary Table 21) and observed a significant association of AMD with the interaction term (OR=1.11 [1.08-1.14]; $p=1.6\times10^{-13}$).

HLA fine-mapping identifies an MHC Class II haplotype

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[43]. Full summary statistics are provided in Supplementary Tables 22-24. Within MVP EA (32,567 cases and 130,444 controls), we replicated previously reported MHC Class II allele associations[44] and identified four novel associations in Class I and Class II alleles reaching genome-wide significance (Supplementary Note 2). 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]; $p=5.2\times10^{-15}$), DQA1*02:01 (1.39 [1.26-1.52]; $p=2.4\times10^{-11}$), and DQB1*02:02 (1.39 [1.27-1.52]; $p=2.1\times10^{-12}$). 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[44] and asparaginase hypersensitivity[45,46]. We then directly tested the risk haplotype with tagging SNPs rs28383172 and rs7775228[47]. In an additive model, the risk haplotype was highly associated with AMD, with OR=1.51 [1.37-1.66; $p=1.8\times10^{-16}$] in AA and replicated with a smaller effect size, OR=1.16 [1.12-1.20; $p=2.1\times10^{-15}$], 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 ($p=1.2\times10^{-26}$) and dry AMD ($p=4.7\times10^{-9}$), which were corroborated in the EA analysis ($p=6.7\times10^{-4}$ and $p=2.4\times10^{-3}$, respectively). Additionally, we replicated the celiac disease association in the EA cohort ($p=1.0\times10^{-12}$). Antibodies directed against HLA-DR showed immunoreactivity in drusen[48], supporting biological plausibility.

Multi-tissue TWAS identifies novel AMD-associated genes

To pair the GWAS signals with functional gene units, predictive models of gene expression[49] were used to perform a multi-tissue transcriptome-wide association study (TWAS), based on our EA meta-analysis. We performed TWAS using the high-powered dorsolateral prefrontal cortex (DLPFC) dataset from the PsychENCODE Consortium (PEC)[50] (Supplementary Table 25), as well as retina[51] (Supplementary Table 26), and 36 other tissue models built from GTEx[52] and STARNET[53]. We then meta-analyzed tissue p-values using ACAT[54] (Supplementary Table 27). AMD-related gene expression was highly correlated across tissues. Consistent with the high protein and gene expression similarity among retina and brain[55] (Human Protein Atlas, proteinatlas.org), we found that the retina TWAS clustered together with the TWAS in DLPFC and other brain tissues (Extended Data Fig. 4). A comparison with previous TWAS studies[51,56] is summarized in Supplementary Table 28.

Applying the DLPFC model, 218 genes and 431 transcripts were significant at the FDR-corrected $p<0.05$ association threshold (Fig. 3a). Novel TWAS genes, arising from either the gene or transcript level analysis, include regulatory genes for protein glycosylation (*MAN2C1*, *MPI*, *MANBA*, and *CTSA*), lipid metabolism (*CHD9*), cell cycle (*FRK*), neurogenesis and neural differentiation (*NTN5* and *NIF3L1*), and complement (CD59) (Supplementary Fig. 7).

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. 7), 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-data-based MR (SMR) experiments (Supplementary Table 29), 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 and the tissue meta-analysis (Supplementary Tables 30-31; Extended Data Fig. 5). 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, as well as their associated molecular functions of transmembrane protein transport and lactonohydrolase activity (Extended Data Fig. 6). Finally, we conducted a gene-based rare variant burden scan of the MVP EA cohort (Supplementary Table 32) using genotyped markers on the MVP 1.0 array[57], which is enriched in protein-altering rare variants. This analysis confirmed the risk-increasing effect of rare *CFH* and *CFI* mutations. Additionally, TWAS genes were significantly enriched for rare variant associations (Extended Data Fig. 7).

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 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. 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[58,59]. 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 an interaction effect between the *CFH* and *ARMS2* risk alleles in EA, suggesting a potential connection between the two principal AMD risk loci (Supplementary Discussion).

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[60].

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[61], 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[36], which can leverage differences in effect size, frequency, and LD to tease apart mechanisms of genetic risk and improve risk prediction.

*Consortium Authors and Affiliations

VA Million Veteran Program

J. Michael Gaziano[28,29], Philip S. Tsao[30,31,32], Saiju Pyarajan[1,40]

International AMD Genomics Consortium

Lars G. Fritsche[46], Wilmar Igl[47], Jessica N. Cooke Bailey[6,10,11], Felix Grassman[48], Sebanti Sengupta[46], Jennifer L. Bragg-Gresham[46,49], Kathryn P. Burdon[50], Scott J. Hebring[51], Cindy Wen[52,53], Mathias Gorski[47], Ivana K. Kim[54], David Cho[55], Donald Zack[56,57,58,59,60], Eric Souied[61], Hendrik P.N. Scholl[56,62], Elisa Bala[5], Kristine E. Lee[63], David J. Hunter[64,65], Rebecca J. Sardell[66], Paul Mitchell[67,68], Joanna E. Merriam[69], Valentina Cipriani[24,70], Joshua D. Hoffman[71], Tina Schick[72], Yara T.E. Lechanteur[73], Robyn H. Guymer[74,75], Matthew P. Johnson[76], Yingda Jiang[77], Chloe M. Stanton[78], Gabriëlle H.S. Buitendijk[79,80], Xiaowei Zhan[46,81,82], Alan M. Kwong[46], Alexis Boleda[83], Matthew Brooks[83], Linn Gieser[83], Rinki Ratnapriya[83], Kari E. Branham[84], Johanna R. Foerster[46], John R. Heckenlively[84], Mohammad I. Othman[84], Brendan J. Vote[50], Helena Hai Liang[75,85], Emmanuelle Souzeau[86], Ian L. McAllister[87], Timothy Isaacs[87], Janette Hall[86], Stewart Lake[86], David A. Mackey[50,75,85,87], Ian J. Constable[87], Jamie E. Craig[86], Terrie Kitchner[51], Zhenglin Yang[88,89,90], Zhiquang Su[91], Hongrong Luo[52,53], Daniel Chen[52,53], Ouyang Ouyang[52,53], Ken Flagg[52,53], Daniel Lin[52,53], Guanping Mao[52,53], Henry Ferreyra[52,53], Klaus Stark[47], Claudia Nvon Strachwitz[92], Armin Wolf[93], Caroline Brandl[47,48,94], Guenther Rudolph[93], Matthias Olden[47], Margaux A. Morrison[95], Deniste J. Morgan[95], Matthew Schu[96,97,98,99,100], Jeeyun Ahn[101], Giuliana Silvestri[102], Evangelia E. Tsironi[103], Kyu Hyung Park[104], Lindsay A. Farrer[96,97,98,99,100], Anton Orlin[105], Alexander Brucker[106], Mingyao Li[107], Christine A. Curcio[108], Saddek Mohand-Saïd[109,110,111,112], José-Alain Sahel[109,110,111,112,113,114,115], Isabelle Audo[109,110,111,116], Mustapha Benchaboune[112], Angela J. Cree[20], Christina A. Rennie[117], Srinivas V. Goverdhan[20], Michelle Grunin[118], Shira Hagbi-Levi[118], Peter Campochiaro[56,58], Nicholas Katsanis[119,120,121], Frank G. Holz[62], Frédéric Blond[109,110,111], Hélène Blanché[122], Jean-François Deleuze[122,123], Robert P. Igo Jr.[45,124], Barbara Truitt[124], Neal S. Peachey[5,41,42], Stacy M. Meuer[63], Chelsea E. Myers[63], Emily L. Moore[63], Ronald Klein[63], Michael A. Hauser[125,126,127], Eric A. Postel[125], Monique D. Courtenay[66], Stephen G. Schwartz[128], Jaclyn L. Kovach[128], William K. Scott[66], Gerald Liew[67,68], Ava G. Tan[67,68], Bamini Gopinath[67,68], John C. Merriam[69], R. Theodore Smith[69,129], Jane C. Khan[87,130,131], Humma Shahidi[131,132], Anthony T. Moore[70,113,133], J. Allie McGrath[71], Reneé Laux[6], Milam A. Brantley Jr.[134], Anita Agarwal[134], Lebriz Ersoy[72], Albert Caramoy[72], Thomas Langmann[72], Nicole T.M. Saksens[73], Eiko K. deJohn[73], Carel B. Hoyng[73], Melinda S. Cain[74,75], Andrea J. Richardson[74,75], Tammy M. Martin[135], John Blangero[76], Daniel E. Weeks[77,136], Bal Dhillon[137],

Cornelia M. van Duijn[80], Kimberly F. Doheny[138], Jane Romm[138], Caroline C.W. Klaver[79,80], Caroline Hayward[78], Michael B. Gorin[139,140], Michael L. Klein[135], Paul N. Baird[74,75], Anneke I. den Hollander[73,141], Sascha Fauser[72], John R.W. Yates[70.113,131], Rando Allikmets[69,142], Jie Jin Wang[67,68], Debra A. Schaumberg[64,143,144], Barbara E.K. Klein[63], Stephanie A. Hagstrom[41], Itay Chowers[118], Andrew J. Lotery[20], Thierry Léveillard[109,110,111], Kang Zhang[52,53,91], Murray H. Brilliant[51], Alex W. Hewitt[50,74,75,87], Anand Swaroop[83], Emily Y. Chew[145], Margaret A. Pericak-Vance[66], Margaret DeAngelis[95], Dwight Stambolian[55], Jonathan L. Haines[124,146], Sudha K. Iyengar[6,10,11,13], Bernhard H.F. Weber[48], Gonçalo R. Abecasis[46], Iris M. Heid[47]

[46]Department of Biostatistics, Center for Statistical Genetics, University of Michigan, Ann Arbor, MI, USA, [47]Department of Genetic Epidemiology, University of Regensburg, Regensburg, Germany, [48]Institute of Human Genetics, University of Regensburg, Regensburg, Germany, [49]Department of Internal Medicine–Nephrology, Kidney Epidemiology and Cost Center, University of Michigan, Ann Arbor, MI, USA, [50]Menzies Research Institute Tasmania, School of Medicine, University of Tasmania, Hobart, Tasmania, Australia, [51]Center for Human Genetics, Marshfield Clinic Research Foundation, Marshfield, WI, USA, [52]Department of Ophthalmology, University of California, San Diego, San Diego, CA, USA, [53]Veterans Affairs San Diego Health System, La Jolla, CA, USA, [54]Department of Ophthalmology, Retina Service, Massachusetts Eye and Ear, Harvard University, Boston, MA, USA, [55]Department of Ophthalmology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA, [56]Department of Ophthalmology, Wilmer Eye Institute, Johns Hopkins University School of Medicine, Baltimore, MD, USA, [57]Department of Molecular Biology and Genetics, Johns Hopkins University School of Medicine, Baltimore, MD, USA, [58]Department of Neuroscience, Johns Hopkins University School of Medicine, Baltimore, MD, USA, [59]Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, MD, USA, [60]Institue de la Vision, Université Pierre et Marie Curie, Paris, France, [61]Hôpital Intercommunal de Créteil, Hôpital Henri Mondor, Université Paris Est Créteil, Créteil, France, [62]Department of Ophthalmology, University of Bonn, Bonn, Germany, [63]Department of Ophthalmology and Visual Sciences, University of Wisconsin, Madison, WI, USA, [64]Department of Epidemiology, Harvard School of Public Health, Boston, MA, USA, [65]Department of Nutrition, Harvard School of Public Health, Boston, MA, USA, [66]John P. Hussman Institute for Human Genomics, Miller School of Medicine, University of Miami, Miami, FL, USA, [67]Centre for Vision Research, Department of Ophthalmology, University of Sydney, Sydney, New South Wales, Australia, [68]Westmead Millennium Institute for Medical Research, University of Sydney, Sydney, New South Wales, Australia, [69]Department of

Ophthalmology, Columbia University, New York, NY, USA, [70]Moorfields Eye Hospital, London, U.K., [71]Center for Human Genetics Research, Vanderbilt University Medical Center, Nashville, TN, USA, [72]Department of Ophthalmology, University Hospital of Cologne, Cologne, Germany, [73]Department of Ophthalmology, Radboud University Medical Centre, Nijmegen, The Netherlands, [74]Centre for Eye Research Australia, University of Melbourne, East Melbourne, Victoria, Australia, [75]Royal Victorian Eye and Ear Hospital, East Melbourne, Victoria, Australia, [76]South Texas Diabetes and Obesity Institute, School of Medicine, University of Texas Rio Grande Valley, Brownsville, TX, USA, [77]Department of Biostatistics, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, PA, USA, [78]Medical Research Council (MRC) Human Genetics Unit, Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh, U.K., [79]Department of Ophthalmology, Erasmus Medical Center, Rotterdam, The Netherlands, [80]Department of Epidemiology, Erasmus Medical Center, Rotterdam, The Netherlands, [81]Department of Clinical Science, Quantitative Biomedical Research Center, University of Texas Southwestern Medical Center, Dallas, TX, USA, [82]Center for the Genetics of Host Defense, University of Texas Southwestern Medical Center, Dallas, TX, USA, [83]Neurobiology, Neurodegeneration and Repair Laboratory, National Eye Institute, US National Institutes of Health, Bethesda, MD, USA, [84]Department of Ophthalmology and Visual Sciences, Kellogg Eye Center, University of Michigan, Ann Arbor, MI, USA, [85]Centre for Eye Research Australia, University of Melbourne, East Melbourne, Australia, [86]Department of Ophthalmology, Flinders Medical Centre, Flinders University, Adelaide, South Australia, Australia, [87]Centre for Ophthalmology and Visual Science, Lions Eye Institute, University of Western Australia, Perth, Western Australia, Australia, [88]Sichuan Provincial Key Laboratory for Human Disease Gene Study, Hospital of the University of Electronic Science and Technology of China, Chengdu, China, [89]Sichuan Provincial People's Hospital, Chengdu, China, [90]Sichuan Translational Medicine Hospital, Chinese Academy of Sciences, Chengdu, China, [91]Molecular Medicine Research Center, State Key Laboratory of Biotherapy, West China Hospital, Sichuan University, Chengdu, China, [92]EyeCentre Southwest, Stuttgart, Germany, [93]University Eye Clinic, Ludwig Maximilians University, Munich, Germany, [94]Department of Ophthalmology, University Hospital Regensburg, Regensburg, Germany, [95]Department of Ophthalmology and Visual Sciences, University of Utah, Salt Lake City, UT, USA, [96]Department of Medicine (Biomedical Genetics), Boston University Schools of Medicine and Public Health, Boston, MA, USA, [97]Department of Ophthalmology, Boston University Schools of Medicine and Public Health, Boston, MA, USA, [98]Department of Neurology, Boston University Schools of Medicine and Public Health, Boston, MA, USA, [99]Department of Epidemiology, Boston University Schools of Medicine and Public

Health, Boston, MA, USA, [100]Department of Biostatistics, Boston University Schools of Medicine and Public Health, Boston, MA, USA, [101]Department of Ophthalmology, Seoul Metropolitan Government Seoul National University Boramae Medical Center, Seoul, Republic of Korea, [102]Centre for Experimental Medicine, Queen's University, Belfast, U.K., [103]Department of Ophthalmology, School of Medicine, University of Thessaly, Larissa, Greece, [104]Department of Ophthalmology, Seoul National University Bundang Hospital, Seongnam, Republic of Korea, [105]Department of Ophthalmology, Weill Cornell Medical College, New York, NY, USA, [106]Department of Ophthalmology, Scheie Eye Institute, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA, USA, [107]Department of Biostatistics and Epidemiology, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA, USA, [108]Department of Ophthalmology, University of Alabama at Birmingham, Birmingham, AL, USA, [109]INSERM, Paris, France, [110]Department of Genetics, Institut de la Vision, Paris, France, [111]Centre National de la Recherche Scientifique (CNRS), Paris, France, [112]Centre Hospitalier National d'Ophtalmologie des Quinze-Vingts, Paris, France, [113]University College London Institute of Ophthalmology, University College London, London, U.K., [114]Fondation Ophtalmologique Adolphe de Rothschild, Paris, France, [115]Académie des Sciences–Institut de France, Paris, France, [116]Department of Molecular Genetics, Institute of Ophthalmology, London, U.K., [117]University Hospital Southampton, Southampton, U.K., [118]Department of Ophthalmology, Hadassah Hebrew University Medical Center, Jerusalem, Israel, [119]Center for Human Disease Modeling, Duke University, Durham, NC, USA, [120]Department of Cell Biology, Duke University, Durham, NC, USA, [121]Department of Pediatrics, Duke University, Durham, NC, USA, [122]Centre d'Etude du Polymorphisme Humain (CEPH) Fondation, Jean Dausset, Paris, France, [123]Commissariat à l'Energie Atomique et aux Energies Alternatives (CEA), Institut de Génomique, Centre National de Génotypage, Evry, France, [124]Department of Epidemiology and Biostatistics, Case Western Reserve University, Cleveland, OH, 44106, USA, [125]Department of Ophthalmology, Duke University Medical Center, Durham, NC, USA, [126]Department of Medicine, Duke University Medical Center, Durham, NC, USA, [127]Duke Molecular Physiology Institute, Duke University Medical Center, Durham, NC, USA, [128]Bascom Palmer Eye Institute, University of Miami Miller School of Medicine, Naples, FL, USA, [129]Department of Ophthalmology, New York University School of Medicine, New York, NY, USA, [130]Department of Ophthalmology, Royal Perth Hospital, Perth, Western Australia, Australia, [131]Department of Medical Genetics, Cambridge Institute for Medical Research, Cambridge, U.K., [132]Department of Ophthalmology, Cambridge University Hospitals National Health Service (NHS) Foundation Trust, Cambridge, U.K., [133]Department of Ophthalmology, University of California San Francisco Medical School,

San Francisco, CA, USA, [134]Department of Ophthalmology and Visual Sciences, Vanderbilt University Medical Center, Nashville, TN, USA, [135]Casey Eye Institute, Oregon Health and Science University, Portland, OR, USA, [136]Department of Human Genetics, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, PA, USA, [137]School of Clinical Sciences, University of Edinburgh, Edinburgh, U.K., [138]Center for Inherited Disease Research (CIDR) Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, MD, USA, [139]Department of Ophthalmology, Stein Eye Institute, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, CA, USA, [140]Department of Human Genetics, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, CA, USA, [141]Department of Human Genetics, Radboud University Medical Centre, Nijmegen, The Netherlands, [142]Department of Pathology and Cell Biology, Columbia University, New York, NY, USA, [143]Center for Translational Medicine, Moran Eye Center, University of Utah School of Medicine, Salt Lake City, UT, USA, [144]Division of Preventive Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA, [145]Division of Epidemiology and Clinical Applications, Clinical Trials Branch, National Eye Institute, US National Institutes of Health, Bethesda, MD, USA, [146]Institute for Computational Biology, Case Western Reserve University, Cleveland, OH, 44106, USA

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 and Rachel Mester (UCLA) 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. This research has been conducted using the UK Biobank Resource under Application Number 2112. Support from the VA Office of Research & Development is acknowledged by N.S.P. (I01BX003364, I01 BX04557, IK6BX005233), P.R. (I01BX004189) and S.J.F. (IK6BX005787). We acknowledge NIH support to P.R. (R01AG050986, R01AG065582, R01AG067025, R01MH125246, U01MH116442), to G.V. (K08MH122911), to H.C. and E.J. (R01EY027004) and to R.B.M., J.Y. and H.C. (RC2AG036607). We acknowledge NIH Core Grants to the Departments of Ophthalmology at Case Western Reserve University (P30EY011373) and Cleveland Clinic School of Medicine at Case Western Reserve University (P30EY025585) and unrestricted support from Research to Prevent Blindness to the Departments of Ophthalmology at Case Western Reserve University, Cleveland Clinic School of Medicine at Case Western Reserve University, and SUNY-Buffalo. The IAMDGC acknowledges support from the NIH (R01EY022310, X01HG006934). This project was supported by the Clinical and Translational Science Collaborative of Cleveland (UL1TR002548). S.K.I. was supported by the Retinal Research Foundation. P.G.H. received support from the BrightFocus Foundation. This research was supported by the Robert Wood Johnson Foundation, Research Program on Genes, the Environment and Health Wayne and Gladys Valley Foundation, the Ellison Medical Foundation and Kaiser Permanente Community Benefit Programs. 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). 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.

Author Contributions

Conceptualization: B.R.G., G.V., R.P.I., P.R., N.S.P., S.K.I. Acquisition, analysis or interpretation of data: B.R.G., G.V., R.P.I., T.K., C.W.H., T.B.B., B.Z., S.V., J.N.C.B., D.C.C., K.M., F.D., P.S., W.Z., T.H., M.D.A., A.D.S., R.B.M., J.Y., H.C., R.K., K.P., P.J.P., B.L.Y., E.J., P.G.H., A.J.L., J.M.G., P.S.T., S.J.F., J.M.S., P.B.G., W.-C.W., T.L.A., S.P., P.R., N.S.P., S.K.I. Drafting of the manuscript: B.R.G., G.V., R.P.I., P.R., N.S.P., S.K.I. Critical revision of the manuscript for important intellectual content: B.R.G., G.V., T.K., T.B.B., J.N.C.B., D.C.C., P.G.H., A.J.L., S.J.F., H.C., B.L.Y., P.B.G., P.R., N.S.P., S.K.I.

Competing interests

A.S. and B.L.Y. are employees of Genentech/Roche and hold stock and stock options in Roche. E.J. is an employee and stockholder of Regeneron Genetics Center. The rest of the authors declare that they have no competing interest.

Figure Legends

Fig. 1. Overview of AMD GWAS meta-analysis primary and secondary analyses. (a) PRS-PheWAS and PRS-LabWAS analyses using the external European-ancestry IAMDGC GWAS of AMD as a base for scoring subjects in MVP. (b) Ancestry-specific GWAS analyses, and multi-ancestry meta-analysis, with corresponding secondary analyses including TWAS and Mendelian randomization (EA-only), and fine-mapping (EA and multi-ancestry). (c) HLA fine-mapping and rare variant analysis. 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 (95% CI) 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 unadjusted association p-values less than 1×10^{-5} are shown. ORs and 95% CIs 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. Full summary statistics are provided in Supplementary Table 3. Phenotypes not directly related to vision are italicized. (c) LabWAS of the AMD PRS with 69 quantitative clinical lab measurements in MVP EA. Beta values and 95% CIs are presented. Unadjusted p-values are shown; only those labs significant after Bonferroni adjustment are plotted. Full summary statistics are provided in Supplementary Table 4.

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 meta-analysis and DLPFC TWAS (*n* = 57,290 cases, 324,430 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 (*n* = 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 (*n* = 1,656 cases, 10,819 controls).

Fig. 4. MR of pigmentation traits on AMD risk in European ancestries (*n* = 57,290 cases, 324,430 controls). Forest plot of OR per unit change in the exposure (scale provided) and 95% CIs 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. Marked cross-ancestry heterogeneity at major AMD risk loci. (a) Comparison of effect sizes (log-OR) of EA and AA index SNPs in the MVP EA GWAS (x-axis; *n* = 32,309 cases, 210,911 controls) and the MVP AA GWAS (y-axis; *n* = 2,302 cases, 29,223 controls). (b) Comparison of effect sizes in the MVP EA GWAS (x-axis; *n* = 32,309 cases, 210,911 controls) and the MVP HA GWAS (y-axis; *n* = 1,656 cases, 10,819 controls) at the same index SNPs. Highly heterogeneous loci are highlighted. Error bars correspond to 95% confidence intervals.

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 model 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 model in HA individuals (top), the EUR ancestry tract from the tractor analysis (middle) and the NAT ancestry tract (NAT) (bottom). (c) Forest plot comparing ORs and 95% CIs of the *CFH* Y402H risk allele in MVP EA (tranches 1 and 2; *n* = 20,064 cases, 162,287 controls), AA (*n* = 2,302 cases, 29,223 controls), and HA populations (*n* = 1,656 cases, 10,819 controls), with Tractor analyses for AA (AFR and EUR tracts), and HA (EUR, NAT, and AFR tracts). (d) Forest plot comparing ORs and 95% CIs of the *ARMS2* A69S risk allele in EA, AA, and HA populations (same sample sizes as above), 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 correspond to whether effect estimates are derived from a standard GWAS (black), or a Tractor GWAS, with blue for AFR tracts, red for EUR tracts, and orange for NAT tracts.

Fig. 7. Regional Miami plot of genetically regulated gene and isoform expression at the 1q32 (*CD46*/*CD55*) locus. GWAS results are above the axis and TWAS results from the gene and transcript isoform DFPLC brain model below the axis. GWAS points are colored according to linkage disequilibrium r^2 with the top SNP using a European-ancestry panel.

References

1. Fleckenstein, M. *et al.* Age-related macular degeneration. *Nat Rev Dis Primers* **7**, 31 (2021).
2. Vanderbeek, B. L. *et al.* 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. Mahr, M. A., Hodge, D. O. & Erie, J. C. 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. Wang, D., Jiang, Y., He, M., Scheetz, J. & Wang, W. 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. Haines, J. L. *et al.* Complement factor H variant increases the risk of age-related macular degeneration. *Science* **308**, 419–421 (2005).
6. Klein, R. J. *et al.* Complement factor H polymorphism in age-related macular degeneration. *Science* **308**, 385–389 (2005).
7. Edwards, A. O. *et al.* Complement factor H polymorphism and age-related macular degeneration. *Science* **308**, 421–424 (2005).
8. Fritsche, L. G. *et al.* A large genome-wide association study of age-related macular degeneration highlights contributions of rare and common variants. *Nat. Genet.* **48**, 134–143 (2016).
9. Gaziano, J. M. *et al.* Million Veteran Program: A mega-biobank to study genetic influences on health and disease. *J. Clin. Epidemiol.* **70**, 214–223 (2016).
10. Ge, T., Chen, C.-Y., Ni, Y., Feng, Y.-C. A. & Smoller, J. W. Polygenic prediction via Bayesian regression and continuous shrinkage priors. *Nat. Commun.* **10**, 1776 (2019).
11. Bulik-Sullivan, B. K. *et al.* LD Score regression distinguishes confounding from polygenicity

in genome-wide association studies. *Nat. Genet.* **47**, 291–295 (2015).

12. Pietzner, M. *et al.* Genetic architecture of host proteins involved in SARS-CoV-2 infection. *Nat. Commun.* **11**, 6397 (2020).
13. Kumar, P., Henikoff, S. & Ng, P. C. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat. Protoc.* **4**, 1073–1081 (2009).
14. Adzhubei, I. A. *et al.* A method and server for predicting damaging missense mutations. *Nat. Methods* **7**, 248–249 (2010).
15. Komatsu, M. & Ruoslahti, E. R-Ras is a global regulator of vascular regeneration that suppresses intimal hyperplasia and tumor angiogenesis. *Nat. Med.* **11**, 1346–1350 (2005).
16. Sawada, J., Li, F. & Komatsu, M. 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. Vähätupa, M. *et al.* Lack of R-Ras leads to increased vascular permeability in ischemic retinopathy. *Invest. Ophthalmol. Vis. Sci.* **57**, 4898–4909 (2016).
18. Surendran, P. *et al.* Trans-ancestry meta-analyses identify rare and common variants associated with blood pressure and hypertension. *Nat. Genet.* **48**, 1151–1161 (2016).
19. Campbell, R. A. *et al.* Placental HTRA1 cleaves α1-antitrypsin to generate a NET-inhibitory peptide. *Blood* **138**, 977–988 (2021).
20. Fawcett, K. A. *et al.* Pleiotropic associations of heterozygosity for the SERPINA1 Z allele in the UK Biobank. *ERJ Open Res* **7**, (2021).
21. Yin, X. *et al.* Genome-wide association studies of metabolites in Finnish men identify disease-relevant loci. *Nat. Commun.* **13**, 1644 (2022).
22. Simcoe, M. J. *et al.* 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. Simcoe, M. *et al.* Genome-wide association study in almost 195,000 individuals identifies

765 50 previously unidentified genetic loci for eye color. *Sci Adv* **7**, (2021).
766 24. Hysi, P. G. *et al.* Genome-wide association meta-analysis of individuals of European
767 ancestry identifies new loci explaining a substantial fraction of hair color variation and
768 heritability. *Nat. Genet.* **50**, 652–656 (2018).
769 25. Liyanage, U. E. *et al.* Combined analysis of keratinocyte cancers identifies novel genome-
770 wide loci. *Hum. Mol. Genet.* **28**, 3148–3160 (2019).
771 26. Visconti, A. *et al.* Genome-wide association study in 176,678 Europeans reveals genetic
772 loci for tanning response to sun exposure. *Nat. Commun.* **9**, 1684 (2018).
773 27. Burgess, S. & Thompson, S. G. Interpreting findings from Mendelian randomization using
774 the MR-Egger method. *Eur. J. Epidemiol.* **32**, 377–389 (2017).
775 28. Lundgren, S. *et al.* Tissue distribution of human gp330/megalin, a putative Ca(2+)-sensing
776 protein. *J. Histochem. Cytochem.* **45**, 383–392 (1997).
777 29. Storm, T., Heegaard, S., Christensen, E. I. & Nielsen, R. Megalin-deficiency causes high
778 myopia, retinal pigment epithelium-macromelanosomes and abnormal development of the
779 ciliary body in mice. *Cell Tissue Res.* **358**, 99–107 (2014).
780 30. Noel, N. C. L. & MacDonald, I. M. RP1L1 and inherited photoreceptor disease: A review.
781 *Surv. Ophthalmol.* **65**, 725–739 (2020).
782 31. Han, B. & Eskin, E. Random-effects model aimed at discovering associations in meta-
783 analysis of genome-wide association studies. *Am. J. Hum. Genet.* **88**, 586–598 (2011).
784 32. Benner, C. *et al.* FINEMAP: efficient variable selection using summary data from genome-
785 wide association studies. *Bioinformatics* **32**, 1493–1501 (2016).
786 33. Zeng, B. *et al.* Multi-ancestry eQTL meta-analysis of human brain identifies candidate
787 causal variants for brain-related traits. *Nat. Genet.* **54**, 161–169 (2022).
788 34. Hormozdiari, F. *et al.* Colocalization of GWAS and eQTL signals detects target genes. *Am.*
789 *J. Hum. Genet.* **99**, 1245–1260 (2016).
790 35. Maples, B. K., Gravel, S., Kenny, E. E. & Bustamante, C. D. RFMix: a discriminative

791 modeling approach for rapid and robust local-ancestry inference. *Am. J. Hum. Genet.* **93**,
792 278–288 (2013).
793 36. Atkinson, E. G. *et al.* Tractor uses local ancestry to enable the inclusion of admixed
794 individuals in GWAS and to boost power. *Nat. Genet.* (2021) doi:10.1038/s41588-020-
795 00766-y.
796 37. Hughes, A. E. *et al.* A common CFH haplotype, with deletion of CFHR1 and CFHR3, is
797 associated with lower risk of age-related macular degeneration. *Nat. Genet.* **38**, 1173–1177
798 (2006).
799 38. Sivakumaran, T. A. *et al.* A 32 kb critical region excluding Y402H in CFH mediates risk for
800 age-related macular degeneration. *PLoS One* **6**, e25598 (2011).
801 39. Holmes, L. V. *et al.* Determining the population frequency of the CFHR3/CFHR1 deletion at
802 1q32. *PLoS One* **8**, e60352 (2013).
803 40. Cipriani, V. *et al.* Increased circulating levels of Factor H-Related Protein 4 are strongly
804 associated with age-related macular degeneration. *Nat. Commun.* **11**, 778 (2020).
805 41. Pappas, C. M. *et al.* Protective chromosome 1q32 haplotypes mitigate risk for age-related
806 macular degeneration associated with the CFH-CFHR5 and ARMS2/HTRA1 loci. *Hum.*
807 *Genomics* **15**, 60 (2021).
808 42. Cipriani, V. *et al.* Beyond factor H: The impact of genetic-risk variants for age-related
809 macular degeneration on circulating factor-H-like 1 and factor-H-related protein
810 concentrations. *Am. J. Hum. Genet.* **108**, 1385–1400 (2021).
811 43. Zheng, X. *et al.* HIBAG--HLA genotype imputation with attribute bagging.
812 *Pharmacogenomics J.* **14**, 192–200 (2014).
813 44. Megiorni, F. & Pizzuti, A. HLA-DQA1 and HLA-DQB1 in Celiac disease predisposition:
814 practical implications of the HLA molecular typing. *J. Biomed. Sci.* **19**, 88 (2012).
815 45. Kutszegi, N. *et al.* HLA-DRB1*07:01-HLA-DQA1*02:01-HLA-DQB1*02:02 haplotype is
816 associated with a high risk of asparaginase hypersensitivity in acute lymphoblastic

leukemia. *Haematologica* **102**, 1578–1586 (2017).

46. Fernandez, C. A. *et al.* HLA-DRB1*07:01 is associated with a higher risk of asparaginase allergies. *Blood* **124**, 1266–1276 (2014).

47. Kutszegi, N. *et al.* 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).

48. Mullins, R. F., Russell, S. R., Anderson, D. H. & Hageman, G. S. 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).

49. Zhang, W. *et al.* Integrative transcriptome imputation reveals tissue-specific and shared biological mechanisms mediating susceptibility to complex traits. *Nat. Commun.* **10**, 3834 (2019).

50. Gandal, M. J. *et al.* Transcriptome-wide isoform-level dysregulation in ASD, schizophrenia, and bipolar disorder. *Science* **362**, (2018).

51. Ratnapriya, R. *et al.* Retinal transcriptome and eQTL analyses identify genes associated with age-related macular degeneration. *Nat. Genet.* **51**, 606–610 (2019).

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

53. Franzén, O. *et al.* Cardiometabolic risk loci share downstream cis- and trans-gene regulation across tissues and diseases. *Science* **353**, 827–830 (2016).

54. Liu, Y. *et al.* ACAT: A fast and powerful p value combination method for rare-variant analysis in sequencing studies. *Am. J. Hum. Genet.* **104**, 410–421 (2019).

55. Sjöstedt, E. *et al.* An atlas of the protein-coding genes in the human, pig, and mouse brain. *Science* **367**, (2020).

56. Strunz, T. *et al.* 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).

57. Hunter-Zinck, H. *et al.* Genotyping array design and data quality control in the Million Veteran Program. *Am. J. Hum. Genet.* **106**, 535–548 (2020).
58. Nédélec, Y. *et al.* Genetic ancestry and natural selection drive population differences in immune responses to pathogens. *Cell* **167**, 657–669.e21 (2016).
59. Randolph, H. E. *et al.* Genetic ancestry effects on the response to viral infection are pervasive but cell type specific. *Science* **374**, 1127–1133 (2021).
60. Klein, B. E. K. *et al.* Sunlight exposure, pigmentation, and incident age-related macular degeneration. *Invest. Ophthalmol. Vis. Sci.* **55**, 5855–5861 (2014).
61. Halladay, C. W. *et al.* 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).

Materials and Methods

Samples

The VA central Institutional Review Board approved the MVP005 and MVP024 study protocols. A general description of the MVP sample may be found in Gaziano et al.[9] Composition of the cohorts studied here is available in Supplementary **Table 1.** 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.[61] 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 no AMD diagnoses on at least two separate eye clinic visits. The three EA cohorts represent three separate releases of the MVP genotype data.

The MVP sample was genotyped for a custom Affymetrix Biobank chip, containing 725,000 variants[62]. 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[63] 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[62] 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 participating in clinical trials for Ranibizumab (NCT00891735, NCT00061594 and NCT00056836) and from patients with geographic atrophy (GA) participating in clinical trials for Lampalizumab (NCT02247479, NCT02247531, NCT01229215) and an observational study (NCT02479386). Patients had to consent for genetic analysis for inclusion eligibility and these cohorts were selected for inclusion based on availability of DNA and phenotypic information.

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 EA.

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

Quality Control. Samples were 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. We then conducted a GWAS in a third tranche of MVP data that later became available to us using REGENIE[64] version 2.2.4 with 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.

MVP AA and HA. Individuals were classified by genetic ancestry using the HARE method[62] which harmonizes self-reported ethnicity with genetic ancestry. We conducted GWAS in MVP AA and HA populations with REGENIE[64] version 1.6.7 using sex and 10 ancestry-specific PCs as covariates. Based on prior literature, we expect mixed-model based approaches such as REGENIE to provide better control of population stratification in diverse ancestries[65–67]. 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 are described in detail elsewhere[8]. Briefly, EA 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[68,69]. Cases and controls were selected from EA 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 EA 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 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 (FE) scheme as implemented in METAL[70]. For the multi-ancestry meta-analysis, we further performed a sensitivity analysis using the Han-Eskin random effects model (RE2) in METASOFT v2.0.1[31]. FE and RE2 p-values at top loci were highly consistent.

Additional Quality Control. 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[71] 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 EA 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 examine 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), using the UK Biobank AA 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[72]. 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[43] 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[73]. 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[47], both of which are genotyped on the MVP array. We phased genotypes using SHAPEIT4[74] 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.[75,76] were used as the NAT reference. We used RFMIX[35] version 2 to generate local ancestry calls for phased genotypes. We then extracted ancestry-specific dosages from the imputed data into PLINK 2.0-compatible files[77] using custom scripts based on the Tractor workflow[36]. 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[78], 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[74] 4.2 and the locus was re-imputed using Minimac4[79] 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[57], which is enriched in protein-altering rare variants, and applied a recent technical advance[80] 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”[81]. 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[64].

Relatedness. For experiments in unrelated samples, we used KING version 2.2.7[82] 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[51] in our European-ancestry meta-analysis using the Summary-based Mendelian Randomization (SMR) software tool[83].

***Two-sample MR*.** We performed MR on selected pigmentation traits and AMD using the TwoSampleMR R package[84]. MR analyses were performed using the following methods: inverse variance-weighted, MR-Egger, weighted median, simple mode, and weighted mode. Horizontal pleiotropy was assessed using the MR-Egger intercept[27]. 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. Eye color was also tested using a selected set of clumped SNPs with reported pleiotropic associations with other pigmentation traits[23].

***Genetic correlation*.** Genetic correlation analyses using the European-ancestry meta-analysis summary statistics were performed using LDSC[11] 1.01 using the provided European-ancestry LD scores derived from the 1000 Genomes project. We did not perform cross-ancestry genetic correlation analyses as summary statistics-based methods such as Popcorn[85] are not well-defined for admixed populations.

***Fine-mapping*.** We performed Bayesian fine-mapping of each genome-wide significant locus in the European-ancestry meta-analysis and multi-ancestry meta-analysis using FINEMAP[32] 1.4. 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[33], leveraging the multi-ancestry RNA-seq-based brain gene expression reference dataset derived from 2,119 donors[33] and the GWAS summary statistics from the AMD multi-ancestry meta-analysis.

PheWAS. We performed PheWAS and LabWAS in unrelated subjects as implemented in the R PheWAS package[86] version 0.12. 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, 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 were constructed as previously described[49,87] for tissues of the GTEx[52] v7, STARNET[53] and PsychENCODE[50,88] cohorts (Supplementary Note 3). The genetic datasets of the GTEx[52], STARNET[53] and PsychENCODE[88] cohorts were uniformly processed for quality control (QC) steps before genotype imputation[49,87]. We restricted our analysis to EA samples[49]. Genotypes were imputed using the University of Michigan server[79] with the Haplotype Reference Consortium (HRC) reference panel[63]. Gene expression information was derived from RNA-seq gene level

counts, which were adjusted for known and hidden confounders, followed by quantile normalization. For GTEx, we used 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[53]. 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[5087]. Finally, we constructed a retinal transcriptomic imputation model based on reference data from Ratnapriya et al.[51] (*n* = 406; Supplementary Note 4).For training, we used PrediXcan[93] 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[49].

Multi-tissue transcriptome-wide association study (TWAS). Briefly, we applied the S-PrediXcan method[94] to integrate the summary statistics and the transcriptomic imputation models constructed above to obtain gene-level association results[50]. Results were corrected for multiple testing with the Benjamini & Hochberg (FDR) method[95]. P-values across tissues were meta-analyzed using ACAT[54] ≤ 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 used gene sets from MsigDB[96] 5.1 and filtered out non-protein coding genes, genes located at MHC as well as genes whose expression could not be reliably imputed. In addition, we assayed enrichment for genesets deriving from the rare variant analysis as above. Statistical significance was evaluated with one-sided Fisher's exact test and the corrected p-values were obtained by the Benjamini-Hochberg (FDR) method[95]. Semantic enrichment analysis of the TWAS results was performed as follows: a) gene ontology (GO) data[97,98] were retrieved on 2021-07-10; b) we ran Fisher-based gene set enrichment analysis as above and retained enrichments with FDR-corrected p-value ≤ 0.1; c) we used GOGO[99] 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[100] using the ward.D2 method[101] according to the Silhouette criterion[102]; e) finally, we built a word cloud for each semantic cluster by mining text terms[103] from its members' GO term titles.

Data availability

The full summary level association data from the multi-ancestry meta-analysis, ancestry-stratified meta-analyses, and individual population association analyses in MVP are available via the dbGaP study accession number phs001672. Polygenic scores based on Euroean-ancestry and African-ancestry GWAS are deposited in the PGS Catalog (https://www.pgscatalog.org/). Data used for brain transcriptome model generation are available from PsychENCODE (http://resource.psychencode.org/); genotypes are controlled data and access instructions are provided at https://www.synapse.org/#!Synapse:syn4921369/wiki/477467. STARNET-based EpiXcan transcriptomic imputation models are available for download at https://labs.icahn.mssm.edu/roussos-lab/resources/. Data used for GTEx-based transcriptomic imputation models are available at https://www.gtexportal.org/home/datasets. MSigDB: http://software.broadinstitute.org/gsea/msigdb. For ancestry-specific linkage disequilibrium r2 with top SNPs in Regional TWAS Miami plots, 1000G Phase3 v5 was used available at http://ftp.1000genomes.ebi.ac.uk/vol1/ftp/release/20130502/. REMC: https://egg2.wustl.edu/roadmap/web_portal/. GO datasets were retrieved from the Gary Bader lab (http://download.baderlab.org/EM_Genesets/current_release/Human/entrezgene/GO/) version Human_GOALL_with_GO_iea_July_01_2021_entrezgene.gmt was used.

Code availability

Software and analytical methods used in data analyses include: PrediXcan (https://github.com/hakyimlab/PredictDB-Tutorial) to generate PrediXcan transcriptomic imputation models,EpiXcan (https://bitbucket.org/roussoslab/epixcan/src/master/) to generate EpiXcan transcriptomic imputation models, S-PrediXcan (https://github.com/hakyimlab/MetaXcan) to perform TWAS, PLINK2 (https://www.cog-genomics.org/plink/2.0/) for on-demand generation of ancestry-specific linkage disequilibrium r2 with top SNPs in Regional TWAS Miami plots,
GOGO (https://github.com/zwang-bioinformatics/GOGO) for semantic enrichment analysis, andR v.4.2.2 for statistical analyses and plotting (https://www.Rproject.org).

Methods-Only References

62. Fang, H. *et al.* Harmonizing genetic ancestry and self-identified race/ethnicity in genome-wide association studies. *Am. J. Hum. Genet.* **105**, 763–772 (2019).

63. McCarthy, S. *et al.* A reference panel of 64,976 haplotypes for genotype imputation. *Nat. Genet.* **48**, 1279–1283 (2016).

64. Mbatchou, J. *et al.* Computationally efficient whole-genome regression for quantitative and binary traits. *Nat. Genet.* **53**, 1097–1103 (2021).

65. Yang, J., Zaitlen, N. A., Goddard, M. E., Visscher, P. M. & Price, A. L. Advantages and pitfalls in the application of mixed-model association methods. *Nat. Genet.* **46**, 100–106 (2014).

66. Loh, P.-R. *et al.* Efficient Bayesian mixed-model analysis increases association power in large cohorts. *Nat. Genet.* **47**, 284–290 (2015).

67. Chen, H. *et al.* Control for population structure and relatedness for binary traits in genetic association studies via logistic mixed models. *Am. J. Hum. Genet.* **98**, 653–666 (2016).

68. Shen, L. *et al.* Differences in the genetic susceptibility to age-related macular degeneration clinical subtypes. *Invest. Ophthalmol. Vis. Sci.* **56**, 4290–4299 (2015).

69. Jorgenson, E. *et al.* Common coding variants in the HLA-DQB1 region confer susceptibility to age-related macular degeneration. *Eur. J. Hum. Genet.* **24**, 1049–1055 (2016).

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

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

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

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

74. Delaneau, O., Zagury, J.-F., Robinson, M. R., Marchini, J. L. & Dermitzakis, E. T.

Accurate, scalable and integrative haplotype estimation. *Nat. Commun.* **10**, 5436 (2019).

75. Mao, X. *et al.* A genomewide admixture mapping panel for Hispanic/Latino populations. *Am. J. Hum. Genet.* **80**, 1171–1178 (2007).

76. Martin, A. R. *et al.* Human demographic history impacts genetic risk prediction across diverse populations. *Am. J. Hum. Genet.* **100**, 635–649 (2017).

77. Chang, C. C. *et al.* Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* **4**, 7 (2015).

78. McInnes, L., Healy, J. & Melville, J. UMAP: Uniform manifold approximation and projection for dimension reduction. *arXiv [stat.ML]* (2018).

79. Das, S. *et al.* Next-generation genotype imputation service and methods. *Nat. Genet.* **48**, 1284–1287 (2016).

80. Man, O. M. *et al.* Novel genotyping algorithms for rare variants significantly improve the accuracy of Applied Biosystems™ Axiom™ array genotyping calls. *bioRxiv* 2021.09.13.459984 (2021) doi:10.1101/2021.09.13.459984.

81. Cirulli, E. T. *et al.* Genome-wide rare variant analysis for thousands of phenotypes in over 70,000 exomes from two cohorts. *Nat. Commun.* **11**, 542 (2020).

82. Manichaikul, A. *et al.* Robust relationship inference in genome-wide association studies. *Bioinformatics* **26**, 2867–2873 (2010).

83. Zhu, Z. *et al.* Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. *Nat. Genet.* **48**, 481–487 (2016).

84. Hemani, G. *et al.* The MR-Base platform supports systematic causal inference across the human phenome. *Elife* **7**, (2018).

85. Brown, B. C., Asian Genetic Epidemiology Network Type 2 Diabetes Consortium, Ye, C. J., Price, A. L. & Zaitlen, N. Transethnic genetic-correlation estimates from summary statistics. *Am. J. Hum. Genet.* **99**, 76–88 (2016).

86. Carroll, R. J., Bastarache, L. & Denny, J. C. R PheWAS: data analysis and plotting tools

for phenome-wide association studies in the R environment. *Bioinformatics* **30**, 2375–2376 (2014).

87. Fullard, J. F. *et al.* Single-nucleus transcriptome analysis of human brain immune response in patients with severe COVID-19. *Genome Med.* **13**, 118 (2021).

88. Wang, D. *et al.* Comprehensive functional genomic resource and integrative model for the human brain. *Science* **362**, (2018).

89. Purcell, S. *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).

90. 1000 Genomes Project Consortium *et al.* A global reference for human genetic variation. *Nature* **526**, 68–74 (2015).

91. Phan, L. *et al.* ALFA: allele frequency aggregator. *National Center for Biotechnology Information, US National Library of Medicine* (2020).

92. Ongen, H., Buil, A., Brown, A. A., Dermitzakis, E. T. & Delaneau, O. Fast and efficient QTL mapper for thousands of molecular phenotypes. *Bioinformatics* **32**, 1479–1485 (2016).

93. Gamazon, E. R. *et al.* A gene-based association method for mapping traits using reference transcriptome data. *Nat. Genet.* **47**, 1091–1098 (2015).

94. Barbeira, A. N. *et al.* Exploring the phenotypic consequences of tissue specific gene expression variation inferred from GWAS summary statistics. *Nat. Commun.* **9**, 1825 (2018).

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

96. Subramanian, A. *et al.* 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).

97. Ashburner, M. *et al.* Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat. Genet.* **25**, 25–29 (2000).

98. Gene Ontology Consortium. The Gene Ontology resource: enriching a GOld mine.

Nucleic Acids Res. **49**, D325–D334 (2021).

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

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

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

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

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