RESEARCH ARTICLE   **Open Access**

Genome-wide association meta-analysis for early age-related macular degeneration highlights novel loci and insights for advanced disease

Thomas W. Winkler[1*†] , Felix Grassmann[2,3,4†], Caroline Brandl[1,2,5], Christina Kiel[2], Felix Günther[1,6], Tobias Strunz[2], Lorraine Weidner[1], Martina E. Zimmermann[1], Christina A. Korb[7], Alicia Poplawski[8], Alexander K. Schuster[7], Martina Müller-Nurasyid[8,9,10,11], Annette Peters[12,13], Franziska G. Rauscher[14,15], Tobias Elze[14,16], Katrin Horn[14,15], Markus Scholz[14,15], Marisa Cañadas-Garre[17], Amy Jayne McKnight[17], Nicola Quinn[17], Ruth E. Hogg[17], Helmut Küchenhoff[6], Iris M. Heid[1†], Klaus J. Stark[1†] and Bernhard H. F. Weber[2,18†]

Abstract

Background: Advanced age-related macular degeneration (AMD) is a leading cause of blindness. While around half of the genetic contribution to advanced AMD has been uncovered, little is known about the genetic architecture of early AMD.

Methods: To identify genetic factors for early AMD, we conducted a genome-wide association study (GWAS) meta-analysis (14,034 cases, 91,214 controls, 11 sources of data including the International AMD Genomics Consortium, IAMDGC, and UK Biobank, UKBB). We ascertained early AMD via color fundus photographs by manual grading for 10 sources and via an automated machine learning approach for > 170,000 photographs from UKBB. We searched for early AMD loci via GWAS and via a candidate approach based on 14 previously suggested early AMD variants.

(Continued on next page)

* Correspondence: thomas.winkler@klinik.uni-regensburg.de
†Thomas W Winkler and Felix Grassmann contributed equally
†Iris M Heid, Klaus J Stark and Bernhard HF Weber jointly supervised this work
[1]Department of Genetic Epidemiology, University of Regensburg, Regensburg, Germany
Full list of author information is available at the end of the article

(Continued from previous page)

Results: Altogether, we identified 10 independent loci with statistical significance for early AMD: (i) 8 from our GWAS with genome-wide significance ($P < 5 \times 10^{-8}$), (ii) one previously suggested locus with experiment-wise significance ($P < 0.05/14$) in our non-overlapping data and with genome-wide significance when combining the reported and our non-overlapping data (together 17,539 cases, 105,395 controls), and (iii) one further previously suggested locus with experiment-wise significance in our non-overlapping data. Of these 10 identified loci, 8 were novel and 2 known for early AMD. Most of the 10 loci overlapped with known advanced AMD loci (near *ARMS2/HTRA1, CFH, C2, C3, CETP, TNFRSF10A, VEGFA, APOE*), except two that have not yet been identified with statistical significance for any AMD. Among the 17 genes within these two loci, in-silico functional annotation suggested *CD46* and *TYR* as the most likely responsible genes. Presence or absence of an early AMD effect distinguished the known pathways of advanced AMD genetics (complement/lipid pathways versus extracellular matrix metabolism).

Conclusions: Our GWAS on early AMD identified novel loci, highlighted shared and distinct genetics between early and advanced AMD and provides insights into AMD etiology. Our data provide a resource comparable in size to the existing IAMDGC data on advanced AMD genetics enabling a joint view. The biological relevance of this joint view is underscored by the ability of early AMD effects to differentiate the major pathways for advanced AMD.

Keywords: Genome-wide association study (GWAS), Meta-analysis, Age-related macular degeneration (AMD), Early AMD, *CD46*, *TYR*, International AMD genomics consortium (IAMDGC), UK biobank (UKBB), Machine-learning, Automated phenotyping

Background

Age-related macular degeneration (AMD) is the leading cause of irreversible central vision impairment in industrialized countries. Advanced AMD presents as geographic atrophy (GA) and/or neovascular (NV) complications [1]. Typically, advanced AMD is preceded by clinically asymptomatic and thus often unrecognized early disease stages. Early AMD is characterised by differently sized yellowish accumulations of extracellular material between Bruch's membrane and retinal pigment epithelium (RPE) or between RPE and the photoreceptors (drusen or subretinal drusenoid deposits, respectively). Other features of early AMD are RPE abnormalities, including depigmentation or increased amount of pigment [1].

Early and advanced AMD can be documented by color fundus imaging of the central retina and/or other multimodal imaging approaches including optical coherence tomography (OCT) [1–3]. While the definition of advanced AMD is reasonably homogeneous across clinical and epidemiological studies, the classification of early AMD is more variable and different studies traditionally apply differing classification systems [4, 5].

Epidemiological studies show that high age is the strongest risk factor for early and advanced AMD onset as well as progression [1, 6–8]. A robust genetic influence was shown for advanced AMD [1, 9–11] with 34 distinct loci at genome-wide significance in a large genome-wide association study (GWAS) for advanced AMD [9]. The genes underneath these advanced AMD loci were found to be enriched for genes in the alternative complement pathway, HDL transport, and extracellular matrix organization and assembly [9].

Exploring the genetics of early AMD offers the potential to understand the mechanisms of early disease processes, but also for the development to advanced AMD when comparing genetic effect sizes for early and advanced stages. Yet there have been few published GWAS searches for early AMD. One meta-analysis on 4089 early AMD patients and 20,453 control persons reported two loci with genome-wide significance, both being well known from advanced AMD, the *CFH* and the *ARMS2/HTRA1* locus [12].

We have thus set out to gather GWAS data for early AMD from 11 sources including own study data, data from the International AMD Genomics Consortium (IAMDGC), dbGaP and UK Biobank to conduct the largest GWAS meta-analysis on early AMD to date.

Methods

GWAS data from 11 sources

We included 11 sources of data with GWAS data and color fundus photography for early AMD phenotyping (Table S1). Our studies were primarily population-based cohort studies, where the baseline survey data were used for this analysis from studies of the authors (GHS, LIFE, NICOLA, KORA, AugUR) as well as for publicly available studies from dbGaP (ARIC, CHS, WHI; accession numbers: phs000090.v5.p1, phs000287.v6.p1, phs000746.v2.p3). We also included data from UK Biobank for participants from baseline and additional participants from the follow-up survey, since the color fundus photography program had started only after the main study onset (application number #33999). The studies captured an age range from 25 to 100 years of age (mean age from 47.5 years to 77.2 years across the 10 population-based studies, AugUR with the very old individuals range from 70.3 years to 95 years). About 50%

of the study participants in each study were male (except for the Women's Health Initiative, WHI), and all demonstrated European ancestry. All studies (except GHS-1, GHS-2 and Life-Adult) excluded any person with at least 2nd degree relationship (Table S1). For each of these cross-sectional data sets, participants with at least one eye gradable for AMD (see below) and with existing GWAS data were eligible for our analysis. We excluded participants with advanced AMD. We used participants with ascertained early AMD as cases and participants being ascertained for not having any signs of AMD as controls (*n* = 7363 cases, 73,358 controls across these population-based studies). Case-control data were also included from IAMD GC (http://amdgenetics.org/). The early AMD GWAS from IAMDGC is based on 24,527 individual participant data from 26 sources [9]. This data includes 17,856 participants with no AMD and 6671 participants with early AMD (excluding the 16,144 participants with advanced AMD). The cases and controls from IAMDGC were 16 to 102 years of age (mean age = 71.7 years). For all of these participants, DNA samples had been gathered and genotyped centrally (see below) [9].

Genotyping and imputation
All population-based studies were genotyped, quality controlled and imputed using similar chip platforms and imputation approaches (Table S2). As the imputation backbone, the 1000 Genomes Phase 1 or Phase 3 reference panel was applied [13], except GHS was imputed based on the Haplotype Reference Consortium (HRC) [14] and UK Biobank was imputed based on HRC and the UK10K haplotype resource [15]. Details on the UK Biobank genotypic resource are described elsewhere [16]. For the IAMDGC case-control data, DNA samples had been gathered across all participants and genotyped on an Illumina HumanCoreExome array and quality controlled centrally. Genotype quality control and imputation to the 1000 Genomes phase 1 version 3 reference panel (> 12 million variants) were conducted centrally. Details on the IAMDGC data were described in detail by Fritsche et al. [9].

Phenotyping
Across all studies included into this analysis, early AMD and the unaffected status was ascertained by color fundus photography. For participants from AugUR and LIFE, "early AMD" was classified according to the Three Continent Consortium (3CC) Severity Scale [4], which separates "mild early" from "moderate" and "severe early" AMD stages depending on drusen size, drusen area, or the presence of pigmentary abnormalities [4]. For the analysis, we collapsed any of these "early" AMD stages into the definition of "early AMD". However, the 3CC Severity Scale was not available for the other

studies. In these, similar early AMD classifications, considering drusen size or area and presence of pigmentary abnormalities, were used (Table S1): For participants from GHS, the Rotterdam Eye Study classification was applied [17]. For participants from NICOLA, the Beckman Clinical Classification was utilized [18]. Participants from the KORA study were classified as "early AMD" based on the AREDS-9 step classification scheme and we defined "early AMD" for this analysis by AREDS-9 steps 2–8 [19]. The ascertainment of IAMDGC study participants is described in detail elsewhere and covers various classification systems [9]. Of note, LIFE and NICOLA phenotyping incorporated OCT information additional to the information from color fundus imaging (Table S1). For UK Biobank participants, color fundus images were received (application number 33999); there was no existing AMD classification available in UK Biobank (see below). The AMD status of a person was derived based on the AMD status of the eye with the more severe AMD stage ("worse eye") when both eyes were gradable, and as the grade of the one available eye otherwise. Eyes were regarded as gradable, if at least one image of the eye fulfilled defined quality criteria allowing for the assessment of AMD (bright image, good color contrast, full macular region captured on images). Images were excluded from AMD grading if they revealed obscuring lesions (e.g. cataract) or lesions considered to be the result of a competing retinal disease (such as advanced diabetic retinopathy, high myopia, trauma, congenital diseases, or photocoagulation unrelated to choroidal neovascularization). Details for IAMDGC are described previously [9]. Persons with gradable images for at least one eye were included in this analysis. Persons with advanced AMD defined as presence of neovascularization or geographic atrophy in at least one eye were excluded for the main GWAS on early AMD.

Automated classification of early AMD in UK biobank
To obtain early AMD phenotype data for UK Biobank participants, we used a pre-trained algorithm for automated AMD classification based on an ensemble of convolutional neural networks [20]. In the UKBB baseline data, fundus images were available for 135,500 eyes of 68,400 individuals with at least one image. Among the additional 38,712 images of 19,501 individuals in the follow-up, there were 17,198 individuals without any image from baseline. For each image (eye) at baseline and follow-up, we predicted the AMD stage on the AREDS-9 step severity scale using the automated AMD classification. We defined a person-specific AMD stage at baseline and follow-up based on the worse eye. Eyes that were classified as ungradable were treated as missing data and, if diagnosis was available for only one eye, the person-specific AMD stage was based on the

classification of the single eye. If we obtained an automated disease classification to an AMD stage (i.e. not "ungradable" for both eyes) at baseline and follow-up, we used the follow-up disease stage (and follow-up age) in the association analysis. By this, we obtained an automated AMD classification for 70,349 individuals (2161 advanced AMD, 3835 early AMD, 64,353 unaffected). Individuals with advanced AMD were excluded from this analysis. Finally, we yielded 57,802 unrelated individuals of European ancestry with valid GWAS data that had either early AMD or were free of any AMD (3105 cases, 54,697 controls). We evaluated the performance of the automated disease classification by selecting 2013 individuals (4026 fundus images) for manual classification based on the 3CC Severity Scale. Details are described elsewhere [21]. We found reasonable agreement between the automated and the manual classification for the four categories of "no AMD", "early AMD", "advanced AMD" and "ungradable" (concordance = 79.5%, Cohen's kappa κ = 0.61, kappa with list-wise exclusion of ungradable individuals κ = 0.47 [22]). We found 305 of the 2013 individuals to be ungradable by the automated approach (i.e. 15.2%), including 257 individuals that have also been ungradable manually (i.e. truly bad image quality for both eyes) and 48 individuals that had been manually gradable, but not with the automated approach (i.e. 2.4%). Further details and further comparisons are provided in the Supplementary Note and Tables S3-S4.

Study-specific association analyses

Study-specific logistic regression analyses (early AMD cases versus controls, excluding advanced AMD cases) were applied by study partners (in Regensburg, Leipzig, Mainz, Belfast) using an additive genotype model and according to a pre-defined analysis plan. All publicly available data from dbGAP (studies ARIC, CHS and WHI) and UK Biobank as well as IAMDGC data was analyzed in Regensburg. All studies inferred the association of each genetic variant with early AMD using a Wald test statistic as implemented in RVTESTS [23]. Age and two principal components (to adjust for population stratification) were included as covariates in the regression models. We conducted sensitivity analyses to evaluate the impact of additionally including sex and 10 principal components in the regression models on the example of the two largest data sets, the IAMDGC and the UKBB data. The IAMDGC analyses were further adjusted for DNA source as done previously [9]. For the IAMDGC data that stemmed from 26 sources, we conducted a sensitivity analysis additionally adjusting for source membership according to previous work highlighting slight differences in effect estimates [24]; we found the same results.

Quality control of study-specific aggregated data

GWAS summary statistics for all data sources were processed through a standardized quality-control (QC) pipeline [25]. This involved QC checks on file completeness, range of test statistics, allele frequencies, population stratification as well as filtering on low quality data. We excluded variants with low minor allele count (MAC < 10, calculated as MAC = $2*N_{eff}*MAF$, with N_{eff} being the effective sample size, $N_{eff} = 4N_{Cases}*N_{Controls}/(N_{Cases} + N_{Controls})$ and MAF being the minor allele frequency), low imputation quality (rsq < 0.4) or large standard error of the estimated genetic effect (SE > 10). Genomic control (GC) correction was applied to each GWAS result to correct for population stratification within each study [26]. The estimation of the GC inflation factor was based on variants outside of the 34 known advanced AMD regions (excluding all variants within < 5 Mb base positions to any of the 34 known advanced AMD lead variants). The GC factors ranged from 1.00 to 1.04 (Table S2). We transferred all variant identifiers to unique variant names consisting of chromosomal, base position (hg19) and allele codes in (e.g. "3:12345:A_C", allele codes in ASCII ascending order).

GWAS meta-analysis across the 11 sources of data for early AMD genetics

For signal detection and effect quantification, study-specific genetic effects were combined using an inverse-variance weighted fixed effect meta-analysis method as implemented in METAL [27]. We performed additional QC on meta-analysis results: We only included variants for identification that were available (i) in at least two of the 11 data sources with a total effective sample size of more than 5000 individuals ($N_{eff} > 5000$) and (ii) for chromosome and position annotation in dbSNP (hg19). A conservative second GC correction (again focusing on variants outside the known advanced AMD regions) was applied to the meta-analysis result, in order to correct for potential population stratification across studies [26]. The GC lambda factor of the meta-analysis was 1.01.

Genome-wide search for early AMD variants, variant selection, and locus definition

In our first approach, we conducted a genome-wide search for variants associated with early AMD and judged at a genome-wide significance level ($P < 5.0 \times 10^{-8}$). Variants identified by this GWAS approach were deemed as established with genome-wide significance in our meta-analysis (*tier 1*). To evaluate the robustness of any novel genome-wide significant AMD locus, we performed leave-one-out (LOO) meta-analyses. We also evaluated heterogeneity between study-specific genetic effect estimates for selected variants using the I^2 measures [28–30].

We combined genome-wide significant variants ($P < 5.0 \times 10^{-8}$) into independent loci by using a locus definition similar to what was done previously [9]: the most significant variant was selected genome-wide, all variants were extracted that were correlated with this lead variant ($r^2 > 0.5$, using IAMDGC controls as reference) and a further 500 kb were added to both sides. All variants overlapping the so-defined locus were assigned to the respective locus. We repeated the procedure until no further genome-wide significant variants were detected. Genes overlapping the so-defined loci were used for biological follow-up analyses (gene region defined from start to end). To identify independent secondary signals at any novel AMD locus, approximate conditional analyses were conducted based on meta-analysis summary statistics using GCTA [31].

Candidate approach

Additionally to the genome-wide search in our meta-analysis of 14,034 cases and 91,214 controls, we adopted a candidate approach based on the 14 reported suggestive variants by Holliday et al. (*P*-values from 8.9×10^{-6} to 1.1×10^{-6} in their meta-analysis, 4089 cases and 20,453 controls) [12]. For this, we analyzed our data without the studies that overlapped with the previously reported data (i.e. removing ARIC, CHS; yielding 13,450 cases and 84,942 controls). We also combined the data from Holliday et al. with this non-overlapping part of our data where possible (i.e. the 14 reported variants) by an inverse-variance weighted meta-analysis (altogether 17,539 cases, 105,395 controls). We judged these 14 variants' association at experiment-wise significance ($P < 0.05/14$) in our non-overlapping data and at genome-wide significance ($P_{Combined} < 5.0 \times 10^{-8}$) in the combined analysis. We considered a candidate-based selected variant as established with both experiment-wise significance ($P < 0.05/14$ in our non-overlapping data) and with genome-wide significance ($P_{Combined} < 5.0 \times 10^{-8}$ in the combined data) (*tier 2*), or as established with experiment-wise significance ($P < 0.05/14$ in our non-overlapping data), but without established genome-wide significance ($P_{Combined} \geq 5.0 \times 10^{-8}$) (*tier 3*).

Gene prioritization at newly identified AMD loci

To prioritize genes and variants at the newly identified AMD loci, we conducted a range of statistical and functional follow-up analyses. The following criteria were used: (1) Statistical evidence; we computed the posterior probability of each variant using Z-scores and derived 95% credible intervals for each locus [32]. The method assumes a single causal signal per locus. (2) Variant effect predictor (VEP) to explore whether any of the credible variants was located in a relevant regulatory gene region [33]. (3) eQTL analysis: We downloaded expression summary statistics for the candidate genes in retina from the EyeGEx database [34] and for 44 other tissues from the GTEx database [35] (both available at www.gtexportal.org/home/datasets) and evaluated whether any of the credible variants showed significant effects on expression levels in the aggregated data. For each significant eQTL in EyeGEx, we conducted colocalization analyses using eCAVIAR [36] to evaluate whether the observed early AMD association signal colocalized with the variants' association with gene expression. (4) Retinal expression: We queried the EyeIntegration database to evaluate genes in the relevant loci for expression in fetal or adult retina or RPE cells [37]. (5) Animal model: We queried the Mouse Genome Informatics (MGI) database (www.informatics.jax.org) for each gene in the relevant loci for relevant eye phenotypes in mice [38]. (6) Human phenotype: The Online Mendelian Inheritance in Man (OMIM)® database was queried for human eye phenotypes (McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University, Baltimore, MD, queried 07/11/2019, www.omim.org).

Phenome-wide association study for newly identified AMD loci

We used 82 other traits and queried reported genome-wide significant ($P < 5.0 \times 10^{-8}$) lead variants and proxies ($r^2 > 0.5$) for any of these traits for overlap with genes underneath our novel loci as done previously [39]. For this, we used GWAS summary results that were previously aggregated from GWAS catalogue [40], GWAS central [41] and literature search.

For the novel early AMD lead variants, we further evaluated their association with 118 non-binary and 660 binary traits from the UK Biobank [42]. The Phenome-wide association study (PheWAS) web browser "GeneATLAS" (www.geneatlas.roslin.ed.ac.uk) was used for the UK Biobank lookup. For each variant, association *P* values were corrected for the testing of multiple traits by the Benjamini-Hochberg false-discovery-rate (FDR) method [43].

Interaction analyses

For the novel early AMD effects and for the 34 known advanced AMD lead variants [9], we investigated whether age modulated early AMD effects by analyzing variant x age interaction in seven data sources for which we had individual participant data available in Regensburg (ARIC, CHS, WHI, IAMDGC, UKBB, AugUR and KORA). For each source, we applied logistic regression and included a variant x AGE interaction term in the model (in addition to the covariates used in the main analysis). We conducted meta-analysis across the seven sources to obtain pooled variant x age interaction effects and applied a Wald test to test for significant interaction (at a Bonferroni-corrected alpha-level). For the novel early AMD effects, we further investigated whether age modulated advanced AMD effects by evaluating

publically available data from IAMDGC [44]. Finally, we investigated whether a novel early AMD lead variant modulated any of the effects of the 34 known AMD variants on advanced AMD [9]. We used the IAMDGC data and applied one logistic regression model for each pair of known advanced AMD variants and novel early AMD variants including the two respective variants and their interaction (and the same other covariates as before).

Comparison of genetic effects on early and advanced stage AMD

We estimated the genetic correlation between early and advanced AMD by utilizing the LDSC tool [45] with the GWAS summary statistics for early and advanced AMD (from the current meta-analysis and the IAMDGC [9], respectively). We used pre-calculated LD scores for European ancestry (https://data.broadinstitute.org/alkesgroup/LDSCORE/eur_w_ld_chr.tar.bz2). We further compared genetic effect sizes between early and advanced AMD for the novel early AMD lead variants and for the 34 known advanced AMD lead variants [9]. For this, we queried the novel early AMD lead variants in the IAMDGC GWAS for advanced AMD [9] and (vice-versa) queried the 34 known advanced AMD lead variants [9] in the early AMD meta-analysis results. We compared effect sizes in a scatter plot and clustered the lead variants by their nominal significant association on advanced and/or early AMD. We classify different types of loci in a similar fashion as done previously for adiposity trait genetics [46]: (1) "advanced-and-early" AMD loci ($P_{early} < 0.05$, $P_{adv} < 0.05$), (2) "advanced-only" AMD loci ($P_{early} \geq 0.05$, $P_{adv} < 0.05$), (3) "early-only" AMD loci ($P_{early} < 0.05$, $P_{adv} \geq 0.05$).

Pathway analysis

To evaluate whether "advanced-and-early" AMD loci versus "advanced-only" AMD loci distinguished the major known pathways for advanced AMD, we performed pathway enrichment analysis separately for these two classes. We used the genes in the gene prioritization for all advanced AMD loci as previously described [9], derived the gene prioritization score and selected the best scored gene in each locus (two genes in the case of ties). We then separated the gene list according to the class of the respective locus, and performed pathway enrichment analysis via Enrichr [47] with default settings searching Reactome's cell signaling pathway database 2016 ($n = 1530$ pathways). P-values were corrected for multiple testing with Benjamini-Hochberg procedure [43].

Results

Eight genome-wide significant loci from a GWAS on early AMD

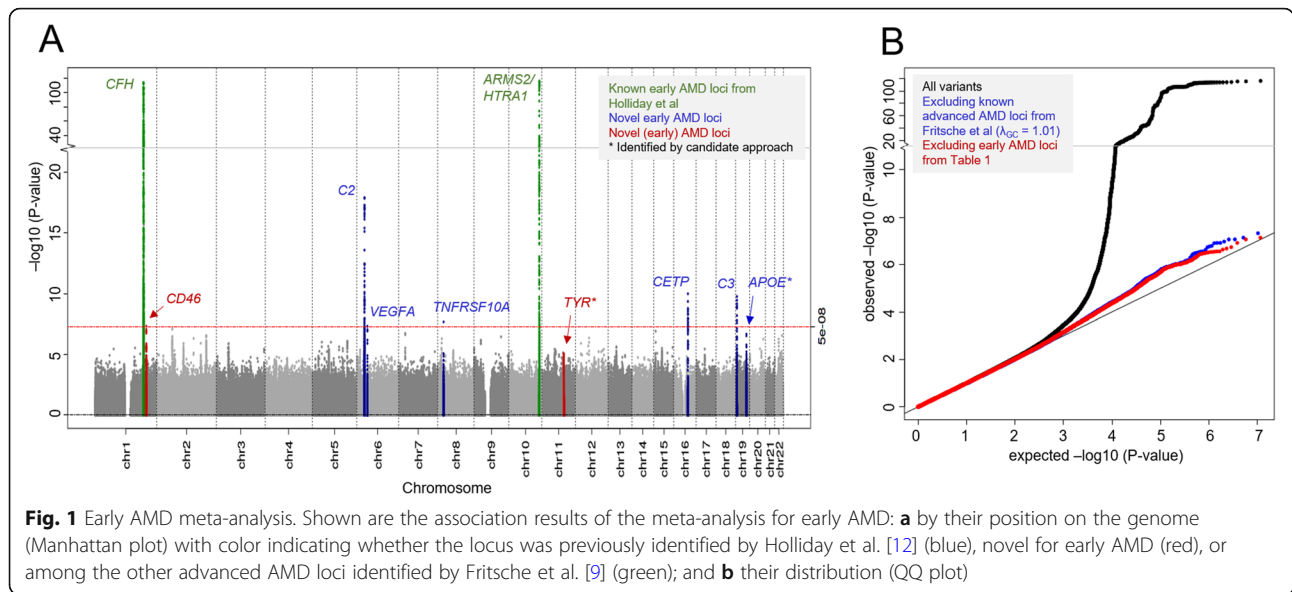We conducted a meta-analysis of genotyped and imputed data from 11 sources (14,034 early AMD cases, 91,214 controls; 11,702,853 variants; for study-specific genotyping, analysis and QC, see Tables S1-S2). For all participants, early AMD or control status (i.e. no early nor late AMD) was ascertained via color fundus photographs (Table S1). This included automated machine-learning based AMD classification of UK Biobank fundus images (application number 33999; 56,699 individuals from baseline, 13,650 additional individuals from follow-up) [16, 20]. Based on logistic regression association analysis in each of the 11 data sets meta-analyzed via fixed effect model, we identified eight distinct loci with genome-wide significance (*tier1*; $P = 1.3 \times 10^{-116}$ to 4.7×10^{-8}, Fig. 1, Table 1; "locus" defined by the lead variant and proxies, $r^2 \geq 0.5$, +/− 500 kb). Six of these loci were novel for early AMD; two loci had been identified for early AMD previously [12].

Most of the eight loci overlap with known loci for advanced AMD [9] (*CFH, ARMS2/HTRA1, C2, C3, CETP, VEGFA, TNFRSF10A*), except one which has not been identified in early or advanced AMD GWAS before ($P = 4.7 \times 10^{-8}$, lead variant rs4844620, near *CD46*, Figure S1). This novel locus around rs4844620 (near *CD46*) is independent of known (advanced) AMD loci: rs4844620 is (i) ∼ 500 K base positions distant from the recently reported AMD locus near *C4BPA/CD55* [11] and uncorrelated to the reported *C4BPA/CD55* lead variant rs11120691 ($r^2 < 0.01$) and (ii) > 10 million base positions distant from the *CFH* locus and uncorrelated to any of the eight reported *CFH* locus variants, $r^2 < 0.01$).

Taken together, we identified eight loci for early AMD, two known for early AMD and six novel, including one novel locus (near *CD46*) for any AMD with genome-wide significance.

Two further significant loci from a candidate-based approach of 14 variants

Subsequently, we applied a candidate-based approach by investigating the 14 variants reported as suggestive by the previous GWAS for early AMD (4089 early AMD cases, 20,453 controls; reported P between 1.1×10^{-6} and 8.9×10^{-6}) [12]. For this, we re-analyzed our data excluding the overlap with the previous GWAS (i.e. excluding ARIC, CHS study, yielding 13,450 early AMD cases and 84,942 controls) and also combined the reported and our non-overlapping data in an inverse-variance weighted meta-analysis where possible (i.e. for the 14 reported variants; altogether 17,539 cases, 105,395 controls). Among the 14 variants, we found four significant variants for early AMD (Table 2): (i) two "*tier 2*" variants (near *PVRL2* and *CD46*) with experiment-wise significance ($P < 0.05/14$) in our non-overlapping data and with genome-wide significance ($P_{Combined} < 5.0 \times 10^{-8}$) in the combined analysis, and (ii) two additional "*tier 3*" variants (near *APOE/TOMM40* and *TYR*) with

Fig. 1 Early AMD meta-analysis. Shown are the association results of the meta-analysis for early AMD: **a** by their position on the genome (Manhattan plot) with color indicating whether the locus was previously identified by Holliday et al. [12] (blue), novel for early AMD (red), or among the other advanced AMD loci identified by Fritsche et al. [9] (green); and **b** their distribution (QQ plot)

experiment-wise significance (P < 0.05/14) in our non-overlapping data.

We compared the evidence from the Bonferroni-corrected analysis judged at experiment-wise significance and the combined analysis judged at genome-wide significance. Both "tier 3" variants showed combined P-values close to the genome-wide significant threshold ($P_{Combined} = 9.0 \times 10^{-8}$ and 1.7×10^{-7}, for the *APOE/TOMM40* and the *TYR* variant, respectively), while the other 10 of the 14 variants showed combined P-values far away from significance ($P_{combined}$ from 4.8×10^{-3} to 0.74). Therefore, the Bonferroni-corrected analysis and the combined analysis yielded similar evidence and separated the 14 suggested variants into four variants with a positive finding (near *CD46, PVRL2, APOE/TOMM40*

and *TYR*) from the 10 variants with no finding (P > 0.05 in our non-overlapping data, Table 2).

Two of the four identified variants were correlated: the tier3-identified variant rs2075650 near *APOE/TOMM40* was located in the same locus as the tier2-identified variant rs6857 near *PVRL2* (r^2 = 0.75 to rs2075650) and thus counted as one locus (near *APOE/TOMM40/PVRL2*). The tier2-identified variant rs1967689 near *CD46* was located in one of the eight loci identified by our GWAS (r^2 = 0.77 to our GWAS lead variant rs4844620). The two other loci (near *APOE/TOMM40/PVRL2* and *TYR*) were identified in addition to our GWAS. These two loci were identified here for the first time with statistical significance as loci for early AMD: one locus known for advanced AMD (near *APOE*) and one identified here for

Table 1 Genome-wide search for early AMD association

Rs identifier	chr:pos [hg19]	EA	OA	EAF	logOR	SE	OR	P	N cases	N controls	Known advanced AMD locus (Fritsche et al.)	Locus name
Novel early AMD loci:												
rs4844620	1:207980901	g	a	0.79	0.095	0.017	1.10	4.7E-08	14,031	91,179	no	*CD46*
rs547154	6:31910938	g	t	0.91	0.218	0.025	1.24	1.3E-18	14,027	91,137	yes	*C2*
rs943080	6:43826627	t	c	0.51	0.080	0.015	1.08	4.7E-08	13,220	85,747	yes	*VEGFA*
rs13278062	8:23082971	t	g	0.52	0.080	0.014	1.08	2.0E-08	13,644	85,908	yes	*TNFRSF10A*
rs5817082	16:56997349	c	ca	0.26	0.108	0.017	1.11	1.0E-10	12,599	81,863	yes	*CETP*
rs11569415	19:6716279	a	g	0.21	0.116	0.018	1.12	1.7E-10	13,115	83,117	yes	*C3*
Known early AMD loci:												
rs4658046	1:196670757	c	t	0.39	0.321	0.014	1.38	2.9E-114	14,034	91,201	yes	*CFH*
rs3750847	10:124215421	t	c	0.22	0.384	0.017	1.47	1.3E-116	14,025	91,171	yes	*ARMS2/HTRA1*

EA effect allele, *OA* other allele, *EAF* effect allele frequency, *logOR* log odds ratio, *SE* standard error of logOR; *OR* odds ratio, *P* double GC corrected early association *P* value from the meta-analysis

The table shows the eight genome-wide significant ($P < 5.0 \times 10^{-8}$) lead variants from the early AMD meta-analysis. The second last column indicates whether the locus was identified by Fritsche et al. for advanced AMD [9]

Table 2 Candidate approach to search for early AMD association

Rs identifier	chr:pos [hg19]	Locus (Holliday et al.)	EA	OA	Holliday et al.			this meta-analysis (excluding ARIC, CHS)				P (combined)	Known advanced AMD locus (Fritsche et al.)
					EAF	OR [CI 95%]	P	EAF	OR [CI 95%]	P	Neff		
Experiment-wise significant association (P < 0.05/14)													
rs1967689	1:208039471	CD34,CD46	C	G	0.25	0.85 [0.80;0.91]	5.1E-06	0.24	0.93 [0.90;0.96]	2.5E-05	42,119	5.5E-09	no
rs621313	11:88913663	TYR	A	G	0.51	0.87 [0.83;0.92]	3.5E-06	0.52	0.95 [0.93;0.98]	6.8E-04	41,661	1.7E-07	no
rs6857	19:45392254	PVRL2	T	C	0.15	0.81 [0.74;0.88]	1.4E-06	0.16	0.92 [0.89;0.96]	7.1E-05	38,938	9.8E-09	yes
rs2075650	19:45395619	APOE/TOMM40	A	G	0.86	1.23 [1.13;1.34]	1.1E-06	0.86	1.08 [1.04;1.13]	2.6E-04	40,405	9.0E-08	yes
No experiment-wise significant association (P ≥ 0.05/14):													
rs16851585	1:177568799	–	C	G	0.92	0.77 [0.69;0.86]	5.0E-06	0.89	1.04 [0.99;1.09]	0.099	42,119	0.74	no
rs6721654	2:121301911	GLI2,INHBB	T	C	0.08	1.26 [1.14;1.4]	6.5E-06	0.08	1.01 [0.96;1.07]	0.65	41,718	0.017	no
rs17586843	4:116924184	–	T	C	0.78	1.18 [1.1;1.27]	1.5E-06	0.78	1.02 [0.98;1.05]	0.31	42,119	0.0048	no
rs7750345	6:106260128	–	A	G	0.75	1.16 [1.09;1.24]	6.8E-06	0.74	1.02 [0.98;1.05]	0.32	42,119	0.0049	no
rs2049622	7:42176282	GLI3	A	G	0.49	0.87 [0.83;0.93]	8.9E-06	0.52	0.99 [0.96;1.02]	0.41	42,060	0.015	no
rs11986011	8:127332657	FAM84B	T	C	0.02	2.5 [1.68;3.71]	5.0E-06	–	–	–	–		no
rs6480975	10:54574996	MBL2	C	G	0.84	1.21 [1.12;1.32]	2.8E-06	0.85	0.99 [0.95;1.03]	0.63	40,651	0.17	no
rs4293143	11:82821382	PCF11,RAB30	T	G	0.69	0.85 [0.79;0.91]	7.8E-06	0.70	0.99 [0.96;1.02]	0.50	42,119	0.0094	no
rs9646096	13:38065446	POSTN,TRPC4	A	C	0.95	0.74 [0.65;0.84]	6.0E-06	0.96	0.99 [0.92;1.06]	0.76	39,782	0.0087	no
rs10406174	19:394240	ITGB1BP3,DAPK3	A	G	0.11	1.24 [1.13;1.36]	5.6E-06	0.11	1.00 [0.93;1.07]	0.92	17,936	0.0057	yes

EA effect allele, *OA* other allele, *EAF* effect allele frequency, *OR* odds ratio, *CI* confidence interval, *P* P value from Holliday et al. or Double GC corrected early association *P* value from this meta-analysis, *Neff* effective sample size

The table shows results for the 14 lead variants reported as suggestive for early AMD by Holliday et al. [12] (effective sample size = 13,631) for their association with early AMD in our non-overlapping data set (P < 0.05/14 = 0.0036, tested at Bonferroni-corrected significance level, effective sample size up to 42,119) and in the combined analysis of the reported and our non-overlapping data ($P_{combined}$ < 5.0 × 10^{-8}, effective sample size up to 55,750). The ARIC and CHS studies were excluded from our meta-analysis data to avoid overlap with the data by Holliday et al. [12]. The last column indicates whether the locus was identified by Fritsche et al. for advanced AMD [9]

the first time for any AMD with statistical significance (rs621313, near *TYR*, $P = 6.8 \times 10^{-4}$, Figure S2).

Together with our GWAS approach, we identified 10 independent loci with statistical significance for early AMD: (i) eight from our GWAS ($P < 5.0 \times 10^{-8}$ in our meta-analysis, *tier 1*), (ii) one additional from the candidate approach with experiment-wise and with genome-wide significance (*tier 2*), and (iii) one additional from the candidate approach with experiment-wise significance (*tier 3*). Among the 10 identified loci (any *tier*), two were reported previously for early AMD (near *CFH*, *ARMS2/HTRA1*) and eight were identified here for the first time for early AMD with statistical significance (near *CD46*, *C2*, *C3*, *CETP*, *TNFRSF10A*, *VEGFA*, *APOE/TOMM40/PVRL2* and *TYR*). The eight loci included two that have not previously been identified with statistical significance for any AMD (near *CD46* and *TYR*). These two loci showed no second signals (GCTA [31], $P_{Cond} > 5.0 \times 10^{-8}$ for *CD46* and $P_{Cond} > 0.05/14$ for *TYR*, Figure S1-2).

Sensitivity analysis on the 10 identified early AMD loci

We conducted various sensitivity analyses to evaluate the robustness of the associations for the 10 identified lead variants: (a) leaving out one of the 11 data sets at a time showed similar effect estimates in most cases, except for the exclusion of the two largest contributing sources, IAMDGC and UKBB, which had the strongest impact on effect sizes due to their relatively large sample size (Figure S3). Exclusion of IAMDGC slightly increased the *CD46* effect size (due to a slightly smaller effect in IAMDGC, Figure S3A), had no impact on the *TYR* effect size (Figure S3B), and decreased effect sizes for *C2*, *CETP*, *C3*, *APOE*, *CFH* and *ARMS2/HTRA1* variants (due to a relatively large association in IAMDGC). Exclusion of UKBB had no impact on the *CD46* variant effect (Figure S3A), slightly diminished the *TYR* effect (due to a relatively strong effect in UKBB, Figure S3B) and increased effect sizes near *C2*, *C3* or *CFH*. (b) Between-study heterogeneity in our meta-analysis was similar to the heterogeneity in any of the leave-one-out meta-analyses, which indicated that the observed heterogeneity was not driven by one single data set (Table S5). We observed low to moderate between-study heterogeneity [28] for the two novel any AMD loci (I^2 between 22 to 48.3% and 45.4 to 63.8% for the *TYR* and *CD46* lead variant, respectively). (c) Effect sizes were robust to additional adjustment for sex or inclusion of additional genetic principal components in the regression analyses (Figure S4). In the following, we were particularly interested in fine-mapping the two loci that have not been identified before for any AMD: the loci near *CD46* and *TYR*.

Gene prioritization at the two novel loci

To prioritize variants and genes at the *CD46* and *TYR* locus, we conducted in silico follow-up analyses for all

variants and overlapping genes for each of these two loci (4451 or 5729 variants, 10 or 7 genes, respectively). We found several interesting aspects (Table 3): (1) When prioritizing variants according to their statistical evidence for being the driver variant by computing 95% credible sets of variants [32], we found 23 and 294 credible set variants for the *CD46* and *TYR* locus, respectively (Table S6). (2) Using the Variant Effect Predictor [33], we assessed overlap of credible set variants with functional regulatory regions and found variants influencing the transcript and/or the protein for four genes (Table S7): variants causing an alternative splice form for *CD46*, a nonsense-mediated mRNA decay (NMD) for *CR1L*, a missense variant for *TYR* (rs1042602, r^2 = 0.56 to the lead variant rs621313), and NMD variants for *NOX4*. (3) We investigated credible set variants for being an expression quantitative trait locus (eQTL) for any of the 17 genes in retina (Eye Genotype Expression database, EyeGEx [34]) or in 44 other tissues (Genotype-Tissue Expression database, GTEx [35]). For the *CD46* locus, we observed significant association of the lead variant and additional 16 credible set variants on *CD46* expression in retina (FDR < 5%, Table S8); the early AMD risk increasing alleles of all 17 variants were associated with elevated *CD46* expression. Importantly, we observed the expression signal to colocalize with the early AMD association signal using eCAVIAR [36] (3 variants with colocalization posterior probability CLPP> 0.01, Table S9, Figure S5-S6). We also found credible variants to be associated with *CD46* expression in 15 other tissues from GTEx, including four brain tissues (FDR < 0.05, Table S10). Among the credible set variants in the two loci, we found no further eQTL for any of the other genes. When extending beyond the credible set, we found one further *CD46* locus variant as eQTL for *CD55*, but without colocalization (Table S9, Figure S5-S6). These findings support the idea that the credible set captures the essential signal. (4) We queried the 17 genes overlapping the two loci for expression in eye tissue and cells in EyeIntegration summary data [37]. We found five and three genes, respectively, expressed in adult retina and adult RPE cells (*CD46*, *PLXNA2*, *CR1*, *CD34*, *CD55*; *TYR*, *GRM5*, *NOX4*; Figure S7-S8). (5) When querying the 17 genes in the Mouse Genome Informatics, MGI [38] or Online Mendelian Inheritance in Man, OMIM®, database, for eye phenotypes in mice or humans, we identified relevant eye phenotypes for five genes in mice (*CD46*, *CR1*, *CR1L*, *PLXNA2*; *TYR*; Table S11) and for one gene in human (*TYR*; Table S12).

While it is debatable how to prioritize evidence for a gene's probability to be causal, one approach is to count any of the following characteristics for each of the 17 genes (Gene Prioritization Score, GPS Table 3): any credible set variant is (i) protein-coding, (ii) involved in

Table 3 Summary of in silico follow-up and gene prioritization score (GPS)

Locus	Candidate gene	Chr	Pos-Start	Pos-End	Number of variants in 95% credible set	GPS	Protein Altering	NMD	Altered splicing	eQTL[a]	Expressed in Eye tissue §	MGI Mouse eye phenotype	OMIM Human eye phenotype
							Annotation for variants in 95% credible set				**Biology of the gene**		
CD46	CD46	1	207,925,382	207,968,861	11	4	0	0	1	1	1	1	0
CD46	CR1L	1	207,818,457	207,897,036	1	1	0	0	0	0	0	1	0
CD46	PLXNA2	1	208,195,587	208,417,665	0	2	0	0	0	0	1	1	0
CD46	CR1	1	207,669,472	207,815,110	0	2	0	0	0	0	1	1	0
CD46	LOC148696	1	207,991,723	207,995,941	1	0	0	0	0	NA	NA	0	0
CD46	CD34	1	208,059,882	208,084,683	0	1	0	0	0	0	1	0	0
CD46	CD55	1	207,494,816	207,534,311	0	1	0	0	0	0	1	0	0
CD46	CR2	1	207,627,644	207,663,240	0	0	0	0	0	0	0	0	0
CD46	MIR29B2	1	207,975,787	207,975,868	0	0	0	0	0	NA	NA	0	0
CD46	MIR29C	1	207,975,196	207,975,284	0	0	0	0	0	NA	NA	0	0
TYR	TYR	11	88,911,039	89,028,927	39	4	1	0	0	0	1	1	1
TYR	NOX4	11	89,057,521	89,322,779	1	3	0	1	1	0	1	0	0
TYR	GRM5	11	88,237,743	88,796,846	109	1	0	0	0	0	1	0	0
TYR	FOLH1B	11	89,392,464	89,431,886	0	0	0	0	0	0	0	0	0
TYR	GRM5-AS1	11	88,237,743	88,257,222	0	0	0	0	0	0	NA	0	0
TYR	TRIM49	11	89,530,822	89,541,743	0	0	0	0	0	0	NA	0	0
TYR	TRIM77	11	89,443,466	89,451,040	0	0	0	0	0	0	NA	0	0

[a] Variants in 95% credible set are a local expression quantitative trait locus for this gene in retina (EyeGEx) or any tissue included in the GTEx database (cis for genes in locus); § Expression in Eye Integration data; *NMD* nonsense-mediated mRNA decay; *OMIM* Online Mendelian Inheritance in Man (https://www.omim.org/); *NA* data not available; The gene start and end positions were extracted from the hg19 gene range list from http://www.cog-genomics.org/plink/1.9/resources
The table summarizes statistical and functional evidence for 10 and seven candidate genes of the novel early AMD loci on chromosome 1 and chromosome 11, respectively. Detailed results on the individual statistical and functional analyses are shown in Tables S6-S12. For the GPS, the sum of cell entries for "annotation" and "biology" was computed per row

NMD, (iii) affecting splice function, (iv) an eQTL for this gene in retina (EyeGEx) or in any other tissue (GTEx), or/and the gene (v) is expressed in retina or RPE, (vi) linked to eye phenotype in mouse or (vii) human. This approach offered *CD46* and *TYR* as the highest scored gene in the respective locus (GPS = 4 for each; Table 3).

Phenome-wide association search for the two novel loci
Co-association of variants in the two novel loci for early AMD with other traits and diseases may provide insights into shared disease mechanisms. We queried different data sets on numerous phenotypes by a gene-based and by a locus-based view.

For the gene-based view, we focused on 82 traits and evaluated reported genome-wide significant ($P < 5.0 \times 10^{-8}$) lead variants (and proxies, $r^2 > 0.5$) for overlap with any of the 17 gene regions (Table S13). For the *CD46* locus, we found significant association corrected for multiple testing (false-discovery rate, FDR < 5%) for schizophrenia (in *CD46* and *CR1L*) and for Alzheimer's disease (in *CR1*, Table S14). For the *TYR* locus, we found significant associations for eye color, skin pigmentation and skin cancer (in *GRM5* and *TYR*, Tables S14).

For the locus-based view, we conducted a phenome-wide association study (PheWAS): we evaluated whether the two lead variants were associated with any of the 778 traits in UK Biobank using GeneAtlas ($n = 452,264$, age-adjusted estimates; Table S15) [42]. For the *CD46* lead variant, we identified 27 significant trait associations (FDR < 5%), including four with particularly strong evidence ($P < 5.0 \times 10^{-8}$; white blood cell, neutrophil, monocyte count and plateletcrit); the early AMD risk increasing allele (G, frequency = 79%) was associated consistently with increased blood cell counts. We did not find a significant association of the *CD46* lead variant with schizophrenia in UK Biobank (FDR > 5%; Alzheimer's disease not available). For the *TYR* lead variant (rs621313, G allele associated with increased early AMD risk, frequency = 48%), we identified 20 significant trait associations including Melanoma (FDR < 5%, G allele associated with increased Melanoma risk) and two with particularly strong evidence for skin color and ease of

skin tanning ($P < 5.0 \times 10^{-8}$, G allele associated with brighter skin color and increased ease of skin tanning).

Advanced AMD association and interaction analyses for the two novel loci

Next, we investigated whether the early AMD loci *CD46* or *TYR* were associated with advanced AMD. We thus queried the two lead variants for early AMD (rs4844620 and rs621313, respectively) for their advanced AMD association in the IAMDGC data (Table S16). We observed nominally significant directionally consistent effects for advanced AMD ($OR_{adv} = 1.05$, 95% confidence interval, CI = [1.01,1.09] and 1.03 [1.00,1.07], $P_{adv} = 0.02$ and 0.05, respectively) that were slightly smaller compared to early AMD effects ($OR_{early} = 1.10$ [1.06,1.14] and 1.05 [1.02, 1.08], $P_{early} = 4.7 \times 10^{-8}$ and 6.8×10^{-4}).

When exploring variant x age interaction for early AMD (in a subset of our meta-analysis of 10,890 early AMD cases and 54,697 controls) or for advanced AMD (IAMDGC data [44]) for the two novel locus lead variants, we found no statistically significant interaction at a Bonferroni-corrected level for early or advanced AMD ($P_{GxAGE} > 0.05/2 = 0.025$, Table S17-S18).

We were interested in whether one of the two novel lead variants showed interaction with any of the 34 known advanced AMD variants for association with advanced AMD (IAMDGC data). We found no significant interaction ($P_{GxG} > 0.05/34/2$, Table S19), which suggests that the known advanced AMD effects are not modulated by the two novel early AMD variants.

Dissecting advanced AMD genetics into shared and distinct genetics for early AMD

We were interested in whether we could learn about advanced AMD genetics from a joint view of advanced and early AMD genetic effects. First, when computing genetic correlation of advanced AMD genetics with early AMD genetics, we found a substantial correlation of 78.8% (based on LD-score regression). Second, we contrasted advanced AMD effect sizes (IAMDGC data [9]) with early AMD effect sizes (our meta-analysis,) for the 34 known advanced AMD lead variants (Fig. 2, Table S16). We found two classes of variants: (1) 25 variants showed nominally significant effects on early AMD ($P < 0.05$; "advanced-and-early-AMD loci"), all directionally consistent and all smaller for early vs. advanced AMD ($OR_{early} = 1.04-1.47$; $OR_{adv} = 1.10-2.81$); (2) nine variants had no nominally significant effect on early AMD ($P \geq 0.05$; "advanced-only AMD loci"). We did not find any variant with early AMD effects into the opposite direction as the advanced AMD effects. Also, we did not find any variant-age interaction on early AMD (Table S18).
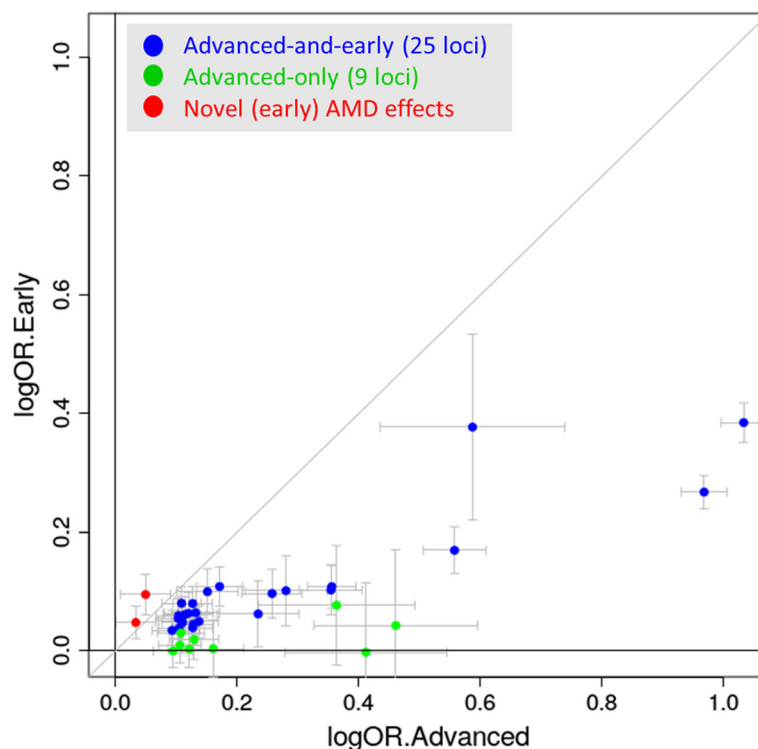
Fig. 2 Advanced vs early AMD effect sizes. Shown are advanced AMD effect sizes contrasted to early AMD effect sizes (effect sizes as log odds ratios) for the 34 known advanced AMD variants [9] (blue or green for $P_{early} < 0.05$ or $P_{early} \geq 0.05$, respectively) and for the two novel (early) AMD variants (red, near *CD46*, *TYR*). Detailed results are shown in Table S16

We observed that complement genes *CFH, CFI, C3, C9,* and *C2* were all included in the 25 advanced-and-early-AMD loci. We were thus interested in whether advanced-and-early-AMD loci suggested different pathways compared to advanced-only-AMD loci. For this, we utilized the GPS from our previous work on advanced AMD [9] to select the best-supported genes in each of these loci (Table S20). We applied Reactome pathway analyses via Enrichr [47] twice: (i) for the 35 genes in the 25 advanced-and-early-AMD loci and (ii) for the nine genes in eight advanced-only-AMD loci (no gene in the "narrow" locus definition of the *RORB* locus). This revealed significant enrichment (corrected P < 0.05) for genes from "complement system" and "lipoprotein metabolism" in the 25 advanced-and-early-AMD loci and enrichment for genes in the pathways "extracellular matrix organization" and "assembly of collagen fibrils" in the 8 advanced-only-AMD loci (Table 4). This suggested that the early AMD effect of advanced AMD variants distinguished the major known pathways for advanced AMD.

Discussion

Based on the largest genome-wide meta-analysis for early AMD to date encompassing ~ 14,000 cases and ~ 91,000 controls, all color fundus photography confirmed and a candidate approach based on 14 suggestive variants from Holliday et al. [12], we identified 10 loci for early AMD including eight novel and two previously identified for early AMD [12]. Eight of the 10 identified loci overlapped with known loci for advanced AMD [9] and two had not been detected by GWAS for early or advanced AMD so far. Our post-GWAS approach highlighted *CD46* and *TYR* as compelling candidate

genes in the two loci. Our joint view on early and advanced AMD genetics allowed us to differentiate between shared and distinct genetics for these two disease stages, which the pathway analyses suggested to be biologically relevant.

We defined three tiers of identified variants that reflected the different levels of data and evidence: tier 1 variants were identified with genome-wide significance in our GWAS meta-analysis data (identifying 8 loci including *CD46*), tier 2 variants were among the 14 candidate-based variants judged at experiment-wise significance in our non-overlapping data ($P < 0.05/14 = 0.0036$) as well as at genome-wide significance in the combined analysis (i.e. previously published summary statistics and ours; identifying one additional locus near *APOE*), and tier 3 variants were among the 14 variants judged at experiment-wise significance, but no genome-wide significance in the combined data (identifying one additional locus near *TYR*). While the establishing of genome-wide significance for a locus is ideal, testing at a statistical significance level controlling the experiment-wise type 1 error with Bonferroni-correction is also an established approach to provide statistical evidence in independent data [48].

Particularly interesting were the two loci near *CD46* and *TYR* that were identified here for early AMD with statistical significance and have not been identified previously, not even for advanced AMD. The locus around *CD46* had been reported as suggestive for early AMD by the previous largest GWAS for early AMD [12] (4089 early AMD cases, 20,453 controls), but not with statistical significance, and had not been identified with statistical significance by the previous largest GWAS for advanced AMD [9] (16,144 advanced AMD cases, 17,832

Table 4 Enriched pathways

Gene group	Reactome pathway	#Genes in gene set	#AMD loci in gene set	Raw P	Corrected P	Genes contributing to enrichment
Effects on early and advanced AMD	Regulation of Complement cascade (R-HSA-977606)	26	5	7.8×10^{-10}	1.2×10^{-6}	C3;CFH;C9;CFI;CFB
	Lipoprotein metabolism (R-HSA-174824)	34	4	3.5×10^{-7}	1.8×10^{-4}	ABCA1;CETP;LIPC;APOE
	Complement cascade (R-HSA-166658)	80	5	2.7×10^{-7}	2.0×10^{-4}	C3;CFH;C9;CFI;CFB
	HDL-mediated lipid transport (R-HSA-194223)	19	3	4.7×10^{-6}	1.8×10^{-3}	ABCA1;CETP;APOE
	Lipid digestion, mobilization, and transport (R-HSA-73923)	71	4	7.0×10^{-6}	2.2×10^{-3}	ABCA1;CETP;LIPC;APOE
	Activation of C3 and C5 (R-HSA-174577)	6	2	4.4×10^{-5}	0.011	C3;CFB
no effects on early AMD	Assembly of collagen fibrils and other multimeric structures (R-HSA-2022090)	54	3	1.5×10^{-6}	2.4×10^{-3}	COL15A1;COL8A1;MMP9
	Collagen formation (R-HSA-1474290)	85	3	6.1×10^{-6}	3.1×10^{-3}	COL15A1;COL8A1;MMP9
	Extracellular matrix organization (R-HSA-1474244)	283	4	4.7×10^{-6}	3.6×10^{-3}	VTN;COL15A1;COL8A1;MMP9

The table shows enriched pathways for highest prioritized genes (from Fritsche et al. 2016 without modifications) in the 25 late AMD loci with early AMD effects (35 genes) versus the 8 loci with no effect on early AMD (9 genes). Pathways with significant corrected *P*-value ($P_{corr} < 0.05$) for each gene group from EnrichR querying human Reactome database 2016 are shown

controls). Our meta-analysis was more than three times larger than the previous early AMD GWAS (effective sample size 48,651 compared to 13,631 [12]) and had a larger power to detect an "any AMD" effect with genome-wide significance than the previous advanced AMD GWAS (e.g. for OR = 1.10, allele frequency 30%: power = 92% compared to 61%, respectively). The *TYR* locus had not been identified with statistical significance in any previous GWAS on early or advanced AMD; it was stated as a locus with suggestive evidence from the previous GWAS on early AMD ($P = 3.5 \times 10^{-6}$) [12] and was identified here with statistical significance at experiment-wise error control. The combined analysis of the Holliday et al. and our non-overlapping data clearly separated the 14 Holliday variants into four with genome-wide or close to genome-wide significance and 10 that were far away from statistical significance.

Prioritization of genes underneath association signals is a known challenge, but highly relevant for selecting promising candidates for functional follow-up. Our systematic approach, scrutinizing all genes underneath our two newly identified loci, highlighted *CD46* and *TYR* as the most supported genes. *CD46* is an immediate compelling candidate as a part of the complement system [49]. Complement activation in retina is thought to have a causal role for AMD [50, 51]. Importantly, we found our *CD46* GWAS signal to colocalize with *CD46* expression with the early AMD risk increasing allele (rs4844620 G) increasing CD46 expression in retinal cells. On the one hand, this contrasts the presumption that a higher *CD46* expression in eye tissue should protect from AMD, based on previous CD46 expression data [52] and a documented AMD risk increasing effect for increased complement inhibition [53]. On the other hand, *CD46* had also been found to have pathogenic receptor properties for human viral and bacterial pathogens (e.g. measles virus) [54] and is known to down-modulate adaptive T helper type 1 cells [55]. Furthermore, a GWAS on neutralizing antibody response to measles vaccine had identified two intronic *CD46* variants (rs2724384, rs2724374) [56]. In our data, these two variants were in the 95% credible set for the *CD46* locus, highly correlated with our lead variant rs4844620 ($r^2 > = 0.95$), and the major alleles (rs2724374 T, rs2724384 A) increased early AMD risk. Interestingly, the rs2724374 G was shown for *CD46* exon skipping resulting in a shorter CD46 isoform with a potential role in pathogen binding [56]. Based on this, one may hypothesize that the observed *CD46* signal in early AMD is related to pathogenic receptor properties rather than complement inactivation.

At the second locus, *TYR* appears as the best supported gene by our systematic scoring. This locus and gene was already discussed by Holliday and colleagues [12]. Briefly,

TYR is important for melanin production and *TYR* variants in human were associated with skin, eye and hair color [57–59]. While we did not identify any *cis* effect between the credible set variants and *TYR* expression, one of our credible set variants in *TYR* (rs1042602) is a missense variant. Interestingly, this variant was a GWAS lead variant not only for skin color [58], but also for macular thickness in UK Biobank [60]; the allele associated with thicker retina showed increased early AMD risk in our data. Since thicker RPE/Bruch's membrane complex was associated with increased early AMD risk in the AugUR study [61], this would be in line with our early AMD risk increasing allele being linked to a process of increased accumulation of drusenoid debris in the RPE/Bruch's membrane complex. Although *CD46* and *TYR* were the most supported genes in the two loci, we could not rule-out the relevance of other genes in the loci.

It is a strength of the current study that early AMD and control status was ascertained by color fundus photography, not relying on health record data. However, the early AMD classification in our GWAS was heterogeneous across the 11 data sets: one study incorporated information from optical coherence tomography (NICOLA), the UK Biobank classification was derived by a machine-learning algorithm [20], and the IAMDGC data was multi-site with different classification approaches [9]. The uncertainty in early AMD classification and the substantial effort required for any manual AMD classification are likely reasons for the sparsity of early AMD GWAS so far. Our sensitivity analysis with the leave-one-out meta-analyses and corresponding heterogeneity estimates showed that effect estimates did not depend on one or the other data source or classification approach.

Our data on early AMD genetics is comparable in size to the existing data on advanced AMD genetics from IAMDGC (summary statistics at http://amdgenetics.org/) and thus provides an important resource (summary statistics at http://genepi-regensburg.de) to enable a joint view. By this joint view, we were able to differentiate the 34 loci known for advanced AMD into 25 "advanced-and-early-AMD loci" and nine "advanced-AMD-only loci". Pathway enrichment analyses conducted separately for these two groups effectively discriminated the major known pathways for advanced AMD genetics [9]: complement complex and lipid metabolism for "advanced-and-early-AMD" loci; extracellular matrix metabolism for "advanced-AMD-only" loci. The two novel loci around *CD46* and *TYR* fit to the definition of "advanced-and-early-AMD" loci and the *CD46* being part of the complement system supports the above stated pathway pattern. The larger effect size for early compared to advanced AMD for the two novel loci may – in part – be winner's curse.

How do our observations relate to potential etiological models? (1) For a genetic variant capturing an underlying mechanism that triggers both early and advanced AMD, we would expect the variant to show association with early and advanced AMD (compared to "healthy") with directionally consistent effects (Fig. 3, Model 1). This would be in line with our observed associations for the 27 "advanced-and-early-AMD" loci (25 known advanced AMD loci, 2 novel loci). This would also suggest that mechanisms of complement system or lipid metabolism trigger both early and advanced disease. (2) For a mechanism that triggers advanced AMD no matter whether the person is "healthy" or has early AMD, we would anticipate a variant effect for advanced AMD, but not for early AMD (Fig. 3, Model 2). This would be in line with our observed associations for the nine "advanced-AMD-only" loci. This would also suggest that mechanisms of extracellular matrix metabolism trigger advanced AMD rather than early AMD. Of note, these include the *MMP9* locus, which is thought to trigger vascularization and wet AMD [9]. (3) Another mechanism is conferred by variants that are purely responsible for progression from early to advanced AMD, but do not increase advanced AMD risk for "healthy" individuals. In such a scenario, the advanced AMD risk increasing allele would be under-represented among persons with early AMD (Fig. 3, Model 3), particularly at older age, and it would be associated with decreased risk of early AMD (compared to "healthy"). None of the identified variants showed this pattern overall or for older age in the variant x age interaction analyses. (4) For a mechanism that triggers early AMD, but has no impact on the progression from early to advanced AMD, we would

have an effect on early AMD, but no effect on advanced AMD (Fig. 3, Model 4). We did not find such a variant.

Our data and joint view on effects for both disease stages support two of the four etiological models. One may hypothesize that the unsupported models are non-existing or unlikely. There are limitations to consider: (1) To reduce complexity, we adopted an isolated view per variant with some accounting for interaction, but ignoring more complex networks. (2) Early AMD effects were estimated predominantly in population-based studies, while advanced AMD effects were from a case-control design. (3) The cut-off of "nominal significance" for separating variants into "advanced-and-early" or "advanced-only" loci is arbitrary and larger data might give rise to re-classification. Still, the power to detect effects for early AMD in our meta-analysis was similar to the power in the advanced AMD data from IAMDGC (for OR = 1.05, allele frequency 30%, nominal significance: power = 94% or 83%, respectively). (4) An improved disentangling of genetic effects for the two chronologically linked disease stages will be an important subject of further research, requiring large-scale population-based studies with long-term follow-up and the estimation of transition probabilities.

Conclusions

In summary, our large GWAS on early AMD identified novel loci, highlighted shared and distinct genetics between early and advanced AMD and provides insights into AMD etiology. The ability of early AMD effects to differentiate the major pathways for advanced AMD underscores the biological relevance of a joint view on early and advanced AMD genetics.

		Expected direction of effect in association data	
		Early AMD (vs. healthy)	Advanced AMD (vs. healthy)
Model 1: „Trigger for early and advanced AMD"	healthy → early → advanced	+	+
Model 2: „Trigger for advanced AMD"	healthy early → advanced	0	+
Model 3: „Trigger for progression"	healthy early → advanced	-	+
Model 4: „Trigger for early AMD"	healthy → early advanced	+	0

Fig. 3 Etiological models and the respective expected association of a variant with early and advanced AMD

Supplementary information

Abbreviations

3CC: Three Continent Consortium; AMD: Age-related macular degeneration; ARIC: The Atherosclerosis Risk in Communities Study; AugUR: Age-related diseases: Understanding Genetic and non-genetic influences - a study at the University of Regensburg; CHS: Cardiovascular Health Study; eQTL: Expression quantitative trait locus; FDR: False-discovery-rate; GCTA: Genome-wide Complex Trait Analysis; GC: Genomic control; GHS: Gutenberg Health Study; GPS: Gene Prioritization Score; GWAS: Genome-wide association study; HRC: Haplotype Reference Consortium; IAMDGC: International AMD Genomics Consortium; KORA: KOoperative Gesundheitsforschung in der Region Augsburg; LIFE: Leipzig Research Centre for Civilization Based Diseases - LIFE Adult population-based study, city of Leipzig, Germany; LOO: Leave-one-out; MAC: Minor allele count; MGI: Mouse Genome Informatics; NICOLA: Northern Ireland Cohort for Longitudinal Study of Ageing; NMD: Nonsense-mediated mRNA decay; OCT: Optical coherence tomography; OMIM: Online Mendelian Inheritance in Man; QC: Quality control; RPE: Retinal pigment epithelium; UKBB: UK Biobank; VEP: Variant effect predictor; WHI: Women's Health Initiative

Acknowledgments

Authors' contributions

TWW, FG1, CB, CK, FG2, IMH, KJS and BHFW designed the study and wrote the manuscript. TWW and FG1 conducted the meta-analysis. FG2 applied the automated grading of UK Biobank fundus images. TWW and CK conducted the PheWAS. TWW, LW, MEZ and KJS conducted analysis for the gene priority scoring. KJS conducted the pathway analyses. FG1, TS and TWW analysed data from the dbGAP studies ARIC, CHS and WHI. FG1 and TWW analysed data from IAMDGC, AugUR and KORA. TWW analyses data from UKBB. CAK, AP and AKS analysed data from the GHS study. MMN and AP analysed data from the KORA study. FGR, TE, KH, and MS analysed data from the LIFE-Adult study. MCG, AJMK, NQ and REH analysed all data within the NICOLA study for this project. FG2, HK, KJS, IMH and TWW wrote the discussion of etiological models. IMH, KJS and BHFW supervised the study. All authors approved the final version of the manuscript.

Funding

Availability of data and materials

The genome-wide meta-analysis summary statistics for early AMD are available for download from www.genepi-regensburg.de/earlyamd. We included three publicly available studies from dbGaP into the meta-analysis: ARIC (accession number: phs000090.v5.p1), CHS (accession number: phs000287.v6.p1), WHI (accession number: phs000746.v2.p3 and phs000200.v11.p3). Data from other studies included in the meta-analysis can be requested from the respective study centers: AugUR (https://www.uni-regensburg.de/medizin/epidemiologie-praeventivmedizin/genetische-epidemiologie/augur), GHS (http://www.gutenberg-gesundheitsstudie.de/), IAMDGC (http://amdgenetics.org/), KORA (https://www.helmholtz-muenchen.de/kora), LIFE-Adult (https://life.uni-leipzig.de/de/erwachsenenkohorten/life_adult.html), NICOLA (https://www.qub.ac.uk/sites/NICOLA/) or UKBB (https://www.ukbiobank.ac.uk/).

Ethics approval and consent to participate

The Institutional Review Board (IRB) of the University of Utah was the umbrella IRB for all other studies contributing data to the International Age-related Macular Degeneration Genomics Consortium (IAMDGC), except for the Beaver Dam Eye Study (BDES). The University of Utah approved and certified each individual study ethic committee's conduct for the data used in this study. Data provided by BDES was approved by the IRB of the University of Wisconsin. Local ethics approval for data access to the studies deposited in dbGAP (WHI, ARIC and CHS) was granted by the IRB of the University of Regensburg. For all other studies, study participants obtained informed consent and local ethics committees approved the study protocols.

Consent for publication

Not applicable.

Competing interests

M.S. receives funding from Pfizer Inc. for a project not related to this research. Retinal grading of the NICOLA study was supported by Novartis (for R.E.H) and Bayer (for Usha Chakravarthy, not a co-author). A.K.S. received financial and research support by Heidelberg Engineering, Novartis, Bayer Vital and Allergen without a link to the content of this work. I.M.H. received funding from Roche Diagnostics for a project not related to this research. None of the other authors have any conflicts of interest.

Author details

[1]Department of Genetic Epidemiology, University of Regensburg, Regensburg, Germany. [2]Institute of Human Genetics, University of Regensburg, Regensburg, Germany. [3]Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden. [4]Institute of Medical Sciences, University of Aberdeen, Aberdeen, Scotland, UK. [5]Department of Ophthalmology, University Hospital Regensburg, Regensburg, Germany. [6]Statistical Consulting Unit StaBLab, Department of Statistics, Ludwig-Maximilians-Universität Munich, Munich, Germany. [7]Department of Ophthalmology, University Medical Center of the Johannes Gutenberg-University Mainz, Mainz, Germany. [8]Institute for Medical Biostatistics, Epidemiology and Informatics, University Medical Center of the Johannes Gutenberg-University Mainz, Mainz, Germany. [9]Institute of Genetic Epidemiology, Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg, Germany. [10]Department of Internal Medicine I (Cardiology), Hospital of the Ludwig-Maximilians-University (LMU) Munich, Munich, Germany. [11]Genetic Epidemiology, IBE, Faculty of Medicine, LMU Munich, Munich, Germany. [12]German Center for Diabetes Research (DZD), Neuherberg, Germany. [13]Institute of Epidemiology, Helmholtz Zentrum München Research Center for Environmental Health, Neuherberg, Germany. [14]Leipzig Research Centre for Civilization Diseases (LIFE), Leipzig University, Leipzig, Germany. [15]Institute for Medical Informatics, Statistics, and Epidemiology (IMISE), Leipzig University, Leipzig, Germany. [16]Schepens Eye Research Institute, Harvard Medical School, Boston, MA, USA. [17]Centre for Public Health, Queen's University of Belfast, Belfast, UK. [18]Institute of Clinical Human Genetics, University Hospital Regensburg, Regensburg, Germany.

References

1. Lim LS, Mitchell P, Seddon JM, Holz FG, Wong TY. Age-related macular degeneration. Lancet Elsevier. 2012;379:1728–38.
2. Garrity ST, Sarraf D, Freund KB, Sadda SR. Multimodal imaging of Nonneovascular age-related macular degeneration. Invest Ophthalmol Vis Sci. 2018;59:AMD48–64.
3. Forte R, Querques G, Querques L, Massamba N, Le Tien V, Souied EH. Multimodal imaging of dry age-related macular degeneration. Acta Ophthalmol. 2012;90:281–7.
4. Klein R, Meuer SM, Myers CE, Buitendijk GHS, Rochtchina E, Choudhury F, et al. Harmonizing the classification of age-related macular degeneration in the three-continent AMD consortium. Ophthalmic Epidemiol. 2014;21:14–23.
5. Brandl C, Zimmermann ME, Günther F, Barth T, Olden M, Schelter SC, et al. On the impact of different approaches to classify age-related macular degeneration: results from the German AugUR study. Sci Rep. 2018;8:1–10.
6. AREDS. Risk factors associated with age-related macular degeneration. A case-control study in the age-related eye disease study: age-related eye disease study report number 3. Age-Related Eye Disease Study Research Group. Ophthalmology. 2000;107:2224–32.
7. Smith W, Assink J, Klein R, Mitchell P, Klaver CC, Klein BE, et al. Risk factors for age-related macular degeneration: pooled findings from three continents. Ophthalmology Elsevier. 2001;108:697–704.
8. Yonekawa Y, Miller J, Kim I. Age-related macular degeneration: advances in management and diagnosis. J Clin Med. 2015;4:343–59.
9. Fritsche LG, Igl W, Bailey JNC, Grassmann F, Sengupta S, Bragg-Gresham JL, et al. A large genome-wide association study of age-related macular degeneration highlights contributions of rare and common variants. Nat Genet. 2016;48:134–43.
10. Grassmann F, Fritsche LG, Keilhauer CN, Heid IM, Weber BHF. Modelling the genetic risk in age-related macular degeneration. PLoS One. 2012;7.
11. Han X, Gharahkhani P, Mitchell P, Liew G, Hewitt AW, MacGregor S. Genome-wide meta-analysis identifies novel loci associated with age-related macular degeneration. J. Hum. Genet. 2020;10:1.

12. Holliday EG, Smith AV, Cornes BK, Buitendijk GHS, Jensen RA, Sim X, et al. Insights into the Genetic Architecture of Early Stage Age-Related Macular Degeneration: A Genome-Wide Association Study Meta-Analysis. PLoS One. 2013;8:e53830.

13. Auton A, Abecasis GR, Altshuler DM, Durbin RM, Bentley DR, Chakravarti A, et al. A global reference for human genetic variation. Nature. 2015;526:68–74.

14. McCarthy S, Das S, Kretzschmar W, Delaneau O, Wood AR, Teumer A, et al. A reference panel of 64,976 haplotypes for genotype imputation. Nat Genet. 2016;48:1279–83.

15. Walter K, Min JL, Huang J, Crooks L, Memari Y, McCarthy S, et al. The UK10K project identifies rare variants in health and disease. Nature. 2015;526:82–9.

16. Bycroft C, Freeman C, Petkova D, Band G, Elliott LT, Sharp K, et al. The UK biobank resource with deep phenotyping and genomic data. Nature. 2018;562:203–9.

17. Korb CA, Kottler UB, Wolfram C, Hoehn R, Schulz A, Zwiener I, et al. Prevalence of age-related macular degeneration in a large European cohort: results from the population-based Gutenberg health study. Graefe's arch. Clin. Exp. Ophthalmol. Springer. Berlin Heidelberg. 2014; 252:1403–11.

18. Ferris FL, Wilkinson CP, Bird A, Chakravarthy U, Chew E, Csaky K, et al. Clinical classification of age-related macular degeneration. Ophthalmology. 2013;120:844–51.

19. Brandl C, Breinlich V, Stark KJ, Enzinger S, Aßenmacher M, Olden M, et al. Features of Age-Related Macular Degeneration in the General Adults and Their Dependency on Age, Sex, and Smoking: Results from the German KORA Study. Thatcher TH, editor. PLoS One. Public Libr Sci. 2016;11:e0167181.

20. Grassmann F, Mengelkamp J, Brandl C, Harsch S, Zimmermann ME, Linkohr B, et al. A deep learning algorithm for prediction of age-related eye disease study severity scale for age-related macular degeneration from color fundus photography. Ophthalmology. 2018;125:1410–20.

21. Guenther F, Brandl C, Winkler TW, et al. Chances and challenges of machine learning-based disease classification in genetic association studies illustrated on age-related macular degeneration. Genet Epidemiol. 2020;10.1002/gepi.22336.

22. De Raadt A, Warrens MJ, Bosker RJ, HAL K. Kappa Coefficients for Missing Data. Educ. Psychol. Meas. 2019.

23. Zhan X, Hu Y, Li B, Abecasis GR, Liu DJ. RVTESTS: an efficient and comprehensive tool for rare variant association analysis using sequence data. Bioinformatics. 2016;32:1423–6.

24. Gorski M, Günther F, Winkler TW, Weber BHF, Heid IM. On the differences between mega- and meta-imputation and analysis exemplified on the genetics of age-related macular degeneration. Genet Epidemiol. 2019;43:559–76.

25. Winkler TW, Day FR, Croteau-Chonka DC, Wood AR, Locke AE, Mägi R, et al. Quality control and conduct of genome-wide association meta-analyses. Nat Protoc. 2014;9:1192–212.

26. Devlin AB, Roeder K, Devlin B. Genomic Control for Association. 2013;55:997–1004.

27. Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. Bioinformatics. 2010;26:2190–1.

28. Harrer M, Cuijpers P, Furukawa T, Ebert DD. Doing meta-analysis in R: a hands-on guide. Lab: Prot; 2019.

29. Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA (editors). Cochrane Handbook for Systematic Reviews of Interventions. 2nd Edition. Chichester: Wiley; 2019.

30. Huedo-Medina TB, Sánchez-Meca J, Marín-Martínez F, Botella J. Assessing heterogeneity in meta-analysis: Q statistic or I 2 index? Psychol Methods. 2006;11:193–206.

31. Yang J, Ferreira T, Morris AP, Medland SE, Madden PAF, Heath AC, et al. Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. Nat Genet Nature Publishing Group. 2012;44:369–75.

32. Kichaev G, Yang WY, Lindstrom S, Hormozdiari F, Eskin E, Price AL, et al. Integrating functional data to prioritize causal variants in statistical fine-mapping studies. PLoS Genet. 2014;10.

33. McLaren W, Gil L, Hunt SE, Riat HS, Ritchie GRS, Thormann A, et al. The Ensembl Variant Effect Predictor. Genome Biol. 2016;17:122.

34. Ratnapriya R, Sosina OA, Starostik MR, Kwicklis M, Kapphahn RJ, Fritsche LG, et al. Retinal transcriptome and eQTL analyses identify genes associated with age-related macular degeneration. Nat Genet. 2019;51:606–10.

35. Aguet F, Brown AA, Castel SE, Davis JR, He Y, Jo B, et al. Genetic effects on gene expression across human tissues. Nature. 2017;550:204–13.

36. Hormozdiari F, van de Bunt M, Segrè AV, Li X, Joo JWJ, Bilow M, et al. Colocalization of GWAS and eQTL signals detects target genes. Am. J Hum Genet. 2016;99:1245–60.

37. Bryan JM, Fufa TD, Bharti K, Brooks BP, Hufnagel RB, McGaughey DM. Identifying core biological processes distinguishing human eye tissues with precise systems-level gene expression analyses and weighted correlation networks. Hum Mol Genet Narnia. 2018;27:3325–39.

38. Bult CJ, Blake JA, Smith CL, Kadin JA, Richardson JE, Anagnostopoulos A, et al. Mouse genome database (MGD) 2019. Nucleic Acids Res Oxford University Press. 2019;47:D801–6.

39. Grassmann F, Kiel C, Zimmermann ME, Gorski M, Grassmann V, Stark K, et al. Genetic pleiotropy between age-related macular degeneration and 16 complex diseases and traits. Genome Med Genome Medicine. 2017;9:1–13.

40. MacArthur J, Bowler E, Cerezo M, Gil L, Hall P, Hastings E, et al. The new NHGRI-EBI catalog of published genome-wide association studies (GWAS catalog). Nucleic Acids Res. 2017;45:D896–901.

41. Beck T, Hastings RK, Gollapudi S, Free RC, Brookes AJ. GWAS central: a comprehensive resource for the comparison and interrogation of genome-wide association studies. Eur J Hum Genet Nature Publishing Group. 2014;22:949–52.

42. Canela-Xandri O, Rawlik K, Tenesa A. An atlas of genetic associations in UK biobank. Nat Genet Springer US. 2018;50:1593–9.

43. oav B, Hochberg Y. Controlling the False Discovery Rate - a Practical and Powerful Approach to Multiple Testing. Journal of the Royal Statistical Society Series B-Methodological 1995.pdf. JR Stat Soc Ser B. 1995.

44. Winkler TW, Brandl C, Grassmann F, Gorski M, Stark K, Loss J, et al. Investigating the modulation of genetic effects on late AMD by age and sex: lessons learned and two additional loci. PLoS One. 2018;13:1–21.

45. Bulik-Sullivan B, Loh PR, Finucane HK, Ripke S, Yang J, Patterson N, et al. LD score regression distinguishes confounding from polygenicity in genome-wide association studies. NatGenet Nature Publishing Group. 2015;47:291–5.

46. Winkler TW, Günther F, Höllerer S, Zimmermann M, Loos RJ, Kutalik Z, et al. A joint view on genetic variants for adiposity differentiates subtypes with distinct metabolic implications. Nat Commun Springer US. 2018;9.

47. Kuleshov MV, Jones MR, Rouillard AD, Fernandez NF, Duan Q, Wang Z, et al. Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. Nucleic Acids Res. 2016;44:W90–7.

48. Gao X, Starmer J, Martin ER. A multiple testing correction method for genetic association studies using correlated single nucleotide polymorphisms. Genet Epidemiol. 2008.

49. Fabregat A, Jupe S, Matthews L, Sidiropoulos K, Gillespie M, Garapati P, et al. The Reactome Pathway Knowledgebase. Nucleic Acids Res. 2018; 46:D649–55.

50. Hageman GS, Luthert PJ, Victor Chong NH, Johnson LV, Anderson DH, Mullins RF. An integrated hypothesis that considers drusen as biomarkers of immune-mediated processes at the RPE-Bruch's membrane interface in aging and age-related macular degeneration. Prog Retin Eye Res. 2001;20:705–32.

51. Johnson LV, Leitner WP, Staples MK, Anderson DH. Complement activation and inflammatory processes in Drusen formation and age related macular degeneration. Exp Eye Res. 2001;73:887–96.

52. Vogt SD, Curcio CA, Wang L, Li CM, McGwin G Jr, Medeiros NE, et al. Retinal pigment epithelial expression of complement regulator CD46 is altered early in the course of geographic atrophy. Exp Eye Res. 2011;93: 413–23.

53. Seya T, Atkinson JP. Functional properties of membrane cofactor protein of complement. Biochem J. 1989;264:581–8.

54. Cattaneo R. Four viruses, two Bacteria, and one receptor: membrane cofactor protein (CD46) as pathogens' magnet. J Virol. 2004;78:4385–8.

55. Cardone J, Le Friec G, Kemper C. CD46 in innate and adaptive immunity: an update. Clin Exp Immunol. 2011;164:301–11.

56. Haralambieva IH, Ovsyannikova IG, Kennedy RB, Larrabee BR, Zimmermann MT, Grill DE, et al. Genome-wide associations of CD46 and IFI44L genetic variants with neutralizing antibody response to measles vaccine. Hum Genet. 2017;136:421–35.

57. Lewis RA. Oculocutaneous Albinism Type 1 [Internet]. GeneReviews®. Seattle: University of Washington; 1993.

58. Galván-Femenía I, Obón-Santacana M, Piñeyro D, Guindo-Martinez M, Duran X, Carreras A, et al. Multitrait genome association analysis identifies new susceptibility genes for human anthropometric variation in the GCAT cohort. J Med Genet. 2018:765–78.

59. Stokowski RP, Pant PVK, Dadd T, Fereday A, Hinds DA, Jarman C, et al. A Genomewide association study of skin pigmentation in a south Asian population. Am J Hum Genet Cell Press. 2007;81:1119–32.

60. Gao XR, Huang H, Kim H. Genome-wide association analyses identify 139 loci associated with macular thickness in the UK biobank cohort. Hum Mol Genet. 2019;28:1162–72.

61. Brandl C, Brücklmayer C, Günther F, Zimmermann ME, Küchenhoff H, Helbig H, et al. Retinal layer thicknesses in early age-related macular degeneration: results from the German AugUR study. Invest Ophthalmol Vis Sci. 2019;60: 1581–94.

Publisher's Note